

## Whittle estimation of EGARCH and other exponential volatility models

Zaffaroni, Paolo

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

[www.peerproject.eu](http://www.peerproject.eu)

### Empfohlene Zitierung / Suggested Citation:

Zaffaroni, P. (2009). Whittle estimation of EGARCH and other exponential volatility models. *Journal of Econometrics*, 151(2), 190-200. <https://doi.org/10.1016/j.jeconom.2009.03.008>

### Nutzungsbedingungen:

Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: <http://www.peerproject.eu> Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

### Terms of use:

This document is made available under the "PEER Licence Agreement". For more information regarding the PEER-project see: <http://www.peerproject.eu> This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der  
  
Leibniz-Gemeinschaft

## Accepted Manuscript

Whittle estimation of EGARCH and other exponential volatility models

Paolo Zaffaroni

PII: S0304-4076(09)00082-7  
DOI: 10.1016/j.jeconom.2009.03.008  
Reference: ECONOM 3192

To appear in: *Journal of Econometrics*

Received date: 24 November 2008

Accepted date: 13 March 2009



Please cite this article as: Zaffaroni, P., Whittle estimation of EGARCH and other exponential volatility models. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.03.008

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Whittle estimation of EGARCH and other exponential volatility models <sup>1</sup>

Paolo Zaffaroni

IMPERIAL COLLEGE BUSINESS SCHOOL  
IMPERIAL COLLEGE LONDON

This draft: 24th November 2008

## Abstract

The strong consistency and asymptotic normality of the Whittle estimate of the parameters in a class of exponential volatility processes are established. Our main focus here are the EGARCH model of Nelson (1991) and other one-shock models such as the GJR model of Glosten, Jaganathan, and Runkle (1993), but two-shock models, such as the SV model of Taylor (1986), are also comprised by our assumptions. The variable of interest might not have finite fractional moment of any order and so, in particular, finite variance is not imposed. We allow for a wide range of degrees of persistence of shocks to conditional variance, allowing for both short and long memory.

*Key words and phrases.* EGARCH, GJR, Stochastic Volatility, Whittle estimation, asymptotics.

## 1 Introduction

Consider an observable satisfying

$$x_t = z_t e^{0.5 h_t}, \quad t \in \mathbb{Z}, \quad (1)$$

$$h_t = \omega_0 + \sum_{k=0}^{\infty} \psi_{0k} \epsilon_{t-k-1} \quad \text{almost surely (a.s.)}, \quad \sum_{j=0}^{\infty} \psi_{0j}^2 < \infty, \quad (2)$$

---

<sup>1</sup>Acknowledgment: I thank two anonymous referees and the Associate Editor (Miguel A. Delgado) for their comments that led to a considerably improved version of the paper. I also thank Claudia Miani for remarkable research assistance.

where  $\mathbb{Z} = \{t : t = 0, \pm 1, \dots\}$ . The  $\{z_t, \epsilon_t\}$  form a sequence of independently identically distributed (*i.i.d.*) unobservable bivariate random variable, although we allow  $z_t$  and  $\epsilon_s$  to be cross-correlated for  $t = s$ . We require, at minimum,  $E\epsilon_0 = 0$ ,  $0 < E\epsilon_0^2 < \infty$  and  $E\log z_0^2 < \infty$  although existence of the moments of the  $z_t$  is not required.

When  $z_t$  and  $\epsilon_t$  are mutually independent and normally distributed, model (1)-(2) becomes the well-known exponential stochastic volatility (SV) model of Taylor (1986). Simple yet successful estimation of such SV models can be carried out by noting that

$$\log x_t^2 = \log z_t^2 + h_t, \quad (3)$$

represents an example of a *linear* signal-plus-noise model. For parameterizations of the  $\psi_{0j}$  that ensure Markovianity, the Kalman filter could be successfully applied to SV models (see Nelson (1988) and Harvey, Ruiz, and Shephard (1994)), where (3) represents the measurement equation and  $h_t$  is the state unobserved variable. An alternative estimation approach for (3) is the Whittle estimator, obtained maximizing the frequency domain approximation of the Gaussian likelihood, so-called Whittle function (see Whittle (1962)).

This paper is prompted by considering that another popular class of exponential volatility models, where  $\epsilon_t = \epsilon(z_t)$  for some instantaneous transformation  $\epsilon(\cdot)$ , also belongs to the class (1)-(2). Such one-shock models, the most important case of which is the exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model of Nelson (1991), can be seen a SV models for which a *singularity* occurs with respect to the joint distribution of the  $z_t$  and  $\epsilon_t$ . For instance EGARCH requires that

$$\epsilon_t = \epsilon(z_t) = \theta_0(z_t - \mu_z) + \delta_0(|z_t| - \mu_{|z|}), \quad (4)$$

for constant parameters  $\theta_0, \delta_0$  with  $\theta_0\delta_0 \neq 0$  where  $\mu_y = Ey$  for any random variable  $y$  with finite first moment. Another one-shock model, popular among practitioners, is the so-called GJR volatility of Glosten, Jaganathan, and

Runkle (1993) for which

$$\epsilon_t = \epsilon(z_t) = \theta_0(z_t - \mu_z) + \delta_0(z_t 1_{(z_t < 0)} - \mu_z 1_{(z < 0)}), \quad (5)$$

where  $1_{(\cdot)}$  denotes the indicator function. Note that  $\epsilon(\cdot)$  is an odd function in both (4) and (5) yielding asymmetric models. Since (1)-(2) holds, the signal-plus-noise representation (3) can be obtained for one-shock models although now the signal  $h_t$  and the noise  $\log z_s^2$  are *correlated* for some  $t, s$ .

It turns out that all the statistical literature of Whittle estimation of linear signal-plus-noise models requires uncorrelated components (see in particular Hosoya (1974), Dunsmuir (1979) and Hosoya and Taniguchi (1982)) and, indeed, the Whittle estimator has been successfully applied to estimation of SV models by Harvey (1998), Breidt, Crato, and deLima (1998) and Deo, Hurvich, and Lu (2006). However, for one-shock exponential models such as the EGARCH or the GJR, one cannot use these results since, by (4) or (5),  $\log z_t^2$  and  $h_s$  could be correlated for some  $t, s$ . The difficulty in estimating the signal-plus-noise model (3) is due to the fact that, despite linearity, the spectral density of the observable  $\log x_t^2$  is not easily factored meaning that it cannot be expressed as  $m_0^2 / (2\pi) |1 + \sum_{k=1}^{\infty} n_{0k} e^{ik\lambda}|^2$  with  $m_0$  and the  $n_{0k}$ ,  $k = 0, 1, \dots$  being respectively function of two disjoint sets of the model parameters (see Hannan (1973)). An alternative approach to this problem is proposed in Robinson (1978), who presents various cases where the model spectral density is not easily factored, other than the signal-plus-noise model. Linearity is nowhere assumed but long memory, except for a mild form, is ruled out and, due to its generality, some assumptions appear un-primitive.

The appeal of the exponential model (1)-(2) is based on the fact that it solves many of the drawbacks characterizing the structure of the ARCH model of Engle (1982). In particular, one needs not to impose non-negativity of the  $\omega_0, \psi_{0k}$  ( $k \geq 0$ ). Second, asymmetric effects, whereby volatility tends to rise in response to ‘bad’ news and to fall in response to ‘bad news’, are easy to parameterize. Third, there is no ambiguity on the interpretation of the persistence of shocks to conditional variance. These were in fact the chief

motivations that led in fact Nelson (1991) to introduce the EGARCH model. Fourth, as the model is observed at a finer and finer time interval, it has a diffusion limit which belongs to the class of continuous time processes frequently used in continuous time mathematical finance. This property holds for both one and two-shock models (see Ghysels, Harvey, and Renault (1995, Section 4.1 and 4.3)). Not surprisingly, EGARCH models are attracting a constantly increasing attention in theoretical (see Duan, Gauthier, Sasseville, and Simonato (2006)) and empirical finance (see Brandt and Jones (2006) among many others).

The main contributions of this paper can be synthesized as follows:

First, we extend the statistical theory of Whittle estimation to cover *correlated* signal-plus-noise models, providing a formal asymptotic distribution theory specifically tailored for parameter estimation of the exponential model (1)-(2), both for the two-shock as well as for the one-shock version. This is relevant since, with the exception of the low-order EGARCH result of Straumann (2005), for general EGARCH, GJR and any other exponential one-shock model, no other estimation approach exists for which we have a complete, formal, understanding of its asymptotic statistical properties.

Second, our theory covers both cases of summable and non-summable  $|\psi_{0j}|$ . Important examples of the latter case, which implies long memory in  $h_t$ , is the fractionally integrated EGARCH (FIEGARCH) of Bollerslev and Mikkelsen (1996), when considering one-shock models, and the long memory SV of Harvey (1998) and Breidt, Crato, and deLima (1998), when considering two-shock models. Note that the statistical literature does cover the case of linear-plus-signal model with long memory (see Hosoya (1997)) but, again, the case of correlated signal and noise is ruled out.

Third, our asymptotic results are based on a set of regularity conditions, easily verifiable with respect to any given choice of the  $\psi_{0k}$ . Our result covers the situation of uncorrelated signal and noise, as a special case. Therefore, even for the case of SV models, one could use our results. This is highly

desirable since the statistical literature of Whittle estimation typically defines regularity conditions in terms of smoothness of the model spectral density and their higher-order mixed derivatives, and checking such conditions can become an arduous task. (See for instance Assumptions C and D of Hosoya (1997).)

Fourth, we present a unified theory which depends on a set of regularity conditions designed to apply to both one and two-shock models that applies to a class of models wider than EGARCH and SV. Meddahi and Renault (2004) firstly pointed out that for many purposes the difference between one-shock GARCH-type and two-shock SV-type models is only apparent rather than substantial. Our results provide a case where this analogy holds with respect to Whittle estimation.

Practical estimation of the exponential model (1)-(2) requires to finite-parameterize the  $\psi_{0k} = \psi_k(\zeta_0)$  for a known set of functions  $\psi_k(\cdot)$  and unknown  $p \times 1$  parameter  $\zeta_0$ , where  $p < \infty$ . Next, let  $\alpha_0 = \alpha(\phi_0) = \text{var}(\log z_0^2)$ ,  $\beta_0 = \beta(\phi_0) = \text{var}(\epsilon_0)$ ,  $\gamma_0 = \gamma(\phi_0) = \text{cov}(\log z_0^2, \epsilon_0)$  for known functions  $\alpha(\cdot)$ ,  $\beta(\cdot)$ ,  $\gamma(\cdot)$  of a  $q \times 1$  unknown vector  $\phi_0$  with  $q < \infty$ . This includes both the case of a known parametric specification for the joint distribution of the  $\{z_t, \epsilon_t\}$ , depending on the unknown  $\phi_0$ , as well as the case of an unspecified distribution in which case  $\phi_0 = (\alpha_0, \beta_0, \gamma_0)'$  with  $q = 3$ . We wish to estimate the  $(p+q) \times 1$  vector  $\vartheta_0 = (\zeta_0', \phi_0')'$ , on the basis of a sample  $(x_1, \dots, x_T)$  of observations. Denote by  $\vartheta = (\zeta', \phi')'$  any admissible value to which corresponds the function

$$f(\lambda; \vartheta) = \frac{\alpha(\phi)}{2\pi} + \frac{\beta(\phi)}{2\pi} |\psi(e^{i\lambda}; \zeta)|^2 + \frac{\gamma(\phi)}{2\pi} (e^{i\lambda}\psi(e^{i\lambda}; \zeta) + e^{-i\lambda}\psi(e^{-i\lambda}; \zeta)), \quad -\pi \leq \lambda < \pi \quad (6)$$

with

$$\psi(z; \zeta) = \sum_{j=0}^{\infty} \psi_j(\zeta) z^j, \quad |z| \leq 1. \quad (7)$$

Note that no truncation of the transfer function  $\psi(z; \zeta)$  is needed here. It can be easily seen that  $f(\lambda) = f(\lambda; \vartheta_0)$  is the spectrum of the  $\log x_t^2$  and,

thus,  $f(\lambda; \vartheta)$  denotes the model-spectrum. Let  $\Theta$  be a prescribed compact subset of  $R^{p+q}$ . The Whittle estimator of  $\vartheta_0$  is

$$\hat{\vartheta}_T = \arg \min_{\vartheta \in \Theta} Q_T(\vartheta),$$

where the discrete Whittle function is

$$Q_T(\vartheta) = \frac{1}{T} \sum_{t=1}^{T-1} \left( \log(f(\lambda_t; \vartheta)) + \frac{I_T(\lambda_t)}{f(\lambda_t; \vartheta)} \right), \quad \lambda_t = \frac{2\pi t}{T}. \quad (8)$$

Hereafter  $I_T(\lambda) = (2\pi T)^{-1} \left| \sum_{t=1}^T \log x_t^2 e^{i\lambda t} \right|^2$ ,  $-\pi \leq \lambda < \pi$ , is the periodogram based on  $T$  consecutive observations of the  $\log x_t^2$  where we can avoid mean correction since  $I_T(\lambda)$  is evaluated at the Fourier frequencies.

The following section lists our assumptions, with discussion. Section 3 presents the main results, namely strong consistency and asymptotic normality of  $\hat{\vartheta}_T$  under conditions that cover a wide variety of parametric specifications, comprising both exponentially and hyperbolically decaying coefficients  $\psi_j(\zeta)$ . Section 4 illustrates how the main results apply to EGARCH and clarifies the advantages and disadvantages of the Whittle estimation approach with respect to other methods. A number of extensions are introduced in Section 5 such as nonstationary  $\log x_t^2$ , arising from non square-summability of the  $\psi_{0j}$ , and filtering and forecasting. Concluding remarks make Section 6. The proofs are reported in the final appendix.

## 2 Assumptions

Denote by  $K$  a generic finite constant, not always the same. Let  $k_1 \geq 0$ ,  $k_2 \in \{0, 1\}$  and  $l$  a non-negative integer. Let  $\sim$  denote asymptotic equivalence:  $a(x) \sim b(x)$  as  $x \rightarrow x_0$  when  $a(x)/b(x) \rightarrow 1$ .

