THE MEMORY OF STOCHASTIC VOLATILITY MODELS*

by

P M Robinson
London School of Economics and Political Science

Contents:

Abstract

A valid asymptotic expansion for the covariance of functions of multivariate normal vectors is applied to approximate autovariances of time series generated by nonlinear transformation of Gaussian latent variates, and nonlinear functions of these, with special reference to long memory stochastic volatility models, serving to identify the roles played by the underlying Gaussian processes and the nonlinear transformation. Implications for simple stochastic volatility models are examined in detail, with numerical and Monte Carlo calculations, and applications to cyclic behaviour, cross-sectional and temporal aggregation, and multivariate models are discussed.

Keywords: Stochastic volatility; long memory; nonlinear functions of Gaussian processes.

JEL No.: C22

1. Introduction

A major theme of nonlinear time series analysis in finance and econometrics concerns the influence of instantaneous nonlinear transformation on measures of memory. One class of measures which has featured frequently in asymptotic theory for time series statistics is mixing numbers, which are known to be essentially invariant to such transformation (see e.g. Ibragimov and Linnik, 1971) in the sense that their rate of decay is unchanged. However, mixing numbers cannot properly be estimated from data, and some empirical evidence about measures that can be estimated prompts further theoretical investigation.

For stationary time series the measures of this type that come most immediately to mind are autocovariances (and autocorrelations, which decay at the same rate). In particular, a well-established empirical finding is that financial time series levels x_t, such as daily asset returns, are apt to exhibit little or no autocorrelation, whereas their squares x_t^2 have noticeable autocorrelation. Attempts to model this phenomenon began with the ARCH(p) model of Engle (1982), followed by its GARCH(p, q) extension (Bollerslev, 1986), and the stochastic volatility model of Taylor (1986), with numerous elaborations on these themes.

The model of Engle (1982), and the bulk of its successors, have the property that $Cov(x_t, x_{t+j}) = 0$, for all nonzero j, whereas $Cov(x_t^2, x_{t+j}^2)$ decays exponentially to zero as $j \to \infty$. Linear processes are incapable of describing this phenomenon, but, mathematically speaking, this divergence in the properties of the levels and squares is not dramatic, and relatively recent empirical studies (see e.g. Ding, Granger and Engle, 1993, Granger and Ding, 1995, 1996, Ding and Granger, 1996, Andersen and Bollerslev, 1997) suggest a degree of persistence in $Cov(x_t^2, x_{t+j}^2)$, which might be modelled in terms of a much slower rate of decay.

In fact models had already been proposed that might explain this behaviour. Robinson (1991) considered extensions of the ARCH(p) and GARCH(p, q) that might entail arbitrarily slow decay of autocorrelations of x_t^2, including long memory, where autocor-

1

relations are not summable, and Whistler (1990) applied relevant tests he developed to financial data. Ding and Granger (1996) and others have developed such models further. On the other hand, Andersen and Bollerslev (1997), Breidt, Crato and de Lima (1998), Harvey (1998) considered a long memory version of Taylor's (1986) stochastic volatility model, Robinson and Zaffaroni (1998) considered an alternative "2-shock" functional form, Robinson and Zaffaroni (1997) considered a nonlinear moving average whose squares have long memory, while Teyssiere (1998) has discussed a variety of ARCH type long memory functional forms involving various forms of nonlinearity.

At the same time, other empirical findings must be borne in mind. A body of opinion asserts that fourth moments of many financial series are infinite, in which case autocovariances of x_t^2 are not well defined. Partly in response, autocovariances of other instantaneous transformations of x_t have been studied, such as $|x_t|^\theta$, for real-valued θ (for example $\theta = 1$), so that for $\theta < 2$ the fourth moment problem is avoided. Ding, Granger and Engle (1993), Ding and Granger (1996), Granger and Ding (1995) found a tendency over a range of series for sample autocorrelations to be relatively large for specific θ, such as $\theta = 1$. Apart from the question of finiteness of moments, absolute powers $|x_t|^\theta$ are mathematically hard to handle in the context of the ARCH models of Engle (1982), Bollerslev (1986), Robinson (1991) when θ is not an even integer.

The present paper demonstrates that, for a wide variety of processes, the autocovariances at long lags of instantaneous nonlinear functions of a general type can be rigorously approximated sufficiently accurately to enable the presence or absence of long memory in the instantaneous function to be determined from the specification of the original process. As well as enabling us to thereby deal with particular, parametrically specified, processes, our results help to explain empirical findings of long memory stochastic volatility in terms of a general class of data generating processes, allowing also the practitioner freedom to choose the form of nonlinear transformation in such a way as to minimize moment conditions, if desired.

A scalar observable (possibly transformed) series y_t will be modelled as

$$y_t = f(\eta_t), \tag{1.1}$$

where η_t is a p-dimensional stationary Gaussian process and $f : R^p \to R$. A particular case of (1.1) that is of interest specifies f to have separable form, such that

$$y_t = f_1(\eta_{1t}) f_2(\eta_{2t}), \tag{1.2}$$

taking $\eta_t = (\eta_{1t}', \eta_{2t}')'$, where η_{it} is $p_i \times 1$, $i = 1, 2$, $p_1 + p_2 = p$, $Cov(\eta_{1t}, \eta_{2t}) = 0$ and $f_i : R^{p_i} \to R$, $i = 1, 2$. If x_t has form (1.2), then so (for a new f_1, f_2) does $|x_t|^\theta$, for any θ. Moreover, subject to finiteness of appropriate moments, if $E f_1(\eta_{1t}) = 0$, then $E y_t = 0$ while if also η_{1t} is an iid sequence and η_{1t} is independent of η_{2s}, $s < t$, then $Cov(y_t, y_{t+j}) = 0$, for all $j \neq 0$, so that the classical zero-autocovariance properties of asset returns $y_t = x_t$ are described, whereas when we take $y_t = |x_t|^\theta$, or its logarithm, then $E f_1(\eta_{1t}) \neq 0$ and moreover we can then get autocorrelation.

The form (1.2) does not cover ARCH-type models (because $p < \infty$) but it covers such models as

$$x_t = \eta_{1t} e^{\alpha + \beta \eta_{2t}}, \tag{1.3}$$

and

$$x_t = \eta_{1t}(\alpha + \beta \eta_{2t}), \tag{1.4}$$

with η_{1t} white noise (having zero mean), η_{2t} either depending only on η_{1s}, $s < t$ or being independent of η_{1s} for all s, t, and $y_t = x_t$ or $y_t = |x_t|^\theta$. The model (1.3) is a standard stochastic volatility model (Taylor, 1986), where Andersen and Bollerslev (1997), Breidt, Crato and de Lima (1998), Harvey (1998) took η_{2t} to have long memory. The model (1.4) is in Robinson and Zaffaroni (1997, 1998) with again η_{2t} having long memory, though they also permitted η_{2t} to be a non-Gaussian linear process. Breidt, Crato and de Lima (1998), Robinson and Zaffaroni (1997, 1998) gave exact formulae for autocovariances of x_t^2 based on (1.3), (1.4) respectively. Andersen and Bollerslev (1997) gave approximate formulae for autocovariances at long lags of certain nonlinear functions of (1.3). Our asymptotic expansion rigorously justifies and

3

refines such asymptotic formulae in a variety of settings. For parametric fitting of a particular parametric form such as (1.3) or (1.4) the autocorrelation in η_t would be parameterized, but this possibility does not concern us here, except insofar as we have to employ a parameterization in generating Monte Carlo observations.

A wide range of models is covered by (1.2), entailing both short and long memory in η_t. We impose no smoothness on f_1 and f_2 and can choose them such that only the second moment requirement

$$Ef_i^2(\eta_{it}) < \infty, \quad i = 1, 2, \tag{1.5}$$

is required, notwithstanding the Gaussianity of η_t. By way of illustration, and also partial motivation for our allowance of $p_i \geq 2$, consider

$$f_1(\eta_{1t}) = \frac{\eta_{1t}^{(1)}(p_1 - 1)^{\frac{1}{2}}}{\left\{\eta_{1t}^{(2)2} + \ldots + \eta_{1t}^{(p_1)2}\right\}^{\frac{1}{2}}}, \tag{1.6}$$

where $\eta_{1t}^{(j)}$ is the jth element of η_{1t} and the $\eta_{1t}^{(j)}$, $1 \leq j \leq p_1$, are mutually independent (see the remark after (2.1) below) but can have autocorrelation. Then (1.6) has a t_{p_1-1} distribution, having finite integer moments up to degree $p_1 - 2$ only, for example when $2 \leq p_1 \leq 5$ it has infinite fourth moment. Of course (1.6) has mean zero so it might be substituted for η_{1t} in (1.3) and (1.4) (cf. Bollerslev, 1987), whence x_t inherits its moment properties. Other random variables expressible as functions of finitely many normals are truncated and censored normals, Beta and F variates. Nonlinear functions of Gaussian processes (1.1) featured in early work on modelling of nonlinear time series (e.g. Kuznetsov, Stratonovich and Tikhonov, 1965, Hannan, 1970, Chapter 2, Hannan and Boston, 1972). They have also played a major role in the asymptotic statistical theory of long memory processes (see e.g. Rosenblatt, 1961, Taqqu, 1975, Ho and Sun, 1987, Sanchez de Naranjo, 1993). However, this literature stresses how certain nonlinear transformations of a long memory process have less memory, even short memory, whereas we, by the product form (1.2), seek to describe a reverse phenomenon, such as when levels have zero autocorrelation but nonlinear transformations have short or long memory autocorrelation.

