ANALYSIS OF THE CORRELATION STRUCTURE OF SQUARE TIME SERIES

BY WILFREDO PALMA AND MAURICIO ZEVALLOS

Pontificia Universidad Católica de Chile Universidad de Santiago de Chile

First Version received March 2001

Abstract. This paper analyses the asymptotic behaviour of the autocorrelation structure exhibited by squares of time series with a Wold expansion where the input error is a sequence of random variables with mean zero and finite kurtosis. Two important cases are discussed: (i) when the errors are independent and, (ii) when the errors are uncorrelated but their squares are correlated. Both situations are addressed when the process exhibits short or long memory. Consequences of these results on certain models widely used in many disciplines are also discussed.

Keywords. Autocorrelation; conditional heteroskedasticity; linear and non-linear time series; long and short memory.

1. INTRODUCTION

Statistical analyses of time series are usually based on the study of the autocorrelation function of the data. Nevertheless, the analysis of the squares or higher powers of the time series may give valuable clues about crucial aspects such as linearity, normality or memory of the process. These three aspects may interact in the nature of the process and their effects may be uncovered by a careful analysis of the autocorrelation of both the original series and its square. This occurs, for example, in financial economics, where several empirical studies have evidenced the presence of peculiar characteristics named *stylized facts*. These include, among others, lack of or very little autocorrelation, possible strong correlation of the squares of the series and non-normality (see for example Baillie, 1996; Shephard, 1996; Lobato and Savin, 1998). Similar features have been observed in data from other fields. In physics, for instance, the presence of strong autocorrelation in the squares of differences in velocity of the mean wind direction has been explored by Barndorff-Nielsen and Shephard (2000), and Mantegna and Stanley (2000).

In order to model the stylized facts in economic time series, Engle (1982) proposed the autoregressive conditional heteroskedasticity, ARCH model. Based on this seminal work a plethora of related models have been introduced: GARCH models (Taylor, 1986; Bollerslev, 1986), EGARCH models (Nelson, 1991), stochastic variance models (Harvey *et al.*, 1994), FIGARCH models (Baillie *et al.*, 1996; Bollerslev and Mikkelsen, 1996), LMGARCH models (Robinson, 1991; Robinson and Henry, 1999; Henry, 2001, among others). A key feature of these models is the time-varying conditional variance which is related to time-dependent squares.

An important issue in the statistical analysis of time series is determining whether the observations or a transformation of them, such as the squares, have short or long memory. A short-memory process has autocorrelations decaying to zero exponentially as the lag increases. In contrast, the autocorrelations of a long-memory process decay to zero at a hyperbolic rate. The autoregressive moving average (ARMA) model described by Box *et al.* (1994) is an example of short-memory process. On the other hand, one of the most well-known long-memory model is the autoregressive fractionally integrated moving average (ARFIMA) (see Granger and Joyeux, 1980; Hosking, 1981). The subject of long-memory modelling has been extensively revised in the time series literature (see for instance, Beran, 1994; Lewis and Ray, 1997; Chan and Palma, 1998; among others). In financial time series, the presence of long memory in some square asset returns was evidenced by Baillie (1996) and Bollerslev and Mikkelsen (1996), to name a few. The analysis of the covariance structure of these series has been advanced by exact expressions for the autocorrelations of the squares obtained by Karanasos (1999) and He and Teräsvirta (1999) for the GARCH model, by Karanasos and Kim (2001) and He *et al.* (2002) for the EGARCH process, by Demos (2002) for a model that nests both the EGARCH and stochastic volatility specifications and by Karanasos *et al.* (2004) for FIGARCH and LMGARCH models. On the other hand, asymptotic expressions for the autocovariance function of squares and other nonlinear transformations of a class of stochastic volatility models have been established by Robinson (2001). Furthermore, model identification by analysing the autocorrelation function of squares is addressed by Bollerslev (1988) for GARCH processes and by Karanasos and Kim (2001) for EGARCH processes. Diagnostics checking of nonlinear models with conditional heteroskedasticity through the squares of residuals is discussed by Li and Mak (1994), Granger and Andersen (1978), Maravall (1983) and McLeod and Li (1983) for the univariate case and Ling and Li (1997b) for the multivariate case, among others.

This paper discusses some key statistical properties of the autocorrelation of the square of a time series with a moving average expansion. First, an exact expression for the autocorrelations of squared values is given. Second, the asymptotic behaviour of the autocorrelations of the squares of the series is investigated, establishing whether they have short or long memory and specifying their decaying rates. The class of models studied include for example, ARMA processes, ARFIMA models and a wide class of conditional heteroskedastic models such as GARCH, EGARCH, FIGARCH, LMARCH, among others. The results presented in this work are highly relevant since they help to identify the nature of the process by analysing the autocorrelation function of both the original and the square of the series. Furthermore, by analysing the behaviour of the autocorrelation of the squares of these processes, it is possible to discard those theoretical models which are incompatible with the data under study. In addition, the autocorrelations of the squares of the observed series can be used to estimate the parameters of the underlying process. For GARCH models, this method was used by Baillie and Chung (2001) to obtain minimum distance estimators.

This paper is organized as follows. A characterization of the autocorrelation function of the squares of a time series having a Wold expansion and its asymptotic behaviour are discussed in Section 2. These theoretical results are applied in Section 3 to the study of the correlation structure of several well known models. Conclusions are presented in Section 4 and proofs of theorems are presented in the Appendix.

2. CHARACTERIZATION OF THE CORRELATION OF SQUARES

The behaviour of the autocorrelation function of the square of a time series with Wold decomposition is discussed in this section. Consider the following process $\{y_t\}$ with expansion

$$y_t = \Psi(B)\varepsilon_t, \tag{1}$$

where

$$\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i, \quad \psi_0 = 1, \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty,$$

and ε_t has finite kurtosis, η. If $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables (strict white noise), then (1) generates a linear process. On the other hand, if the sequence $\{\varepsilon_t\}$ is white noise but not strict, then the resulting process may be nonlinear (see for example Mills, 1999, p. 28). Therefore, two situations are distinguished in this study: when $\{y_t\}$ is a linear process and when $\{y_t\}$ is a nonlinear process. In addition, these two cases are addressed for both short and long memory filter, $\Psi(B)$. Note that for a short-memory filter, $\psi_i \sim v^i$ for some $|v| < 1$ whereas for a long-memory filter $\psi_i \sim i^{-\beta}$ for some $\beta \in (1/2, 1)$.

This section contains two parts. An explicit expression for the autocorrelation function of the squared process, $\rho_{y^2}(n)$, is given in Theorem 1 in Section 2.1. A consequence of this result, Corollary 1, concerns the specific case of a linear process obtained when the input sequence, $\{\varepsilon_t\}$, is independent. In addition, the asymptotic behaviour of $\rho_{y^2}(n)$ is analysed in Section 2.2 for both linear and nonlinear processes and the results are summarized in Theorems 2, 3 and 4. These two cases are discussed for both short- and long-memory filters $\Psi(B)$ in (1).

2.1. *Autocorrelation function of squared processes (ACFSq)*

Assume that $\{\varepsilon_t\}$ are random variables with zero mean, finite kurtosis, $\eta = E(\varepsilon_t^4)/[E(\varepsilon_t^2)]^2$, uncorrelated but not necessarily independent with

$$E(\varepsilon_u\varepsilon_v) = \begin{cases} \sigma^2 & \text{if } u = v \\ 0 & \text{elsewhere,} \end{cases} \tag{2}$$

$$E(\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v) = \begin{cases} [1 + (\eta - 1)\rho_{\varepsilon^2}(s - v)]\sigma^4 & \text{if } s = t,\ u = v \text{ or } s = u,\ t = v \\ [1 + (\eta - 1)\rho_{\varepsilon^2}(s - t)]\sigma^4 & \text{if } s = v,\ t = u \\ 0 & \text{elsewhere.} \end{cases} \tag{3}$$

The next theorem establishes an explicit expression for the autocorrelation function of $\{y_t^2\}$ for linear and nonlinear processes satisfying both (2) and (3). This formula plays a key role in the theorems presented in the next subsection dealing with the asymptotic behaviour of the autocorrelation function of the square observations.