**Assumption**  $A(k_1)$ . *The  $\{z_t, \epsilon_t\}$  are i.i.d. variates with  $E\epsilon_0 = 0$  and  $E |\log z_0^2|^{k_1} < \infty$ ,  $E |\epsilon_0|^{k_1} < \infty$ .*

**Assumption B.**  $\Theta = Z \times \Phi$  where  $Z$  is a compact subspace of  $R^p$  and  $\Phi$  is a compact subspace of  $R^q$  with  $s = p + q < \infty$ .  $\vartheta_0$  is an interior point of  $\Theta$ .

**Assumption C**( $k_2, l$ ). For any  $\vartheta \in \Theta$ ,  $|\psi_k(\zeta)| \leq K |\psi_j(\zeta)|$  for  $1 \leq j \leq k$ , all  $k \geq 1$ , and  $\psi_j(\zeta)$  has continuous  $l$ th derivative such that for boundedly differentiable functions  $d(\zeta) \in (-\infty, 1/2)$  and  $e(\zeta) \in (-1, 1)$ , both not function of  $l$ ,

$$\frac{\partial^r \psi_j(\zeta)}{\partial \zeta_{i_1} \dots \partial \zeta_{i_r}} \sim k_2 E_r(j; \zeta) e^j(\zeta) + (1 - k_2) D_r(j; \zeta) j^{d(\zeta)-1} \quad \text{as } j \rightarrow \infty$$

for all  $i_h = 1, \dots, p$ ,  $h = 1, \dots, r$ ,  $r = 0, \dots, l$ , where  $|E_r(j; \zeta)| \leq K j^r$  and  $D_r(j; \zeta)$  is measurable slowly varying at infinity:  $D_r(tx; \zeta)/D_r(x; \zeta) \rightarrow 1$  as  $x \rightarrow \infty$  for any  $t > 0$  (see Yong (1974, Def. I-7)).

**Assumption D**( $k_2, l$ ). For any  $\vartheta \in \Theta$ ,  $\alpha(\phi)$ ,  $\beta(\phi)$ ,  $\gamma(\phi)$  and all the  $\psi_j(\zeta)$  have continuous  $l$ th derivative and

$$\left| \frac{\partial^r \psi_j(\zeta)}{\partial \zeta_{i_1} \dots \partial \zeta_{i_r}} - \frac{\partial^r \psi_{j+1}(\zeta)}{\partial \zeta_{i_1} \dots \partial \zeta_{i_r}} \right| \leq K (k_2 + |1 - k_2| j^{-1}) \left| \frac{\partial^r \psi_j(\zeta)}{\partial \zeta_{i_1} \dots \partial \zeta_{i_r}} \right| \quad \text{for any } j > J,$$

for some constant  $J < \infty$  and all  $i_h = 1, \dots, s$ ,  $h = 1, \dots, r$ ,  $r = 0, \dots, l$ .

**Assumption E.** For any  $\vartheta \in \Theta$ ,  $\alpha(\phi)$ ,  $\beta(\phi)$ ,  $\gamma(\phi)$  and all the  $\psi_j(\zeta)$  are continuously differentiable and there exist integers  $j_i(\vartheta)$ ,  $i = 2, \dots, s$ , such that  $1 \leq j_2(\vartheta) < \dots < j_s(\vartheta) < \infty$  and the  $s \times s$  matrix

$$\left( \frac{\partial}{\partial \vartheta} c_0(\vartheta), \frac{\partial}{\partial \vartheta} c_{j_2}(\vartheta), \dots, \frac{\partial}{\partial \vartheta} c_{j_s}(\vartheta) \right)' \quad (9)$$

has full rank, setting

$$c_u(\vartheta) = 1_{(u=0)} \alpha(\phi) + \beta(\phi) \sum_{j=0}^{\infty} \psi_j(\zeta) \psi_{j+u}(\zeta) + 1_{(u \neq 0)} \gamma(\phi) \psi_{|u|-1}(\zeta), \quad u = 0, \pm 1, \dots \quad (10)$$

**Assumption F.** For any  $\vartheta \in \Theta$ ,  $|\gamma(\phi)| < (\alpha(\phi) \beta(\phi))^{\frac{1}{2}} < \infty$ .

**Remarks.**

(i). We require Ass. A(2) for consistency and A(4) for asymptotic normality of  $\hat{\vartheta}_T$ . For two-shock models the conditional expectation  $E(x_t^{2\delta} \mid \mathcal{F}_{t-1})$ ,  $\delta > 0$ , might not be bounded, where  $\mathcal{F}_{t-1}$  defines the sigma-algebra induced by the  $\{z_s, \epsilon_s; s \leq t-1\}$ . Instead, for one-shock model a stronger moment condition for the  $z_t$  is implied, since  $E(x_t^2 \mid \mathcal{F}_{t-1}) < \infty$  when (4) holds. However, the unconditional moment  $E|x_t|^\delta$  need not to be bounded for any  $\delta > 0$ . Important examples of this case are when the  $z_t$  have a Student-t distribution with  $\nu > 2$  degrees of freedom or a generalized exponential distribution (henceforth GED) with tail thickness parameter  $\nu \leq 1$ . See Nelson (1991, p.453).

(ii). The  $x_t$  are strictly stationary and ergodic under Ass. A(2) and square summability of the  $\psi_{0j}$  whereas  $|\log x_t^2| = \infty$  a.s. when the  $\psi_{0j}$  are not square summable .

(iii). Ass. B implies that there exist constant  $0 < \alpha_L < \alpha_U < \infty$ ,  $0 < \beta_L < \beta_U < \infty$ ,  $-1 < e_L < e_U < 1$  and  $-\infty < d_L < d_U < 1/2$  such that  $\alpha_L \leq \alpha(\phi) \leq \alpha_U$ ,  $\beta_L \leq \beta(\phi) \leq \beta_U$  and  $e_L \leq e(\zeta) \leq e_U$ ,  $d_L \leq d(\zeta) \leq d_U$  for any  $\vartheta \in \Theta$ .

(iv). The parameter  $\omega_0$  is not identified by the Whittle function, since enters linearly in  $\log x_t^2$  and it disappears when calculating the empirical autocovariances of the  $\log x_t^2$ . Nevertheless alternative estimation methods do exist and will be discussed in the sequel.

(v). We are concerned here with two cases: exponentially decaying (case  $k_2 = 1$ ) and hyperbolically decaying (case  $k_2 = 0$ ) coefficients  $\psi_j(\zeta)$ . The functions  $E_r(\cdot)$ ,  $D_r(\cdot)$  arise as a result of differentiation in most cases of interest. For example, typically  $|D_r(j; \zeta)| < K(\log(j+1))^r$ . When  $d(\zeta) = 0$  one has to distinguish the two cases of summable and non-summable  $|D_r(j; \zeta)| j^{-1}$ . In the latter case the model spectral density is still unbounded at zero frequency, although it diverges very slowly as the zero frequency is approached. In the former case we allow for the possibility that  $D_r(j; \zeta) = 0$  when  $d(\zeta) = 0$ .

(vi). Ass.  $B$  and  $C(k_2, 0)$  imply that for any  $\vartheta \in \Theta$

$$\sup_{\vartheta \in \Theta} \psi_j(\zeta) \leq K j^{d_\nu - 1}, \quad j \geq 1, \quad (11)$$

and thus  $\sup_{\vartheta \in \Theta} \sum_{j=0}^{\infty} \psi_j^2(\zeta) \leq \sum_{j=0}^{\infty} \sup_{\vartheta \in \Theta} \psi_j^2(\zeta) < \infty$ .

(vii). Ass.  $D(k_2, l)$  implies that the  $\psi_j(\zeta)$  and its derivatives converge toward zero in a sufficiently smooth manner, implying quasi monotonic convergence toward zero and pure bounded variation (see Yong (1974, Definitions I-2 and I-4)). As we will see, these, together with the exact rate condition of Ass.  $C(k_2, l)$ , define unambiguously the behaviour near the origin of the model spectral density and its derivatives, as well as a form of uniform continuity away from zero frequency. When  $k_2 = 1$  this is already implied by Ass.  $C(1, l)$ , which imparts absolute summability of  $\psi_j(\zeta)$  and of their derivatives.

(viii). Ass.  $E$  is a *rank* identification assumption. It is easy to see that  $c_u = c_u(\vartheta_0)$ ,  $u = 0, \pm 1, \dots$  defines the autocovariance function of the  $\log x_t^2$ . By simple calculations, the left hand side of (9) can be expressed

$$\left( \begin{array}{c} e_1 \frac{\partial \alpha(\phi)}{\partial \phi'} + \begin{pmatrix} \Delta_0(\zeta) \\ \Delta_{j_2}(\zeta) \\ \vdots \\ \Delta_{j_s}(\zeta) \end{pmatrix} \frac{\partial \beta(\phi)}{\partial \phi'} + \begin{pmatrix} 0 \\ \psi_{j_2-1}(\zeta) \\ \vdots \\ \psi_{j_s-1}(\zeta) \end{pmatrix} \frac{\partial \gamma(\phi)}{\partial \phi'} \\ \vdots \\ \vdots \\ \vdots \end{array} \right) \begin{array}{c} \vdots \\ e_1 \Phi'_0(\vartheta) + e_2 \Phi'_{j_2}(\vartheta) + \dots e_s \Phi'_{j_s}(\vartheta) \\ \vdots \\ \vdots \end{array} \quad (12)$$

setting  $\Delta_u(\zeta) = \sum_{j=0}^{\infty} \psi_j(\zeta) \psi_{j+u}(\zeta)$   $u = 0, \pm 1, \dots$ ,  $\Phi_0(\vartheta) = \beta(\phi) \frac{\partial \Delta_0(\zeta)}{\partial \zeta}$ ,  $\Phi_j(\vartheta) = \left( \beta(\phi) \frac{\partial \Delta_j(\zeta)}{\partial \zeta} + \gamma(\phi) \frac{\partial \psi_{j-1}(\zeta)}{\partial \zeta} \right)$   $j \geq 1$ , and  $e_j$  is the  $s \times 1$  vector with all zeros but 1 in the  $j$ th entry. Simple inspection of (12) shows that a necessary *order* condition for identification is  $q \leq 3$ . Moreover, when  $q = 3$  the identification condition must include  $\partial c_0(\zeta) / \partial \vartheta'$  for otherwise  $\phi$  is not identified.

(ix). Ass.  $F$  is a sufficient condition for strict positivity of the model spectral density at all frequencies. This is required for asymptotic normality of  $\hat{\vartheta}_T$  although not for consistency. When  $\{z_t, \epsilon_t\}$  have a parametric distribution depending on  $\phi_0$ , then we can assume that for any  $\vartheta \in \Theta$ , there exists a

collection of *i.i.d.* bivariate stochastic processes  $\{z_t(\phi), \epsilon_t(\phi)\}$ , indexed by  $\phi$ , such that

$$\alpha(\phi) = \text{var}(\log z_0^2(\phi)), \beta(\phi) = \text{var}(\epsilon_0(\phi)), \gamma(\phi) = \text{cov}(\log z_0(\phi), \epsilon_0(\phi)). \quad (13)$$

Ass.  $F$  is violated whenever the Schwarz inequality holds with the equality sign: either perfectly collinear  $\log z_t^2(\phi)$  and  $\epsilon_t(\phi)$  or, alternatively, degenerate  $\log z_t^2(\phi)$  or  $\epsilon_t(\phi)$ . For instance, perfect collinearity arises for a version of the EGARCH model where (4) is substituted by  $\epsilon_t = \epsilon(z_t) = \theta_0(\log z_t^2 - \mu_{\log z^2})$  where  $\delta_0 = 0$ .

### 3 Main Results

We present the asymptotic results for the Whittle estimator  $\hat{\vartheta}_T$ .

**Theorem 1** *Under Assumptions A(2), B, C(k<sub>2</sub>, 0), D(k<sub>2</sub>, 0), E, F, as  $T \rightarrow \infty$ ,*

$$\hat{\vartheta}_T \xrightarrow{a.s.} \vartheta_0.$$

**Theorem 2** *Under Assumptions A(4), B, C(k<sub>2</sub>, 2), D(k<sub>2</sub>, 2), E, F, as  $T \rightarrow \infty$ ,*

$$T^{\frac{1}{2}}(\hat{\vartheta}_T - \vartheta_0) \rightarrow_d \mathcal{N}_{p+q}(0, M^{-1}VM^{-1}),$$

where

$$M(\vartheta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} N(\lambda; \vartheta) N(\lambda; \vartheta)' d\lambda, \quad M = M(\vartheta_0), \quad N(\lambda; \vartheta) = \frac{\partial \ln f(\lambda; \vartheta)}{\partial \vartheta}, \quad N(\lambda) = N(\lambda; \vartheta_0),$$

and

$$V = \frac{1}{\pi} \int_{-\pi}^{\pi} N(\omega) N'(\omega) d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{N(\omega_1) N(\omega_2)'}{f(\omega_1) f(\omega_2)} Q(-\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2, \quad (14)$$

$Q(\omega_1, \omega_2, \omega_3)$  denoting the trispectrum of the  $\log x_t^2$  (the Fourier transform of the fourth-order cumulants of the  $\log x_t^2$ ).

Under Assumptions  $A(4)$ ,  $B$ ,  $C(k_2, 3)$ ,  $D(k_2, 3)$ ,  $F$  as  $T \rightarrow \infty$ ,  $M_T(\hat{\vartheta}_T)$ ,  $V_T(\hat{\vartheta}_T)$ , defined respectively in (31) and (32), are consistent estimates of  $M, V$ .