The following section develops our expansion for the covariance between functions of two sets of normal vectors. The expansion is shown to be a proper asymptotic one as the covariance between the normal vectors tends to zero, while its form, and in particular its leading terms, are governed by the nature of the nonlinear functions involved. Approximations of autocovariances for models (1.3) and (1.4) are derived in Section 3, with numerical calculations and Monte Carlo evidence of the accuracy of the approximations. Section 4 outlines extensions to multivariate observable processes, cyclic phenomena and temporal and cross-sectional aggregation. Section 5 contains some final remarks.

2. Covariance of Nonlinear Functions of Gaussian Variates

The present section makes no reference to time series applications, and we drop t subscripts and consider the covariance between $f(\eta)$ and $g(\zeta)$, where $f : R^p \to R$ and $g : R^q \to R$, η and ζ are respectively p- and q-dimensional normal vectors, and

$$Ef^2(\eta) + Eg^2(\zeta) < \infty. \tag{2.1}$$

We take η and ζ to individually be spherical normal, that is vectors of independent standard normal variates; no generality is thereby lost, as noted by, e.g. Ho and Sun (1987), Sanchez de Naranjo (1993). We refer to the above specification as Assumption A.

Write

$$R = E(\eta\zeta')$$

where R has (k, ℓ)th element $\rho_{k\ell}$, $k = 1, ..., p$, $\ell = 1, ..., q$. Denote by $\phi(\cdot)$ the standard normal density function and by $H_j(\cdot)$ the kth Hermite polynomial, given by

$$H_j(s)\phi(s) = \frac{1}{\sqrt{2\pi}} \int_R (it)^j \phi(t)e^{-ist}dt. \tag{2.2}$$

Let c_h, $1 \le h \le p$, d_j, $1 \le j \le q$, be nonnegative integers and define

$$c = (c_1, ..., c_p)', \quad d = (d_1, ..., d_q)', \tag{2.3}$$

5

and

$$F_c = \int\limits_{R^p} f(u) \prod_{h=1}^{p} \left\{ H_{c_h}(u_h)\phi(u_h)du_h \right\}, \tag{2.4}$$

$$G_d = \int\limits_{R^q} g(u) \prod_{j=1}^{q} \left\{ H_{d_j}(v_j)\phi(v_j)dv_j \right\}, \tag{2.5}$$

taking $u = (u_1, ..., u_p)$, $v = (v_1, ..., v_q)$. Now define

$$\xi_h = \sum_{\ell=1}^{q} |\rho_{h\ell}|, \, 1 \leq h \leq p, \, \chi_j = \sum_{k=1}^{p} |\rho_{kj}|, \, 1 \leq j \leq q,$$

$$\tau = \sum_{h=1}^{p} \xi_h = \sum_{j=1}^{q} \chi_j$$

and denote by

$$m_\xi = (m_1, ..., m_p), \quad n_\chi = (n_1, ..., n_q),$$

the vectors of non-negative integers such that

$$m_\xi = \arg\max_c \left\{ \prod_{h=1}^{p} \xi_h^{c_h} \mid F_c \neq 0 \right\},$$

$$n_\chi = \arg\max_d \left\{ \prod_{j=1}^{q} \chi_j^{d_j} \mid F_d \neq 0 \right\}.$$

The m_ξ, n_χ are extensions of the Hermite rank introduced by Taqqu (1975) in scalar problems. In case $\rho_{k\ell} \equiv \rho$, say, for $1 \leq k \leq p$, $1 \leq \ell \leq q$, we write

$$m_* = m_\xi, \quad n_* = n_\chi,$$

where

$$m_\xi = \min_c \left\{ \sum_{h=1}^{p} c_h \mid F_c \neq 0 \right\}, \quad n_\chi = \min_d \left\{ \sum_{j=1}^{q} d_j \mid F_d \neq 0 \right\},$$

which do not depend on the ξ_h ($\equiv p\rho$) or χ_j ($\equiv q\rho$).

In the statement of the following theorem, and its proof, the sums are over all non-negative integers satisfying the indicated conditions, 1_s denotes the $s \times 1$ vector of 1's, and $r_{h.} = (r_{h1}, ..., r_{hq})'$, $r_{.j} = (r_{1j}, ..., r_{pj})'$.

Theorem 1 Let Assumption A hold and

$$\tau < 1. \tag{2.6}$$

Then

$$Cov(f(\eta), g(\zeta)) = \sum_{i=1}^{\infty} a_i, \tag{2.7}$$

where

$$a_i = \sum_{c_h, d_j : 1 \le h \le p, 1 \le j \le q, 1_p' c + 1_q' d = 2i} F_c G_d \left\{ \sum_{r_{k\ell} : 1 \le k \le p, 1 \le \ell \le q, 1_q' r_{h.} = c_h, 1_p' r_{.j} = d_j} \frac{\rho_{k\ell}^{r_{k\ell}}}{r_{k\ell}!} \right\}, \tag{2.8}$$

where the sum in (2.7) converges absolutely, indeed for all $i \ge 1$

$$|a_i| \le \left\{ Ef^2(\eta) Eg^2(\zeta) \right\}^{\frac{1}{2}} \min\left(\tau^i, \left\{ \prod_{h=1}^{p} \frac{\xi_h^{m_h}}{1 - \xi_h} \prod_{j=1}^{q} \frac{\chi_j^{n_j}}{1 - \chi_j} \right\}^{\frac{1}{2}} \right) \tag{2.9}$$

for all $i \ge 1$, so that for all $n \ge 1$

$$\sum_{i=n}^{\infty} |a_i| \le \left\{ Ef^2(\eta), Eg^2(\tau) \right\}^{\frac{1}{2}} \frac{\tau^n}{1 - \tau}. \tag{2.10}$$

Proof The theorem is similar to a number of others in the literature, following from work of Kendall (1941), and, in the more recent long memory literature, Taqqu (1975), but we present a brief proof. We have

$$
\begin{aligned}
Ef(\eta)g(\zeta) &= (2\pi)^{-(p+q)/2} \int_{R^{p+q}} f(u)g(v) \left\{ \int_{R^{p+q}} \exp\left(-is'u - it'v \right) \right. \\
&\quad \left. \times \exp\left(-\frac{1}{2}(s', t') \begin{bmatrix} I_p & R \\ R' & I_q \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} \right) dsdt \right\} dudv, \tag{2.11}
\end{aligned}
$$

where here s and t denote the vectors $s = (s_1, ..., s_p)'$, $t = (t_1, ..., t_q)'$. The second exponential factor can be written

$$\exp\left(-\frac{s's}{2} - \frac{t't}{2} \right) \prod_{k=1}^{p} \prod_{\ell=1}^{q} \exp(-s_k t_\ell \rho_{k\ell}),$$

in which the double product can be written

$$\prod_{k=1}^{p}\prod_{\ell=1}^{q}\sum_{r_{k\ell}=0}^{\infty}\frac{(-s_k t_\ell \rho_{k\ell})^{r_{k\ell}}}{r_{k\ell}!}$$

$$=\sum_{r_{k\ell}:1\le k\le p,1\le \ell\le q}\left\{\prod_{k=1}^{p}\prod_{\ell=1}^{q}\frac{\rho_{k\ell}^{r_{k\ell}}}{r_{k\ell}!}\right\}\prod_{h=1}^{p}(is_h)^{c_h}\prod_{j=1}^{q}(it_j)^{d_j},$$

where c_h and d_j are as in (2.8). Thus from (2.2), the factor in braces in (2.11) is

$$(2\pi)^{(p+q)/2}\sum_{r_{k\ell}:1\le k\le p,1\le \ell\le q}\left\{\prod_{k=1}^{p}\prod_{\ell=1}^{q}\frac{\rho_{k\ell}^{r_{k\ell}}}{r_{k\ell}!}\right\}$$

$$\times\prod_{h=1}^{p}\left\{H_{c_h}(u_h)\phi(u_h)\right\}\prod_{j=1}^{q}\left\{H_{d_j}(v_j)\phi(v_j)\right\}$$

so that, from (2.4) and (2.5), (2.11) is

$$\sum_{r_{k\ell}:1\le k\le p,1\le \ell\le q}\left\{\prod_{k=1}^{p}\prod_{\ell=1}^{q}\frac{\rho_{k\ell}^{r_{k\ell}}}{r_{k\ell}!}\right\}F_c G_d,$$

with c and d given in (2.3). Noting that $Ef(\eta)=F_0$, $Eg(\eta)=G_0$, where the zero subscripts here denote vectors of zeros, substraction and rearrangement produces (2.7). To prove (2.9), the factor in braces in (2.8) is bounded in absolute value by