THEOREM 1. *For the process defined by* (1) *with errors satisfying* (2) *and* (3) *with finite kurtosis* η, *the autocorrelation function of the squared process*, ρ_{y^2}, *is given by*

$$\rho_{y^2}(n) = \frac{2}{\kappa - 1}\rho_y^2(n) + \frac{\kappa - 3}{\kappa - 1}\alpha(n) + \frac{\eta - 1}{\kappa - 1}[\tau(n) + 2\Delta(n) - 3\Delta(0)\alpha(n)], \tag{4}$$

where the autocorrelation function of the process $\{y_t\}$ *is defined as*

$$\rho_y(n) = \left(\sum_{i=0}^{\infty}\psi_i^2\right)^{-1}\sum_{i=0}^{\infty}\psi_i\psi_{i+n}, \tag{5}$$

and

$$\alpha(n) = \left(\sum_{i=0}^{\infty}\psi_i^4\right)^{-1}\sum_{i=0}^{\infty}\psi_i^2\psi_{i+n}^2, \tag{6}$$

$$\Delta(n) = \left(\sum_{i=0}^{\infty}\psi_i^2\right)^{-2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\psi_{i+n}\psi_{j+n}\rho_{\varepsilon^2}(i - j), \tag{7}$$

$$\tau(n) = \left(\sum_{i=0}^{\infty}\psi_i^2\right)^{-2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i^2\psi_j^2\rho_{\varepsilon^2}(n + j - i), \tag{8}$$

$$\kappa = 3 - 2\eta\left(\sum_{i=0}^{\infty}\psi_i^2\right)^{-2}\sum_{i=0}^{\infty}\psi_i^4 + 3(\eta - 1)\Delta(0), \tag{9}$$

where κ *is the kurtosis of* y_t *and* ρ_{ε^2} *is the autocorrelation function of* $\{\varepsilon_t^2\}$.

Observe that if the sequence $\{\varepsilon_t\}$ is a strict white noise then

$$\tau(n) = \Delta(n) = \Delta(0)\alpha(n) = \left(\sum_{i=0}^{\infty} \psi_i^2\right)^{-2} \sum_{i=0}^{\infty} \psi_i^2 \psi_{i+n}^2,$$

and therefore $\tau(n) + 2\Delta(n) - 3\Delta(0)\alpha(n) = 0$. Replacing this in expression (4), given by Theorem 1, yields the following result for a linear process which was first obtained by Taylor (1986).

Corollary 1. (*Linear process*): *Assume that* $\{\varepsilon_t\}$ *are i.i.d. random variables with zero mean and finite kurtosis* η. *Then*,

$$\rho_{y^2}(n) = \frac{2}{\kappa - 1} \rho_y^2(n) + \frac{\kappa - 3}{\kappa - 1} \alpha(n), \tag{10}$$

where κ *is the kurtosis of* y_t *given by*

$$\kappa = (\eta - 3) \left(\sum_{i=0}^{\infty} \psi_i^2\right)^{-2} \sum_{i=0}^{\infty} \psi_i^4 + 3. \tag{11}$$

Expression (4) given by Theorem 1 is highly relevant to theoretical and applied aspects of the analysis of times series. On the theoretical side, (4) establishes an explicit relationship between the autocorrelation of y_t^2 and y_t and extends expression (10) to nonlinear processes by adding a new term involving ρ_{ε^2}. Moreover, it gives a framework for finding an exact expression for ρ_{y^2} by evaluating (5)–(9) and, as shown in Section 2.2, it allows for the analysis of the asymptotic behaviour of the autocorrelations of the squares. On the application side, it helps the identification of the nature of the process by discarding those models incompatible with the data. In this sense, Baillie and Chung (2001) observe that 'several previous articles dealing with financial market data have commented on the behaviour of the autocorrelations of the squared returns series, and the desirability of having a model which comes close to replicating the autocorrelations of the square returns'. Accordingly, Theorem 1 is a tool to analyse whether a particular model can replicate the correlation structure of the data. Furthermore, (4) can be used to obtain the minimum distance estimator of the parameters of heteroskedastic models (cf. Baillie and Chung, 2001).

2.2. *Asymptotic behaviour of ACFSq*

Based on Corollary 1, the next theorem establishes the asymptotic behaviour of the autocorrelation function of squares for linear process. In what follows, $x_n = O(\omega_n)$ means that $|x_n/\omega_n| \leq c$ for all n, where c is a positive constant.

Theorem 2. (*Independent input*): *Let* $\{y_t\}$ *be a process satisfying* (1) *with* $\{\varepsilon_t\}$ *a strict white noise sequence with finite kurtosis* η. *Then*

(a) *Short-memory filter: If* $\psi_i \sim \upsilon^i$ *for some* $|\upsilon| < 1$, *then both* $\alpha(n)$ *and* $\rho_{y^2}(n)$ *are* $O(\upsilon^{2n})$.

(b) *Long-memory filter: If* $\psi_i \sim i^{-\beta}$ *for some* $\beta \in (1/2,1)$, *then* $\alpha(n)$ *is* $O(n^{-2\beta})$ *and* $\rho_{y^2}(n)$ *is* $O(n^{2-4\beta})$.

Hosking (1996) points out, without formal proof, Theorem 2(b) for the case $\beta \in (1/2, 3/4)$. In addition, for a short-memory linear process $\rho_y(n) = O(v^n)$ and $\rho_y(n) = O(n^{1-2\beta})$ in the long-memory case, cf. Brockwell and Davis (1991, p. 520).

Observe that Theorem 2(a) states that the decaying rate of the autocorrelation function of the square of a linear short-memory time series is twice as fast as the decaying rate of the autocorrelation function of the original series. That is, $\rho_y(n)$ requires twice as many lags, n, to achieve a similar value of $\rho_{y^2}(n)$ $[\rho_y(2n) \cong \rho_{y^2}(n)]$. On the other hand, by virtue of Theorem 2(b), for a linear long-memory time series, the decaying rate of the square of the series equals the square root of the decaying rate of the original series, i.e. $\rho_y(n) \cong \rho_{y^2}(\sqrt{n})$, cf. Table I for the independent input case.

In the context of nonlinear processes, comparing expressions (4) and (10) we note the presence of an additional term for the nonlinear case which involves two quantities, $\tau(n)$ and $\Delta(n)$. As a consequence of Lemma 5 in the Appendix, $\Delta(n)$ and $\rho_y^2(n)$ have the same order, however $\tau(n)$ may have a different asymptotic order. The following two theorems establish the asymptotic behaviour of the autocorrelation function of squares for two important situations

$$\rho_{\varepsilon^2}(m) \sim a^{|m|}, \quad 0 < a < 1, \tag{12}$$

$$\rho_{\varepsilon^2}(m) \sim |m|^{-\lambda}, \quad 0 < \lambda < 1. \tag{13}$$

TABLE I

Asymptotic Behaviour of Autocorrelation Function of the Squares of the Process $y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ where ε_t has Finite Kurtosis η, $0 < a < 1$, $0 < \lambda < 1$, $|v| < 1$, $\beta \in (\frac{1}{2}, 1)$ and $\lambda + 2\beta > 2$

	Short-memory filter ($\psi_i \sim v^i$)	Long-memory filter ($\psi_i \sim i^{-\beta}$)
Linear: Independent input		
$\alpha(n)$	$O(v^{2n})$	$O(n^{-2\beta})$
$\rho_y(n)$	$O(v^n)$	$O(n^{1-2\beta})$
$\rho_{y^2}(n)$	$O(v^{2n})$	$O(n^{2-4\beta})$
Nonlinear: Short memory input: $\rho_{\varepsilon^2}(n) \sim a^n$		
$\Delta(n)$	$O(v^{2n})$	$O(n^{2-4\beta})$
$\alpha(n)$	$O(v^{2n})$	$O(n^{-2\beta})$
$\tau(n)$	$O([\max\{v^2, a\}]^n)$	$O(n^{-2\beta})$
$\rho_y(n)$	$O(v^n)$	$O(n^{1-2\beta})$
$\rho_{y^2}(n)$	$O([\max\{v^2, a\}]^n)$	$O(n^{2-4\beta})$
Nonlinear: Long memory input: $\rho_{\varepsilon^2}(n) \sim n^{-\lambda}$		
$\Delta(n)$	$O(v^{2n})$	$O(n^{2-4\beta})$
$\alpha(n)$	$O(v^{2n})$	$O(n^{-2\beta})$
$\tau(n)$	$O(n^{-\lambda})$	$O(n^{2-2\beta-\lambda})$
$\rho_y(n)$	$O(v^n)$	$O(n^{1-2\beta})$
$\rho_{y^2}(n)$	$O(n^{-\lambda})$	$O(n^{2-2\beta-\lambda})$

Note that under condition (12), the square of the input sequence $\{\varepsilon_t\}$ has short memory. On the other hand, under condition (13) this sequence has long memory. In what follows, *short-memory input* means that ε_t^2 has short memory. Besides, *long-memory input* indicates that ε_t^2 is a long-memory process.