**Remarks.**

(i) Both the rate of convergence and asymptotic normality do not depend on whether  $d(\zeta)$  is zero or not. This result represents one of the finest features of the Whittle estimator, due to the automatic compensation, characterizing the Whittle function, for possible lack of square integrability of the model spectral density, occurring when  $1/4 \leq d(\zeta) < 1/2$ . On the other hand, obviously, the asymptotic covariance matrix depends on the assumed parametric choice made for  $\psi_j(\zeta)$ . But the Whittle estimator displays other advantages. For one, the discrete Whittle function does not require estimation of the mean which might otherwise affect the small sample performance of the Whittle estimator (see Diebold and Cheung (1994)). Second, the Whittle function does not require any truncation, such as substituting  $\log x_t^2$  by  $\log z_t^2 + \omega_0 + \sum_{k=0}^{t-2} \psi_{0k} \epsilon_{t-k-1}$ . This is because the spectral density (6) is a function of all the  $\psi_{0k}$ ,  $k = 0, 1, \dots$ , through the transfer function (15). This contrasts with the maximum likelihood estimator (henceforth MLE) where typically one needs to distinguish the observable pseudo log likelihood, function of a sample  $(x_1, x_2, \dots, x_T)$ , from the unobservable pseudo log likelihood, function of all  $\{x_s; s = \dots, -1, 0, 1, \dots, T\}$ . Such truncation is usually asymptotically negligible but might not be so for long memory parameterizations, and can induce an asymptotic bias, such as for the pseudo MLE (henceforth PMLE) of ARCH( $\infty$ ), as shown by Robinson and Zaffaroni (2006).

(ii) The *mean* parameter  $\omega_0$  is not identifiable by the autocorrelation function, and thus it cannot be estimated by the Whittle estimator. A simple estimate is based on the sample mean  $\widehat{\log x^2}_T = 1/T \sum_{t=1}^T \log x_t^2$ , which is a consistent estimate of  $E \log x_0^2 = \omega_0 + E \log z_0^2$  under Ass.  $A(2)$ . Thus, a  $\sqrt{T}$ -consistent estimate of  $\omega_0$  is obtained by subtracting the Whittle estimate

of  $E \log z_0^2$ , which depends on  $\hat{\phi}_T$ , from  $\widehat{\log x^2}_T$ .

(iii) For practical use of the asymptotic distribution result, a consistent estimate of asymptotic covariance matrix is required. Here we prove consistency of plug-in estimators of  $M$  and  $V$ , respectively in Lemma 8 and 9. Alternatively, for  $V$  one can use Taniguchi (1982), although it is unclear whether Taniguchi's result extends to case  $d(\zeta) \in (0, 1/2)$ .

## 4 Implications for EGARCH

We first check that the regularity conditions requested by Theorem 1 and 2 are satisfied by EGARCH. For the same model, we discuss how the properties of the Whittle estimator relates to the ones of the MLE.

Define the generating function

$$\psi(z; \zeta) = \sum_{j=0}^{\infty} \psi_j(\zeta) z^j, \quad |z| \leq 1, \quad (15)$$

and consider the class of functions

$$\psi(z; \zeta) = \frac{a(z; \zeta)}{b(z; \zeta)} (1 - z)^{-d(\zeta)} \quad (16)$$

where  $d(\zeta) < 1/2$  is a known function of  $\zeta$ , and  $a(z; \zeta)$  and  $b(z; \zeta)$  are polynomials in  $z$  of known degrees  $m$  and  $n$  respectively, whose coefficients are known functions of  $\zeta$ , which have no zeros in common:

$$a(z; \zeta) = 1 + \sum_{j=1}^m \zeta_j z^j \quad \text{for } m \geq 0, \quad a(z; \zeta) \neq 0, |z| \leq 1, \quad (17)$$

$$b(z; \zeta) = 1 - \sum_{j=1}^n \zeta_{j+m} z^j \quad \text{for } n \geq 0, \quad b(z; \zeta) \neq 0, |z| \leq 1, \quad (18)$$

setting  $\sum_{j=1}^n c_j = 0$  when  $n < 1$  and with  $\zeta_i$  denoting the  $i$ th element of  $\zeta$ . Hereafter assume that  $E(z_0) = 0$ ,  $\text{var}(z_0) = 1$ . Then, the EGARCH( $m, n$ ) model is defined by (4) and (16)-(17)-(18) with  $d(\zeta) \equiv 0$ . The differentiability

and the rate of convergence of the  $\psi_j(\zeta)$  follow since the former are well-known analytic functions of the roots of the polynomials  $a(z; \zeta)$ ,  $b(z; \zeta)$  (see Brockwell and Davis (1987, section 3.6)). Ass.  $C(1, l)$  and  $D(1, l)$  are then satisfied with  $e(\zeta)$  equal to the maximum of the inverse modulus of the roots of  $b(z; \zeta)$ . The function  $E_r(j; \zeta)$  could either be monotonically non decreasing as well as a trigonometric function, the latter case arising in case of dominant complex conjugates roots. Concerning the parameter  $\phi$ , (4) implies that the distribution of the  $z_t$  can at most depend on one parameter  $\phi_{03}$  due to Ass.  $E$ , since  $q \leq 3$ , and imposing  $Ez_0 = 0$  and  $\text{var}(z_0) = 1$  one gets

$$\alpha(\phi_0) = \text{var}(\log z_0^2), \quad \beta(\phi_0) = \phi_{01} + \phi_{02}^2(1 - \mu_{|z|}^2), \quad \gamma(\phi_0) = \phi_{02} \text{cov}(\log z_0^2, |z_0|),$$

where  $\phi_{01} = \theta_0^2$ ,  $\phi_{02} = \delta_0$ . The sign of  $\theta_0$  is not identifiable through the (univariate) model spectral density although it will be using a bivariate extension of the Whittle estimator, as discussed below. Identification requires that  $\phi_{03} = \text{var}(\log z_0^2)$  only can be left as a free parameter, and one must then set ‘a priori’ both  $\mu_{|z|}$  and  $\text{cov}(\log z_0^2, |z_0|)$  in such a way that Ass.  $F$  is satisfied. This represents the most general (semi-parametric) specification of the EGARCH model, in terms of the distribution of the  $z_t$ . Considering parametric specifications, meaning that the distribution of the  $z_t$  depends on an unknown parameter  $\phi_{03}$ , then  $\text{var}(\log z_0^2)$ ,  $\mu_{|z|}$  and  $\text{cov}(\log z_0^2, |z_0|)$  are jointly determined. Ass.  $F$  will be automatically satisfied, whenever the  $z_t$  have a non-degenerate (parametric) distribution. For instance, when  $z_t$  are *i.i.d* GED with tail thickness parameter  $0 < \phi_{03} \leq \infty$

$$\alpha_0 = \left(\frac{2}{\phi_{03}}\right)^2 \Psi\left(\frac{1}{\phi_{03}}\right), \quad \beta_0 = \phi_{01} + \phi_{02}^2(1 - \mu_{|z|}^2), \quad \gamma_0 = \frac{2\phi_{02}}{\phi_{03}} \mu_{|z|} \left( \psi\left(\frac{2}{\phi_{03}}\right) - \psi\left(\frac{1}{\phi_{03}}\right) \right), \quad (19)$$

with  $\mu_{|z|} = \Gamma(2/\phi_{03}) / (\sqrt{\Gamma(3/\phi_{03})\Gamma(1/\phi_{03})})$  where  $\psi(z)$  is the digamma function (the derivative of  $\log\Gamma(z)$  with  $\Gamma(z)$  being the Gamma function) and  $\Psi(z)$  is the trigamma function (the derivative of  $\psi(z)$ ); see Gradshteyn and Ryzhik (1994, sections 6.3 and 6.4). A distribution with fatter tails than the normal is obtained whenever  $\phi_{03} < 2$  and with thinner tails whenever

$\phi_{03} > 2$  (see Nelson (1991, p.353)). The GED nests the normal distribution, for  $\phi_{03} = 2$ , and the uniform distribution, for  $\phi_{03} \rightarrow \infty$ . Let us discuss now Ass.  $E$  more in depth. Consider first the left hand side sub matrix of (12), made up of the first  $q$  columns. It is evident that linear independence of the columns of this sub matrix requires  $\partial\alpha(\phi)/\partial\phi$ ,  $\partial\beta(\phi)/\partial\phi$ ,  $\partial\gamma(\phi)/\partial\phi$  to be linearly independent. Since for EGARCH  $\partial\alpha(\phi)/\partial\phi_1 = \partial\alpha(\phi)/\partial\phi_2 = \partial\gamma(\phi)/\partial\phi_1 = 0$  it is easy to see that this is achieved when

$$\frac{\partial\alpha(\phi)}{\partial\phi_3} \frac{\partial\beta(\phi)}{\partial\phi_1} \frac{\partial\gamma(\phi)}{\partial\phi_2} \neq 0. \quad (20)$$

For the normal case (20) is not satisfied since  $\alpha(\phi)$  is independent from  $\phi$  and, indeed, we can identify only  $\phi'_{01} = \phi_{01}\phi_{03}$  and  $\phi'_{02} = \phi_{02}\sqrt{\phi_{03}}$ , implying  $q = 2$ . However, both for the GED and Students' $t$  case (20) is satisfied. Under this circumstance, full rank of the left hand side sub matrix of (12) is then guaranteed if also  $\Delta_0(\zeta) \Delta_{j_a}(\zeta) \neq 0$  for at least one  $j_a$ ,  $a \in \{2, \dots, s\}$ . Let us now consider the right hand sub matrix made up by the last  $p$  columns of (12), and consider for sake of simplicity the EGARCH(1, 1) case, yielding  $p = 2$  and  $\zeta = (\zeta_1, \zeta_2)'$ . It is well known that for this case  $\psi_0 = 1$ ,  $\Delta_0(\zeta) = (1 + \zeta_1^2 + 2\zeta_1\zeta_2)/(1 - \zeta_2^2)$  and  $\psi_u = \zeta_2^{u-1}(\zeta_1 + \zeta_2)$ ,  $\Delta_u(\zeta) = \zeta_2^{u-1}(\zeta_1 + \zeta_2)(1 + \zeta_1\zeta_2)/(1 - \zeta_2^2)$ ,  $u \geq 1$ . Simple yet tedious calculations yield

$$\begin{aligned} \Phi_0(\vartheta) &= 2\beta(\phi) \frac{\zeta_1 + \zeta_2}{1 - \zeta_2^2} \begin{pmatrix} 1 \\ \frac{1 + \zeta_1\zeta_2}{1 - \zeta_2^2} \end{pmatrix}, \quad \Phi_1(\vartheta) = \frac{\beta(\phi)}{1 - \zeta_2^2} \begin{pmatrix} 1 + 2\zeta_1\zeta_2 + \zeta_2^2 \\ \frac{(1 + \zeta_1\zeta_2)^2 + (\zeta_1 + \zeta_2)^2}{1 - \zeta_2^2} \end{pmatrix} \\ \Phi_u(\vartheta) &= \zeta_2^{u-2} \left( \beta(\phi) \zeta_2 \frac{\partial\Delta_1(\zeta)}{\partial\zeta} + \gamma(\phi) \frac{\partial\psi_1(\zeta)}{\partial\zeta} \right) \\ &+ \zeta_2^{u-3} \begin{pmatrix} 0 \\ (u-1)\zeta_2\beta(\phi)\Delta_1(\zeta) + (u-2)\gamma(\phi)\psi_1(\zeta) \end{pmatrix}, \quad u \geq 2, \end{aligned}$$

which form a nonsingular basis in  $R^2$  whenever  $|\zeta_1| < 1$ ,  $|\zeta_2| < 1$ ,  $\zeta_1 + \zeta_2 \neq 0$ . Under the same conditions, full rank of matrix (12) follows since each of the first  $q$  column vector is linearly independent from the last  $p$  column vectors.

The asymptotic covariance matrix of  $\hat{\vartheta}_T$  involve the trispectrum  $Q(\omega_1, \omega_2, \omega_3)$ , defined in (33), which in turn depends on various mixed fourth-order cu-

mulants for the  $\log z_t^2, \epsilon_t$ . Since  $\text{cum}(a, b, c, d) = E(abcd) - E(ab)E(cd) - E(ac)E(bd) - E(ad)E(bc)$ , assuming symmetric  $z_t$ , for EGARCH one gets:

$$\begin{aligned}\kappa_{\log z^2 \log z^2 \log z^2 \epsilon} &= \delta_0 \left( E(\left( (\log z_0^2)^3 \mid z_0 \right) - 3\mu_{(\log z^2)^2} E \log z_0^2 \mid z_0) \right), \\ \kappa_{\log z^2 \log z^2 \epsilon \epsilon} &= (\theta_0^2 + \delta_0^2) E(\log z_0^2)^2 z_0^2 - (\theta_0^2 + \delta_0^2) \mu_{(\log z^2)^2} - 2\delta_0^2 (E \log z_0^2 \mid z_0)^2, \\ \kappa_{\log z^2 \epsilon \epsilon \epsilon} &= (\delta_0^3 + 3\theta_0^2 \delta_0) E \log z_0^2 \mid z_0|^3 - 3\delta_0 (\theta_0^2 + \delta_0^2) E \log z_0^2 \mid z_0, \\ \kappa_{\epsilon \epsilon \epsilon \epsilon} &= (\theta_0^4 + \delta_0^4 + 6\theta_0^2 \delta_0^2) \mu_{z^4} - 3(\theta_0^2 + \delta_0^2)^2.\end{aligned}$$

We now discuss the analogies of our results with respect to the MLE of EGARCH. Nelson (1991, p.93) proposed estimation of the EGARCH by ML based on the GED with tail thickness parameter  $\nu$ , but its asymptotic properties remain unknown even when  $\nu$  is assumed known ( $\nu = 2$  yields the Gaussian likelihood). ML estimation requires invertibility of the model, that is the possibility to express  $z_t$  as a convergent (in some norm) function of all the  $x_s$  ( $s \leq t$ ). Invertibility is necessary for the ‘observed’ likelihood, function of a sample  $(x_1, \dots, x_T)$  of size  $T$ , to be well-behaved asymptotically without exploding nor converging toward zero for any admissible parameter value. Establishing invertibility is typically a formidable task to be achieved for nonlinear moving average models (see Granger and Andersen (1978)). Recently Straumann and Mikosch (2006, eq.(3.18) and Remark 3.13) provide a sufficient condition for invertibility of EGARCH(0, 1) but also suggest that such condition is practically infeasible, except when  $\zeta_{01} = 0$ , that is for the EGARCH(0, 0) which, in turn, implies  $\psi_{00} = 1$ ,  $\psi_{0j} = 0$ ,  $j \geq 1$ . This means, for instance, that  $h_t$  is *i.i.d.* and  $x_t$  is independent of  $x_{t-j}$  for all  $j > 1$ . Based on this invertibility condition, the asymptotics properties of the MLE of the EGARCH(0, 0) are then established (see Straumann (2005, Theorem 5.7.9).