$$\{\prod_{h=1}^{p}\sum_{r_{h\ell}:1\le \ell\le q,1_q'r_{h.}=c_h}\prod_{\ell=1}^{q}\frac{|\rho_{h\ell}|^{r_{h\ell}}}{r_{h\ell}!}\}^{\frac{1}{2}}\{\prod_{j=1}^{q}\sum_{r_{kj}:1\le k\le p,1_p'r_{.j}=d_j}\sum^{\infty}\prod_{k=1}^{p}\frac{|\rho_{kj}|^{r_{kj}}}{r_{kj}!}\}^{\frac{1}{2}}$$

$$=\left\{\prod_{k=1}^{p}\frac{\xi_h^{c_h}}{c_h!}\right\}^{\frac{1}{2}}\left\{\prod_{j=1}^{q}\frac{\chi_j^{d_j}}{d_j!}\right\}^{\frac{1}{2}}$$

by the multinomial theorem. Thus, writing

$$S=\left\{c:c_h\ge m_h,1\le h\le p\right\},\quad T=\left\{d:d_j\ge n_j,1\le j\le q\right\},$$

noting that $F_c=0$, $c\notin S$, $F_d=0$, $d\notin T$, and applying the Schwarz inequality,

$$|a_i|\ \le\ \{\sum_{c_h:1\le h\le p}\frac{F_c^2}{\prod_{h=1}^{p}c_h!}\sum_{d_j:1\le j\le q}^{\infty}\frac{G_d^2}{\prod_{j=1}^{q}d_j!}\}^{\frac{1}{2}}$$

$$\times\{\sum_{c_h,d_j:1\le h\le p,1\le j\le q,1_p'c_h+1_q'd_j=2i,c\in S,d\in T}\prod_{h=1}^{p}\xi_h^{c_h}\prod_{j=1}^{q}\chi_j^{d_j}\}^{\frac{1}{2}}.$$

The first term in braces is bounded by $Ef^2(\eta)Eg^2(\zeta) < \infty$ from (2.1), whereas the second is bounded by

$$\sum_{c_h,d_j:1\leq h\leq p,1\leq j\leq q,1'_p c_h+1'_q d_j=2i}\prod_{h=1}^{p}\xi_h^{c_h}\prod_{j=1}^{q}\chi_j^{d_j}=\tau^{2i}$$

by the multinomial theorem, and, on the other hand, also by

$$\sum_{c_h,d_j:1\leq h\leq p,1\leq j\leq q,c\in S,d\in T}\prod_{h=1}^{p}\xi_h^{c_h}\prod_{j=1}^{q}\chi_j^{d_j}\leq\prod_{h=1}^{p}\frac{\xi_h^{m_h}}{1-\xi_h}\prod_{j=1}^{q}\frac{\chi_j^{n_j}}{1-\chi_j},$$

by summation of geometric series, to prove (2.9). Then (2.10) follows immediately.

$\square$

Note that if the first element (say) of η is independent of ζ, and $\int_R f(u)\phi(u_1)du_1 = 0$ for all $(u_2,...,u_p)$, then $Cov(f(\eta),g(\zeta)) = 0$. The Theorem's bounds are then sharp, $\sum_{i=1}^{\infty}|a_i| = 0$, because it follows that $\xi_1 = 0$ and $m_1 > 0$. Note also that by the inequality between geometric and arithmetic means and the inequality $\prod_1^n(1-x_j)\geq 1-\sum_1^n x_j$, for $0\leq x_j\leq 1$, $1\leq j\leq n$, we have

$$\left\{\prod_{h=1}^{p}\left(\frac{\xi_h^{m_h}}{1-\xi_h}\right)\prod_{j=1}^{q}\left(\frac{\chi_j^{n_j}}{1-\chi_j}\right)\right\}^{\frac{1}{2}}\leq\left(\frac{\max_h m_h}{\sum m_h}\right)^{\frac{\sum m_h}{2}}\left(\frac{\max_j n_j}{\sum n_j}\right)^{\frac{\sum n_j}{2}}\frac{\tau^{\frac{1}{2}(\sum m_h+\sum n_j)}}{1-\tau}.$$

For τ small enough this provides a sharper bound for $|a_i|$ than the one of order τ^i when $i < \frac{1}{2}(\sum m_h+\sum n_j)$, but it is the latter which is eventually important and ensures validity of the expansion

$$Cov\{f(\eta),g(\zeta)\}=(1+o(1))\sum_{i=1}^{n-1}a_i$$

for all $n\geq 2$, if

$$\tau^n = o(|\rho_{k\ell}|^{n-1}), \qquad 1\leq k\leq p, 1\leq\ell\leq q,$$

as these $\rho_{k\ell}$ all tend to zero, which entails τ tending to zero and thence that (2.6) is eventually satisfied. Note that (2.6) implies $\xi_h < 1$, $1\leq h\leq p$, and $\chi_j < 1$, $1\leq j\leq q$.

We have presented the Theorem in the form (2.7), (2.8), because a_i involves powers of degree i in the $\rho_{k\ell}$, and is thus of order τ^i. In particular, denoting by $c(k)$, $d(\ell)$

9

the values of c and d such that $c_i = \delta_{ik}$, $1 \le i \le p$, $d_j = \delta_{j\ell}$, $1 \le j \le p$, where δ is the Kronecker delta, we have

$$a_1 = \sum_{k=1}^{p} \sum_{\ell=1}^{q} F_{c(k)} G_{d(\ell)} \rho_{k\ell}, \tag{2.12}$$

and denoting by $c(k,m)$, $d(\ell,n)$ the values of c and d such that $c_i = \delta_{ik} + \delta_{im}$, $1 \le i \le p$, $d_j = \delta_{j\ell} + \delta_{jn}$, $1 \le j \le q$, we have

$$
\begin{aligned}
a_2 &= \frac{1}{2} \sum_{k=1}^{p} \sum_{\ell=1}^{q} F_{c(k,k)} G_{d(\ell,\ell)} \rho_{k\ell}^2 \\
&+ \sum_{\substack{k=1 \\ (k,\ell) \ne (m,n)}}^{p} \sum_{\ell=1}^{q} \sum_{m=1}^{p} \sum_{n=1}^{q} F_{c(k,m)} G_{d(\ell,n)} \rho_{k\ell} \rho_{mn},
\end{aligned}
\tag{2.13}
$$

and so on. It is thus the question of whether the relevant $F_c G_d$ is zero or not that determines the presence or absence of $\Pi_{k=1}^{p} \Pi_{\ell=1}^{q} \rho_{k\ell}^{r_{k\ell}}$ for a particular $\{r_{k\ell}; 1 \le k \le p, 1 \le \ell \le q\}$, while in our applications it is the lowest order powers that are not thereby eliminated that tend to dominate.

3. Autocovariances for Simple Stochastic Volatility Models

We shall investigate autocovariance properties of y_t in (1.2) by applying the Theorem in the simple case $p_1 = p_2 = 1$, which is motivated by the specifications (1.3) and (1.4), though the results are not restricted to these models. To apply the Theorem we take

$$
\begin{aligned}
f &= f_1 f_2, \quad g = f_1 f_2, \\
\eta_t &= (\eta_{1t}, \eta_{2t})', \quad \zeta = (\eta_{1,t+j}, \eta_{2,t+j})' \\
\rho_{k\ell} &= Cov(\eta_{kt}, \eta_{\ell,t+j}) \stackrel{def}{=} \gamma_{k\ell}(j), \ \text{k},\ell = 1,2.
\end{aligned}
$$

We then have

$$F_c = G_c = F_{1c_1} F_{2c_2},$$

10

where we define

$$F_{ij} = \int_{\mathbb{R}} f_i(u)h_j(u)\phi(u)du.$$

For brevity, write $f_{it} = f_i(\eta_{it})$. Condition (2.1) is equivalent to

$$Ef_{1t}^2 + Ef_{2t}^2 < \infty.$$

Define $\gamma(j) = Cov(\eta_t, \eta_{t+j})$. We deduce from (2.12) and (2.13) that the two "leading" terms in the expansion of $\gamma(j)$ are

$$
\begin{aligned}
a_1 &= \gamma_{11}(j)F_{11}^2F_{20}^2 + \gamma_{12}(j)F_{11}F_{20}F_{10}F_{21} + \gamma_{22}(j)F_{10}^2F_{21}^2, \\
a_2 &= \frac{1}{2}\left\{\gamma_{11}^2(j)F_{12}^2F_{20}^2 + \gamma_{12}^2(j)F_{12}F_{20}F_{10}F_{22} + \gamma_{22}^2(j)F_{10}^2F_{22}^2\right\} \\
&\quad +2\left\{\gamma_{11}(j)\gamma_{12}(j)F_{12}F_{20}F_{11}F_{21} + \gamma_{11}(j)\gamma_{22}(j)F_{11}^2F_{21}^2 \right. \\
&\quad \left. +\gamma_{12}(j)\gamma_{22}(j)F_{11}F_{21}F_{10}F_{22}\right\}.
\end{aligned}
$$

Further, for $K < \infty$, the Theorem gives $\sum_{j=3}^{\infty}|a_i| \le K\delta(j)^3$, for $\delta(j) = |\gamma_{11}(j)| + |\gamma_{12}(j)| + |\gamma_{22}(j)| < 1$. If η_t is ergodic, so $\delta(j) \to 0$ as $j \to \infty$, then

$$\gamma(j) = a_1(1 + o(1))$$

if $\delta(j)^2 = o(|\gamma_{k\ell}(j)|)$ for $(k, \ell) = (1, 1), (1, 2), (2, 2)$; further, we have the refinement

$$\gamma(j) = (a_1 + a_2)(1 + o(1))$$

if $\delta(j)^3 = o(\gamma_{k\ell}^2(j))$ for $(k, \ell) = (1, 1), (1, 2), (2, 2)$.

