PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 31, Fasc. 1 (2011), pp. 47–59

NONLINEARITY OF ARCH AND STOCHASTIC VOLATILITY MODELS AND BARTLETT'S FORMULA

BY

PIOTR S. KOKOSZKA* (LOGAN) AND DIMITRIS N. POLITIS** (SAN DIEGO)

Abstract. We review some notions of linearity of time series and show that ARCH or stochastic volatility (SV) processes are not only non-linear: they are not even weakly linear, i.e., they do not even have a martingale representation. Consequently, the use of Bartlett's formula is unwarranted in the context of data typically modeled as ARCH or SV processes such as financial returns. More surprisingly, we show that even the squares of an ARCH or SV process are not weakly linear. Finally, we discuss an alternative estimator for the variance of sample autocorrelations that is applicable (and consistent) in the context of financial returns data.

2000 AMS Mathematics Subject Classification: Primary: 62M10; Secondary: 60G42.

Key words and phrases: ARCH processes, GARCH processes, linear time series, stochastic volatility.

1. INTRODUCTION

In the theory and practice of time series analysis, an often used assumption is that a time series $\{X_t, t \in \mathbf{Z}\}$ of interest is *linear* [17], i.e., that

$$X_t = a + \sum_{i=-\infty}^{\infty} \alpha_i \xi_{t-i}, \quad \text{where } \xi_t \sim \text{ i.i.d. } (0,1), \tag{1.1}$$

which means the ξ_ts are independent and identically distributed with mean zero and variance one.[1]

Recall that a linear time series $\{X_t\}$ is called *causal* if $\alpha_k = 0$ for $k < 0$, that is, if

$$X_t = a + \sum_{i=0}^{\infty} \alpha_i \xi_{t-i}, \quad \text{where } \xi_t \sim \text{ i.i.d. } (0,1). \tag{1.2}$$

* Partially supported by NSF grants DMS-0804165 and DMS-0931948.

** Partially supported by NSF grant DMS-0706732.

[1] When writing an infinite sum as in (1.1), it will be tacitly assumed throughout the paper that the coefficients α_i are (at least) square-summable, i.e., that $\sum_i \alpha_i^2 < \infty$.

Equation (1.2) should not be confused with the Wold decomposition that *all* purely nondeterministic time series have [5]. In the Wold decomposition the 'error' series $\{\xi_t\}$ is only assumed to be a white noise, i.e., uncorrelated, and not i.i.d. A weaker form of (1.2) amounts to relaxing the i.i.d. assumption on the errors to the assumption of a martingale difference, i.e., to assume that

$$X_t = a + \sum_{i=0}^{\infty} \alpha_i \nu_{t-i}, \tag{1.3}$$

where $\{\nu_t\}$ is a stationary martingale difference adapted to $\mathcal{F}_t$, the σ-field generated by $\{X_s, s \leqslant t\}$, i.e., that

$$E[\nu_t|\mathcal{F}_{t-1}] = 0 \quad \text{and} \quad E[\nu_t^2|\mathcal{F}_{t-1}] = 1 \quad \text{for all } t. \tag{1.4}$$

For conciseness, we propose to use the term *weakly linear* for a time series $\{X_t, t \in \mathbf{Z}\}$ that satisfies (1.3) and (1.4).

Many asymptotic theorems in the literature have been proven under the assumption of linearity or weak linearity [5]. In the last thirty years, however, there has been a surge of research activity on *non*linear time series models. One of the first such examples is the family of bilinear models [16], that is a subclass of the family of ARCH/GARCH models introduced in the 1980s (see [3] and [10]) to model financial returns. A popular alternative to ARCH/GARCH models is the family of stochastic volatility models introduced by Taylor [24] around the same time.

An important result shown to hold under weak linearity, [12], [17], is the celebrated Bartlett's formula for the asymptotic variance of the sample autocorrelations, which has recently been modified in [11] for processes (1.3) whose innovations satisfy $E\nu_{t_1}\nu_{t_2}\nu_{t_3}\nu_{t_4} = 0$ if $t_1 \neq t_2, t_1 \neq t_3, t_1 \neq t_4$. Very early on, Granger and Andersen [16] warned against the use of Bartlett's formula in the context of bilinear time series; Diebold [9] repeated the same warning for ARCH data. Despite additional such warnings ([2], [19], [20], [22]), even to this day practitioners often give undue credence to the Bartlett $\pm 1.96/\sqrt{n}$ bands – that many software programs automatically overlay on the correlogram – in the context of financial returns data. A possible reason for such misuse of Bartlett's formula is its abovementioned validity for weakly linear series. However, as will be apparent from the main developments of this paper, ARCH/GARCH processes and their squares are not only non-linear: they are not even weakly linear. The intuition that these models are somehow not linear is not new, but precise results stating assumptions under which it holds have not been formulated. One of the goals of this note is to formulate such general results. It is however a fairly common intuition that the squares of ARCH models can be treated as linear, for example, the popular ARMA representation of the squares of GARCH processes is often taken to imply that results proven for weakly linear time series extend to such models. We show that this is not the case and, in particular, that Bartlett's cannot be used. We first illustrate these issues with the following motivating example which is continued in Section 3.