However, it is easy to establish the central limit theorem (CLT) for the first derivative of the log likelihood function evaluated at the true parameter value for a general EGARCH( $m, n$ ). This represents a necessary step to establish the asymptotic distribution of the PMLE which, however, does not require invertibility of the model. It turns out that  $E z_0^2 + E \mid z_0 \mid^{2\nu} < \infty$  is required, when adopting a GED log likelihood for  $z_t$ , with tail thickness

parameter  $0 < \nu < \infty$ . (Obviously, when  $z_t$  is truly GED distributed then all its moments exist.) Since  $\nu > 1$  is likely to be required (see Nelson (1991, last paragraph of p.106)), this simplifies to  $E |z_0|^{2\nu} < \infty$ , which equals  $Ez_0^4 < \infty$  when adopting a Gaussian likelihood. Concerning the Whittle estimator, Theorem 2 simply requires  $Ez_0^4 < \infty$  and  $E(\log z_0^2)^2 < \infty$ . (The latter is implied by the former for most distributions of interest.) Therefore, our moment condition equals the one required by Gaussian PMLE and will only be slightly stronger for GED PMLE with  $\nu < 2$ . This contrasts with estimation of ARCH( $\infty$ ) model whereas Whittle estimation requires  $Ex_0^8 < \infty$  (see Giraitis and Robinson (2001)). Conditions required for asymptotic normality of Gaussian PMLE of ARCH( $\infty$ ) are instead much weaker, since finite variance of the  $x_t$  is not even required (see Robinson and Zaffaroni (2006)). Linton and Mammen (2005) consider semiparametric estimation of a class of asymmetric ARCH( $\infty$ ), with a nonparametric specification of the effect of past squares on the conditional variance. Their asymptotic results require  $E|x_0|^\delta < \infty$ ,  $\delta > 4$ . Zaffaroni (2003) considers Whittle estimation of a class of nonlinear moving average processes, which however does not admit a signal-plus-noise representation, treating  $x_t^2$  as the observable. A bounded fourth moment condition is postulated.

## 5 Extensions

We now present a number of relevant extensions, the additional proofs of which follow by suitably adapting the corresponding proofs of Theorem 1 and 2, and are not reported for sake of simplicity. Further extensions that lead to efficiency improvements and applicability to high-frequency return and duration data are discussed in Zaffaroni (2008).

### 5.1 Non-stationarity

The very first empirical application of EGARCH, based on the value-weighted market index from the CRSP tapes, indicates a great deal of persistence, with

the largest estimated autoregressive roots approximately equal to 0.99962, yielding a  $t$  statistic for testing the unit root hypothesis of  $-0.448$  Nelson (1991, Table 5.2 and p.99). As another example, Bollerslev and Mikkelsen (1996, Table 5) report the results of an empirical application of FIEGARCH, based on the Standard & Poor's 500 composite stock index, yielding a point estimate of the long memory parameter  $d_0 = d(\zeta_0)$  equal approximately to 0.633. The  $t$  statistic for testing the hypothesis  $d = 0.5$  is 2.111. Both examples suggest that condition  $\sum_{k=0}^{\infty} \psi_{0k}^2 < \infty$  might be too strong in certain circumstances. Let us consider for instance the case where  $\psi_{0j} \sim c j^{d_0-1}$  as  $j \rightarrow \infty$ , for  $1/2 \leq d < 3/2$ . Then square summability of the  $\psi_{0j}$  clearly fails but  $\sum_{k=1}^{\infty} (\psi_{0k} - \psi_{0k-1})^2 < \infty$ . Several approaches can be considered in such circumstance. First, one can adapt the idea of using data tapering used by Velasco and Robinson (2000) for parametric Whittle estimation of nonstationary linear time models. Second, one can simply differentiate  $\log x_t^2$  in order to achieve stationarity. For the case just described, this implies considering as observable the series

$$\tilde{x}_t = \frac{x_t}{x_{t-1}} = \frac{z_t}{z_{t-1}} e^{0.5(h_t - h_{t-1})} = \tilde{z}_t e^{0.5(h_t - h_{t-1})},$$

setting  $\tilde{z}_t = z_t/z_{t-1}$ , and then consider Whittle estimation based on the  $\log \tilde{x}_t^2$ . The differenced process will satisfy the conditions listed in Section 2, and the corresponding spectral density will be  $|1 - e^{i\lambda}|^2 f(\lambda)$  where  $f(\lambda)$  is the (pseudo) spectral density of the  $\log x_t^2$ . The Whittle estimator of the model parameters will then be strongly consistent and asymptotically normal but the limit covariance matrix will have a different form from the un-differenced case reported in Theorem 2, since a different model spectral density is considered. Moreover, an efficiency loss is likely to occur. In fact,  $\log \tilde{x}_t^2 = \log \tilde{z}_t^2 + (h_t - h_{t-1})$ , and the noise term  $\log \tilde{z}_t^2$  has a larger variance, in fact double, than  $\log z_t^2$ . Note that  $\log \tilde{z}_t^2$  is autocorrelated and moreover correlated with  $h_t - h_{t-1}$ , except for the simple case of mutually uncorrelated  $\log z_t^2, \epsilon_t$ .

For Gaussian  $z_t$ , the  $\tilde{z}_t$  will be Cauchy distributed so the first moment

does not exist and the first absolute moment is unbounded. This suggests that applying the same idea to PMLE could be problematic, stressing instead the great flexibility enhanced by the Whittle estimator. The log-difference transformation mainly affects the proof of the CLT (cf. Lemma 7 in the appendix), which nevertheless could be generalized to the case where the observable is the log-square of  $\left(\prod_{j=0}^{m-1} z_{t-j}^{c_j}\right) e^{0.5(\omega_0 + \sum_{k=0}^{\infty} \psi_{0k} \epsilon_{t-k-1})}$ , with known  $c_j = 0, \pm 1, \dots$  and square summable  $\psi_{0j}$ . This implies that any finite number of log-square differentiations is permitted in order to achieve stationarity. By differencing one might obtain that the spectral density of the differenced series is  $O(|\lambda|^\delta), \lambda \rightarrow 0$  for some  $\delta > 0$ . This affects the proof of asymptotic normality, unless  $\delta < 1$ , but not of consistency. Therefore, in order to construct confidence intervals, two estimations are required with the first one aiming at finding the suitable degree of differentiation to possibly ensure  $\delta < 1$ .

## 5.2 Filtering and forecasting

Consider one-shock models such as the EGARCH and the GJR models. If the (nonlinear) autoregressive representation of  $x_t$  was available, it could be used for both forecasting  $h_{T+s}, s \geq 0$ , based on a sample of  $x_1, \dots, x_{T-1}$  or, alternatively, to back out the within-sample (yet unobserved) volatility  $h_s, 1 \leq s \leq T-1$ . Establishing the existence of such representation is analog to establish invertibility of the  $x_t$ , which appears a formidable task to be achieved, as discussed before. However, the frequency domain set up allows us to make use of the Wiener-Kolmogorov (WK) theory of forecasting and signal extraction for linear models (see Whittle (1983)). Harvey (1998) proposed to use the WK filter for long memory SV models when  $\gamma_0 = 0$ . It is relatively less known, though, that the WK theory naturally allows for correlated signal and noise and it can then be applied to one-shock models, once a consistent parameters estimate is available. Therefore, in what follows, we simply describe how to implement the WK theory within our framework, when correlation between signal and noise is allowed for.

Assume for sake of simplicity that  $E\log x_0^2 = E\log z_0^2 + \omega_0 = 0$ . In the more realistic case of non-zero mean, a consistent estimate of the mean (see Remark (ii) of Section 3) will be subtracted from the data prior to filtering and forecasting yielding asymptotically equivalent results.

For forecasting purposes, the minimum mean-square linear forecast of  $\log x_{T+h}^2, h = 0, 1, \dots$  based on the infinite past of the series  $\{\log x_{T-j}^2, j = 1, \dots\}$  is

$$\widehat{\log x_{T+h}^2} = - \sum_{j=1}^h a_j \widehat{\log x_{T+h-j}^2} - \sum_{j=1}^{\infty} a_{j+h} \log x_{T-j}^2, \quad h = 0, 1, \dots \quad (21)$$

Forecast of  $h_{T+h-1}$  easily follow by (3). The coefficients  $a_j$  are obtained by means of the canonical factorization of  $f(\lambda)$  (see Whittle (1983, p.26)) which exists by Lemma 4, part (ii). A consistent estimate of the  $a_j$  is  $\hat{a}_u = (2N)^{-1} \sum_{j=-N+1}^N \hat{A}_j e^{iu\mu_j}$ ,  $u = 0, 1, \dots, N-1$ , where  $\hat{A}_u = \exp \left\{ - \sum_{j=1}^{N-1} \hat{c}_j e^{-ij\mu_u} \right\} = \bar{\hat{A}}_{-u}$ ,  $\hat{c}_u = \frac{1}{N} \sum_{k=1}^{N-1} \log(f(\mu_k; \hat{\nu}_T)) \cos(u\mu_k)$  for  $\mu_n = \pi n/N$ ,  $n = 0, \pm 1, \dots, \pm N$ ,  $N = [T^{\frac{1}{2}}/4]$ ,  $[\cdot]$  is the integer part of its argument and  $\bar{a}$  is the conjugate of the complex number  $a$ . Computation of the  $\hat{a}_u$  implies that (21) is replaced by the finite sum  $\widehat{\log x_{T+h}^2} = - \sum_{j=1}^h \hat{a}_j \widehat{\log x_{T+h-j}^2} - \sum_{j=1}^{N-h-1} \hat{a}_{j+h} \log x_{T-j}^{2*}$ ,  $h = 0, 1, \dots$ . Hereafter  $\{x_{t-j}^*, j = 1, 2, \dots, N-1\}$  represents a sample of the  $x_t$  not used in the estimation of  $\vartheta_0$ .

Signal extraction of  $h_s$ ,  $1 \leq s \leq T-1$  based on the infinite sequence  $\log x_t^2$ ,  $t = 0, \pm 1, \dots$  can be carried out evaluating the best linear predictor

$$\hat{h}_t = \sum_{j=-\infty}^{\infty} g_j \log x_{t-j}^2 \quad (22)$$

where the  $g_j = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-ij\lambda} f^{-1}(\lambda) (f_{hh}(\lambda) + f_{h\log z^2}(\lambda)) d\lambda$  and  $f_{hh}(\lambda) = \beta_0 / (2\pi) \sum_{u=-\infty}^{\infty} \Delta_{|u|} e^{iu\lambda}$ ,  $f_{h\log z^2}(\lambda) = \gamma_0 / (2\pi) \sum_{u=1}^{\infty} \psi_{0u-1} e^{iu\lambda}$ ,  $-\pi \leq \lambda < \pi$ , setting  $\Delta_u = \Delta_u(\zeta_0) = \sum_{j=0}^{\infty} \psi_{0j} \psi_{0j+u}$ . Similarly to before, a consistent estimate of  $g_j$  is  $\hat{g}_j = \frac{1}{2N} \sum_{p=-N+1}^{N-1} f^{-1}(\mu_p; \hat{\nu}_T) \left( f_{hh}(\mu_p; \hat{\nu}_T) + f_{h\log z^2}(\mu_p; \hat{\nu}_T) \right) e^{-ij\mu_p}$ ,  $j = 0, \pm 1, \dots$ . Truncating suitably (22) and plugging in the  $\hat{g}_j$  yields  $\hat{h}_t^* = \sum_{j=t-T}^{t-1} \hat{g}_j \log x_{t-j}^{2*}$ ,  $t = 1, \dots, T$ . Theoretical justification for  $\widehat{\log x_{T+h}^2}$  and  $\hat{h}_t^*$ , as  $T \rightarrow \infty$ , can be obtained adapting Hidalgo and Yajima (2002) and Bhansali (1974).

## 6 Final remarks

We have established the asymptotic distribution theory of the Whittle estimate of a class of exponential volatility models (1)-(2) the most notable element of which is the EGARCH model of Nelson (1991) and the FIEGARCH of Harvey (1998) and Breidt, Crato, and deLima (1998). Perez and Zaffaroni (2008) present a finite-sample comparison of the Whittle estimator with MLE for EGARCH and FIEGARCH models. We have discussed generalizations, in particular when dealing with nonstationarity as well as with filtering and forecasting. Other generalizations are of interest. The frequency domain set-up easily allows to estimate seasonality effects with various degrees dependence. Second, we focused on estimation of the conditional variance parameters by assuming martingale difference observations, but one can consider simultaneous estimation of both conditional mean and conditional variance parameters. Third, multivariate extensions of the exponential model, along the lines of Harvey, Ruiz, and Shephard (1994) but also considering one-shock models, can be developed and estimated based on a multivariate version of the Whittle estimator. Fourth, the exponential model can be generalized by leaving the news impact curve  $\epsilon(z_t)$  unspecified, similarly to the semiparametric ARCH( $\infty$ ) model of Linton and Mammen (2005). Similarly to them, the estimation procedure of this semiparametric exponential model will then combine aspects of parametric and nonparametric estimation.