Before describing special cases, we note that $Ey_t = Ef_{1t}Ef_{2t}$, which is zero if $Ef_{it} = F_{i0} = 0$ for $i = 1$ or 2 as is true for $y_t = x_t$ in (1.3) and (1.4).

Case I :

$$F_{10} = 0, \tag{3.1}$$

so $Ef_{1t} = 0$. Then

$$\gamma(j) = \gamma_{11}(j)F_{11}^2F_{20}^2(1 + o(1)), \tag{3.2}$$

so the autocorrelation in η_{1t} dominates. If

$$\gamma_{11}(j) = 0, \ j \ne 0, \tag{3.3}$$

11

we deduce exactly

$$\gamma(j) = 0, \, j \neq 0,$$

from the Theorem (because $\xi_1 = 0, m_1 > 0$) or from direct calculation. This is the familiar outcome of white noise levels, (3.1) holding for (1.3) and (1.4) if $y_t = x_t$.

Case II : (3.3) is true but (3.1) is not, so f_{1t} is white noise with non-zero mean. This is the case for many nonlinear transformations of x_t given by (1.3), (1.4). We have

$$\gamma(j) = \left\{ \gamma_{12}(j) F_{11} F_{20} F_{10} F_{21} + \gamma_{22}(j) F_{10}^2 F_{21}^2 \right\} (1 + o(1)), \text{ as } j \to \infty. \quad (3.4)$$

Suppose also that

$$\gamma_{12}(j) = 0, \, j > 0, \quad (3.5)$$

so η_{1s} and η_{2t} are independent for all s, t. This is the usual specification in (1.3), and is imposed in (1.4) by Robinson and Zaffaroni (1998), but not by Robinson and Zaffaroni (1997). Then

$$\gamma(j) = \gamma_{22}(j) F_{10}^2 F_{21}^2 (1 + o(1)), \text{ as } j \to \infty, \quad (3.6)$$

so now the autocorrelation in η_{2t} dominates.

Case III: (3.3) and

$$\eta_{2t} = \sum_{j=0}^{\infty} \alpha_j \eta_{1, t+1-j}, \quad (3.7)$$

where

$$\alpha_j \sim c j^{d-1}, \, 0 < d < \frac{1}{2}, \quad (3.8)$$

for a nonzero constant c, which will take different values in the sequel, and with '$\sim$' indicating that the ratio of left- and right-hand sides tends to 1. We do not impose (3.1) or (3.5). The specification (3.7) is just a consequence of Gaussianity, independence of η_{1t}, η_{2t}, and (3.3), but with also (3.8) it follows that

$$\gamma_{12}(j) \sim c j^{d-1}, \, \gamma_{22}(j) \sim c j^{2d-1}, \text{ as } j \to \infty, \quad (3.9)$$

the latter relation indicating that η_{2t} has long memory; for example, η_{2t} could be a fractionally integrated autoregressive moving average (FARIMA) process. Then (3.4) and (3.8) imply that we again get (3.6), and indeed

$$\gamma(j) \sim cj^{2d-1}, \text{ as } j \to \infty, \tag{3.10}$$

so y_t inherits the long memory of η_{2t}; this property was derived more heuristically for (1.3) by Andersen and Bollerslev (1997).

Case IV: as in Case III but with (3.8) replaced by

$$\alpha_j \sim -cj^{d-1}, \quad -\frac{1}{2} < d < 0, \tag{3.11}$$

$$\alpha_j - \alpha_{j+1} = O\left(\frac{|\alpha_j|}{j}\right), \tag{3.12}$$

and

$$\sum_{j=0}^{\infty} \alpha_j = 0. \tag{3.13}$$

Then for $j > 0$,

$$\gamma_{22}(j) = \sum_{i=0}^{\lfloor j/2 \rfloor} \alpha_i \alpha_{i+j} + \sum_{i=\lfloor j/2 \rfloor+1}^{\infty} \alpha_i \alpha_{i+j},$$

and the second sum is $O(j^{2d-1})$ by (3.10) and the Cauchy inequality, while the first is, by summation by parts,

$$\sum_{i=j}^{\lfloor j/2 \rfloor-1} (\alpha_{i+j} - \alpha_{i+j+1}) \sum_{h=0}^{i} \alpha_h + \alpha_{\lfloor j/2 \rfloor} \sum_{i=0}^{\lfloor j/2 \rfloor} \alpha_i$$

$$= -\sum_{i=j}^{\lfloor j/2 \rfloor-1} (\alpha_{i+j} - \alpha_{i+j+1}) \sum_{h=i+1}^{\infty} \alpha_i - \alpha_{\lfloor j/2 \rfloor} \sum_{i=\lfloor j/2 \rfloor}^{\infty} \alpha_i$$

$$= O(j^{d-1} \cdot j^d) = O(j^{2d-1}), \text{ as } j \to \infty.$$

Also, $\sum_{j=-\infty}^{\infty} \gamma_{22}(j) = 0$ from (3.13), so that the spectral density of η_{2t} is zero at frequency zero. Thus η_{2t} has negative dependence or antipersistence; the conditions (3.11) - (3.13) are again satisfied by FARIMA models, (3.12) being

a quasi-monotonicity condition (see Yong, 1974). Because also $\gamma_{12}(j) \sim -cj^{d-1}$ we deduce from (3.4) that, instead of (3.6) and (3.10),

$$\begin{aligned} \gamma(j) &= \gamma_{12}(j)F_{11}F_{20}F_{10}F_{20}(1 + o(1)), \text{ as } j \to \infty \\ &\sim cj^{d-1}. \end{aligned} \tag{3.14}$$

Case V:

$$F_{11} = 0, \tag{3.15}$$

as is true when $f_{1t} = |\eta_{1t}|^{\theta}$, as can arise from (1.3) and (1.4). There can be autocorrelation in η_{1t}, and we have

$$\begin{aligned} \gamma(j) &= \gamma_{22}(j)F_{10}^2 F_{21}^2 + \frac{1}{2} \left\{ \gamma_{11}^2(j)F_{12}^2 F_{20}^2 \right. \\ &\quad \left. + \gamma_{12}^2(j)F_{12}F_{20}F_{10}F_{22} \right\} (1 + o(1)). \end{aligned} \tag{3.16}$$

Under (3.5) or the long memory specification (3.7), (3.8), we deduce (3.6), (3.10) again, unless η_{1t} has sufficiently long memory for the $\gamma_{11}^2(j)$ term to dominate.

Case VI: (3.3) and either (3.5) or (3.7) and (3.8), as well as

$$F_{21} = 0,$$

which holds if f_2 is an even function, as in $f_{2t} = |\beta\eta_{2t}|^{\theta}$, as follows from (1.4) with $\alpha = 0$ and $y_t = |x_t|^{\theta}$. Then

$$\gamma(j) = \frac{1}{2}\gamma_{22}^2(j)F_{10}^2 F_{22}^2(1 + o(1)). \tag{3.17}$$

This is $cj^{4d-2}(1 + o(1))$ under (3.7), (3.8) so that long memory in η_{2t} produces long memory in y_t only when $\frac{1}{4} < d < \frac{1}{2}$, and even then y_t has less memory than η_{2t}. The same result for (1.4), when $\theta = 2$, was deduced by Robinson and Zaffaroni (1997, 1998) from exact formulae for $\gamma(j)$.

We can calculate the factors F_{ij} arising in such approximations as (3.2), (3.4), (3.6) (3.14), (3.16) and (3.17) in special cases. In view of the earlier discussion we consider

as "leading cases" (3.6) and (3.17), under models prompted by (1.3) and (1.4) which lead to analytic formulae.