THEOREM 3. (*Short-memory input*): *Consider a process satisfying* (1) *and assume conditions* (2) *and* (3) *with error sequence satisfying* (12) *and finite kurtosis* η. *Then*

(a) *Short-memory filter: If* $\psi_i \sim \upsilon^i$ *for some* $|\upsilon| < 1$, *then* $\rho_y(n) = O(\upsilon^n)$, $\alpha(n)$ *and* $\Delta(n)$ *are* $O(\upsilon^{2n})$, *and* $\tau(n) = O([\max\{\upsilon^2, a\}]^n)$. *Therefore,* $\rho_{y^2}(n)$ *is* $O([\max\{\upsilon^2, a\}]^n)$.

(b) *Long-memory filter: If* $\psi_i \sim i^{-\beta}$ *for some* $\beta \in (\frac{1}{2}, 1)$, *then* $\rho_y(n) = O(n^{1-2\beta})$, $\alpha(n)$ *and* $\tau(n)$ *are* $O(n^{-2\beta})$, *and* $\Delta(n)$ *is* $O(n^{2-4\beta})$. *Consequently,* $\rho_{y^2}(n)$ *is* $O(n^{2-4\beta})$.

THEOREM 4. (*Long-memory input*): *Consider a process satisfying* (1). *Under conditions* (2) *and* (3) *with error sequence satisfying* (13) *and finite kurtosis* η. *Then*

(a) *Short-memory filter: If* $\psi_i \sim \upsilon^i$ *for some* $|\upsilon| < 1$, *then* $\rho_y(n) = O(\upsilon^n)$, $\alpha(n)$ *and* $\Delta(n)$ *are* $O(\upsilon^{2n})$, *and* $\tau(n) = O(n^{-\lambda})$. *Consequently,* $\rho_{y^2}(n)$ *is* $O(n^{-\lambda})$.

(b) *Long-memory filter: If* $\psi_i \sim i^{-\beta}$ *for some* $\beta \in (\frac{1}{2}, 1)$ *and* $0 < \lambda < 1$, *then* $\rho_y(n) = O(n^{1-2\beta})$, $\alpha(n)$ *is* $O(n^{-2\beta})$, $\Delta(n)$ *is* $O(n^{2-4\beta})$, *and* $\tau(n)$ *is* $O(n^{2-\lambda-2\beta})$ *for* $\lambda + 2\beta > 2$. *Therefore,* $\rho_{y^2}(n)$ *is* $O(n^{2-\lambda-2\beta})$.

These results are summarized in Table I. From this table, the square of a linear process has an autocorrelation structure similar to the square of the autocorrelation of the original series, no matter what the distribution of the process is or the memory introduced by the Wold expansion. On the other hand, the autocorrelation function of the square of a nonlinear process depends on both the memory of the squares of the input error sequence and the structure of the Wold expansion. For instance, in the case that the squares of the input sequence $\{\varepsilon_t\}$ have short memory, if the filter presents long memory then the process $\{y_t^2\}$ has long memory. The same occurs for the nonlinear case with short-memory filter and squares of the input sequence with long memory. Besides, if both the filter and the input sequence are short memory, then $\{y_t^2\}$ has short memory. Finally, if the filter and the input sequence have long memory, then $\{y_t^2\}$ has long memory. Applications of these results to specific models are discussed in the next section.

3. APPLICATIONS

An important contribution of the theorems discussed in Section 2 is that they allow us to establish, for each particular model satisfying the Wold decomposition

(1), whether it has the ability to reproduce key features exhibited by the data or not. These characteristics include, for example, time series with very little autocorrelation but with strongly dependent squares; time series with strong dependency but weakly dependent squares; strongly dependent processes and their squares or weakly dependent processes and weakly dependent squares.

In what follows we apply the results from the previous section to the statistical analysis of some well-known linear and nonlinear time series models. We focus our attention on the class of models specified by the Wold decomposition in (1) and by the following expression for the error sequence $\{\varepsilon_t\}$:

$$\varepsilon_t = \sqrt{h_t}\epsilon_t, \tag{14}$$

where $\{\epsilon_t\}$ is a strict white noise sequence with zero mean, and unit variance, h_t is the conditional variance of ε_t given $\mathcal{F}_{t-1}$, the σ-field generated by the past information $\{\varepsilon_{t-1}, \varepsilon_{t-2},\ldots\}$. Observe that by setting h_t equal to a positive constant for all t, the class of linear processes is obtained. On the other hand, by specifying an evolution of h_t in terms of $\mathcal{F}_{t-1}$ we obtain many other nonlinear time series, including the conditional heteroskedasticity processes. These models are widely used, among other areas, for modelling financial market volatility (see for example, the surveys by Bollerslev *et al.*, 1992; Baillie, 1996; Ghysels *et al.*, 1996; Shephard, 1996). For clarity, we discuss linear and nonlinear processes separately.

3.1. *Linear processes*

A well-known class of linear strongly dependent processes is the ARFIMA model (see for example, Hosking, 1981), defined by the discrete-time equation

$$\Phi(B)(1-B)^d y_t = \Theta(B)\varepsilon_t, \tag{15}$$

where $|d| < 1/2$, $\{\varepsilon_t\}$ is a strict white noise sequence with zero mean and variance σ_ε^2, B is the backshift operator

$$By_t = y_{t-1}, \quad \Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \quad \text{and} \quad \Theta(B) = 1 + \theta_1 + \cdots + \theta_q B^q$$

are polynomials of degrees p and q, respectively, with no common zeroes and all their roots outside the unit circle, and $(1 - B)^d$ is the fractional difference operator defined by the binomial series

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{(k-d-1)!}{k!(-d-1)!} B^k.$$

ARFIMA processes have been widely used for modelling long-range dependency (see for example, Beran, 1994; Palma and Del Pino, 1999, among others).

If $d = 0$, (15) corresponds to a short-memory ARMA process (cf. Brockwell and Davis, 1991, p. 524), and exact expressions for the autocorrelations of this process, ρ_y, are given in Zinde-Walsh (1988). For $0 < d < 1/2$, (15) is a long-memory process and Sowell (1992) gives exact formulae for the autocorrelation function, ρ_y. In this case, $\psi_i \sim i^{-\beta}$ for large i, where $\beta = 1 - d$ and $\rho_y(n) = O(n^{1-2\beta})$ for large lag n.

It is well known that if $\{\varepsilon_t\}$ is Gaussian, then $\rho_{y^2}(n) = \rho_y^2(n) = O(n^{2-4\beta})$ (cf. Beran, 1994, p. 71). However, as a consequence of Theorem 2(b), $\rho_{y^2}(n) = O(n^{2-4\beta})$ for any distribution of the error sequence $\{\varepsilon_t\}$ with finite kurtosis. Hence, if $1/4 < d < 1/2$, then the squares have long memory. If $0 < d < \frac{1}{4}$, then the squares have *intermediate memory* as defined in Brockwell and Davis (1991, p. 520), i.e. they have absolutely summable autocorrelations decaying to zero at an hyperbolic rate.

A class of linear processes with uncorrelated observations but with possibly correlated squares, is the so-called all-pass models discussed by Breidt *et al.* (2001). A causal all-pass time series is an ARMA process $\{y_t\}$ satisfying the difference equation

$$\phi_0(B)y_t = \frac{B^p \phi_0(B^{-1})}{-\phi_{0r}} \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a strict white noise with continuously differentiable distribution in a neighborhood of zero and has median zero,

$$\phi_0(z) = 1 - \phi_{01}z - \cdots - \phi_{0p}z^p,$$

where $\phi_0(z) \neq 0$ for $|z| \leq 1$, $\phi_{00} = 1$, $\phi_{0r} \neq 0$ for some $r = 0, 1,\ldots, p$ and $\phi_{0j} = 0$ for $j = r + 1,\ldots, p$. One interesting property of this model is that if the error sequence $\{\varepsilon_t\}$ is Gaussian then the process $\{y_t\}$ is strict white noise and therefore its squares are also strict white noise. On the other hand, if the error sequence is non-Gaussian, then $\{y_t\}$ is a nonstrict white noise and $\{y_t^2\}$ may exhibit some degree of correlation. However, since the linear filter has exponentially decaying coefficients, $\psi_i \sim v^i$, an application of Theorem 2(a) indicates that the process $\{y_t^2\}$ has fast decaying autocorrelations and therefore it cannot display long-memory behaviour.