## Appendix: mathematical proofs

$K$  defines a non zero constant, not always the same, and  $K_0$  a constant that could be eventually zero.  $\rightarrow_{a.s.}$ ,  $\rightarrow_d$  define *a.s.* convergence and convergence in distribution respectively. For any  $i_h = 1, \dots, p$ ,  $h = 0, \dots, l$ , let  $U_{i_1, i_2, \dots, i_l}(\lambda; \zeta) = \frac{\partial^l}{\partial \zeta_{i_1} \dots \partial \zeta_{i_l}} |\psi(e^{i\lambda}; \zeta)|^2$ ,  $V_{i_1, i_2, \dots, i_l}(\lambda; \zeta) = \frac{\partial^l}{\partial \zeta_{i_1} \dots \partial \zeta_{i_l}} (e^{i\lambda} \psi(e^{i\lambda}; \zeta) + e^{-i\lambda} \psi(e^{-i\lambda}; \zeta))$  and for any  $j_h = 1, \dots, s$ ,  $h = 0, \dots, l$ , let  $f_{j_1, j_2, \dots, j_l}(\lambda; \vartheta) = \frac{\partial^l}{\partial \vartheta_{j_1} \dots \partial \vartheta_{j_l}} f(\lambda; \vartheta)$ , any  $-\pi \leq \lambda < \pi$  and  $\vartheta \in \Theta$ . When  $l = 0$  we mean that no differentiation occurs.

**Lemma 1** Under Assumptions  $C(0, l)$  and  $D(0, l)$ :

(i) As  $\lambda \rightarrow 0^+$

$$V_{i_1, i_2, \dots, i_r}(\lambda; \zeta) \sim \begin{cases} K D_r(\lambda^{-1}; \zeta) \lambda^{-d(\zeta)}, & d(\zeta) \in (0, 1/2), \\ K D_r'(\lambda^{-1}; \zeta), & d(\zeta) = 0 \text{ and non summable } |D_r(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) = 0 \text{ and summable } |D_r(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) < 0, \end{cases}$$

$$U_{i_1, i_2, \dots, i_r}(\lambda; \zeta) \sim \begin{cases} K D_r''(\lambda^{-1}; \zeta) \lambda^{-2d(\zeta)}, & d(\zeta) \in (0, 1/2), \\ K D_r'''(\lambda^{-1}; \zeta), & d(\zeta) = 0 \text{ and non summable } |D_r(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) = 0 \text{ and summable } |D_r(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) < 0, \end{cases}$$

for all  $i_h = 1, \dots, s$ ,  $h = 1, \dots, r$ ,  $r \leq l$  where  $D_r'(x; \zeta)$ ,  $D_r''(x; \zeta)$ ,  $D_r'''(x; \zeta)$  are slowly varying functions.

(ii) For all  $\vartheta$  and  $\lambda \neq 0$   $U_{i_1, i_2, \dots, i_r}(\lambda; \zeta)$  and  $V_{i_1, i_2, \dots, i_r}(\lambda; \zeta)$  satisfy an approximate Lipschitz continuous conditions  $\text{Lip}(\delta)$  with  $\delta \geq \min[1, 1 - 2d(\zeta)]$  (see Zygmund (1977, Section 3)).

**Proof.** (i) Consider case  $l = 0$  and  $U(\lambda; \zeta)$ . Case  $l > 0$  easily follows. When  $d(\zeta) < 0$  the  $\psi_j(\zeta)$  are summable implying that the  $c_j(\vartheta)$  are summable. Hence  $f(\lambda; \vartheta)$  is continuous for all  $\lambda \in [-\pi, \pi]$  including  $\lambda = 0$ . The same applies when  $d(\zeta) = 0$  with summable  $|D_0(j; \zeta)| j^{-1}$ .

We discuss case  $d(\zeta) \in [(0, 1/2)$  with non summable  $|D_0(j; \zeta)| j^{-1}$ . It is easy to see that our assumptions imply

$$|\Delta_u(\zeta) - \Delta_{u+1}(\zeta)| \leq K \frac{|\Delta_u(\zeta)|}{u} \quad \text{as } u \rightarrow \infty. \quad (23)$$

This implies that the  $c_u(\vartheta)$  are quasi-monotonically convergent to zero and also satisfy the pure-bounded variation condition:  $\sum_{k=u}^{\infty} |\Delta_k(\zeta) - \Delta_{k+1}(\zeta)| = O(|\Delta_u(\zeta)|)$  as  $u \rightarrow \infty$ . See, for example, Robinson (1994, Lemma 11) when  $d \in (0, 1/2)$ . Moreover  $\Delta_u(\zeta) = \sum_{j=0}^u \psi_j(\zeta)\psi_{j+u}(\zeta) + \sum_{j=u+1}^{\infty} \psi_j(\zeta)\psi_{j+u}(\zeta) \sim \psi_u \sum_{j=0}^u \psi_j + \sum_{j=u+1}^{\infty} \psi_j^2(\zeta) \sim D_0''(j; \zeta) u^{2d(\zeta)-1}$ , as  $u \rightarrow \infty$ , setting  $D_0''(j; \zeta) = D_0^2(j; \zeta)$ , where the last  $\sim$  relationship follows by Yong (1974, Lemma I-11, (1-32') and (1-32'') and Lemma I-16). It easily follows that same properties apply to the  $c_u(\vartheta)$ . When  $d(\zeta) = 0$ ,  $\Delta_u(\zeta) = \sum_{j=0}^u \psi_j(\zeta)\psi_{j+u}(\zeta) +$

$\sum_{j=u+1}^{\infty} \psi_j(\zeta) \psi_{j+u}(\zeta) \sim \psi_u \sum_{j=0}^u \psi_j + \sum_{j=u+1}^{\infty} \psi_j^2(\zeta) \sim D_0(u; \zeta) u^{-1} \int_1^u D_0(t; \zeta) t^{-1} dt + D_0^2(u; \zeta) u^{-1} \sim D_0'''(u; \zeta) u^{-1}$  as  $u \rightarrow \infty$ , by Yong (1974, Lemma III-7, (i), (ii) and (iii)) where  $D_0'''(\cdot; \zeta) = D_0(u; \zeta) \int_1^u D_0(t; \zeta) t^{-1} dt$  is slowly varying. Note that the latter always dominates  $D_0^2(\cdot; \zeta)$  in all cases. By the same arguments (23) follows. Therefore, the proof is completed using Yong (1974, Lemma III-12) when  $d(\zeta) \in (0, 1/2)$  and Yong (1974, Lemma III-22, (i)) when  $d(\zeta) = 0$  with non summable  $|D_0(j; \zeta)| j^{-1}$ . The proof for  $V(\lambda; \zeta)$  follows directly.

(ii) The result follows by Robinson (1994, Lemma 8). ■

**Lemma 2** Under Assumptions  $C(0, a)$ ,  $D(0, a)$  for any  $i_m = 1, \dots, q$  and  $j_n = q + 1, \dots, q + p$  with  $0 \leq m \leq b \leq a$ ,  $0 \leq n \leq c \leq a$ ,

$$f_{i_1, \dots, i_b, j_1, \dots, j_c}(\lambda; \vartheta) = \frac{\alpha_{i_1, \dots, i_b}(\phi)}{2\pi} + \frac{\beta_{i_1, \dots, i_a}(\phi)}{2\pi} U_{j_1, \dots, j_c}(\lambda; \zeta) + \frac{\gamma_{i_1, \dots, i_b}(\phi)}{2\pi} V_{j_1, \dots, j_c}(\lambda; \zeta).$$

**Proof.** Straightforward. ■

**Lemma 3** Under Assumptions  $C(k_2, a)$ ,  $D(k_2, a)$ , for any  $i_m = 1, \dots, q$ ,  $0 \leq m \leq b \leq a$  and  $j_n = q + 1, \dots, q + p$ ,  $0 \leq n \leq c \leq a$ ,  $f_{i_1, \dots, i_b, j_1, \dots, j_c}(\lambda; \vartheta)$  is continuous for all  $\vartheta \in \Theta$  and  $\lambda \neq 0 \pmod{2\pi}$ .

**Proof.** When the  $c_u(\vartheta)$  are absolutely summable then continuity holds for any  $\vartheta \in \Theta$  and any  $\lambda \in (-\pi, \pi]$ . Summability holds for  $k_2 = 1$ ,  $k_2 = 0$ ,  $d(\zeta) < 0$  and  $k_2 = 0$ ,  $d(\zeta) = 0$ ,  $\sum_{j=1}^{\infty} j^{-1} |D_a(j; \zeta)| < \infty$ . When  $k_2 = 0$ ,  $d(\zeta) \in (0, 1/2)$  and  $k_2 = 0$ ,  $d(\zeta) = 0$ , non summable  $j^{-1} |D_a(j; \zeta)|$  then Robinson (1994, Lemma 8) applies, in the second case also using Lemma 1 together with Yong (1974, Lemma III-7, (i), (ii) and (iii)) and Yong (1974, Lemma III-22, (i)). ■

**Lemma 4**

(i) Under Assumption E,  $f(\lambda; \vartheta) \neq f(\lambda; \vartheta_0)$  for all  $\vartheta \neq \vartheta_0$ ,  $\vartheta \in \Theta$ .

(ii) Under Assumption F,  $f(\lambda; \vartheta) \geq K > 0$  for all  $\vartheta \in \Theta$  and  $-\pi \leq \lambda < \pi$ .

**Proof.** (i) We proceed by contradiction. Let us assume that there exist  $\vartheta \in \Theta$  such that  $\vartheta \neq \vartheta_0$  and  $f(\lambda; \vartheta) = f(\lambda; \vartheta_0) = f(\lambda)$  for  $-\pi \leq \lambda < \pi$ . However, since  $f(\lambda; \vartheta)$  is uniquely identified by its Fourier transform, it also follows that  $c_u(\vartheta) = c_u(\vartheta_0)$ ,  $u = 0, \pm 1, \dots$ . By Assumption  $E$  and the mean value theorem

$$\begin{pmatrix} c_0(\vartheta) \\ c_{j_2}(\vartheta) \\ \vdots \\ c_{j_s}(\vartheta) \end{pmatrix} = \begin{pmatrix} c_0(\vartheta_0) \\ c_{j_2}(\vartheta_0) \\ \vdots \\ c_{j_s}(\vartheta_0) \end{pmatrix} + \begin{pmatrix} \frac{\partial}{\partial \vartheta'} c_0(\tilde{\vartheta}) \\ \frac{\partial}{\partial \vartheta'} c_{j_2}(\tilde{\vartheta}) \\ \vdots \\ \frac{\partial}{\partial \vartheta'} c_{j_s}(\tilde{\vartheta}) \end{pmatrix} (\vartheta - \vartheta_0),$$

where  $\|\tilde{\vartheta} - \vartheta\| \leq \|\vartheta_0 - \vartheta\|$  and  $\|\cdot\|$  is the Euclidian norm. But this implies  $\vartheta = \vartheta_0$  since the  $s \times s$  matrix of derivatives is full rank.

(ii) By Assumption  $F$  we can always find a sufficiently small  $\delta(\phi) > 0$  such that  $|\gamma(\phi)| < (\alpha'(\phi))^{\frac{1}{2}}(\beta(\phi))^{\frac{1}{2}}$  with  $\alpha'(\phi) = \alpha(\phi) - \delta(\phi)$ . Then

$$f(\lambda; \vartheta) = \frac{\delta(\phi)}{2\pi} + \left( \frac{\alpha'(\phi)}{2\pi} + \frac{\beta(\phi)}{2\pi} |\psi(e^{i\lambda}; \zeta)|^2 + \frac{\gamma(\phi)}{2\pi} (e^{i\lambda}\psi(e^{i\lambda}; \zeta) + e^{-i\lambda}\psi(e^{-i\lambda}; \zeta)) \right),$$

and the term in brackets on the right hand side is nonnegative for any  $-\pi \leq \lambda < \pi$ , since it represents a well-defined model spectral density. When (13) holds, one can take  $\delta(\phi) = \text{var}(\log z_0^2(\phi) - E(\log z_0^2(\phi) | \epsilon_0(\phi)))$ . Finally take  $K = \inf_{\vartheta \in \Theta} \delta(\phi)/(2\pi)$ . ■

**Lemma 5** *Under Assumption A(2) and (2) the  $x_t$  satisfy  $|x_t| < \infty$  a.s. and are ergodic and strictly stationary.*

**Proof.** Almost sure boundedness, ergodicity and strict stationarity follow easily adapting Nelson (1991, Theorem 2.1). ■

**Lemma 6**

(i) *Under Assumptions A(2),  $C(k_2, 0)$ ,  $D(k_2, 0)$ ,  $F$ ,*

$$Q_T(\vartheta) \rightarrow_{a.s.} Q(\vartheta) \quad \text{as } T \rightarrow \infty,$$

*uniformly in  $\vartheta \in \Theta$ , where  $Q(\vartheta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f(\lambda; \vartheta)) d\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) f^{-1}(\lambda; \vartheta) d\lambda$ .*

(ii) *Under Assumption E, F,  $Q(\vartheta) \geq Q(\vartheta_0)$  for any  $\vartheta \in \Theta$ .*

**Proof.** (i) All the convergences below hold as  $T \rightarrow \infty$ . Uniform convergence of  $1/T \sum_{t=1}^{T-1} I_T(\lambda_t)/f(\lambda_t; \vartheta)$  is obtained by Hannan (1973, Lemma 1). Let us consider the non random term  $1/T \sum_{t=1}^{T-1} \log(f(\lambda_t; \vartheta))$ . When  $f(\lambda; \vartheta)$  is continuous for all  $\lambda \in [-\pi, \pi)$  then uniform convergence also follows adapting Hannan (1973, Lemma 1) since, uniformly on  $\Theta$ ,  $|\log(f(\lambda; \vartheta)) - g_M(\lambda; \vartheta)| < \delta$ , for some arbitrary  $\delta > 0$  taking large enough  $M$ , where  $g_M(\lambda; \vartheta)$  is the  $M$ -terms Cesaro sum of the Fourier series of  $\log(f(\lambda; \vartheta))$ . When  $f(\lambda; \vartheta)$  is not continuous at zero frequency the result follows by adapting Zaffaroni (2003, Lemma 10) bearing in mind that  $\sup_{\vartheta \in \Theta} f(\lambda; \vartheta) \leq K |\lambda|^{-2d_U}$ ,  $\inf_{\vartheta \in \Theta} f(\lambda; \vartheta) \geq K > 0$ ,  $\lambda \in [-\pi, \pi)$ , with  $d_U < 1/2$ .