First consider (3.6), when $y_t = |x_t|^\theta$, $\theta > 0$, under (1.3). With $\Gamma(\cdot)$ denoting the Gamma function,

$$F_{10} = \int |s|^\theta \phi(s)ds = \frac{2^{\theta/2}}{\sqrt{\pi}}\Gamma\left(\frac{\theta}{2}+\frac{1}{2}\right), \tag{3.18}$$

$$F_{21} = e^{\alpha\theta}\int se^{\beta\theta s}\phi(s)ds = e^{\alpha\theta}\beta\theta\exp\left(\frac{\beta^2\theta^2}{2}\right),$$

so that

$$F_{10}^2 F_{21}^2 = \frac{2^\theta}{\pi}e^{2\alpha\theta}\beta^2\theta^2\Gamma\left(\theta+\frac{1}{2}\right)\exp(\beta^2\theta^2).$$

We can also develop corresponding approximations for autocorrelations, as is relevant because the variance will also depend on θ, α and β. We have

$$Var(|x_t|^\theta) = \frac{2^\theta e^{2\alpha\theta}}{\sqrt{\pi}}\exp(\beta^2\theta^2)\left\{\Gamma(\theta+\frac{1}{2}) - \frac{\Gamma^2\left(\frac{\theta}{2}+\frac{1}{2}\right)\exp(\beta^2\theta^2)}{\sqrt{\pi}}\right\}.$$

After rearrangement, we deduce from (3.7) that, for all α,

$$\rho_\theta(j) \stackrel{def}{=} Corr\left(|x_t|^\theta, |x_{t+j}|^\theta\right) = C(\theta,\beta)\gamma_{22}(j)(1+o(1)), \qquad \text{as } j \to \infty, \tag{3.19}$$

where

$$C(\theta,\beta) = \frac{\frac{1}{2}\beta^2\theta^2 B\left(\frac{\theta}{2}+\frac{1}{2},\frac{\theta}{2}+\frac{1}{2}\right)}{\exp(\beta^2\theta^2)B\left(\frac{\theta}{2}+\frac{1}{2},\frac{1}{2}\right) - B\left(\frac{\theta}{2}+\frac{1}{2},\frac{\theta}{2}+\frac{1}{2}\right)},$$

$B(.,.)$ being the Beta function. Ding, Granger and Engle (1993), Ding and Granger (1996) reported empirical evidence suggesting stronger autocorrelation in $|x_t|^\theta$ when $\theta = 1$ in case of asset returns, and when $\theta = \frac{1}{4}$ in case of exchange rate series, than for other values of θ which they tried, including $\theta = 2$. Our results can only be capable of explaining such phenomena for large enough j, and, if (1.3) is a reasonable model for such data, (3.19) indicates that variation in $\rho_\theta(j)$ over θ is due solely to variation in $C(\theta,\beta)$. In Table 1 we give $C(\theta,\beta)$ for $\theta = 0.1, 0.5, 1.0, 1.5, 2.0, 4.0$, and $\beta = 0.1$, 0.2, 0.3, 0.5, 0.7. The mode of $C(\theta,\beta)$ with respect to θ varies with β, which itself is an indicator of the departure of x_t from an iid sequence. For $\beta = 0.1, 0.2$, the mode is at $\theta = 2$, for $\beta = 0.3$ at $\theta = 1.5$, for $\beta = 0.5, 0.6$ at $\theta = 1$, and for $\beta = 1$ at $\theta = 0.5$.

To illustrate (3.17), take $f_{1t} = |\eta_{1t}|^\theta$, $f_{2t} = |\eta_{2t}|^{\theta\psi}$, $\theta > 0$, $\psi > 0$, so that $y_t = |x_t|^\theta$ when

$$x_t = \eta_{1t} |\beta\eta_{2t}|^\psi, \tag{3.20}$$

or also, when $\psi = 1$,

$$x_t = \beta\eta_{1t}\eta_{2t}. \tag{3.21}$$

The model (3.21) comes from (1.4) of Robinson and Zaffaroni (1997, 1998), with $\alpha = 0$, whereas (3.20) is a partial extension of Robinson and Zaffaroni's (1997, 1998) models. We have

$$F_{22} = \frac{1}{2} \int \left(s^2 - 1\right) |s|^{\theta\psi} \phi(s)ds = \frac{2^{\frac{1}{2}\theta\psi - 1}}{\sqrt{\pi}} \theta\psi\Gamma\left(\frac{\theta\psi}{2} + \frac{1}{2}\right),$$

so that from (3.15)

$$F_{10}^2 F_{22}^2 = |\beta|^{2\theta\psi} \frac{2^{\theta+\theta\psi}}{\pi^2} \theta^2\psi^2\Gamma^2\left(\frac{\theta}{2} + \frac{1}{2}\right)\Gamma^2\left(\frac{\theta\psi}{2} + \frac{1}{2}\right).$$

We can also derive

$$Var(|x_t|^\theta) = |\beta|^{2\theta\psi} 2^{\theta+\theta\psi}\frac{\Gamma(\theta + \frac{1}{2})\Gamma(\theta\psi + \frac{1}{2})}{\pi}\left\{1 - \frac{\Gamma(\theta + \frac{1}{2})\Gamma(\theta\psi + \frac{1}{2})}{\pi}\right\},$$

whence we have, for all $\beta \neq 0$

$$\rho_\theta(j) = D(\theta, \psi)\gamma_{22}^2(j)(1 + o(1)), \qquad \text{as } j \to \infty, \tag{3.22}$$

where

$$D(\theta, \psi) = \frac{\frac{1}{4}\theta^2\psi^2 B\left(\frac{\theta}{2} + \frac{1}{2}, \frac{\theta}{2} + \frac{1}{2}\right) B\left(\frac{\theta\psi}{2} + \frac{1}{2}, \frac{\theta\psi}{2} + \frac{1}{2}\right)}{B\left(\frac{\theta}{2} + \frac{1}{2}, \frac{1}{2}\right) B\left(\frac{\theta\psi}{2} + \frac{1}{2}, \frac{1}{2}\right) - B\left(\frac{\theta}{2} + \frac{1}{2}, \frac{\theta}{2} + \frac{1}{2}\right) B\left(\frac{\theta\psi}{2} + \frac{1}{2}, \frac{\theta\psi}{2} + \frac{1}{2}\right)}.$$

The factor $D(\theta, \psi)$ is tabulated in Table 2 for the same θ as in Table 1, and $\psi = 0.1, 0.5, 1.0, 1.5, 2.0$ and 4.0. Overall, except for $\psi = 0.1$, $D(\theta, \psi)$ tends to be notably larger than $C(\theta, \beta)$, over the range of parameter values considered, though it must be remarked that $D(\theta, \psi)$ is a factor of $\gamma_{22}^2(j)$ not $\gamma_{22}(j)$, so it by no means follows that the corresponding $\rho_\theta(j)$ will have the same ordering. The shape of $D(\theta, \psi)$ is very similar to that of $C(\theta, \beta)$, with the mode in θ falling off from $\theta = 2$ to $\theta = 0.5$ as ψ, an index of nonlinearity of x_t, increases.

16

Broader classes of models than (3.20) and (3.21) are, respectively,

$$x_t = \eta_{1t} |\alpha + \beta \eta_{2t}|^\psi, \qquad (3.23)$$

and (1.4), with $\alpha \neq 0$ in each case. We stressed (3.20) and (3.21) because when $\alpha \neq 0$ we are in Case II and deduce (3.6), which is similar to the outcome for (1.3); because unlike for (1.3) the scale factor in the approximation for $\rho_\theta(j)$ for (3.23) varies with α; and because a closed-form expression for the scale factor is only available for $\alpha \neq 0$ when θ is an even integer under (1.4), and when $\theta\psi$ is an even integer under (3.23).

It is of interest to numerically compare our approximations with actual autocovariances. Only in very special cases are exact formulae for the latter available, so we employ Monte Carlo simulations. We consider $y_t = |x_t|^\theta$ with x_t given by (3.20) for $\beta = 1$ and with η_{1t} white noise and η_{2t} the simple fractionally integrated model $(1 - L)^d \eta_{2t} = \varepsilon_t$, where L is the lag operator, ε_t is white noise and $0 < d < \frac{1}{2}$ (see Adenstedt, 1974). Then $\gamma_{22}(j)$ satisfies (3.9) and we are in the setting of Case VI. For selected values of d, ψ and θ we wish to compare the approximation with the actual autocorrelations, which are calculated by simulation. For given d, the series η_{2t}, $t = 1, ..., n = 1000$, was generated by the algorithm of Davies and Harte (1987). Then the y_t, $t = 1, ..., n$, were calculated for given ψ, θ, and their sample autocorrelations at lags $j = 1, ..., m = 100$ were computed. For the given d, ψ, θ, this process was repeated $r = 5000$ times, and compared with the leading term in (3.22). We employed each combination of $d = 0.15, 0.3, 0.45$, $\psi = 1, 2$ and $\theta = 0.5, 1, 2$, that is 18 cases in all, but only plot, in Figures 1-6, the cases in which $d = 0.3$. The approximations, across all ψ, θ, appear poor for lags less than 20, but then seem satisfactory, and very good for lags greater than 45. There is little sensitivity to either the nonlinearity parameter ψ, or the transformation parameter θ.

4. Further Applications

We briefly describe how the Theorem can be used to derive informative approximations in more elaborate circumstances than those discussed so far, namely in the

context of cyclic variation, cross-sectional aggregation, temporal aggregation, and multivariate models.