3.2. *Non-linear processes*

When the conditional variance h_t is no longer constant, a wide class of nonlinear processes arises. In what follows, we analyse a number of specifications for the filter $\Psi(B)$ in (1) and the evolution of h_t in (14). As described below, there are four combinations of these two elements: short- or long-memory filter and short- or long-memory input ε_t^2 which memory depends on the function h_t. Of these four combinations, only the first, i.e. short-memory filter and short-memory input produces a short-memory squared data, $\{y_t^2\}$, and all the other combinations produce long-memory output, cf. Table I. In what follows, ARMA and ARFIMA filters correspond to

$$\Psi(B) = \Theta(B)\Phi(B)^{-1} \quad \text{and} \quad \Psi(B) = \Theta(B)\Phi(B)^{-1}(1-B)^{-d},$$

respectively, where $\Theta(B)$ and $\Phi(B)$ were defined in (15).

3.2.1. *Short-memory input, short-memory filter*

One of the simplest ways to specify a short-memory filter in the context of (1) and (14) is by setting $\Psi(B) = 1$. In this case, there are several models containing squared errors with short-memory behaviour as in (12) including, for example, the ARCH(q) process introduced by Engle (1982) where

$$h_t = \omega + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}^2;$$

the GARCH(p,q) model proposed by Bollerslev (1986) and Taylor (1986) where

$$h_t = \omega + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \phi_j h_{t-j};$$

the exponential GARCH, EGARCH(p,q) model proposed by Nelson (1991) where

$$\log(h_t) = \omega + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right)\left(1 - \sum_{j=1}^{p} \phi_j B^j\right)^{-1} g(\epsilon_{t-1})$$

with $g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E|\epsilon_t|]$ and the N-component GARCH(1,1) (cf. Ding and Granger, 1996) where

$$h_t = \sum_{i=1}^{N} \omega_i h_{i,t} \quad \text{with } h_{i,t} = \sigma^2(1 - \alpha_i - \beta_i) + \alpha_i \varepsilon_{t-1}^2 + \beta_i h_{i,t-1}.$$

Another approach for modelling volatility is the stochastic variance (SV) model discussed by Harvey *et al.* (1994) where

$$\log(h_t) = \gamma \ \log(h_{t-1}) + \xi_t,$$

ϵ_t has standard normal distribution and ξ_t is a normal variable with zero mean and variance σ_ξ^2, independent of ϵ_t.

Exact expressions for the autocorrelation function of squared errors, namely ε_t^2, of some of the previously mentioned models have been derived in the literature. For instance, Karanasos (1999) and He and Teräsvirta (1999) establish exact formulae for ρ_{ε^2} in GARCH(p,q) models and study conditions for the existence of the fourth moment of the errors. All these results still hold when the distribution of ϵ_t is non-Gaussian. Similar results are found for the N-component GARCH(1, 1) model, which can be expressed as a GARCH(N, N) process, and the two-component GARCH(N, N) (see Karanasos, 1999). Furthermore, in the context of exponential GARCH models, exact expressions for ρ_{ε^2} are given by He *et al.* (2002) for the particular case EGARCH(1, 1); by Karanasos and Kim

(2001) for the EGARCH(p, q) model; and by Demos (2002) for a model that nests both EGARCH and stochastic volatility specifications.

By specifying an ARMA filter in (1) combined with a short-memory input sequence, more complex models for the data $\{y_t\}$ are produced. Thus, if $\{\varepsilon_t\}$ follows a GARCH process then the class of ARMA–GARCH models is obtained (see Weiss, 1986). Another example is the ARMA–EGARCH process studied by Karanasos and Kim (2001) and Demos (2002), obtained when $\{\varepsilon_t\}$ follows an EGARCH process.

By an application of Theorem 3(a), we observe that, under some conditions, $\{y_t^2\}$ has short memory in all the previous models. On the contrary, the processes investigated in the following three categories have long-memory output, $\{y_t^2\}$, as shown in Table I.

3.2.2. *Short-memory input, long-memory filter*

In order to incorporate long-memory in the sequence of observations of (1), Ling and Li (1997a) propose a fractionally integrated autoregressive model with conditional heteroskedasticity, ARFIMA(p, d, q)-GARCH(r, s). This is a discrete time process with $\Psi(B)$ as in (15), ϵ_t as in (14) with standard normal distribution, and

$$h_t = \omega + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j}.$$

If $|d| < 0.5$, all roots of $\Phi(B)$ and $\Theta(B)$ are outside the unit circle and

$$\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j < 1,$$

then $\{y_t\}$ is invertible, strictly stationary and ergodic. From (2), it can be shown that the autocorrelation function of ARFIMA–GARCH and ARFIMA models is the same. Since $\{\varepsilon_t^2\}$ has short memory, from Theorem 3(b) and following the discussion in Section 3.1 for ARFIMA models, we conclude that in ARFIMA–GARCH models the observations have long-memory if $0 < d < 1/2$; however, the squares of observations have intermediate-memory if $0 < d < 1/4$ and the squares have long-memory if $1/4 < d < 1/2$. Similar conclusions are reached by an application of Theorem 3(b) to ARFIMA–EGARCH models.

3.2.3. *Long-memory input, short-memory filter*

As in the previous cases, if $\Psi(B)$ in (1) corresponds to an ARMA filter, we obtain a large class of processes with short-memory filter and long-memory input. Among these models with $\Psi(B) = 1$, we find: the FIEGARCH(p, d, q) process proposed by Bollerslev and Mikkelsen (1996), where

$$(1-B)^d \log(h_t) = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \alpha(B)^{-1} g(\epsilon_{t-1}),$$

with $\alpha(B)$ satisfying

$$1 - \sum_{j=1}^{p} \phi_j B^j = \alpha(B)(1 - B)^d$$

and $g(\cdot)$ defined in Section 3.2.1; and the LMGARCH model examined in Robinson (1991), Robinson and Henry (1999) and Henry (2001), where

$$h_t = [1 - (1 - B)^d a(B) b(B)^{-1}]\varepsilon_t^2, \quad \text{with } a(B) = 1 - \sum_{j=1}^{r} a_j B^j \text{ and } b(B) = 1 + \sum_{i=1}^{s} b_i B^i.$$

For the latter model, exact expressions for ρ_{ε^2} was found by Karanasos *et al.* (2004). These models can be extended by applying an ARMA filter $\Psi(B)$ to FIEGARCH and LMGARCH errors producing the ARMA–FIEGARCH and ARMA–LMGARCH processes, respectively see for example (Bollerslev and Mikkelsen, 1996; Robinson and Henry, 1999; Henry, 2001, among others). Observe that, under conditions of Theorem 4(a), the squared observations in all aforementioned models have long memory.

Another interesting example of long-memory sequence error is the FIGARCH model discussed by Baillie *et al.* (1996) and Bollerslev and Mikkelsen (1996), where $h_t = \omega + [1 - (1 - B)^d a(B) b(B)^{-1}]\varepsilon_t^2$, with positive ω. This model is strictly stationary but not covariance stationary. However, as pointed out by Henry (2001), for this class of models the autocorrelation function of squares can be well defined even though the fourth moment $\{\varepsilon_t^4\}$ is not finite. In fact, Karanasos *et al.* (2004) have shown that, under certain conditions, ρ_{ε^2} is the same for FIGARCH and LMGARCH models.