EXAMPLE 1. Consider a simple ARCH(1) process, i.e., let

$$X_t = \varepsilon_t \sqrt{\beta_0 + \beta_1 X_{t-1}^2}, \tag{1.5}$$

where $\varepsilon_t \sim$ i.i.d. $(0, 1)$. If $\beta_1 < 1$, then $EX_t^2 = \beta_0/(1-\beta_1)$. Write $Y_t = X_t^2$ and let $\mathcal{F}_{t-1}$ be the σ-field generated by $Y_{t-1}, Y_{t-2}, \ldots$ Let $\sigma_t^2 = \beta_0 + \beta_1 LY_t$ be the volatility function, where L denotes the lag-operator, i.e., $LY_t = Y_{t-1}$. Since $\beta_1 LY_t = \sigma_t^2 - \beta_0$, it follows that

$$\begin{aligned}(1-\beta_1 L)(Y_t - EY_t) &= Y_t - \beta_1 LY_t - (1-\beta_1)EY_t \\ &= \sigma_t^2\varepsilon_t^2 - (\sigma_t^2 - \beta_0) - (1-\beta_1)\frac{\beta_0}{1-\beta_1} = \sigma_t^2\varepsilon_t^2 - \sigma_t^2.\end{aligned}$$

Setting $\nu_t = \sigma_t^2(\varepsilon_t^2 - 1)$, by the above calculation we obtain

$$(1-\beta_1 L)(Y_t - EY_t) = \nu_t, \tag{1.6}$$

and hence

$$Y_t = \frac{\beta_0}{1-\beta_1} + \sum_{i=0}^{\infty} \beta_1^i \nu_{t-i}. \tag{1.7}$$

In view of the fact that the innovations ν_t consitute a *white noise* [15], equation (1.6) is simply the recursive equation of an AR(1) model with nonzero mean. In this light, equation (1.7) is the usual MA representation of an AR(1) process, thereby giving an allusion toward linearity.

Nevertheless, this allusion is false: linearity does not hold true here, not even in its weak form; this is a consequence of the fact that the innovations ν_t do not have a constant *conditional* variance as required in the martingale representations (1.3) and (1.4). To see this, just note that

$$E[\nu_t^2|\mathcal{F}_{t-1}] = E[\sigma_t^4(\varepsilon_t^2-1)^2|\mathcal{F}_{t-1}] = \sigma_t^4 E[(\varepsilon_t^2-1)^2] = \sigma_t^4 \mathrm{Var}[\varepsilon_t^2].$$

The above simple example shows that the common intuition that the squares of an ARCH process are weakly linear is inaccurate. We will show in Section 2 that neither the general ARCH(∞) nor stochastic volatility (SV) models are weakly linear; more surprisingly, we show that this negative result also extends to their squares. As a consequence, using Bartlett's formula on a correlogram of financial returns *or* their squares is unjustified, as made clear in Section 3 where an alternative estimator for the variance of sample autocorrelations for ARCH/GARCH *or* SV processes is also discussed. The proposed estimator is consistent, and *model-free*; as such, it may become a useful tool to practitioners working with financial returns data.

2. ARCH AND SV PROCESSES AND THEIR SQUARES ARE NOT WEAKLY LINEAR

Consider a time series $\{X_t, t \in \mathbf{Z}\}$ that obeys the often-used model:

$$X_t = \sigma_t \varepsilon_t, \quad \text{where } \varepsilon_t \sim \text{ i.i.d. } (0,1). \tag{2.1}$$

Two general models of interest can be put in this type of framework:

• $ARCH(\infty)$ models where

$$\sigma_t^2 = \beta_0 + \sum_{j=1}^{\infty} \beta_j X_{t-j}^2; \tag{2.2}$$

this class includes all ARCH(p) and (invertible) GARCH(p,q) models.

• *Stochastic volatility* models where $L_t := \log \sigma_t$ satisfies the independent AR(p) equation

$$L_t = \phi_0 + \sum_{j=1}^{p} \phi_j L_{t-j} + u_t, \tag{2.3}$$

where $u_t \sim$ i.i.d. $(0, \tau^2)$ and $\{u_t, t \in \mathbf{Z}\}$ is independent of $\{\varepsilon_t, t \in \mathbf{Z}\}$.

We introduce the following conditions:

(ia) For each t, σ_t is $\mathcal{F}_{t-1}$-measurable and square integrable.

(ib) The sequences $\{\sigma_t\}$ and $\{\varepsilon_t\}$ are independent, and σ_t is square integrable for each t.

(iia) There is t for which σ_t^2 is not equal to a constant.

(iib) There is t for which $E[\sigma_t^2 | \mathcal{F}_{t-1}]$ is not equal to a constant.

(iii) For each t, ε_t is independent of $\mathcal{F}_{t-1}$, and it is square integrable with $E\varepsilon_t = 0, E\varepsilon_t^2 > 0$.

DEFINITION 2.1. We say that $\{X_t, t \in \mathbf{Z}\}$ is an *ARCH-type process* if (2.1) holds together with conditions (ia), (iia) and (iii). Similarly, $\{X_t, t \in \mathbf{Z}\}$ is called an *SV-type process* if (2.1) holds together with conditions (ib), (iib) and (iii).

The processes satisfying Definition 2.1 form a broader class than the usual ARCH and SV processes. We selected only the assumptions needed to establish the results of this paper.