(i) Cyclic Behaviour

Formulae such as (3.16) suggest that 'linear' terms, when present, might sometimes be dominated by 'quadratic' or higher order terms, depending on the autocorrelation of the various latent variates. Even when this is not the case it is important to stress that first-order approximations may only be useful for very large j, and for more moderate j can be improved by including terms of smaller order. The qualitative impact of such terms is notable in series with a cyclic component. Andersen and Bollerslev (1997) have found evidence of strong intraday periodicity in return volatility in foreign exchange and equity model markets. To model this kind of phenomenon, note that a lag-j autocovariance proportional to $r(j; \omega, d) = \cos(j\omega)j^{2d-1}$ as $j \to \infty$, for $0 < d < \frac{1}{2}, 0 < \omega < \pi$, has the long memory property of non-summability, but also oscillates, changing sign every π/ω lags; a class of parametric models with this property was studied by Gray, Zheng and Woodward (1989). Suppose y_t is given by (1.2) with η_{2t} having two elements, $\eta_{2t}^{(1)}$ with autocorrelation decaying like $r(j; 0, d_1) = j^{2d_1-1}$, and $\eta_{2t}^{(2)}$ with autocorrelation decaying like $r(j; \omega, d_2)$, for $0 < \omega < \pi$. Then if, for example, η_{1t} is white noise and independent of η_{2s} for all t, s, then from the Theorem $\gamma(j) = Cov(y_t, y_t)$ has linear terms in j^{2d_1-1} and $\cos(j\omega)j^{2d_2-1}$, as $j \to \infty$. If $d_1 > d_2$ then the second is of smaller order, but nevertheless its exclusion is liable to give a misleading picture of $\gamma(j)$. However, nonlinear modelling of cyclic phenomena by (1.2) needs more careful thought. For example, if two elements of η_t have lag-j autocovariance decaying like $r(j; \omega_1, d_1)$, $r(j; \omega_2, d_2)$, respectively, the Theorem implies in general a contribution to $\gamma(j)$ decaying like

$$\prod_{\ell=1}^{2} r(j; \omega_1, d_1) = \frac{1}{2} \left[\cos\left\{ j(\omega_1 + \omega_2) \right\} + \cos\left\{ j(\omega_1 - \omega_2) \right\} \right] j^{2d_1 + 2d_1 - 2}. \tag{1}$$

When $d_1 + d_2 > \frac{1}{2}$, a second-order approximation displays cycles at frequencies $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$, along with linear terms with cycles at ω_1 and ω_2 (when (4.1) corresponds

to "squared" terms, so $\omega_1 = \omega_2$, $d_1 = d_2$, there is a contribution to non-cyclic long memory). Inclusion of terms of even higher order complicates matters further; the absence of such terms in (1.4) might make it preferable to (1.3) in this context. As an alternative approach, Andersen and Bollerslev (1997) modelled periodicity in returns by means of deterministic weights.

(ii) Cross-Sectional Aggregation

For series that result from cross-sectional aggregation, such as stock indices, one might prefer to model the underlyng micro-series. Suppose there are m of these $\{x_{it}\}$, $i = 1, ..., m$, and the data available are

$$x_t = \sum_{i=1}^{m} x_{it}, \quad t = 1, 2, \ldots .$$

Taking $x_{it} = f_i(\eta_t^{(i)})$, where $\eta_t^{(i)}$ is a $p^{(i)} \times 1$ vector, it follows that $y_t = g(x_t)$ is of form (1.1) with $\eta_t = \left(\eta_t^{(i)'}, ..., \eta_t^{(m)'} \right)'$ if the stationary $\eta_t^{(i)}$ are spherical normal and mutually independent. In this set-up the units x_{it} are independent across i, but can be heterogeneous owing to possible variability with i of f_i and $\Gamma^{(i)}(j) = Cov(\eta_t^{(i)}, \eta_{t+j}^{(i)})$. We approximate $\gamma(j) = Cov(y_t, y_{t+j})$ by applying the Theorem with R a block-diagonal matrix with i-th diagonal block $\Gamma^{(i)}(j)$. In the event of some long memory in the $\eta_t^{(i)}$, linear terms in the $\Gamma^{(i)}(j)$ will generally dominate. Units with the strongest autocorrelation will determine asymptotic behaviour, but these may give a misleading impression of $\gamma(j)$ at moderate j, as they need not be the largest or most numerous. Notice that, unless $y_t = x_t^\theta$, for integer θ, y_t cannot be represented as a sum of terms of the form (1.2), even if individual x_{it} have such product form, except for example if one of the functions of x_{it} is constant across i.

(iii) Temporal Aggregation and Skip-Sampling

Data can be time-aggregated or skip-sampled, especially in view of the processing problems posed by the extremely long, finely-sampled financial series nowadays available. The effect of temporal aggregation on the model (1.3) in case of long memory

19

has been considered by Andersen and Bollerslev (1997), Bollerslev and Wright (1999). More generally, consider series z_t, $t = 1, 2, \ldots$. Define, for $m \geq 1$, the temporally aggregated series

$$x_t = \sum_{s=1}^{m-1} z_{t+s-1}, \quad t = 1, 2, \ldots.$$

Suppose that $z_t = g(\zeta_t)$ for an $s \times 1$ stationary Gaussian vector process ζ_t, such that $E\zeta_t = 0$, $E\zeta_t\zeta_t' = I_s$ (so that ζ_t has the same basic form as η_t in (1.1)). Denote the rank of the covariance matrix Ω of $\zeta_t^{(m)} = (\zeta_t', \ldots, \zeta_{t+m-1}')'$ by r, and define an $r \times 1$ vector $\tau = A\zeta_t^{(m)}$, where A is an $r \times ms$ matrix such that $A\Omega A' = I_r$. When $s > 1$ it is possible that $r < ms$; for example, take $s = 2$, $\zeta_t = (1 - \rho^2)^{-\frac{1}{2}}(\xi_t - \rho\xi_{t-1}, (1 - \rho^2)^{\frac{1}{2}}\xi_{t-1})'$ for a scalar Gaussian process ξ_t with zero mean, unit variance and lag-one autocorrelation ρ, so that we can write $z_t = g(\zeta_t) = g^*(\xi_t, \xi_{t-1})$ for some g^* and think of the two latent variates generating z_t as consecutive variates from the same process. Now denote by $y_t = h(x_t)$ the instantaneous function of interest of the aggregated series x_t. Thus we have

$$y_t = h\left(g(\zeta_t) + \ldots + g(\zeta_{t+m-1})\right) = f(\eta_t)$$

for some function f, so that we are back to precisely the situation of (1.1). On applying the Theorem, $\gamma(j) = Cov(y_t, y_{t+j})$ can be expanded in terms of the autocovariance matrix of ζ_t at lags $j + 1 - m, \ldots, j$ as $j \to \infty$ (with m fixed), but the rate of decay will be unaffected by the aggregation. The case of skip-sampling of a process $y_t = f(\eta_t)$, temporally aggregated or not, is trivially handled. If η_t is observed at intervals $n > 1$, we deduce from the Theorem approximations for $\gamma(nj) = Cov(y_t, y_{t+nj})$, as $nj \to \infty$; this can be interpreted for fixed j with the sampling becoming coarser ($n \to \infty$), but n is regarded as fixed in case $n = m$ with m as above, where we replace each consecutive block of m z_t by an aggregate.

(iv) Jointly Dependent Series

It is routine to extend (1.1) to the r jointly dependent processes $y_{it} = f_i(\eta_t)$, $1 \leq i \leq r$. Approximations for cross autocovariances between y_{it} and $y_{k,t+j}$, $1 \leq i, k \leq r$,

are then readily deduced from the Theorem. It is of interest, however, to view this general setup in the context of a model for underlying observables x_{it}, $i = 1 \leq i \leq s$. If $x_{it} = g_i(\eta_t)$, $1 \leq i \leq s$, and $y_{it} = g_i(x_{1t}, ..., x_{st})$ is an instantaneous function of these x_{it}, then indeed we can write $y_{it} = f_i(\eta_t)$. We may wish to consider some particular structure for the x_{it}, such as

$$x_{it} = \sum_{k=1}^{p_1} \eta_{1t}^{(k)} h_{ik}(\eta_{2t}), \quad i = 1, ..., s$$

where $\eta_{1t}^{(k)}$ is the k-th element of η_{1t}, to cover multivariate extensions of (1.3) and (1.4). Notice that in order to allow for general contemporaneous correlation in x_{it}, $\eta_{1t}^{(k)}$ for any k will in general be common to two or more x_{it}, so that processes $y_{it} = |x_{it}|^\theta$ will not have the product structure (1.2). However in one case of empirical interest (see Granger and Ding, 1996) there is a single underlying observable, $s = 1$, but two or more functions y_{it}, for example $y_{1t} = x_t$ and $y_{2t} = |x_t|^\theta$, when product structure in x_t implies the same for the y_{it}.

5. Final Comments

As much of our discussion indicates, cases when we can derive analytic formulae for scale factors of terms in our expansions are the exception rather than the rule. In simple models such as (1.4), or (3.23) with $\alpha \neq 0$ and $y_t = |x_t|^\theta$, the typical absence of analytic functions is not a major disadvantage as numerical integration is entirely feasible. However, this may not be the case when multidimensional integrals are involved, and in any case the ability to contemplate either analytic or numerical integration presupposes a rather precise specification of the observable process. Our results are still useful in their ability to explain how long memory can arise in a variety of circumstances, and to elucidate the role of particular properties of the functional form (determining whether various terms in the expansion are eliminated). Further, they provide some justification for 'semiparametric' statistical inference on long memory. Asymptotic relations like (3.10) form the basis of estimates of d based on long lags (see Robinson, 1994), and are equivalent under a mild additional condition to asymptotic power law behaviour for the spectral density near frequency zero,

which for the basis of estimates based on low-frequency periodograms (see Geweke and Porter-Hudak, 1983, Robinson, 1995). It is also possible to obtain an approximation for the spectral density of a nonlinear function of Gaussian processes by applying the techniques of Hannan (1970, pp.82-88) to the leading terms of our autocovariance expansion.

References

Adenstedt, R., 1974, On large-sample estimation for the mean of a stationary random sequence. Annals of Statistics 2, 1095-1107.