3.2.4. *Long-memory input, long-memory filter*

This class of double long-memory models includes, for example, combinations of ARFIMA filters and conditionally heteroskedastic input with long-range dependency such as the aforementioned FIEGARCH or LMGARCH processes. As a result, the ARFIMA–FIEGARCH and ARFIMA–LMGARCH processes are obtained. In this context, suppose that the sequence $\{\varepsilon_t^2\}$ is strongly dependent with $\lambda = 1 - 2d_\varepsilon$ in (13); the linear filter in (1) has long memory, i.e. $\psi_i \sim i^{-\beta}$ with $\beta = 1 - d_y$; $0 < d_\varepsilon, d_y < 1/2$ and $d_\varepsilon + d_y < 1/2$. Then, under conditions of Theorem 4(b), $\rho_{y^2}(n)$ is $O(n^{2(d_\varepsilon + d_y) - 1})$, i.e. the squares behave like a long-memory process with parameter $d^* = d_\varepsilon + d_y < 1/2$ (see Robinson and Hidalgo, 1997 for a similar type of result in the linear regression context). Moreover, given that under some conditions the squares of FIGARCH and LMGARCH processes share the same autocorrelation function (cf. Karanasos *et al.*, 2004), the squared ARFIMA-FIGARCH also has long memory.

In order to incorporate long range dependency in the squared errors, Harvey (1998) introduces a modification of the stochastic variance models (SV) described in Section 3.2.1: the so-called long-memory stochastic volatility (LMSV) processes defined by $(1-B)^{d_\varepsilon} \log(h_t) = \xi_t$. In this case, if σ_ξ^2 is small or $\rho_{\log(h)}$ is close to one, then $\rho_{\varepsilon^2}(n) \cong \rho_{\log(h)}(n)$. Thus, the squared error sequence has long memory. In addition, if we consider a long-memory filter in (1), for example $\Psi(B) = (1 - B)^{-d_y}$, then by virtue of Theorem 4(b) we conclude that $\{y_t^2\}$ is a long-range-dependent process with long-memory parameter d^* as above.

4. CONCLUSIONS

The results discussed in this paper help to determine the asymptotic behaviour of the autocorrelation function of the squares of a wide variety of linear and nonlinear time series models. For illustration purposes, we have applied these results to some well-known processes; however, they may be used to evaluate many other models which can be expressed by the expansion (1). Furthermore, equation (4) gives an exact expression for the autocorrelation function of the squares and quantifies the departure of this from the autocorrelation function of the original process. These results still hold when the underlying distribution of the input error sequence is non-Gaussian but has finite kurtosis.

APPENDIX

Proof of Theorems

We start this section with the proof of Theorem 1 and then proceed showing five technical lemmas that are needed in the proof of Theorems 2–4.

PROOF OF THEOREM 1. Without loss of generality we consider $n \geq 1$. Let,

$$S_1 = E(y_t^2 y_{t-n}^2) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} \psi_i\psi_j\psi_l\psi_m E(\varepsilon_{t-i}\varepsilon_{t-j}\varepsilon_{t-n-l}\varepsilon_{t-n-m}). \tag{A.1}$$

From (2) and (3) we have the following situations for the terms of S_1. If $i = j$, $l = m$ and $i = n + l$ then we have

$$S_{11} = \eta\sigma^4 \sum_{l=0}^{\infty} \psi_l^2 \psi_{l+n}^2.$$

If $i = j$, $l = m$ and $i \neq n + l$ we have

$$S_{12} = \sigma^4 \sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \psi_i^2\psi_l^2[1 + (\eta - 1)\rho_{\varepsilon^2}(n + l - i)]I_{[i\neq n+l]},$$

and since $I_{[i\neq n+l]} = 1 - I_{[i=n+l]}$,

$$S_{12} = \sigma^4 \left(\sum_{i=0}^{\infty} \psi_i^2\right)^2 + \sigma^4(\eta - 1)\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \psi_i^2\psi_l^2\rho_{\varepsilon^2}(n + l - i) - \sigma^4\eta\sum_{l=0}^{\infty} \psi_l^2\psi_{l+n}^2.$$

If $i = n + m \neq j = n + l$ we have

$$S_{13} = \sigma^4 \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \psi_l \psi_m \psi_{n+l} \psi_{n+m} [1 + (\eta - 1)\rho_{\varepsilon^2}(l - m)] I_{[m \neq l]},$$

and since $I_{[m \neq l]} = 1 - I_{[m=l]}$,

$$S_{13} = \sigma^4 \left(\sum_{l=0}^{\infty} \psi_l \psi_{l+n} \right)^2 + \sigma^4 (\eta - 1) \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \psi_l \psi_m \psi_{l+n} \psi_{m+n} \rho_{\varepsilon^2}(m - l) - \sigma^4 \eta \sum_{l=0}^{\infty} \psi_l^2 \psi_{l+n}^2.$$

The same expression S_{13} is obtained for the case $i = n + l \neq j = n + m$ and the other terms of the sum S_1 are zero. Therefore (A.1) can be expressed as $S_1 = S_{11} + S_{12} + 2S_{13}$, i.e.

$$\begin{aligned} S_1 = E(y_t^2 y_{t-n}^2) = \sigma^4 \Bigg\{ & -2\eta \sum_{i=0}^{\infty} \psi_i^2 \psi_{i+n}^2 + \left(\sum_{i=0}^{\infty} \psi_i^2 \right)^2 + (\eta - 1) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i^2 \psi_j^2 \rho_{\varepsilon^2}(n + j - i) \\ & + 2 \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+n} \right)^2 + 2(\eta - 1) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \psi_{i+n} \psi_{j+n} \rho_{\varepsilon^2}(j - i) \Bigg\}. \end{aligned} \tag{A.2}$$

On the other hand, from (2) and (3),

$$S_2 = E(y_t^2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j E(\varepsilon_{t-i} \varepsilon_{t-j}) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2. \tag{A.3}$$

Observe that $\mathrm{cov}(y_t^2, y_{t-n}^2) = S_1 - S_2^2$. Hence, from (A.2) and (A.3) we obtain

$$\begin{aligned} \gamma(n) = \mathrm{cov}(y_t^2, y_{t-n}^2) = \sigma^4 \Bigg\{ & -2\eta \sum_{i=0}^{\infty} \psi_i^2 \psi_{i+n}^2 + (\eta - 1) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i^2 \psi_j^2 \rho_{\varepsilon^2}(n + j - i) \\ & + 2 \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+n} \right)^2 + 2(\eta - 1) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \psi_{i+n} \psi_{j+n} \rho_{\varepsilon^2}(j - i) \Bigg\}, \end{aligned}$$

$$\gamma(0) = \sigma^4 \left\{ -2\eta \sum_{i=0}^{\infty} \psi_i^4 + 2 \left(\sum_{i=0}^{\infty} \psi_i^2 \right)^2 + 3(\eta - 1) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i^2 \psi_j^2 \rho_{\varepsilon^2}(j - i) \right\}.$$

Since $\rho_{y^2}(n) = \gamma(n)/\gamma(0)$, and taking into account $\rho_y(n)$, $\alpha(n)$, $\Delta(n)$, $\tau(n)$, defined in (5), (6), (7), (8), respectively, we have,

$$\begin{aligned} \rho_{y^2}(n) = & \left\{ -2\eta\alpha(n) \frac{\sum_{i=0}^{\infty} \psi_i^4}{(\sum_{i=0}^{\infty} \psi_i^2)^2} + (\eta - 1)\tau(n) + 2\rho_y^2(n) + 2(\eta - 1)\Delta(n) \right\} \Bigg/ \\ & \left\{ -2\eta \frac{\sum_{i=0}^{\infty} \psi_i^4}{(\sum_{i=0}^{\infty} \psi_i^2)^2} + 2 + 3(\eta - 1)\Delta(0) \right\}. \end{aligned} \tag{A.4}$$

Furthermore, given that κ is the kurtosis of y_t, then from (A.2) with $n = 0$ and (A.3) we obtain,

$$\kappa = E(y_t^4)\{E(y_t^2)\}^{-2}$$
$$= \left\{-2\eta\sum_{i=0}^{\infty}\psi_i^4 + 3\left(\sum_{i=0}^{\infty}\psi_i^2\right)^2 + 3(\eta-1)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i^2\psi_j^2\rho_{\varepsilon^2}(j-i)\right\}\left\{\sum_{i=0}^{\infty}\psi_i^2\right\}^{-2}$$
$$= 3 - 2\eta\left\{\sum_{i=0}^{\infty}\psi_i^4\right\}\left\{\sum_{i=0}^{\infty}\psi_i^2\right\}^{-2} + 3(\eta-1)\Delta(0).$$