PROPOSITION 2.1. *If $\{X_t\}$ is either an ARCH-type or an SV-type process, then it is not weakly linear.*

Proof. Suppose *ad absurdum* that $\{X_t\}$ is weakly linear. Since by (2.1) we have $EX_t = 0$, the constant term a in (1.3) must be zero, and the representation would have to be

$$X_t = \sum_{i=0}^{\infty} \alpha_i \nu_{t-i}. \tag{2.4}$$

The square summability of the α_j implies that $X_t = \lim_{m\to\infty} \sum_{i=1}^m \alpha_i \nu_{t-i}$ in L^2. Since the conditional expectation of an L^2 random variable with respect to a σ-field $\mathcal{F}$ coincides with the orthogonal projection on $L^2(\mathcal{F})$, and this projection is a continuous operator in L^2, we have $E[X_t|\mathcal{F}_{t-1}] = \sum_{i=0}^\infty \alpha_i E[\nu_{t-i}|\mathcal{F}_{t-1}]$. Since $E[\nu_t|\mathcal{F}_{t-1}] = 0$, we further obtain

$$E[X_t|\mathcal{F}_{t-1}] = \sum_{i=1}^\infty \alpha_i \nu_{t-i}. \tag{2.5}$$

We will now show that for both ARCH-type and SV-type processes

$$E[X_t|\mathcal{F}_{t-1}] = 0. \tag{2.6}$$

If $\{X_t\}$ is of ARCH-type, then $E[X_t|\mathcal{F}_{t-1}] = E[\sigma_t\varepsilon_t|\mathcal{F}_{t-1}] = \sigma_t E\varepsilon_t = 0$. Similarly, if $\{X_t\}$ is of SV-type, then

$$E[X_t|\mathcal{F}_{t-1}] = E\{[\sigma_t\varepsilon_t|\sigma(\sigma_t, \mathcal{F}_{t-1})]|\mathcal{F}_{t-1}\} = E\varepsilon_t E[\sigma_t|\mathcal{F}_{t-1}] = 0.$$

By (2.4)–(2.6), $X_t = \alpha_0\nu_t$; thus, by (1.4),

$$E[X_t^2|\mathcal{F}_{t-1}] = E[\alpha_0^2\nu_t^2|\mathcal{F}_{t-1}] = \alpha_0^2. \tag{2.7}$$

For an ARCH-type process, we obtain, on the other hand,

$$E[X_t^2|\mathcal{F}_{t-1}] = E[\sigma_t^2\varepsilon_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 E\varepsilon_t^2. \tag{2.8}$$

Equations (2.7) and (2.8) imply that, for each t, $\sigma_t^2 E\varepsilon_t^2 = \alpha_0^2$. Since $E\varepsilon_t^2 > 0$, this contradicts assumption (iia) of Definition 2.1.

Similarly, for an SV-type process, $E[X_t^2|\mathcal{F}_{t-1}] = E\varepsilon_t^2 E[\sigma_t^2|\mathcal{F}_{t-1}]$, and we obtain a contradiction with condition (iib). ∎

We next turn to the squares $Y_t = X_t^2$ of ARCH and SV processes. We put

$$\mathcal{F}_{t-1}^X = \sigma\{X_{t-1}, X_{t-2}, \ldots\}, \quad \mathcal{F}_{t-1}^Y = \sigma\{Y_{t-1}, Y_{t-2}, \ldots\}.$$

Giraitis et al. [13] showed that if (2.1) and (2.2) hold, and

$$(E\varepsilon_0^4)^{1/2} \sum_{j=1}^\infty \beta_j < \infty,$$

then the series Y_t admits the representation

$$Y_t = a + \sum_{i=0}^\infty \alpha_i \nu_{t-i} \tag{2.9}$$

in which the ν_t are martingale differences in the sense that $E[\nu_t|\mathcal{F}^Y_{t-1}] = 0$ and $E\nu_t^2 =: v^2$ is a finite constant. Nevertheless, the conditional variance

$$E[\nu_t^2|\mathcal{F}^Y_{t-1}] = \sigma_t^4 \mathrm{Var}[\varepsilon_0^2]$$

is *not* constant. Proposition 2.2 below shows that in general the squares of ARCH and SV processes are not weakly linear because they do not admit representation (2.9) with innovations ν_t having nonzero *constant* conditional variance. First, consider two technical assumptions to be used in connection with a series $\{X_t\}$ obeying model (2.1), and its squares $Y_t = X_t^2$.

ASSUMPTION 2.1 (ARCH). *Each σ_t^2 is $\mathcal{F}^Y_{t-1}$-measurable, ε_t^2 is independent of $\mathcal{F}^Y_{t-1}$, and $\{\sigma_t^2\}$ is not a.s. constant as a function of t, i.e., the event $\{\sigma_t^2 = \sigma_k^2$ for all $t, k\}$ has probability zero.*

ASSUMPTION 2.2 (SV). *The sequences $\{\sigma_t^2\}$ and $\{\varepsilon_t^2\}$ are independent, and the following two conditions hold:*

(i) *Y_t is not $\mathcal{F}^Y_{t-1}$-measurable.*

(ii) *The function of t, $v_t^2 := E[\sigma_t^4|\mathcal{F}^Y_{t-1}] - (E[\sigma_t^2|\mathcal{F}^Y_{t-1}])^2$, is not equal to a positive constant.*