Andersen, T.G. and T. Bollerslev, 1997, Heterogeneous information arrivals of return volatility dynamics: increasing the long run in high frequency returns. Journal of Finance 52, 975-1006.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 302-327.

Bollerslev, T., 1987, A conditional heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69, 542-547.

Bollerslev, T. and J. Wright, 1997, Semiparametric estimation of long-memory volatility dependencies: the role of high frequency data. Journal of Econometrics, forthcoming.

Breidt, F.J., Crato, N. and P. de Lima, P., 1998, On the detection and estimation of long memory in stochastic volatility. Journal of Econometrics 73, 325-34.

Davies, R.B. and D.S. Harte, 1987, Tests for Hurst effect. Biometrika 74, 95-101.

Ding, Z., and C.W.J. Granger, 1996, Modelling volatility persistence of speculative returns: a new approach. Journal of Econometrics 73, 185-215.

Ding, Z., C.W.J. Granger and R.F. Engle, 1993, A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

Engle, R.F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdon inflation. Econometrica 50, 987-1007.

Geweke, J. and S. Porter-Hudak, 1985, The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221-238.

Granger, C.W.J. and Z. Ding, 1995, Some properties of absolute returns, an alternative measure of risk. Annales d'Economie et de Statistique 40, 67-91.

Granger, C.W.J. and Z. Ding, 1996, Varieties of long memory models. Journal of Econometrics 73, 61-77.

Gray, H.L., N.F. Zheng and W.A. Woodward, 1989, On generalized fractional processes. Journal of Time Series Analysis 10, 15-29.

Hannan, E.J., 1970, Multiple time series. Wiley, New York.

Hannan, E.J. and R.C. Boston, 1972, The estimation of a nonlinear system. In: Optimization (R.S. Andersson, L.S. Jennings and D.M. Ryan, eds.), University of Queensland Press, St. Lucia, 69-85.

Harvey, A.C., 1998, Long memory in stochastic volatility. In: Forecasting volatility in financial markets (J. Knight and S. Satchell, eds.), Butterworth-Heinemann, London.

Ho, H.-C. and T.-C. Sun, 1987, A central limit theorem for non-instantaneous filters of a stationary Gaussian process. Journal of Multivariate Analysis 22, 144-155.

Ibragimov, I.A. and Yu.V. Linnik, 1971, Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen.

Kendall, M.G., 1941, Proof of relations connected with tetrachoric series and its generalization. Biometrika 32, 196-198.

Kuznetsov, P.I., R.L. Stratonovich and V.I. Tikhonov, 1965, Nonlinear transformations of stochastic processes. Pergamon Press, New York.

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

Robinson, P.M., 1994, Time series with strong dependence. In: Advances in econometrics, vol.1 (C.A. Sims, ed.) Cambridge University Press, Cambridge, 47-95.

Robinson, P.M., 1995, Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048-1072.

Robinson, P.M. and P. Zaffaroni, 1997, Modelling nonlinearity and long memory in time series. Fields Institute Communications 11, 161-170.

Robinson, P.M. and P. Zaffaroni, 1998, Nonlinear time series with long memory: a model for stochastic volatility. Journal of Statistical Planning and Inference 68, 359-371.

Rosenblatt, M., 1961, Independence and dependence. In: Proceedings of the 4th Berkeley symposium on mathematical statistical and probability, University of California Press, Berkeley, pp.411-443.

Sanchez de Naranjo, M.V., 1993, Non-central limit theorems for non-linear functionals of k Gaussian fields. Journal of Multivariate Analysis 44, 227-258.

Taqqu, M., 1975, Weak convergence to fractional Brownian motion and to the Rosenblatt process. Zeitschrift für Wahrscheinlichkeitstheorie 31, 287-302.

Taylor, S.J., 1986, Modelling financial time series. Chichester, UK.

Teyssiere, G., 1998, Nonlinear and semiparametric long-memory ARCH. Preprint.

Whistler, D.E.N., 1990, Semiparametric models of daily and intra-daily exchange rate volatility. Ph.D. thesis, University of London.

Yong, C.H., 1974, Asymptotic behaviour of trigonometric series. Chinese University of Hong Kong, Hong Kong.

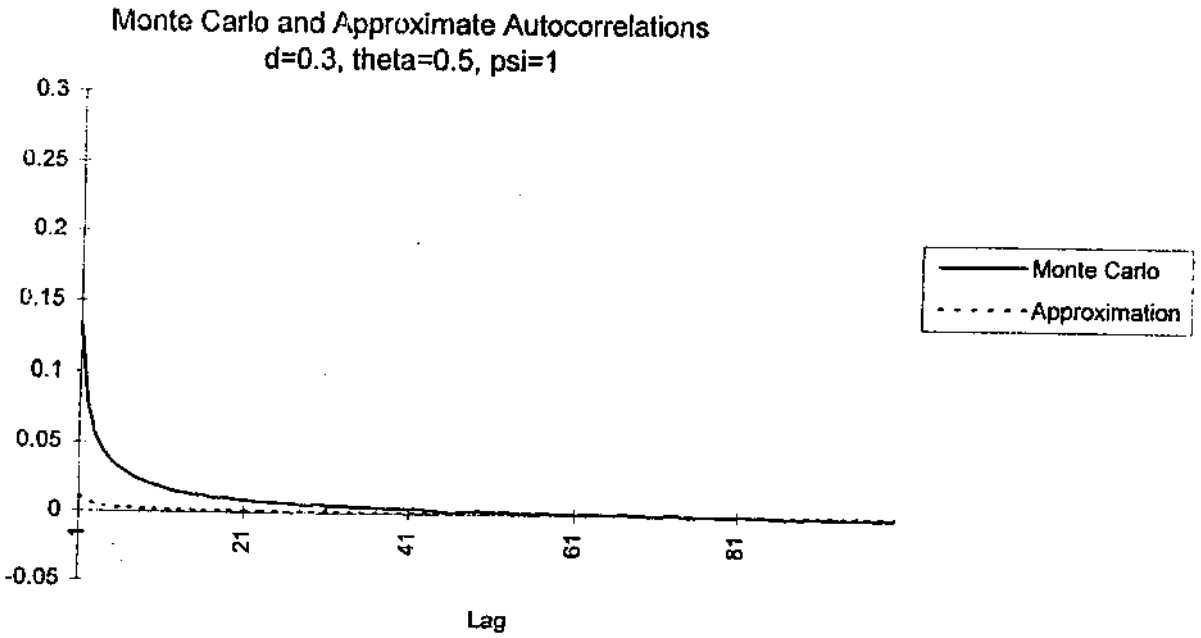Table 1: $C(\theta, \beta)$

		0.1	0.2	0.3	0.5	0.7	1.0
	0.1	.00068	.00277	.00680	.01679	.03180	.06057
	0.5	.00300	.01168	.02520	.06188	.10234	.15321
θ	1.0	.00490	.01849	.03787	.07972	.10819	.11270
	1.5	.00594	.02148	.04110	.07095	.07332	.04354
	2.0	.00632	.02166	.03803	.05064	.03586	.00920
	4.0	.00450	.01091	.01082	.00229	.00010	.00000

The β header spans across the column values (0.1, 0.2, 0.3, 0.5, 0.7, 1.0).

Table 2: $D(\theta, \psi)$

		ψ					
		0.1	0.5	1.0	1.5	2.0	4.0
	0.1	.00032	.00508	.01569	.02973	.04326	.09657
	0.5	.00135	.02292	.06250	.10202	.13673	.21474
θ	1.0	.00219	.03418	.08333	.12079	.14286	.12903
	1.5	.00263	.03474	.08034	.10162	.10404	.05000
	2.0	.00278	.03572	.06667	.07258	.06349	.01564
	4.0	.00199	.01587	.01569	.00880	.00391	.00006