Hence, the denominator in (A.4) is equal to $\kappa-1$. Now, by replacing

$$-2\eta\frac{\sum_{i=0}^{\infty}\psi_i^4}{(\sum_{i=0}^{\infty}\psi_i^2)^2} = \kappa - 3 - 3(\eta-1)\Delta(0)$$

into the numerator of (A.4) and arranging terms, expression (4) is obtained. QED

LEMMA 1. *Let* $\delta \in (1, 2)$, $n \geq 1$. *Then*

$$\sum_{j=1}^{\infty}[j(j+n)]^{-\delta} = O(n^{-\delta}). \tag{A.5}$$

PROOF. Since $(1+\frac{j}{n})^{-\delta} \leq 1$ for $n, j \geq 1$ then

$$\sum_{j=1}^{\infty}[j(j+n)]^{-\delta} = n^{-\delta}\sum_{j=1}^{\infty}j^{-\delta}\left(1+\frac{j}{n}\right)^{-\delta} \leq n^{-\delta}\sum_{j=1}^{\infty}j^{-\delta}.$$

But $\sum_{j=1}^{\infty} j^{-\delta} < \infty$ for $\delta \in (1,2)$, hence the result follows. QED

LEMMA 2. *For* $a < 1$,

$$\lim_{m\to\infty} m^{2\beta}\sum_{i=1}^{m-1}a^i(m-i)^{-2\beta} = \frac{a}{1-a}.$$

PROOF. Let

$$b_m = m^{2\beta}\sum_{i=1}^{m-1}a^i(m-i)^{-2\beta} = m^{2\beta}\sum_{j=1}^{m-1}a^{m-j}j^{-2\beta}$$

then $(m+1)^{-2\beta}b_{m+1} = am^{-2\beta}b_m + am^{-2\beta}$ and $b_{m+1} = a(1+m^{-1})^{2\beta}b_m + a(1+m^{-1})^{2\beta}$.
Thus, the result follows as $m \to \infty$. QED

LEMMA 3. *Let* $\psi_i \sim i^{-\beta}$ *with* $\beta \in (\frac{1}{2}, 1)$,

(a) $\sum_{i=0}^{\infty}\psi_i\psi_{i+n} = O(n^{1-2\beta})$.
(b) $\sum_{i=0}^{\infty}\psi_i^2\psi_{i+n}^2 = O(n^{-2\beta})$.
(c) If ρ_{ε^2} *satisfies (12) then* $\sum_{i=0}^{\infty}\psi_i^2\rho_{\varepsilon^2}(m-i) = O(m^{-2\beta})$ *for* $m \geq 1$.
(d) If ρ_{ε^2} *follows (13) with* $\lambda < 1$ *and* $\lambda + 2\beta > 2$, *then* $\sum_{i=0}^{\infty}\psi_i^2\,\rho_{\varepsilon^2}(m-i) = O(m^{2-\lambda-2\beta})$ *for* $m \geq 1$.

PROOF. (a) Direct from Hosking (1981). (b) Follows directly from Lemma 1 with $\delta = 2\beta$. (c) Let

$$A_m = \sum_{i=0}^{\infty} \psi_i^2 \rho_{\varepsilon^2}(m-i).$$

Thus,

$$A_m \sim a^m + \sum_{i=1}^{\infty} i^{-2\beta} a^{|m-i|}$$

and then,

$$A_m \sim a^m + \sum_{i=1}^{m-1} a^i (m-i)^{-2\beta} + m^{-2\beta} + \sum_{i=1}^{m-1} a^i (m+i)^{-2\beta} + a^m \sum_{i=0}^{\infty} a^i (2m+i)^{-2\beta}. \quad \text{(A.6)}$$

Let

$$P_1 = \sum_{i=1}^{m-1} a^i (m-i)^{-2\beta}, \quad P_2 = \sum_{i=1}^{m-1} a^i (m+i)^{-2\beta} \quad \text{and} \quad P_3 = a^m \sum_{i=0}^{\infty} a^i (2m+i)^{-2\beta}.$$

By applying Lemma 2 to P_1 we conclude that $P_1 = O(m^{-2\beta})$. On the other hand, since

$$(m+i)^{-2\beta} \leq (m-i)^{-2\beta} \quad \text{for} \quad i \geq 1 \quad \text{then } P_2 \leq P_1 \quad \text{and} \quad P_2 = O(m^{-2\beta}).$$

Besides, given that

$$(2m)^{-\beta} \geq (2m+i)^{-\beta} \quad \text{for } i \geq 0$$

we obtain

$$P_3 \leq a^m (2m)^{-2\beta} \sum_{i=0}^{\infty} a^i$$

and since $0 < a < 1$, then $P_3 = O(m^{-2\beta})$. Finally, since $a^m = O(m^{-2\beta})$ and by taking into account the orders of P_1, P_2, P_3 in (A.6) the Lemma 3(c) follows. (d) Let

$$A_m = \sum_{i=0}^{\infty} \psi_i^2 \rho_{\varepsilon^2}(m-i).$$

Then

$$A_m \sim \rho_{\varepsilon^2}(m) + \sum_{i=1}^{\infty} i^{-2\beta} \rho_{\varepsilon^2}(m-i).$$

Thus,

$$A_m \sim \rho_{\varepsilon^2}(m) + \sum_{i=1}^{m-1} i^{-2\beta} \rho_{\varepsilon^2}(m-i) + m^{-2\beta} \rho_{\varepsilon^2}(0) + \sum_{i=m+1}^{\infty} i^{-2\beta} \rho_{\varepsilon^2}(m-i).$$

Now, by taking into account the expression (13) with $m \geq 1$

$$A_m \sim m^{-\lambda} + m^{-2\beta} + \sum_{i=1}^{m-1} i^{-2\beta} (m-i)^{-\lambda} + \sum_{i=m+1}^{\infty} i^{-2\beta} (i-m)^{-\lambda}. \quad \text{(A.7)}$$

Let

$$S = \sum_{i=1}^{m-1} i^{-2\beta}(m-i)^{-\lambda} \quad \text{and} \quad T = \sum_{i=m+1}^{\infty} i^{-2\beta}(i-m)^{-\lambda}.$$

Since $\frac{1}{m} \leq \frac{i}{m}$ for $i \geq 1$ and by Polya and Szegö (1972, p. 52–3) we have

$$S \leq m^{2-\lambda-2\beta} \sum_{i=1}^{m-1} \left(\frac{i}{m}\right)^{1-2\beta} \left(1 - \frac{i}{m}\right)^{-\lambda} \left(\frac{1}{m}\right) \sim m^{2-\lambda-2\beta} \int_0^1 x^{1-2\beta}(1-x)^{-\lambda} \mathrm{d}x.$$

This expression is equal to $m^{2-\lambda-2\beta}\mathrm{Beta}(2 - 2\beta, 1 - \lambda)$ if $\lambda < 1$ and $\beta < 1$. Thus

$$S = O(m^{2-\lambda-2\beta}) \quad \text{if } \lambda + 2\beta > 2, \ \lambda < 1 \quad \text{and} \quad \beta < 1.$$

On the other hand,

$$T = m^{1-\lambda-2\beta} \sum_{i=m+1}^{\infty} \left(\frac{i}{m}\right)^{-2\beta} \left(\frac{i}{m} - 1\right)^{-\lambda} \left(\frac{1}{m}\right).$$

Now, from Polya and Szegö (1972, p. 52–3) and with the change of variable $y = x^{-1}$ we obtain

$$T \sim m^{1-\lambda-2\beta} \int_1^{\infty} x^{-2\beta}(x-1)^{-\lambda} \mathrm{d}x = m^{1-\lambda-2\beta} \int_0^1 y^{\lambda+2\beta-2}(1-y)^{-\lambda} \mathrm{d}y$$
$$= m^{1-\lambda-2\beta}\mathrm{Beta}(\lambda + 2\beta - 1, 1 - \lambda)$$

if $\lambda < 1$ and $\lambda + 2\beta > 1$. Thus $T = O(m^{1-\lambda-2\beta})$ if $\lambda < 1$ and $\lambda + 2\beta > 1$. Then, by considering the orders of S and T in (A.7) we conclude that $A_m = O(m^{2-\lambda-2\beta})$ for $\lambda < 1$, $\beta < 1$ and $\lambda + 2\beta > 2$. QED

Lemma 4. *Let $\psi_i \sim v^i$ with $|v| < 1$ and ρ_{ε^2} satisfies* (13), *then*

$$A_m = \sum_{i=0}^{\infty} \psi_i^2 \rho_{\varepsilon^2}(m-i) = O(m^{-\lambda}) \ \textit{for } m \geq 1.$$