Conditions (i) and (ii) automatically hold for ARCH models. Indeed, if Y_t were $\mathcal{F}^Y_{t-1}$-measurable, then $\varepsilon_t^2 = \sigma_t^{-2} Y_t$ would be also $\mathcal{F}^Y_{t-1}$-measurable, and so $E[\varepsilon_t^2|\mathcal{F}^Y_{t-1}] = \varepsilon_t^2$. Since, in the ARCH case, ε_t^2 is independent of $\mathcal{F}^Y_{t-1}$, we also have $E[\varepsilon_t^2|\mathcal{F}^Y_{t-1}] = E\varepsilon_t^2$, implying $\varepsilon_t^2 = E\varepsilon_t^2$. Thus, unless ε_t^2 is a.s. constant, condition (i) holds in the ARCH case. Condition (ii) holds in the ARCH case because the $\mathcal{F}^Y_{t-1}$-measurability of σ_t^2 implies that $v_t^2 = \sigma_t^4 - \sigma_t^4 = 0$.

Condition (i) practically always holds in SV models because $Y_t = \sigma_t^2 \varepsilon_t^2$ need not be a function of $\sigma_{t-1}^2 \varepsilon_{t-1}^2, \sigma_{t-2}^2 \varepsilon_{t-2}^2, \ldots$ Because of condition (i), v_t^2 is in general a random variable, not a constant, so (ii) also practically always holds in the SV case.

PROPOSITION 2.2. *Suppose $Y_t = X_t^2$, where X_t follows equation* (2.1), *and either Assumption* 1 *or* 2 *is satisfied. Then Y_t is not weakly linear.*

P r o o f. To lighten the notation we put $\mathcal{F}_{t-1} = \mathcal{F}^Y_{t-1}$ and suppose

$$Y_t = a + \sum_{i=0}^{\infty} \alpha_i \nu_{t-i}, \quad \sum_{i=0}^{\infty} \alpha_i^2 < \infty \tag{2.10}$$

and (1.4) holds. Conditioning on $\mathcal{F}_{t-1}$, we obtain

$$E[Y_t|\mathcal{F}_{t-1}] = a + \sum_{i=1}^{\infty} \alpha_i \nu_{t-i}. \tag{2.11}$$

Subtracting (2.11) from (2.9), we thus obtain $Y_t - E[Y_t|\mathcal{F}_{t-1}] = \alpha_0 \nu_t$.

If $\alpha_0 = 0$, then $Y_t = E[Y_t|\mathcal{F}_{t-1}]$ would be $\mathcal{F}_{t-1}$-measurable, which would contradict assumption (i). Thus for each t

$$\nu_t = \alpha_0^{-1}\{Y_t - E[Y_t|\mathcal{F}_{t-1}]\}. \tag{2.12}$$

The proof will then be complete if we show that the sequence $\{Y_t - E[Y_t|\mathcal{F}_{t-1}]\}$ does not have a constant positive conditional variance. Let

$$\xi_t = \varepsilon_t^2, \quad \lambda = E\xi_t, \quad \rho_t = \sigma_t^2.$$

Under Assumption 1 (ARCH case), $E[Y_t|\mathcal{F}_{t-1}] = E[\xi_t\rho_t|\mathcal{F}_{t-1}] = \lambda\rho_t$, and so $E\{(Y_t - E[Y_t|\mathcal{F}_{t-1}])^2|\mathcal{F}_{t-1}\} = E[(\xi_t\rho_t - \lambda\rho_t)^2|\mathcal{F}_{t-1}] = \rho_t^2\mathrm{Var}[\xi_o]$, which is not a constant sequence.

Under Assumption 2 (SV case),

$$E[Y_t|\mathcal{F}_{t-1}] = E\big[\xi_t\rho_t|\mathcal{F}_{t-1}\big] = E\big[E[\xi_t\rho_t|\sigma\{\rho_t, \mathcal{F}_{t-1}\}]|\mathcal{F}_{t-1}\big] = \lambda E[\rho_t|\mathcal{F}_{t-1}].$$

Therefore,

$$\begin{aligned}
&E\{(Y_t - E[Y_t|\mathcal{F}_{t-1}])^2|\mathcal{F}_{t-1}\} = E\{(\xi_t\rho_t - \lambda E[\rho_t|\mathcal{F}_{t-1}])^2|\mathcal{F}_{t-1}\}\\
&= E\{\xi_t^2\rho_t^2|\mathcal{F}_{t-1}\} - 2\lambda E\{\xi_t\rho_t E[\rho_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-1}\} + \lambda^2\{(E[\rho_t|\mathcal{F}_{t-1}])^2|\mathcal{F}_{t-1}\}\\
&= \lambda^2 E\{\rho_t^2|\mathcal{F}_{t-1}\} - 2\lambda^2(E[\rho_t|\mathcal{F}_{t-1}])^2 + \lambda^2(E[\rho_t|\mathcal{F}_{t-1}])^2,
\end{aligned}$$

and so we obtain $E[\nu_t^2|\mathcal{F}_{t-1}] = \alpha_0^{-2}\lambda^2\left\{E\{\rho_t^2|\mathcal{F}_{t-1}\} - (E[\rho_t|\mathcal{F}_{t-1}])^2\right\}$. By condition (ii), the ν_t are not martingale differences with constant conditional variance, and the proof is complete. ■

3. BARTLETT'S FORMULA DOES NOT WORK FOR ARCH AND SV MODELS: AN ALTERNATIVE ESTIMATOR

The results of the previous section show, in particular, that using Bartlett's formula to approximate the variance of sample autocorrelations from financial returns modeled as ARCH/GARCH or SV processes is unjustified. More surprisingly,[2] the same is true for the squared returns since – as shown above – these too are not weakly linear.