Figure 1

Monte Carlo and Approximate Autocorrelations
d=0.3, theta=0.5, psi=1

Figure 2

Monte Carlo and Approximate Autocorrelations
d=0.3, theta=0.5, psi=2

Figure 3

Monte Carlo and Approximate Autocorrelations
d=0.3, theta=1, psi=1

Figure 4
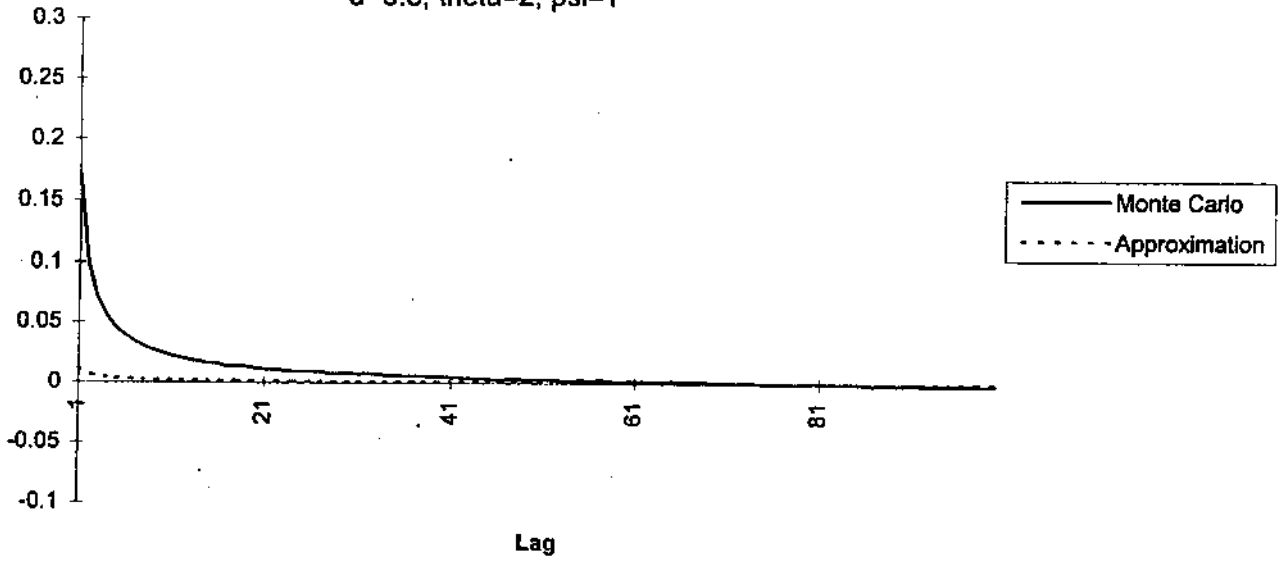

Monte Carlo and Approximate Autocorrelations
d=0.3, theta=1, psi=2
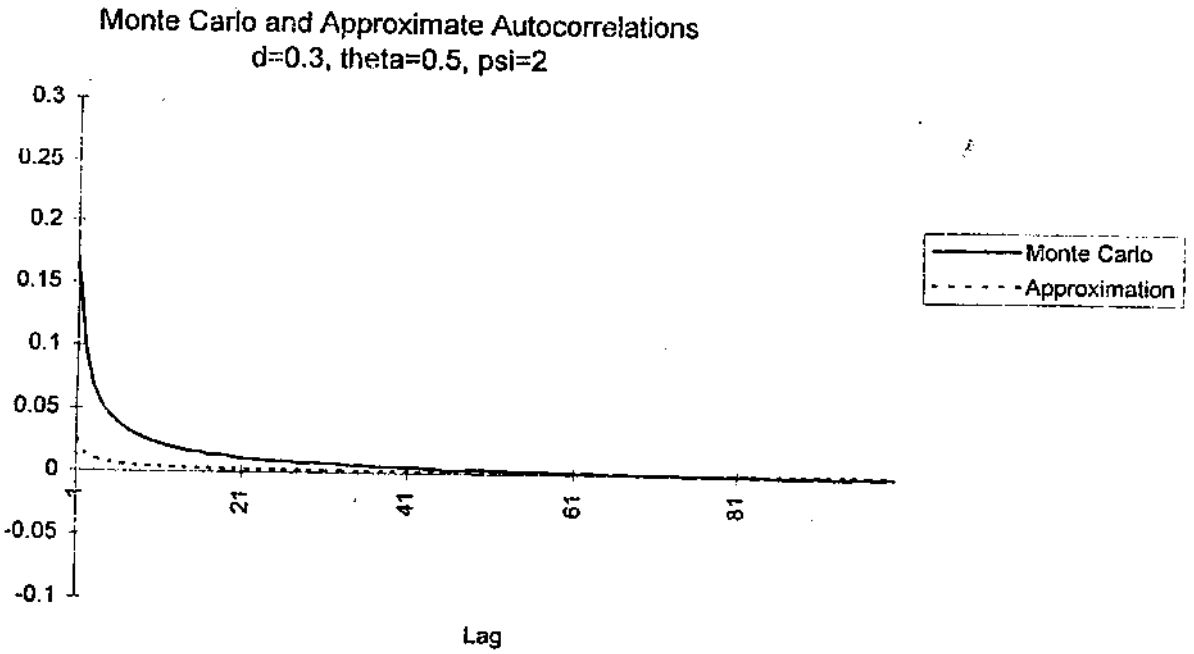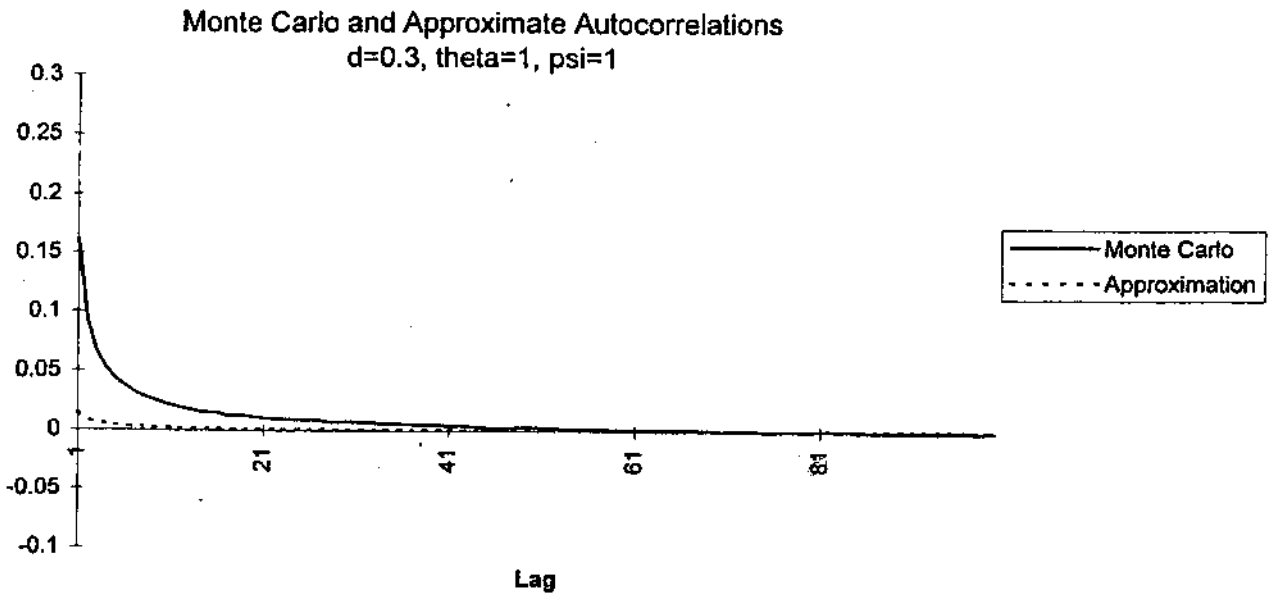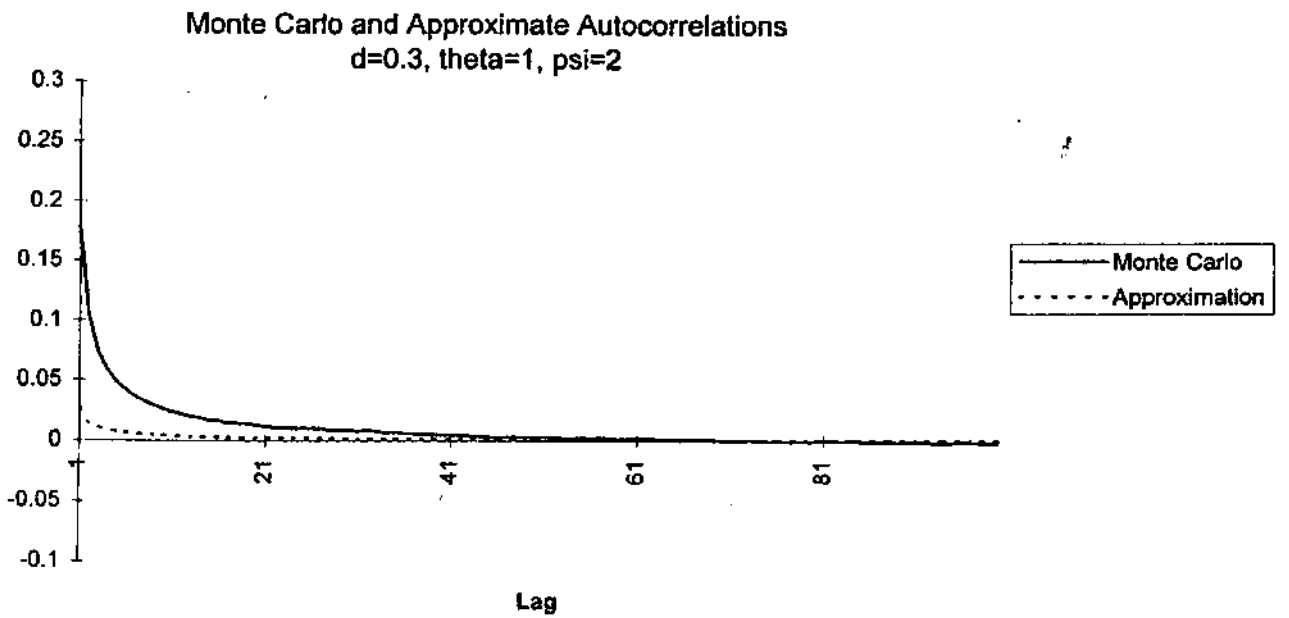
Figure 5

Monte Carlo and Approximate Autocorrelations
d=0.3, theta=2, psi=1

Figure 6

Monte Carlo and Approximate Autocorrelations
d=0.3, theta=2, psi=2