(ii) By Assumption  $F$  one gets  $f(\lambda)/f(\lambda; \vartheta) - 1 \geq -\log(f(\lambda; \vartheta)/f(\lambda))$ , equality holding only for  $\vartheta = \vartheta_0$  by Lemma 4-(i). ■

**Lemma 7** Under Assumptions  $A(4)$ ,  $B$ ,  $C(k_2, 1)$ ,  $D(k_2, 1)$ ,  $F$ ,

$$T^{\frac{1}{2}} \left( \int_{-\pi}^{\pi} g(\lambda) (I_T(\lambda) - EI_T(\lambda)) d\lambda \right) \rightarrow_d N(0, \tilde{V}), \quad \text{as } T \rightarrow \infty, \quad (24)$$

where  $g(\lambda) = g(\lambda; \vartheta_0)$ ,  $g(\lambda; \vartheta) = \frac{\partial}{\partial \vartheta} f^{-1}(\lambda; \vartheta)$ ,  $-\pi \leq \lambda < \pi$ , and  $\tilde{V}$  defines an  $s \times s$  positive semi definite matrix.

**Proof.** At first we characterize the local behaviour of  $g(\lambda; \vartheta)$  near the zero frequency. As  $\lambda \rightarrow 0^+$  by Lemma 1

$$f^{-2}(\lambda; \vartheta) \frac{\partial f(\lambda; \vartheta)}{\partial \phi_i} \sim \begin{cases} K \frac{\beta_{\phi_i}}{D_0''(\lambda^{-1}; \zeta)} \lambda^{2d(\zeta)}, & d(\zeta) \in (0, 1/2), \\ K \frac{\beta_{\phi_i}}{D_0'''(\lambda^{-1}; \zeta)}, & d(\zeta) = 0 \text{ and non summable } |D_0(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) = 0 \text{ and summable } |D_0(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) < 0, \end{cases}$$

for  $i = 1, \dots, q$ , and

$$f^{-2}(\lambda; \vartheta) \frac{\partial f(\lambda; \vartheta)}{\partial \zeta_j} \sim \begin{cases} K \frac{D_1''(\lambda^{-1}; \zeta)}{D_0''^2(\lambda^{-1}; \zeta)} \lambda^{2d(\zeta)}, & d(\zeta) \in (0, 1/2), \\ K \frac{D_1'''(\lambda^{-1}; \zeta)}{D_0'''^2(\lambda^{-1}; \zeta)}, & d(\zeta) = 0 \text{ and non summable } |D_1(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) = 0 \text{ and summable } |D_1(j; \zeta)| j^{-1}, \\ K_0, & d(\zeta) < 0, \end{cases}$$

for  $j = 1, \dots, p$ . Just considering the first case  $d(\zeta) > 0$ , it easily follows since  $f^2(\lambda; \vartheta) \sim KD_0''^2(\lambda^{-1}; \zeta)\lambda^{-4d(\zeta)}$  and  $\frac{\partial}{\partial \phi_i} f(\lambda; \vartheta) \sim K\beta_{\phi_i} D_0''(\lambda^{-1}; \zeta)\lambda^{-2d(\zeta)}$ . The other cases can be obtained in the same manner, noting that the terms that involve  $V(\lambda, \zeta)$  are dominated by terms in  $U(\lambda, \zeta)$ .

All the convergences below hold as  $T \rightarrow \infty$ . Let  $h_u(\vartheta) = (2\pi)^{-1} \int_{-\pi}^{\pi} g(\lambda; \vartheta) e^{iu\lambda} d\lambda$ ,  $u = 0, \pm 1, \pm 2, \dots$  be the Fourier coefficients of  $g(\lambda; \vartheta)$  and write  $h_u(\vartheta) = (h_{u\phi_1}(\vartheta), \dots, h_{u\phi_q}(\vartheta), h_{u\zeta_1}(\vartheta), \dots, h_{u\zeta_p}(\vartheta))'$ . Set  $h_u = h_u(\vartheta_0)$ ,  $h_{u\vartheta_i} = h_{u\vartheta_i}(\vartheta_0)$ ,  $i = 1, \dots, s$  and  $y_t = \log x_t^2 - \mu_{\log x^2}$ , where  $\mu_{\log x^2} = \omega_0 + \mu_{\log z^2}$ . We follow the approach put forward by Giraitis and Surgailis (1990) and approximate  $\int_{-\pi}^{\pi} g(\lambda) I_T(\lambda) d\lambda = \frac{1}{4\pi^2 T} \sum_{t,s=1}^T h_{\lfloor t-s \rfloor}(\log x_t^2 - \hat{\mu}_{\log x^2})(\log x_s^2 - \hat{\mu}_{\log x^2})$  (here  $\hat{\mu}_{\log x^2} = \sum_{t=1}^T \log x_t^2 / T$ ) by another quadratic form, which shares the same asymptotic distribution but is nevertheless much easier to handle. First note that by Fox and Taqqu (1987, Lemma 8.1) no change in the asymptotic distribution of the previous quadratic form occurs when substituting  $\hat{\mu}_{\log x^2}$  with the population mean  $\mu_{\log x^2}$ . Next, set

$$P_T = \sum_{t,s=1}^T h_{\lfloor t-s \rfloor} y_t y_s, \quad P_T(N) = \sum_{t,s=1}^T h_{\lfloor t-s \rfloor} y_t(N) y_s(N)$$

with  $y_t(N) = \log z_t^2 - \mu_{\log z^2} + \sum_{j=0}^N \psi_{0j} \epsilon_{t-j-1}$ ,  $0 < N < \infty$ . The main part of the proof is devoted to establish  $\text{var}(P_T - P_T(N)) = O(T \delta_N)$  for a sequence of positive terms satisfying  $\delta_N \rightarrow 0$  as  $N \rightarrow \infty$ .

We first show that  $\text{var}(P_T) = O(T)$ . In fact

$$\text{var}(P_T) = \sum_{t_1, s_1, t_2, s_2=1}^T h_{\lfloor t_1-s_1 \rfloor} h'_{\lfloor t_2-s_2 \rfloor} (c_{\lfloor t_1-t_2 \rfloor} c_{\lfloor s_1-s_2 \rfloor} + c_{\lfloor s_1-t_2 \rfloor} c_{\lfloor t_1-s_2 \rfloor}) \quad (25)$$

$$+ \sum_{t_1, s_1, t_2, s_2=1}^T h_{\lfloor t_1-s_1 \rfloor} h'_{\lfloor t_2-s_2 \rfloor} \text{cum}(y_{t_1}, y_{t_2}, y_{s_1}, y_{s_2}), \quad (26)$$

where  $\text{cum}(\cdot, \cdot, \cdot, \cdot)$  defines the fourth-order cumulant operator. Setting  $Q_T(f)$ ,  $Q_T(g)$  equal to the  $T \times T$  Toeplitz matrix based on the Fourier transforms of  $f(\lambda)$  and  $g(\lambda)$  respectively, the two terms on the right hand side of (25) can be written as  $\text{tr}(Q_T(f)Q_T(g)Q_T(f)Q_T(g))$ . When  $|f(\lambda)| = O(\lambda^{-a-\delta})$ ,  $|g(\lambda)| =$

$O(\lambda^{-b-\delta})$  as  $\lambda \rightarrow 0$ , any  $\delta > 0$ , and satisfying the regularity condition of Fox and Taqqu (1987, p.215), then  $\text{tr}(Q_T(f)Q_T(g)Q_T(f)Q_T(g)) = O(T)$  when  $2(a+b) < 1$  by Fox and Taqqu (1987, Theorem 1). Since any slowly varying function satisfies  $|L(u)| = O(u^\delta)$ ,  $u \rightarrow \infty$ , any  $\delta > 0$ , and for all possible of our cases  $-b \geq a \geq 0$  the result follows. For (26), it can be easily seen that  $\text{cum}(y_{t_1}, y_{t_2}, y_{s_1}, y_{s_2})$  is made by the sum of sixteen terms of the form:

$$\begin{aligned}
 & \kappa_{\log z^2 \log z^2 \log z^2 \log z^2} \mathbf{1}_{t_1=t_2=s_1=s_2} && \text{(one term),} \\
 & \kappa_{\log z^2 \log z^2 \log z^2 \epsilon} \psi_{s_2-t_1-1} \mathbf{1}_{t_1=t_2=s_1} && \text{(four terms),} \\
 & \kappa_{\log z^2 \log z^2 \epsilon \epsilon} \mathbf{1}_{t_1=t_2} \psi_{s_1-t_2-1} \psi_{s_2-t_1-1} && \text{(six terms),} \\
 & \kappa_{\log z^2 \epsilon \epsilon \epsilon} \psi_{t_2-t_1-1} \psi_{s_1-t_1-1} \psi_{s_2-t_1-1} && \text{(four terms),} \\
 & \kappa_{\epsilon \epsilon \epsilon \epsilon} \sum_{k=0}^{\infty} \psi_k \psi_{k+t_2-t_1} \psi_{k+s_1-t_1} \psi_{k+s_2-t_1} && \text{(one term),}
 \end{aligned} \tag{27}$$

where we set  $\kappa_{abcd} = \text{cum}(a_0, b_0, c_0, d_0)$  for *i.i.d.*  $\{a_t, b_t, c_t, d_t\}$  with bounded fourth moment. Consider the last term. Since  $\psi_j = 0$ ,  $j < 0$  then  $0 \leq s_2 - t_1 + k$  implies  $s_1 - s_2 \leq s_1 - t_1 + k$ . Likewise  $s_2 - s_1 \leq s_2 - t_1 + k$ , yielding  $|\psi_{k+s_1-t_1} \psi_{k+s_2-t_1}| = \mathbf{1}_{s_2-s_1 > 0} |\psi_{k+s_1-t_1} \psi_{k+s_2-t_1}| + \mathbf{1}_{s_1-s_2 \geq 0} |\psi_{k+s_1-t_1} \psi_{k+s_2-t_1}| \leq \mathbf{1}_{s_2-s_1 > 0} K |\psi_{s_2-s_1}| + \mathbf{1}_{s_1-s_2 \geq 0} K |\psi_{s_1-s_2}| \leq K |\psi_{|s_2-s_1|}|$  which in turn implies  $\kappa_{\epsilon \epsilon \epsilon \epsilon} \sum_{k=0}^{\infty} |\psi_k \psi_{k+t_2-t_1} \psi_{k+s_1-t_1} \psi_{k+s_2-t_1}| \leq K |\psi_{|s_2-s_1|}| \sum_{k=0}^{\infty} |\psi_k \psi_{k+|t_2-t_1|}| \leq K \sum_{k=0}^{\infty} |\psi_k \psi_{k+|s_2-s_1|}| \sum_{k=0}^{\infty} |\psi_k \psi_{k+|t_2-t_1|}|$ . Therefore we can apply Fox and Taqqu (1987, Theorem 1) precisely as we have done above, yielding that the term of (26) involving  $\kappa_{\epsilon \epsilon \epsilon \epsilon} \sum_{k=0}^{\infty} \psi_k \psi_{k+t_2-t_1} \psi_{k+s_1-t_1} \psi_{k+s_2-t_1}$  is  $O(T)$ . Along the same lines, one can easily show that all the other terms of (26) are  $O(T)$ , implying  $\text{var}(P_T) = O(T)$ . Write  $\text{var}(P_T - P_T(N)) = A_1 + A_2 + A_3$  where  $A_1 = \text{var}\left(\sum_{t,s=1}^T h_{|t-s|} y_t (y_s - y_s(N))\right)$ ,  $A_2 = \text{var}\left(\sum_{t,s=1}^T h_{|t-s|} y_s(N) (y_t - y_t(N))\right)$ ,  $A_3 = 2\text{cov}\left(\sum_{t,s=1}^T h_{|t-s|} y_t (y_s - y_s(N)), \sum_{t,s=1}^T h_{|t-s|} y_s(N) (y_t - y_t(N))\right)$ . The same bound apply to  $A_1$  and  $A_2$  and therefore, by Schwarz inequality, to  $A_3$  as well, so we just consider  $A_1$ . By the cumulants theorem (see Leonov and Shiryaev (1959)) one obtains  $A_1 = B_1 + B_2 + B_3$  with

$$\begin{aligned}
 B_1 &= \sum_{t_1, s_1, t_2, s_2=1}^T h_{|t_1-s_1|} h'_{|t_2-s_2|} \text{cov}(y_{t_1}, y_{t_2}) \text{cov}(y_{s_1} - y_{s_1}(N), y_{s_2} - y_{s_2}(N)) \\
 B_2 &= \sum_{t_1, s_1, t_2, s_2=1}^T h_{|t_1-s_1|} h'_{|t_2-s_2|} \text{cov}(y_{t_1}, y_{s_2} - y_{s_2}(N)) \text{cov}(y_{t_2}, y_{s_1} - y_{s_1}(N)), \\
 B_3 &= \sum_{t_1, s_1, t_2, s_2=1}^T h_{|t_1-s_1|} h'_{|t_2-s_2|} \text{cum}(y_{t_1}, y_{t_2}, y_{s_1} - y_{s_1}(N), y_{s_2} - y_{s_2}(N)).
 \end{aligned}$$

By Assumption A(2),  $\text{cov}(y_t, y_s) = \alpha_0 \mathbf{1}_{t=s} + \beta_0 \Delta_{|t-s|}(\zeta_0) + \gamma_0 \mathbf{1}_{t \neq s} \psi_{0|s-t-1}$ ,