Therefore, the only valid use of Bartlett's formula on a correlogram of squared returns is under the null hypothesis of conditional homoscedasticity, i.e., for testing the hypothesis that $\beta_j = 0$ for $j > 0$ in the ARCH model or that $\phi_j = 0$ for $j > 0$ in the SV model. However, it should be stressed that Bartlett's formula *cannot* be

[2]For example, in the context of bilinear series, Granger and Andersen [16] recommended using Bartlett's formula on the correlogram of the squared data. Nevertheless, a bilinear model of order one is a special case of an ARCH(1) model with $b = 0$ and $\beta_j = 0$ for $j > 1$ in equation (2.2), and therefore falls under the premises of Proposition 2.2.

used to test a hypothesis other than the above, and it certainly *cannot* be used to construct a confidence interval for the autocorrelation of squared returns since a confidence interval must be valid for all possible values of the parameters (not just under a trivial null hypothesis). We now address these issues in greater detail.

For a second order stationary sequence $\{X_t\}$, define the population and sample autocovariances at lag k by

$$R_k = \mathrm{Cov}(X_1, X_{1+k}) \quad \text{and} \quad \hat{R}_k = n^{-1} \sum_{i=1}^{n-k} (X_i - \bar{X}_n)(X_{i+k} - \bar{X}_n),$$

respectively, where $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Similarly, we define the corresponding autocorrelations $\rho_k = R_0^{-1} R_k$ and $\hat{\rho}_k = \hat{R}_0^{-1} \hat{R}_k$, as well as the p-dimensional vectors $\boldsymbol{\rho} = [\rho_1, \rho_2, \ldots, \rho_p]^T$ and $\hat{\boldsymbol{\rho}} = [\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_p]^T$.

If $\{X_t\}$ is a linear process (1.1) with i.i.d. innovations ξ_t having finite fourth moment, then [5]:

$$\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) \xrightarrow{d} N(\mathbf{0}, \mathbf{W}) \quad \text{as } n \to \infty. \tag{3.1}$$

The linearity also implies that the entries of the $p \times p$ matrix $\mathbf{W}$ are given by Bartlett's formula [1]:

$$w_{ij} = \sum_{k=-\infty}^{\infty} [\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} + 2\rho_i\rho_j\rho_k^2 - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i}].$$

However, as we have seen in Section 2, the ARCH and SV processes typically used to model financial returns are not even weakly linear, so Bartlett's formula cannot be expected to hold with such data. To illustrate, note that for all white noise data, i.e., when $\rho_k = 0$ for all $k \geqslant 1$, Bartlett's formula implies $\mathrm{var}(\sqrt{n}\hat{\rho}_1) \to 1$. This convergence, however, breaks down with ARCH data despite the fact that they are uncorrelated.

EXAMPLE 1 (CONTINUED). Suppose $\{X_t\}$ is the ARCH(1) process (1.5), and assume for simplicity that $\varepsilon_t \sim$ i.i.d. $N(0,1)$, and that $\beta_1 < 1/\sqrt{3}$. Then, $\{X_t\}$ is a strictly stationary sequence with finite $(4+\delta)$th moment with

$$EX_1^2 = \frac{\beta_0}{1-\beta_1}, \quad EX_1^4 = \frac{3\beta_0^2(1+\beta_1)}{(1-\beta_1)(1-3\beta_1^2)}, \quad R_0 = \frac{\beta_0}{1-\beta_1}; \tag{3.2}$$

see Section 3 of [3]. By Theorem 2.1 of [22], $\sqrt{n}\hat{\rho}_1 \xrightarrow{d} N(0, \tau_1^2)$, where[3]

$$\tau_1^2 = R_0^{-2} \big[\mathrm{Var}(X_1X_2) + 2\sum_{i=1}^{\infty} \mathrm{Cov}(X_1X_2,\ X_{1+i}X_{2+i})\big]. \tag{3.3}$$

[3]The mixing condition in Theorem 2.1 of [22] holds since ARCH(1) processes are even β-mixing with exponential rate; see e.g. [6].

But

$$\begin{aligned}\mathrm{Var}(X_1X_2) &= E[X_1^2X_2^2] = E[X_1^2(\beta_0+\beta_1\sigma_1^2\varepsilon_1^2)\varepsilon_2^2]\\ &= E[X_1^2(\beta_0+\beta_1\sigma_1^2\varepsilon_1^2)] = \beta_0E[X_1^2]+\beta_1E[X_1^4]\\ &= \beta_0\frac{\beta_0}{1-\beta_1}+\beta_1\frac{3\beta_0^2(1+\beta_1)}{(1-\beta_1)(1-3\beta_1^2)} = \frac{\beta_0^2(1+3\beta_1)}{(1-\beta_1)(1-3\beta_1^2)}.\end{aligned}$$

Since

$$\mathrm{Cov}(X_1X_{1+k},\ X_{1+i}X_{1+k+i}) = 0 \quad \text{for all } i \geqslant 1,\ k \neq 0, \tag{3.4}$$

it follows that

$$\tau_1^2 = R_0^{-2}\mathrm{Var}(X_1X_2) = (1-\beta_1)(1+3\beta_1)/(1-3\beta_1^2). \tag{3.5}$$

Note that τ_1^2 increases monotonically from 1 to ∞ as β_1 increases from 0 to $1/\sqrt{3}$; not surprisingly, $\tau_1^2 \to \infty$ as $EX_1^4 \to \infty$.