Proof. Since

$$A_m \sim \rho_{\varepsilon^2}(m) + \sum_{i=1}^{m-1} v^{2i}\rho_{\varepsilon^2}(m-i) + v^{2m}\rho_{\varepsilon^2}(0) + \sum_{i=m+1}^{\infty} v^{2i}\rho_{\varepsilon^2}(m-i),$$

we have

$$A_m \sim m^{-\lambda} + v^{2m} + \sum_{i=1}^{m-1} v^{2i}(m-i)^{-\lambda} + v^{2m} \sum_{i=1}^{\infty} v^{2i} i^{-\lambda}. \tag{A.8}$$

Let

$$P_1 = \sum_{i=1}^{m-1} v^{2i}(m-i)^{-\lambda} \quad \text{and} \quad P_2 = v^{2m} \sum_{i=1}^{\infty} v^{2i} i^{-\lambda}.$$

By applying Lemma 2 with $a = v^2$ we obtain $P_1 = O(m^{-\lambda})$. Now, since $i^{-\lambda} \leq 1$ for $i \geq 1$ we have

$$P_2 \leq v^{2m} \sum_{i=1}^{\infty} v^{2i} \quad \text{and} \quad P_2 = O(v^{2m}).$$

Thus, by taking into account the orders of P_1 and P_2 in (A.8) we conclude that $A_m = O(m^{-\lambda})$. QED

LEMMA 5. *$\Delta(n)$ and $\rho_y^2(n)$ have the same asymptotic order as $n \to \infty$.*

PROOF. Given that $\rho(x) \leq 1$, then from (7) we have

$$\Delta(n) \sim \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \psi_{i+n} \psi_{j+n} \rho_{\varepsilon^2}(i-j) \leq \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+n} \right)^2.$$

Besides, from (5) $\rho_y(n) \sim \sum_{i=0}^{\infty} \psi_i \psi_{i+n}$, hence the result follows. QED

PROOF OF THEOREM 2. (a) Straightforward. (b) From (10) $\rho_y^2(n) = O\{\rho_y^2(n)\} + O\{\alpha(n)\}$. Now, from Lemma 3(a)

$$\rho_y \sim \sum_{i=0}^{\infty} \psi_i \psi_{i+n} = O(n^{1-2\beta}).$$

Furthermore, from Lemma 3(b)

$$\alpha(n) \sim \sum_{i=0}^{\infty} \psi_i^2 \psi_{i+n}^2 = O(n^{-2\beta})$$

then the result follows. QED

PROOF OF THEOREM 3. (a) It is straightforward to prove that $\rho_y(n) = O(v^n)$ and $\alpha(n) = O(v^{2n})$. Besides, from Lemma 5, $\Delta(n) = O(v^{2n})$. On the other hand,

$$\tau(n) \sim \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i^2 \psi_j^2 \rho_{\varepsilon^2}(n+j-i) = \sum_{j=0}^{\infty} \psi_j^2 A_{(n+j)} \quad \text{with } A_{(n+j)} = \sum_{i=0}^{\infty} \psi_i^2 \rho_{\varepsilon^2}(n+j-i).$$

Let $m = n + j$, then

$$A_m \sim \sum_{i=0}^{\infty} v^{2i} a^{|m-i|} = S + T \quad \text{where } S = \sum_{i=0}^{m-1} v^{2i} a^{m-i}$$

and

$$T = \sum_{i=m}^{\infty} v^{2i} a^{i-m}.$$

But, since $|v^2 a| < 1$ then $T = v^{2m} \sum_{i=0}^{\infty} (v^2 a)^i = O(v^{2m})$. Now, in order to find the order of S we consider two cases. If $v^2 > a$, then

$$\frac{S}{v^{2m}} = \sum_{i=1}^{m} \left(\frac{a}{v^2}\right)^i = \frac{v^2}{v^2 - a}\left\{1 - \left(\frac{a}{v^2}\right)^{m+1}\right\} \le c,$$

where c is a finite constant. Thus $S = O(v^{2m})$ in this case. If $v^2 < a$, then

$$\frac{S}{a^m} = \sum_{i=0}^{m-1} \left(\frac{v^2}{a}\right)^i = \frac{a}{a - v^2}\left\{1 - \left(\frac{v^2}{a}\right)^m\right\} \le c,$$

where c is a finite constant. Thus, in this case $S = O(a^m)$. Therefore, S is $O([\max\{v^2, a\}]^m)$ and since $T = O(v^{2m})$ we conclude that $A_m = O([\max\{v^2, a\}]^m)$. Thus

$$\tau(n) \sim \sum_{j=0}^{\infty} v^{2j}[\max\{v^2, a\}]^{n+j}.$$

Now, if $v^2 > a$,

$$\tau(n) \sim v^{2n} \sum_{j=0}^{\infty} v^{4j} = O(v^{2n})$$

and if $v^2 < a$,

$$\tau(n) \sim a^n \sum_{j=0}^{\infty} (v^2 a)^j = O(a^n)$$

because $|v^2 a| < 1$. Then $\tau(n) = O([\max\{v^2, a\}]^n)$ as required. (b) From Lemma 3(a) we have

$$\rho_y(n) \sim \sum_{i=0}^{\infty} \psi_i \psi_{i+n} = O(n^{1-2\beta}).$$

By Lemma 3(b),

$$\alpha(n) \sim \sum_{i=0}^{\infty} \psi_i^2 \psi_{i+n}^2 = O(n^{-2\beta})$$

and from Lemma 5,

$$\Delta(n) = O(n^{2-4\beta}).$$

On the other hand,

$$\tau(n) \sim \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i^2 \psi_j^2 \rho_{\varepsilon^2}(n+j-i) \sim \sum_{j=0}^{\infty} \psi_j^2 A_{(n+j)} \quad \text{with } A_{(n+j)} = \sum_{i=0}^{\infty} \psi_i^2 \rho_{\varepsilon^2}(n+j-i).$$

Furthermore, from Lemma 3(c) with $m = n + j$ we conclude $A_{(n+j)} = O\{(n + j)^{-2\beta}\}$. Then an application of Lemma 1 with $\delta = 2\beta$ in (A.1) yields

$$\tau(n) \sim n^{-2\beta} + \sum_{j=1}^{\infty} [j(n+j)]^{-2\beta} = O(n^{-2\beta}).$$

Thus, the result follows. QED

Proof of Theorem 4. (a) Similarly to the proof of Theorem 3(a), it can be shown that $\rho_y(n) = O(v^n)$ and $\alpha(n) = O(v^{2n})$. Besides, from Lemma 5, $\Delta(n) = O(v^{2n})$. On the other hand,

$$\tau(n) \sim \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i^2\psi_j^2\rho_{\varepsilon^2}(n+j-i) = \sum_{j=0}^{\infty}\psi_j^2 A_{(n+j)} \quad \text{with } A_{(n+j)} = \sum_{i=0}^{\infty}\psi_i^2\rho_{\varepsilon^2}(n+j-i).$$

But, from Lemma 4 $A_{(n+j)} = O\{(n + j)^{-\lambda}\}$, then

$$\tau(n) \sim \sum_{j=0}^{\infty} v^{2j}(n+j)^{-\lambda} = S + T$$

where

$$S = \sum_{j=0}^{n} v^{2j}(n+j)^{-\lambda} \quad \text{and } T = v^{2n}\sum_{j=1}^{\infty} v^{2j}(2n+j)^{-\lambda}.$$