$\text{cov}(y_t - y_t(N), y_s - y_s(N)) = \beta_0 \sum_{j=N+1}^{\infty} \psi_{0j} \psi_{0j+t-s}, \text{cov}(y_t, y_s - y_s(N)) = \gamma_0 \psi_{0s-t-1} \mathbf{1}_{(s-t > N+2)} + \beta_0 \sum_{j=N+1}^{\infty} \psi_{0j} \psi_{0j+t-s},$  and by Assumption A(4)

$$\begin{aligned} \text{cum}(y_{t_1}, y_{t_2}, y_{s_1} - y_{s_1}(N), y_{s_2} - y_{s_2}(N)) &= \quad (28) \\ &\kappa_{\log z^2 \log z^2 \epsilon \epsilon} \psi_{s_1-t_2-1} \psi_{s_2-t_1-1} \mathbf{1}_{t_1=t_2} \mathbf{1}_{s_1-t_2 \geq N+1} \mathbf{1}_{s_2-t_1 \geq N+1} \\ + &\kappa_{\log z^2 \epsilon \epsilon \epsilon} \psi_{|t_2-t_1|-1} \psi_{s_1-t_1-1} \psi_{s_2-t_1-1} \mathbf{1}_{s_1-t_1 \geq N+1} \mathbf{1}_{s_2-t_1 \geq N+1} \\ + &\kappa_{\log z^2 \epsilon \epsilon \epsilon} \psi_{|t_1-t_2|-1} \psi_{s_1-t_2-1} \psi_{s_2-t_2-1} \mathbf{1}_{s_1-t_2 \geq N+1} \mathbf{1}_{s_2-t_2 \geq N+1} \\ + &\kappa_{\epsilon \epsilon \epsilon \epsilon} \sum_{k=0}^{\infty} \mathbf{1}_{k \geq \max\{0, N+1-s_1+t_1, N+1-s_2+t_1\}} \psi_k \psi_{k+t_2-t_1} \psi_{k+s_1-t_1} \psi_{k+s_2-t_1}. \end{aligned}$$

For  $B_1$ , for some  $0 < \eta < 1$

$$|\text{cov}(y_t, y_s)| \leq K \left\{ \sum_{j=0}^{\infty} |\psi_{0j}|^{1-\eta/2} |\psi_{0j+t-s}|^{1-\eta/2} \right\}, \quad (29)$$

$$\begin{aligned} |\text{cov}(y_t - y_t(N), y_s - y_s(N))| &\leq \beta_0 \left| \sum_{j=N+1}^{\infty} |\psi_{0j}|^{\eta/2} |\psi_{0j}|^{1-\eta/2} |\psi_{0j+t-s}|^{\eta/2} |\psi_{0j+t-s}|^{1-\eta/2} \right| \\ &\leq K |\psi_{0N}|^{\eta} \left\{ \sum_{j=0}^{\infty} |\psi_{0j}|^{1-\eta/2} |\psi_{0j+t-s}|^{1-\eta/2} \right\}. \quad (30) \end{aligned}$$

Consider Ass.  $C(0, 1)$  and  $d = d(\theta_0) \in (0, 1/2)$  and set  $D_r(u) = D_r(u; \vartheta_0)$ ,  $r = 0, \pm 1, \dots$ . As  $u \rightarrow \infty$ ,  $\sum_{k=0}^{\infty} |\psi_k|^{1-\eta/2} |\psi_{k+u}|^{1-\eta/2} \sim |\psi_u|^{1-\eta/2} \sum_{k=0}^u \psi_k + \sum_{k=u+1}^{\infty} |\psi_k|^{2-\eta} \sim K |D_0(u)|^{2-\eta} u^{(1-\eta/2)(2d-1)}$ . Take  $j$  large enough so that  $\psi_i > 0$ ,  $j \leq i < \infty$ . The same applies when  $\psi_i < 0$ ,  $j \leq i < \infty$ . This is because  $D_0(\cdot)$  is slowly varying at infinity and thus does not change sign asymptotically. Then for any  $\eta > 0$ ,  $|(\psi_{0j}^{\eta} - \psi_{0j+1}^{\eta}) \psi_{0j}^{1-\eta}| \leq |(\psi_{0j}^{\eta} - \psi_{0j+1}^{\eta}) \psi_{0j}^{1-\eta} + \psi_{0j+1}^{\eta} (\psi_{0j}^{1-\eta} - \psi_{0j+1}^{1-\eta})| = |\psi_{0j}^{\eta} \psi_{0j}^{1-\eta} - \psi_{0j+1}^{\eta} \psi_{0j+1}^{1-\eta}| = |\psi_{0j} - \psi_{0j+1}| \leq K \psi_{0j}/j$  implying that  $\psi_{0j}^{\eta}$  is quasi-monotonically convergent to zero, any  $\eta > 0$ . Therefore, by Yong (1974, Theorem III-23 and III-33 (ii)) the Fourier transform of the terms in  $\{\}$ -brackets on left hand side of (29) and (30) is  $O(\lambda^{-(1+(1-\eta/2)(2d-1))-\delta})$ , any  $\eta, \delta > 0$ , as  $\lambda \rightarrow 0$ . Therefore  $B_1 = O(|\psi_{0N}|^{\eta} T)$  by Fox and Taquq (1987, Theorem 1) as long as  $2(-2d + 1 + (1 - \eta/2)(2d - 1)) < 1$  implied by  $\eta < 1/(1 - 2d)$ . The same

arguments carry through to show that  $B_2 = O(|\psi_{0N}|^\eta T)$ . For the cumulant term  $B_3$ , consider the last term on the left hand side of (28). When both  $s_1 - t_1 > N + 1$ ,  $s_2 - t_1 > N + 1$  this is bounded by  $K |\psi_{0N}|^\eta \sum_{k=0}^{\infty} |\psi_k \psi_{k+|s_2-s_1|}|^{1-\eta/2} \sum_{k=0}^{\infty} |\psi_k \psi_{k+|t_2-t_1|}|^{1-\eta/2}$  by following the same arguments used for the corresponding term of (26). When either  $s_1 - t_1 > N + 1$  or  $s_2 - t_1 > N + 1$  then the bound is  $K |\psi_{0N}|^{\eta/2} \sum_{k=0}^{\infty} |\psi_k \psi_{k+|s_2-s_1|}|^{1-\eta/2} \sum_{k=0}^{\infty} |\psi_k \psi_{k+|t_2-t_1|}|^{1-\eta/2}$ . Finally when both  $s_1 - t_1 \leq N + 1$ ,  $s_2 - t_1 \leq N + 1$  the bound is  $K |\psi_{0N}| \sum_{k=0}^{\infty} |\psi_k \psi_{k+|s_2-s_1|}| \sum_{k=0}^{\infty} |\psi_k \psi_{k+|t_2-t_1|}|$ . A tighter bound applies to the other terms of (28) yielding  $B_3 = O(|\psi_{0N}|^{\eta/2} T)$ . Cases  $C(0, 1)$ ,  $D(0, 1)$ , with  $d \leq 0$ , and  $C(1, 1)$  follow along the same lines, with an even simpler proof.

The proof ends considering that  $P_T(N)/\sqrt{T}$  is a quadratic form in  $N$ -dependent variates. This implies that it is  $\phi$ -mixing with arbitrarily fast decreasing mixing coefficients and Ibragimov and Linnik (1971, Theorem 18.5.1) applies. ■

**Lemma 8** Under Assumptions  $A(2)$ ,  $B$ ,  $C(k_2, 3)$ ,  $D(k_2, 3)$ ,  $F$ , for  $M_T(\vartheta) = \frac{\partial^2}{\partial \vartheta \partial \vartheta'} Q_T(\vartheta)$ ,

$$\sup_{\vartheta \in \Theta} \|M_T(\vartheta) - M(\vartheta)\| \rightarrow_{a.s.} 0 \quad \text{as } T \rightarrow \infty.$$

**Proof.** All the convergences below hold as  $T \rightarrow \infty$ . We first establish pointwise convergence of  $M_T(\vartheta)$  to  $M(\vartheta)$  *a.s.* for each  $\vartheta \in \Theta$ . Let

$$M_T(\vartheta) = A_{1T}(\vartheta) + A_{2T}(\vartheta) + A_{3T}(\vartheta) + A_{4T}(\vartheta), \quad (31)$$

with

$$\begin{aligned} A_{1T}(\vartheta) &= \frac{1}{T} \sum_{t=1}^{T-1} \frac{\partial^2 f(\lambda_t; \vartheta)}{\partial \vartheta \partial \vartheta'} \frac{1}{f(\lambda_t; \vartheta)}, & A_{2T}(\vartheta) &= -\frac{1}{T} \sum_{t=1}^{T-1} \frac{\partial f(\lambda_t; \vartheta)}{\partial \vartheta} \frac{\partial f(\lambda_t; \vartheta)}{\partial \vartheta'} \frac{1}{f^2(\lambda_t; \vartheta)}, \\ A_{3T}(\vartheta) &= -\frac{1}{T} \sum_{t=1}^{T-1} \frac{\partial^2 f(\lambda_t; \vartheta)}{\partial \vartheta \partial \vartheta'} \frac{I_T(\lambda_t)}{f^2(\lambda_t; \vartheta)}, & A_{4T}(\vartheta) &= \frac{2}{T} \sum_{t=1}^{T-1} \frac{\partial f(\lambda_t; \vartheta)}{\partial \vartheta} \frac{\partial f(\lambda_t; \vartheta)}{\partial \vartheta'} \frac{I_T(\lambda_t)}{f^3(\lambda_t; \vartheta)}. \end{aligned}$$

By Hannan (1973, Lemma 1), given Lemmas 3 and 4-(ii), uniformly in  $\vartheta \in \Theta$ ,  $A_{3T}(\vartheta) \rightarrow_{a.s.} A_3(\vartheta) = -(2\pi)^{-1} \int_{-\pi}^{\pi} f(\lambda)/f^2(\lambda; \vartheta) (\frac{\partial^2}{\partial \vartheta \partial \vartheta'} f(\lambda; \vartheta)) d\lambda$ ,  $A_{4T}(\vartheta) \rightarrow_{a.s.}$

$A_4(\vartheta) = (\pi)^{-1} \int_{-\pi}^{\pi} f(\lambda)/f^3(\lambda; \vartheta) (\frac{\partial}{\partial \vartheta} f(\lambda; \vartheta)) (\frac{\partial}{\partial \vartheta'} f(\lambda; \vartheta)) d\lambda$ . When  $d(\zeta) \in [0, 1/2)$ , much in the same way as in Zaffaroni (2003, Lemma 14),  $A_{1T}(\vartheta) \rightarrow_{a.s.} A_1(\vartheta) = (2\pi)^{-1} \int_{-\pi}^{\pi} f^{-1}(\lambda; \vartheta) (\frac{\partial^2}{\partial \vartheta \partial \vartheta'} f(\lambda; \vartheta)) d\lambda$ ,  $A_{2T}(\vartheta) \rightarrow_{a.s.} A_2(\vartheta) = -(2\pi)^{-1} \int_{-\pi}^{\pi} f^{-2}(\lambda; \vartheta) (\frac{\partial}{\partial \vartheta} f(\lambda; \vartheta)) (\frac{\partial}{\partial \vartheta'} f(\lambda; \vartheta)) d\lambda$ , whereas when  $d(\zeta) < 0$  we adapt Hannan (1973, Lemma 1) idea, using the decomposition  $|1/T \sum_{t=1}^T g(\lambda_t) - 1/(2\pi) \int_{-\pi}^{\pi} g(\lambda) d\lambda| \leq |1/T \sum_{t=1}^T g(\lambda_t) - 1/T \sum_{t=1}^T g_T(\lambda_t)| + |1/T \sum_{t=1}^T g_T(\lambda_t) - 1/(2\pi) \int_{-\pi}^{\pi} g(\lambda) d\lambda|$  for a generic Lip-continuous function  $g(\lambda)$  with Cesaro sum  $g_T(\lambda)$ . It remains to show that the following equicontinuity property holds:  $\sup_{\tilde{\vartheta}: \|\tilde{\vartheta} - \vartheta\| < \epsilon} \|A_{iT}(\tilde{\vartheta}) - A_{iT}(\vartheta)\| \rightarrow 0$  as  $\epsilon \rightarrow 0$ , and that  $A_i(\vartheta)$  are continuous, for  $i = 1, 2$ . These are implied by  $\|\frac{\partial A_{iT}(\vartheta)}{\partial \vartheta}\| + \|\frac{\partial A_i(\vartheta)}{\partial \vartheta}\| < \infty$ , for  $i = 1, 2$ , and the latter easily follow by the same arguments used to establish pointwise convergence, given the smoothness of the third-order derivatives of  $f(\lambda; \vartheta)$  away from zero frequency, and making use of Lemma 1 regarding their local behaviour around zero frequency. ■