Similarly, by Theorem 3.2 of [22], the asymptotic variance of $\sqrt{n}\hat{\rho}_k$ equals

$$\tau_k^2 = R_0^{-2}[c_{k+1,k+1} - 2\rho_k c_{1,k+1} + \rho_k^2 c_{1,1}],$$

where $c_{i+1,j+1} = \sum_{d=-\infty}^{\infty}\mathrm{Cov}(X_0X_i, X_dX_{d+j})$. Since $\rho_k = 0$ and $E[X_dX_{d+k}] = 0$, we have

$$\begin{aligned}\tau_k^2 &= R_0^{-2}E[X_0^2X_k^2]\\ &\quad + R_0^{-2}\sum_{d\geqslant 1}E[X_0X_kX_{-d}X_{-d+k}] + R_0^{-2}\sum_{d\geqslant 1}E[X_0X_kX_dX_{d+k}].\end{aligned}$$

In the first sum ε_k is independent of the other variables, and in the second sum ε_{d+k} is independent of the other variables. Thus, the two infinite sums above vanish, and $\tau_k^2 = R_0^{-2}E[X_0^2X_k^2]$ as before. Now, by induction,

$$\begin{aligned}E[X_0^2X_k^2] &= \beta_0(1+\beta_1+\ldots+\beta_1^{k-1})EX_0^2 + \beta_1^kEX_0^4\\ &= \beta_0\frac{1-\beta_1^k}{1-\beta_1}EX_0^2 + \beta_1^kEX_0^4.\end{aligned}$$

Finally, using (3.2), we obtain the general expression

$$\begin{aligned}\tau_k^2 &= \left(\frac{1-\beta_1}{\beta_0}\right)^2\left[\beta_0\frac{1-\beta_1^k}{1-\beta_1}\frac{\beta_0}{1-\beta_1} + \beta_1^k\frac{3\beta_0^2(1+\beta_1)}{(1-\beta_1)(1-3\beta_1^2)}\right]\\ &= (1-\beta_1)\left[\frac{1-\beta_1^k}{1-\beta_1} + \beta_1^k\frac{3(1+\beta_1)}{1-3\beta_1^2}\right].\end{aligned} \tag{3.6}$$

The calculation of τ_k^2 in an ARCH(1) model was previously given in Berlinet and Francq [2] in the case $k = 1$. For general ARCH(∞) processes (2.2), expressing τ_k^2 in terms of the coefficients β_j in closed form is difficult and not necessarily useful. However, the salient properties established in the above example do carry over to the general case, as the following proposition shows.

PROPOSITION 3.1. *Suppose* $\{X_t\}$ *is the ARCH-type process of Definition* 2.1 *with exponentially decaying* α*-mixing coefficients and finite* $(4+\delta)$*th moment. Then convergence* (3.1) *holds with*

$$w_{ij} = \delta_{ij} \frac{E[X_1^2 X_{1+i}^2]}{(EX_1^2)^2}, \tag{3.7}$$

where δ_{ij} *is the Kronecker delta.*

Moreover, if $\{X_t\}$ *admits representation* (2.2), *then*

(a) $w_{ii} \geqslant 1$, *and* $w_{ii} > 1$ *if* $\mathrm{var}[\varepsilon_1^2] > 1$;

(b) *if* $\beta_i > 0$, *then* $w_{ii} \to \infty$ *as* $EX_1^4 \to \infty$.

P r o o f. Since (3.4) holds for any ARCH-type process, formula (3.7) follows directly from Theorems 3.1 and 3.2 of [22].

We now prove the statements for the $\{X_t\}$ admitting representation (2.2).

(a) Observe that $E[X_1^2 X_{1+i}^2] - (EX_1^2)^2 = \mathrm{Cov}(X_1^2, X_{1+i}^2)$. By Lemma 2.1 of [13], $\mathrm{Cov}(X_1^2, X_{1+i}^2) \geqslant 0$, so $w_{ii} \geqslant 1$. By formula (2.11) of [13], $w_{ii} > 1$ if $\mathrm{var}[\varepsilon_1^2] > 1$.

(b) Let $\lambda = E\varepsilon_1^2$ and note that

$$\begin{aligned} E[X_1^2 X_{1+i}^2] &= \lambda E\big[X_1^2 \big(\beta_0 + \sum_{j=1}^{\infty} \beta_j X_{1+i-j}^2\big)\big] \\ &= \lambda \beta_0 E X_1^2 + \lambda \sum_{j=1}^{\infty} \beta_j E[X_1^2 X_{1+i-j}^2] \geqslant \lambda \beta_i E[X_1^4]. \end{aligned}$$

Thus,

$$w_{ii} \geqslant \lambda \beta_i \frac{EX_1^4}{(EX_1^2)^2}. \quad \blacksquare$$

All ARCH and GARCH models used in practice have exponentially decaying α-mixing coefficients; see Sections 3 and 4 of [6]. Convergence (3.1) for GARCH processes with finite fourth moment also follows from Theorem 1 of [26] where the so-called physical dependence measure decays exponentially fast; see Section 5 of [23].