Now, since $(1 + \frac{j}{n})^{-\lambda} \leq 1$ for $j \geq 1$, $n \geq 1$, then $S \leq n^{-\lambda}\sum_{j=0}^{n} v^{2j}$ and therefore $S = O(n^{-\lambda})$. Given that $(2n + j)^{-\lambda} \leq 1$ for $j \geq 1$, $n \geq 1$ we have $T \leq v^{2n}\sum_{j=1}^{\infty} v^{2j}$ and then $T = O(v^{2n})$. Consequently, $\tau(n) = O(n^{-\lambda})$ as required. (b) From Lemmas 3(a), (b) and 5 we obtain $\rho_y(n) = O(n^{1-2\beta})$, $\alpha(n) = O(n^{-2\beta})$ and $\Delta(n) = O(n^{2-4\beta})$, respectively. Besides,

$$\tau(n) \sim \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i^2\psi_j^2\rho_{\varepsilon^2}(n+j-i) \sim \sum_{j=0}^{\infty}\psi_j^2 A_{(n+j)}$$

with

$$A_{(n+j)} = \sum_{i=0}^{\infty}\psi_i^2\rho_{\varepsilon^2}(n+j-i).$$

From Lemma 3(d) with $m = n + j$ we have $A_{(n+j)} = O\{(n + j)^{2-\lambda-2\beta}\}$ if $\lambda + 2\beta > 2$, $\lambda < 1$ and $\beta < 1$. Then

$$\tau(n) \sim \sum_{j=0}^{\infty}\psi_j^2(n+j)^{2-\lambda-2\beta} \sim n^{2-\lambda-2\beta} + \sum_{j=1}^{\infty} j^{-2\beta}(n+j)^{2-\lambda-2\beta} \leq n^{2-\lambda-2\beta}\left[1 + \sum_{j=1}^{\infty} j^{-2\beta}\right]$$

because $(1 + \frac{j}{n})^{2-\lambda-2\beta} \leq 1$. Since $2\beta > 1$, $\sum_{j=1}^{\infty} j^{-2\beta}$ is a finite constant and therefore $\tau(n) = O(n^{2-\lambda-2\beta})$ for $\lambda + 2\beta > 2$, $\lambda < 1$ and $\beta < 1$. Consequently, Theorem 4(b) follows. QED

ACKNOWLEDGEMENTS

We thank an anonymous referee for extremely helpful comments which greatly improved this paper. This research was supported in part by grant 8000004 from Fondecyt.

REFERENCES

BAILLIE, R. T. (1996) Long memory processes and fractional integration in econometrics. *J. Economet.* 73, 5–59.

BAILLIE, R. T. and CHUNG, H. (2001) Estimation of GARCH models from the autocorrelations of the squares of a process. *J. Time Ser. Anal.* 22, 631–50.

BAILLIE, R. T., BOLLERSLEV, T. and MIKKELSEN, H. (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Economet*. 74, 3–30.

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2000) *Modelling by Levy process for financial econometrics*. D.P. No 2000-W3, Nuffield College, Oxford University.

BERAN, J. (1994) *Statistics for Long-Memory Processes*. New York: Chapman and Hall.

BOLLERSLEV, T. (1986) Generalized autoregressive conditional heteroscedasticity. *J. Economet*. 31, 307–27.

BOLLERSLEV, T. (1988) On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *J. Time Ser. Anal*. 9, 121–31.

BOLLERSLEV, T. and MIKKELSEN, H. (1996) Modeling and pricing long memory in stock market volatility. *J. Economet*. 73, 151–84.

BOLLERSLEV, T., CHOU, R. Y. and KRONER, K. F. (1992) ARCH modeling in finance. *J. Economet*. 52, 5–59.

BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994) *Time Series Analysis: Forecasting and Control*. London: Prentice Hall.

BREIDT, F. J., DAVIS, R. A. and TRINDADE, A. (2001) Least absolute deviation estimation for all-pass time series models. *Ann. Stat*. 29, 919–46.

BROCKWELL, P. and DAVIS, R. (1991) *Time Series Theory and Methods* (2nd edn). New York: Springer-Verlag.

CHAN, N. H. and PALMA, W. (1998) State–space modeling of long-memory processes. *Ann. Stat*. 26, 719–40.

DEMOS, A. (2002) Moments and dynamic structure of a time-varying parameter stochastic volatility in mean model. *Econom. J*. 5, 345–57.

DING, Z. and GRANGER, C. W. J. (1996) Modeling volatility persitence of speculative returns: a new approach. *J. Economet*. 73, 185–215.

ENGLE, R. (1982) Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

GHYSELS, E., HARVEY, A. and RENAULT, E. (1996) Stochastic volatility. In *Handbook of Statistics*, Vol. 14 (eds G. S. Maddala and C. R. Rao). New York: Elsevier.

GRANGER, C. W. J. and ANDERSEN, A. P. (1978) *An Introduction to Bilinear Time Series Models*. Gottingen: Vandenhoek and Ruprecht.

GRANGER, C. W. J. and JOYEUX, R. (1980) An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal*. 1, 15–29.

HARVEY, A. C. (1998) Long memory in stochastic volatility. In *Forecasting Volatility in Financial Markets* (eds J. Knight and S. Satchell). Oxford: Butterworth-Heineman.

HARVEY, A. C., RUIZ, E. and SHEPHARD, N. (1994) Multivariate stochastic variance models. *Rev. Econ. Stud*. 61, 247–64.

HE, C. and TERÄSVIRTA, T. (1999) Fourth moment structure of the GARCH(p, q) process. *Economet. Theory* 15, 824–46.

HE, C. TERÄSVIRTA, T. and MALMSTEN, H. (2002) Moment structure of a family of first-order exponential GARCH models. *Economet. Theory* 18, 868–85.

HENRY, M. (2001) Averaged periodogram spectral estimation with long-memory conditional heteroscedasticity. *J. Time Ser. Anal*. 22, 431–59.

HOSKING, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165–76.

HOSKING, J. R. M. (1996) Asymptotic distributions of the sample mean, autocovariance and autocorrelation of long-memory time series. *J. Economet*. 73, 261–84.

KARANASOS, M. (1999) The second moment and the autocovariance function of the squared errors of the GARCH models. *J. Economet*. 90, 81–7.

KARANASOS, M. and KIM, J. (2001) *Moments of the ARMA-EGARCH model*. Department of Economics and Related Studies, University of York. Unpublished paper.

KARANASOS, M., PSARADAKIS, Z. and SOLA, M. (2004) On the autocorrelation properties of long memory GARCH processes. *J. Time Ser. Anal*. 25, 265–82.

LEWIS, P. A. W. and RAY, B. K. (1997) Modeling long-range dependence, nonlinearity, and periodic phenomena in sea surface temperatures using TSMARS. *J. Am. Stat. Assoc*. 92, 881–93.

LI, W. K. and MAK, T. K. (1994) On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *J. Time Ser. Anal*. 15, 627–36.

LING, S. and LI, W. K. (1997a) On fractionally integrated autoregressive time series models with conditionally heteroscesdaticity. *J. Am. Stat. Assoc*. 92, 1184–94.

LING, S. and LI, W. K. (1997b) Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *J. Time Ser. Anal.* 18, 447–64.

LOBATO, I. N. and SAVIN, N. E. (1998) Real and spurious long-memory properties of stock-market data. *J. Bus. Econ. Stat.* 16, 261–83.

MANTEGNA, R. N. and STANLEY, H. E. (2000) *An Introduction to Econophysics. Correlations and Complexity in Finance.* Cambridge: Cambridge University Press.

MARAVALL, A. (1983) An application of nonlinear time series forecasting. *J. Bus. Econ. Stat.* 1, 66–74.

MCLEOD, A. I. and LI, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Ser. Anal.* 4, 269–73.

MILLS, T. C. (1999) *The Econometric Modelling of Financial Time Series.* Cambridge: Cambridge University Press.

NELSON, D. B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–70.

PALMA, W. and DEL PINO, G. (1999) Statistical analysis of incomplete long range dependent data. *Biometrika* 86, 965–72.

POLYA, G. and SZEGÖ, G. (1972) *Problems and Theorems in Analysis I.* Heildelberg: Springer-Verlag.

ROBINSON, P. M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Economet.* 47, 67–84.

ROBINSON, P. M. (2001) The memory of stochastic volatility models. *J. Economet.* 101, 195–218.

ROBINSON, P. M. and HENRY, M. (1999) Long and short memory conditional heteroscedasticity in estimating the memory parameters of levels. *Economet. Theory* 15, 299–336.

ROBINSON, P. M. and HIDALGO, F. J. (1997) Time series regression with long-range dependence. *Ann. Stat.* 25, 77–104.

SHEPHARD, N. (1996) Statistical aspects of ARCH and stochastic volatility. In *Time Series Models: In Econometrics, Finance and Other Fields* (eds D. R. Cox, D. B. Hinkley and O. E. Barndorff-Nielsen). London: Chapman Hall.

SOWELL, F. (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Economet.* 53, 165–88.

TAYLOR, S. J. (1986) *Modelling Financial Time Series.* New York: Wiley.

WEISS, A. A. (1986) ARCH and bilinear time series models: comparison and combination. *J. Bus. Econ. Stat.* 4, 59–70.

ZINDE-WALSH, V. (1988) Some exact formulae for autoregressive moving average processes. *Economet. Theory* 4, 384–402.