**Lemma 9** Under Assumptions  $A(4)$ ,  $B$ ,  $C(k_2, 2)$ ,  $D(k_2, 2)$ ,  $F$ ,

$$\sup_{\vartheta \in \Theta} \|V_T(\vartheta) - V(\vartheta)\| \rightarrow 0 \quad \text{as } T \rightarrow \infty,$$

where

$$\begin{aligned}
 V_T(\vartheta) &= \frac{2}{T} \sum_{t=1}^{T-1} N(\lambda_t; \vartheta) N'(\lambda_t; \vartheta) + \frac{2\pi}{T^2} \sum_{t_1=1}^{T-1} \sum_{t_2=1}^{T-1} \frac{N(\lambda_{t_1}; \vartheta)}{f(\lambda_{t_1}; \vartheta)} \frac{N(\lambda_{t_2}; \vartheta)'}{f(\lambda_{t_2}; \vartheta)} Q(-\lambda_{t_1}, \lambda_{t_2}, -\lambda_{t_2}; \vartheta), \quad (32) \\
 V(\vartheta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} N(\omega; \vartheta) N'(\omega; \vartheta) d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{N(\omega_1; \vartheta)}{f(\omega_1; \vartheta)} \frac{N(\omega_2; \vartheta)'}{f(\omega_2; \vartheta)} Q(-\omega_1, \omega_2, -\omega_2; \vartheta) d\omega_1 d\omega_2
 \end{aligned}$$

and where the trispectrum  $Q(\omega_1, \omega_2, \omega_3) = Q(\omega_1, \omega_2, \omega_3; \vartheta_0)$  is defined by

$$\begin{aligned}
Q(\omega_1, \omega_2, \omega_3; \vartheta) &= \frac{\kappa_{\log z^2 \log z^2 \log z^2 \log z^2}}{(2\pi)^3} \tag{33} \\
&+ \frac{\kappa_{\log z^2 \log z^2 \log z^2 \log z^2 \epsilon}}{(2\pi)^3} \left( e^{i\omega_1} \psi(e^{i\omega_1}; \zeta) + e^{i\omega_2} \psi(e^{i\omega_2}; \zeta) + e^{i\omega_3} \psi(e^{i\omega_3}; \zeta) + e^{-i(\omega_1+\omega_2+\omega_3)} \psi(e^{-i(\omega_1+\omega_2+\omega_3)}; \zeta) \right) \\
&+ \frac{\kappa_{\log z^2 \log z^2 \log z^2 \epsilon \epsilon}}{(2\pi)^3} \left( e^{i(\omega_2+\omega_3)} \psi(e^{i\omega_2}; \zeta) \psi(e^{i\omega_3}; \zeta) + e^{i(\omega_1+\omega_3)} \psi(e^{i\omega_1}; \zeta) \psi(e^{i\omega_3}; \zeta) + e^{i(\omega_1+\omega_2)} \psi(e^{i\omega_1}; \zeta) \psi(e^{i\omega_2}; \zeta) \right. \\
&\quad \left. + \psi(e^{-i(\omega_1+\omega_2+\omega_3)}; \zeta) \{ e^{-i(\omega_1+\omega_3)} \psi(e^{i\omega_2}; \zeta) + e^{-i(\omega_2+\omega_3)} \psi(e^{i\omega_1}; \zeta) + e^{-i(\omega_1+\omega_2)} \psi(e^{i\omega_3}; \zeta) \} \right) \\
&+ \frac{\kappa_{\log z^2 \epsilon \epsilon \epsilon}}{(2\pi)^3} \left( e^{i(\omega_1+\omega_2+\omega_3)} \psi(e^{i\omega_1}; \zeta) \psi(e^{i\omega_2}; \zeta) \psi(e^{i\omega_3}; \zeta) \right. \\
&\quad \left. + \psi(e^{-i(\omega_1+\omega_2+\omega_3)}; \zeta) \{ e^{-i\omega_1} \psi(e^{i\omega_2}; \zeta) \psi(e^{i\omega_3}; \zeta) + e^{-i\omega_2} \psi(e^{i\omega_1}; \zeta) \psi(e^{i\omega_3}; \zeta) + e^{-i\omega_3} \psi(e^{i\omega_1}; \zeta) \psi(e^{i\omega_2}; \zeta) \} \right) \\
&+ \frac{\kappa_{\epsilon \epsilon \epsilon \epsilon}}{(2\pi)^3} \psi(e^{i(\omega_1+\omega_2+\omega_3)}; \zeta) \psi(e^{-i\omega_1}; \zeta) \psi(e^{-i\omega_2}; \zeta) \psi(e^{-i\omega_3}; \zeta).
\end{aligned}$$

**Proof.** We first establish (33) evaluated at  $\vartheta_0$ . The general case  $\vartheta \neq \vartheta_0$  follows by substituting the  $\psi_{0k}$  with the  $\psi_k(\zeta)$ . Set  $u_1 = t_2 - t_1$ ,  $u_2 = s_1 - t_1$ ,  $u_3 = s_2 - t_1$ . We need to evaluate the Fourier transform of each of the sixteen terms which make  $\text{cum}(y_0, y_{u_1}, y_{u_2}, y_{u_3})$ , listed in (27). The first term is trivial. Let us consider the four terms which involve  $\kappa_{\log z^2 \log z^2 \log z^2 \log z^2 \epsilon}$ . These correspond to the four cases  $u_1 = u_2 = 0$ ,  $u_1 = u_3 = 0$ ,  $u_2 = u_3 = 0$  and  $u_1 = u_2 = u_3$ . For the first, neglecting constants,  $\sum_{u_1=u_2=0, u_3=-\infty}^{\infty} \psi_{u_3-1} e^{i(u_1\omega_1+u_2\omega_2+u_3\omega_3)} = e^{i\omega_3} \sum_{u_3=1}^{\infty} \psi_{0u_3-1} e^{i(u_3-1)\omega_3} = e^{i\omega_3} \psi(e^{i\omega_3}; \zeta_0)$ , and likewise for the second and third term. For the fourth  $\sum_{u_1=u_2=u_3=u, u=-\infty}^{\infty} \psi_{-u-1} e^{i(u_1\omega_1+u_2\omega_2+u_3\omega_3)} = \sum_{u=-\infty}^{-1} \psi_{0-u-1} e^{iu(\omega_1+\omega_2+\omega_3)} = e^{-i(\omega_1+\omega_2+\omega_3)} \psi(e^{-i(\omega_1+\omega_2+\omega_3)}; \zeta_0)$ . Evaluation of the other terms of (33) follows along the same lines and details are skipped for sake of simplicity. Pointwise convergence of  $V_T(\vartheta)$  to  $V(\vartheta)$  easily follow by the same arguments used to establish convergence for terms  $A_{iT}(\vartheta)$ ,  $i = 1, 2$  in Lemma 8, since the factorization of  $Q(-\omega_1, \omega_2, -\omega_2; \vartheta)$  permits to evaluate the double sum term of  $V_T(\vartheta)$  by looking at each sum separately. Finally, uniform convergence follows by equicontinuity of  $V_T(\vartheta)$  and continuity of  $V(\vartheta)$  which in turn hold since by the same arguments  $\|\frac{\partial V_T(\vartheta)}{\partial \vartheta}\| + \|\frac{\partial V(\vartheta)}{\partial \vartheta}\| < \infty$ . ■

**Proof of Theorem 1.** In view of Lemmas 4-6, the result follows adapting the proof of Hannan (1973, Theorem 1). ■

**Proof of Theorem 2.** The result follow by the Delta method, in view of Lemmas 7 and 8, once we show that

$$\left( \int_{-\pi}^{\pi} g(\lambda) E(I_T(\lambda)) d\lambda + \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \vartheta_0)}{\partial \vartheta} d\lambda \right) = o(T^{-\frac{1}{2}}). \quad (34)$$

By Parseval's relation the left hand side of (34) equals  $\left( \sum_{u=-T+1}^{T-1} (1 - u/T) c_u \int_{-\pi}^{\pi} g(\lambda) e^{iu\lambda} d\lambda - \sum_{u=-\infty}^{\infty} h_u c_u \right)$ , and the norm of the above expression is bounded by  $O(\|\sum_{u=-T+1}^{\infty} h_u c_u + \sum_{u=T-1}^{\infty} \frac{u}{T} h_u c_u\|) = O(|L(T)| T^{-1})$ . Expression (32) of  $V$  follows directly as a by-product from the proof of Lemma 7, for the part just involving the spectral density, whereas for the part involving the trispectrum one can relatively easily adapt Hosoya (1997, Lemma 3.2). A consistent estimate of asymptotic covariance is obtained by Lemma 8 and 9 by plugging  $\hat{\vartheta}_T$  into  $M_T(\vartheta)$  and  $V_T(\vartheta)$ . To obtain the estimator of  $V$ , we also require a consistent estimate of the joint cumulants in the  $\log z_t^2$  and  $\epsilon_t$ , listed in (33), which can be easily obtained since the cumulants, as any mixed moment of  $(z_t, \epsilon_t)$ , are known functions of  $\phi_0$ . A consistent estimate is simply obtained by plugging  $\hat{\phi}_T$  into the expression for the theoretical cumulants. ■

## References

- BHANSALI, R. (1974): "Asymptotic properties of the Wiener-Kolmogorov predictor.I.," *Journal of the Royal Statistical Society Series B*, 36, 61–73.
- BOLLERLSEV, T., AND H. MIKKELSEN (1996): "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, 73/1, 151–184.
- BRANDT, M. W., AND C. S. JONES (2006): "Volatility Forecasting With Range-Based EGARCH Models," *Journal of Business & Economic Statistics*, 24, 470–486.

- BREIDT, F., N. CRATO, AND P. DELIMA (1998): “The detection and estimation of long memory in stochastic volatility,” *Journal of Econometrics*, 83, 325–348.
- BROCKWELL, P., AND R. DAVIS (1987): *Time series: theory and methods*. New York: Springer Verlag.
- DEO, R., C. HURVICH, AND Y. LU (2006): “Forecasting realized volatility using a long memory stochastic volatility model: estimation, prediction and seasonal adjustment,” *Journal of Econometrics*, 131, 29–58.
- DIEBOLD, F., AND Y. CHEUNG (1994): “Maximum Likelihood Estimation of Fractionally Integrated Noise with Unknown Mean,” *Journal of Econometrics*, 62, 301–316.
- DUAN, J., G. GAUTHIER, C. SASSEVILLE, AND J. SIMONATO (2006): “Approximating the GJR-GARCH and the EGARCH option pricing models analytically,” *Journal of Computational Finance*, 9.
- DUNSMUIR, W. (1979): “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.
- ENGLE, R. F. (1982): “Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom,” *Econometrica*, 50, 987–1007.
- FOX, R., AND M. TAQQU (1987): “Central limit theorems for quadratic forms in random variables having long-range dependence,” *Probability Theory and related Fields*, 74, 213–240.
- GHYSELS, E., A. HARVEY, AND E. RENAULT (1995): “Stochastic Volatility,” in *Handbook of Statistics vol 14*. Amsterdam: North Holland.
- GIRAITIS, L., AND P. M. ROBINSON (2001): “Whittle estimation of ARCH models,” *Econometric Theory*, 17, 608–631.

- GIRAITIS, L., AND D. SURGAILIS (1990): “A central limit theorem for quadratic forms in strongly dependent linear variabls and its application to asymptotically normality of Whittle’s estimate,” *Probability Theory and Related Fields*, 86, 87–104.
- GLOSTEN, L., R. JAGANATHAN, AND D. RUNKLE (1993): “On the relation between the expected value and the volatility of the nominal excess returns on stocks,” *Journal of Finance*, 48, 1779–1801.
- GRADSHTEYN, I., AND I. RYZHIK (1994): *Table of integrals, series and products*. San Diego: Academic Press, fifth edn.
- GRANGER, C., AND A. ANDERSEN (1978): “On the invertibility of time series models,” *Stochastic Processes and their Applications*, 8, 87–92.
- HANNAN, E. (1973): “The asymptotic theory of linear time series models,” *Journal of Applied Probability*, 10, 130–145.
- HARVEY, A. (1998): “Long memory in stochastic volatility,” in *Forecasting volatility in financial markets*, ed. by J. Knight, and S. Satchell. London: Butterworth-Heineman.
- HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): “Multivariate Stochastic Variance Models,” *Review of Economic Studies*, 61, 247–264.
- HIDALGO, J., AND Y. YAJIMA (2002): “Prediction and signal extraction of strongly dependent processes in the frequency domain,” *Econometric Theory*, 18, 584–624.
- HOSOYA, Y. (1974): “Estimation problems on stationary time series,” Unpublished Ph.D. Dissertation, Yale University.
- (1997): “A limit theory for long-range dependence and statistical inference on related models,” *Annals of Statistics*, 25, 105–137.

- HOSOYA, Y., AND M. TANIGUCHI (1982): “A central limit theorem for stationary processes and the parameter estimation of linear processes,” *Annals of Statistics*, 10, 132–153.
- IBRAGIMOV, I., AND J. LINNIK (1971): *Independent and stationary sequences of random variables*. Groningen: Walters-Noordhoff.
- LEONOV, V., AND A. SHIRYAEV (1959): “On a method of calculation of semi-invariants,” *Theory of Probability Applications*, 4, 319–329.
- LINTON, O., AND E. MAMMEN (2005): “Estimating semiparametric ARCH( $\infty$ ) models by kernel smoothing methods,” *Econometrica*, 73, 771–836.
- MEDDAHI, N., AND E. RENAULT (2004): “Temporal aggregation of volatility models,” *Journal of Econometrics*, 119, 355–379.
- NELSON, D. (1988): “The time series behaviour of stock market volatility and returns,” Unpublished Ph.D. Dissertation, Massachusetts Institute of Technology.
- NELSON, D. (1991): “Conditional heteroscedasticity in asset pricing: a new approach,” *Econometrica*, 59, 347–370.
- PEREZ, A., AND P. ZAFFARONI (2008): “Finite Sample Properties of Maximum Likelihood and Whittle Estimators in EGARCH and FIEGARCH Models,” Imperial College London and Universidad de Valladolid.
- ROBINSON, P., AND P. ZAFFARONI (2006): “Pseudo maximum likelihood estimation of ARCH( $\infty$ ) models,” *Annals of Statistics*, 34, 1049–1074.
- ROBINSON, P. M. (1978): “Alternative models for stationary stochastic processes,” *Stochastic Processes and their Applications*, 8, 141–152.
- (1994): “Rates of convergence and optimal bandwidth for long-range dependence,” *Probability Theory and Related Fields*, 99, 443–473.

- STRAUMANN, D. (2005): *Estimation in conditionally heteroskedastic time series models. Lecture notes in statistics*, no. 181. Berlin: Springer.
- STRAUMANN, D., AND T. MIKOSCH (2006): “Quasi-maximum likelihood estimation in heteroskedastic time series: a stochastic recurrence equations approach,” *Annals of Statistics*, 34, 2449–2495.
- TANIGUCHI, M. (1982): “On estimation of the integrals of the fourth order cumulant spectral density,” *Biometrika*, 69, 117–122.
- TAYLOR, S. (1986): *Modelling Financial Time Series*. Chichester (UK): Wiley.
- VELASCO, C., AND P. ROBINSON (2000): “Whittle pseudo-maximum likelihood estimation for nonstationary time series,” *Journal of the American Statistical Association*, 95, 1229–1243.
- WHITTLE, P. (1962): “Gaussian estimation in stationary time series,” *Bulletin of the International Statistical Institute*, 39, 105–129.
- (1983): *Prediction and regulation by linear least-squares method*. Oxford: Basil Blackwell, second edn.
- YONG, C. H. (1974): *Asymptotic Behaviour of Trigonometric Series*. Hong Kong: Chinese University of Hong Kong.
- ZAFFARONI, P. (2003): “Gaussian inference on certain long-range dependent volatility models,” *Journal of Econometrics*, 115, 199–258.
- (2008): “Whittle estimation of exponential volatility models,” Imperial College London.
- ZYGMUND, A. (1977): *Trigonometric series*. Cambridge (UK): Cambridge University Press.