PROPOSITION 3.2. *Suppose* $\{X_t\}$ *is the SV-type process of Definition* 2.1 *with exponentially decaying* α*-mixing coefficients and finite* $(4+\delta)$*th moment. Then convergence* (3.1) *and formula* (3.7) *also hold.*

P r o o f. This follows by direct application of Theorems 3.1 and 3.2 of [22]. $\blacksquare$

Note that if an SV process is of the form $X_t = \exp(L_t)\varepsilon_t$ with L_t defined by (2.3), and the errors u_t have a density, then $\{L_t\}$ is α-mixing with exponential rate; see Section 6 of [4]. Since multiplying by an i.i.d. sequence ε_t does not affect mixing, $\{X_t\}$ is then also α-mixing with exponential rate.

Propositions 3.1 and 3.2 suggest a simple method-of-moments (MOM) estimator of the asymptotic variance of $\sqrt{n}\hat{\rho}_i$ for ARCH or SV processes, namely

$$\frac{1}{n-i}\sum_{d=1}^{n-i} X_d^2 X_{d+i}^2 \Big/ \left(\frac{1}{n}\sum_{d=1}^{n} X_d^2\right)^2.$$

More generally, consider the estimator

$$\hat{w}_{ii} = nc_{in}\frac{\sum_{d=1}^{n-i} X_d^2 X_{d+i}^2}{\left(\sum_{d=1}^{n} X_d^2\right)^2}, \tag{3.8}$$

where c_{in} denotes a (deterministic) correction factor that is asymptotically equal to one. The case $c_{in} = n/(n-i)$ corresponds to the above-mentioned MOM estimator, while the case $c_{in} = 1$ corresponds to an estimator proposed by Taylor [25] using a totally different – and rather ingenious – line of arguments based on symmetry. Estimator (3.8) is a general alternative to Bartlett's formula in the case of data that are either ARCH/GARCH or SV processes with mean zero and finite fourth moments. For ease of reference, we state its consistency properties as our final proposition.

PROPOSITION 3.3. *Suppose that, for any fixed* $i \geqslant 1$, $c_{in} \to 1$ *as* $n \to \infty$.

(i) *If* $\{X_t\}$ *is a strictly stationary and ergodic sequence with* $EX_t^4 < \infty$, *then* $\hat{w}_{ii} \xrightarrow{a.s.} w_{ii}$, *where* w_{ii} *was defined in equation* (3.7).

(ii) *Under the conditions of Proposition* 3.1 *or those of Proposition* 3.2, $\hat{w}_{ii}$ *is an a.s. consistent estimator of the variance of the limit distribution of* $\sqrt{n}\hat{\rho}_i$.

Proof. Part (i) follows from the ergodic theorem for stationary sequences; see e.g. Theorem 9.6 of [18]. Ergodicity is a very weak property that is implied by any form of mixing [4]; hence part (ii) is immediate. ∎

The consistency of $\hat{w}_{ii}$ and its general applicability to ARCH, GARCH or SV models should make it a useful tool for practitioners working with financial returns data. Estimator (3.8) is even easier to use than Bartlett's formula which requires the difficult choice of a bandwidth-type parameter (the truncation point for the summation) for practical application. Concerning the correction factor: it is apparent that using the MOM choice $c_{in} = n/(n-i)$ would result in bigger variance estimates, and thus to more conservative inference as compared to Taylor's $c_{in} = 1$. Of course, if i is small compared to n, the difference between the two choices of correction factor is negligible.

A small simulation of the finite-sample behavior of estimator (3.8) with $c_{in} = n/(n-i)$ was run based on 999 stretches of size n from an ARCH(1) model. As Table 1 shows, with the exception of the case $\beta_1 = 0.55$, $\hat{w}_{11}$ appears to have bias and variance that both decrease as the sample size increases as consistency demands. Recall that when $\beta_1 \to 1/\sqrt{3} \approx 0.577$, then $EX_1^4 \to \infty$ and the assumptions of Proposition 3.3 break down. Therefore, the case $\beta_1 = 0.55$ is understandably problematic as it is so close to this threshold. Furthermore, the presence of a slight negative bias in $\hat{w}_{11}$ indicates that the MOM correction factor $c_{in} = n/(n-i)$ might be preferable to Taylor's $c_{in} = 1$ even for $i = 1$.

TABLE 1. Empirically found mean and standard deviation of the estimator $\hat{w}_{11}$ for the ARCH(1) process (1.5) with $\beta_0 = 0.01$ and $\varepsilon_t \sim$ i.i.d. $N(0,1)$

β_1	0.05	0.15	0.25	0.35	0.45	0.55
τ_1^2 of equation (3.5)	1.101	1.322	1.615	2.107	3.293	12.89
$[n=500]$ $E\hat{w}_{11}$	1.091	1.306	1.556	1.965	2.638	3.566
standard deviation$(\hat{w}_{11})$	0.110	0.187	0.319	0.787	2.311	3.277
$[n=2000]$ $E\hat{w}_{11}$	1.097	1.324	1.585	2.036	2.975	4.610
standard deviation$(\hat{w}_{11})$	0.055	0.098	0.185	0.544	2.237	6.777

Note that if $EX_t^4 = \infty$, then it is pointless to talk about the variance of $\hat{\rho}_k$. Nevertheless, as long as $EX_t^2 < \infty$, $\hat{\rho}_k$ is still consistent, and has a distribution typically falling in the domain of attraction of a stable law (see [7] and [8]). Subsampling [21] could then be employed to estimate the quantiles of this distribution, and to construct confidence intervals and tests for ρ_k that are valid whether EX_t^4 is finite or not.

Finally, note that the squares of ARCH and SV processes have nonvanishing correlations at any lags, so an analog of (3.4) no longer holds. Consequently, a simple formula such as (3.7) is unavailable, and to approximate the distribution of $\hat{\rho}_k$ as computed from the squared returns it is again recommended to use a resampling and/or subsampling approach [20]–[22].

REFERENCES

[1] M. S. Bartlett, *On the theoretical specification and sampling properties of autocorrelated time series*, J. Roy. Statist. Soc. Supplement 8 (1946), pp. 27–41. [Correction: 10 (1948), p. 200.]

[2] A. Berlinet and C. Francq, *On Bartlett's formula for non-linear processes*, J. Time Ser. Anal. 18 (1997), pp. 535–552.

[3] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, J. Econometrics 31 (1986), pp. 307–327.

[4] R. C. Bradley, *Basic properties of strong mixing conditions*, in: *Dependence in Probability and Statistics*, E. Eberlein and M. S. Taqqu (Eds.), Birkhäuser, 1986, pp. 165–192.

[5] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, second edition, Springer, New York 1991.

[6] M. Carasco and X. Chen, *Mixing and moment properties of various GARCH and stochastic volatility models*, Econometric Theory 18 (2002), pp. 17–39.
[7] R. A. Davis and T. Mikosch, *The sample autocorrelations of heavy-tailed processes with applications to ARCH*, Ann. Statist. 26 (1998), pp. 2049–2080.
[8] R. A. Davis and T. Mikosch, *Point process convergence of stochastic volatility processes with application to sample autocorrelation*, J. Appl. Probab. 38A (2001), pp. 93–104.
[9] F. Diebold, *Testing for serial correlation in the presence of ARCH*, in: *American Statistical Association: 1986 Proceedings of the Business and Economic Statistics Section*, 1986, pp. 323–328.
[10] R. Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation*, Econometrica 50 (1982), pp. 987–1008.
[11] C. Francq and J.-M. Zakoian, *Bartlett's formula for a general class of non-linear processes*, J. Time Ser. Anal. 30 (2009), pp. 449–465.
[12] W. A. Fuller, *Introduction to Statistical Time Series*, second edition, Wiley, New York 1996.
[13] L. Giraitis, P. Kokoszka and R. Leipus, *Stationary ARCH models: dependence structure and Central Limit Theorem*, Econometric Theory 16 (2000), pp. 3–22.
[14] L. Giraitis, P. Robinson and D. Surgailis, *A model for long memory conditional heteroskedasticity*, Ann. Appl. Probab. 10 (2000), pp. 1002–1024.
[15] C. Gouriéroux, *ARCH Models and Financial Applications*, Springer, New York 1997.
[16] C. Granger and A. Andersen, *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen 1978.
[17] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*, Wiley, New York 1988.
[18] O. Kallenberg, *Foundations of Modern Probability*, Springer, New York 1997.
[19] I. N. Lobato, J. C. Nankervis and N. E. Savin, *Testing for zero autocorrelation in the presence of statistical dependence*, Econometric Theory 18 (2002), pp. 730–743.
[20] D. N. Politis, *The impact of bootstrap methods on time series analysis*, Statist. Sci. 18 (2) (2003), pp. 219–230.
[21] D. N. Politis, J. P. Romano and M. Wolf, *Subsampling*, Springer, New York 1999.
[22] J. P. Romano and L. Thombs, *Inference for autocorrelations under weak assumptions*, J. Amer. Statist. Assoc. 91 (1996), pp. 590–600.
[23] X. Shao and W. B. Wu, *Asymptotic spectral theory for nonlinear time series*, Ann. Statist. 35 (2007), pp. 1773–1801.
[24] S. J. Taylor, *Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961–1979*, in: *Time Series Analysis: Theory and Practice I*, O. D. Anderson (Ed.), North Holland, Amsterdam 1982.
[25] S. J. Taylor, *Estimating the variances of autocorrelations calculated from financial time series*, Applied Statistics 33 (3) (1984), pp. 300–308.
[26] W. B. Wu, *An asymptotic theory for sample covariances of Bernoulli shifts*, Stochastic Process. Appl. 119 (2009), pp. 453–467.

Utah State University
Department of Mathematics and Statistics
Logan, UT 84322-3900, USA
E-mail: Piotr.Kokoszka@usu.edu

University of California
Department of Mathematics
La Jolla, CA 92093-0112, USA
E-mail: dpolitis@ucsd.edu

Received on 19.8.2010;
revised version on 12.10.2010