Econometric Theory, **16**, 2000, 692–728. Printed in the United States of America.

ESTIMATING WEAK GARCH REPRESENTATIONS

Christian Francq
Université du Littoral–Côte d'Opale

Jean-Michel Zakoïan
GREMARS, Université de Lille 3 and CREST

The classical definitions of GARCH-type processes rely on strong assumptions on the first two conditional moments. The common practice in empirical studies, however, has been to test for GARCH by detecting serial correlations in the squared regression errors. This can be problematic because such autocorrelation structures are compatible with severe misspecifications of the standard GARCH. Numerous examples are provided in the paper. In consequence, standard (quasi-) maximum likelihood procedures can be inconsistent if the conditional first two moments are misspecified. To alleviate these problems of possible misspecification, we consider weak GARCH representations characterized by an ARMA structure for the squared error terms. The weak GARCH representation eliminates the need for correct specification of the first two conditional moments. The parameters of the representation are estimated via two-stage least squares. The estimator is shown to be consistent and asymptotically normal. Forecasting issues are also addressed.

1. INTRODUCTION

In the past 15 years, there have been rapid developments in the field of modeling time-varying conditional variances in both applied and theoretical econometrics. Since the introduction of autoregressive conditional heteroskedasticity by Engle (1982), and its generalization by Bollerslev (1986), GARCH models have been the most widely used (see the review by Bollerslev, Engle, and Nelson, 1994). However, a plethora of alternative models has emerged in recent years. First, a number of specifications of the conditional variance generalizing the basic formulation (based on squared innovations) have been proposed. These alternatives are, in general, motivated by the need to capture some empirical stylized facts of financial time series (such as asymmetry). Second, some researchers have focused on new classes of processes that do not belong to the GARCH family. These are mainly the so-called stochastic volatility processes (see the review by Ghysels, Harvey, and Renault, 1996). In these models, by

An earlier version of this paper was presented at the Econometric Society European Meeting 1997, Toulouse. We are grateful to Bruce Hansen and two referees for very helpful comments. Address correspondence to: Jean-Michel Zakoïan, CREST, 15 Bd. Gabriel Péri, 92245 Malakoff Cedex, France; e-mail: zakoian@ensae.fr.

  0266-4666/00 $9.50

contrast with GARCH models, the scaling process is not measurable with respect to past observables: the standard representation takes the form of an ARMA whose innovations are scaled by an unobservable autoregression. The close association between stochastic volatility specifications and the time-varying volatility diffusion processes commonly used in the finance theory has generated great interest in these alternatives to the GARCH models. See Andersen (1994) for a connection between GARCH and stochastic volatility processes. Finally, following Hamilton (1989), some recent papers have introduced models where the conditional variance changes according to an unobserved Markov chain.

A common feature of models of conditional heteroskedasticity is the existence of a univariate process of the general form $\epsilon_t = \sigma_t Z_t$, where σ_t is strictly positive and measurable with respect to some σ-field I_{t-1}, (I_t) being a filtration; the Z_t process is mean zero and unit variance conditionally on I_{t-1}. In the GARCH context $I_{t-1} = \sigma(\epsilon_{t-1}, \dots)$, whereas I_{t-1} includes unobserved variables in the stochastic volatility framework. Therefore, a strong assumption imposed by standard models of changing variances is that (ϵ_t) is a martingale difference sequence (with respect to (I_t)). A natural question is how restrictive this requirement is. The fact that σ_t is not observable makes the answer very difficult for stochastic volatility models. In the GARCH context, the notion that some financial series might violate the martingale difference assumption can be seriously entertained. Casual examination of plots of empirical correlations between ϵ_t and some (nonlinear) functions of its past suggest the martingale difference assumption is likely to be too strong. Moreover, some important issues such as modeling time-varying skewness and kurtosis, conditional on I_t, are ruled out by the classical assumptions.[1]

The martingale difference assumption in the GARCH framework involves other important shortcomings in terms of temporal aggregation. In an important paper, Drost and Nijman (1993) have shown that "the classical (semi)-strong GARCH assumptions [i.e., the innovation is a martingale difference with a specified conditional variance] on the available data frequency are arbitrary. Generally a (semi-)strong GARCH process aggregates to some weak GARCH process [i.e., in which only projections of the noise and its square are considered] that is not semi-strong GARCH." In addition they have shown that the class of weak GARCH processes is closed under temporal aggregation: more precisely, the low frequency model that is implied by an assumed high frequency GARCH model can be derived. In the same spirit, Drost and Werker (1996) derive explicit relations expressing the weak GARCH parameters at arbitrary frequencies in terms of an underlying GARCH diffusion, whereas Nijman and Sentana (1996) obtain results on contemporaneous aggregation and marginalization of vector processes. See also Meddahi and Renault (1996).

Finally, the martingale difference assumption has crucial importance for asymptotic theory in statistical inference. The literature on the estimation of GARCH-type processes is now quite substantial, and much theoretical analysis assuming a time-varying conditional variance uses a quasi–maximum likelihood estimator (QMLE).[2] On the other hand, Weiss (1984, 1986) can be credited with first having analyzed the asymptotic properties of two-stage least-squares (LS) estimation. Again, the standard proof for the asymptotic normality of the LS estimator (based on a central limit theorem for martingale differences) does not extend to the situation where only white noise assumptions can be made. The estimation of stochastic volatility models entails additional difficulties because, unlike GARCH, the conditional likelihood cannot be computed in closed form. A variety of alternative procedures have been proposed to fit these models (see the review by Shephard, 1996). For similar reasons, Markov-switching models are generally less tractable than GARCH. Asymptotic properties of ML estimation remain in large part to be uncovered.[3]

This paper presents a unified statistical treatment of a wide class of conditionally heteroskedastic processes. Our approach is based on a general two-stage representation including the weak GARCH proposed by Drost and Nijman (1993). It consists of two ARMA equations, the first one on the observable process X, the second one on the square of its linear innovation. Such a representation is well known to hold for GARCH models. To anticipate the results that follow, the representation is remarkably robust to certain types of misspecification in GARCH models.[4]

In this paper, first of all, we show that in various situations where the GARCH model is not the correct data-generating process (DGP), an underlying ARMA representation for the squared innovations holds. In particular, it offers the possibility of dealing with several types of misspecifications of the conditional variance in the GARCH framework. Moreover, a striking feature of the proposed representation is that it nests not only the standard GARCH model but also most of the popular specifications of the literature, e.g., the standard GARCH, some asymmetric GARCH, some stochastic volatility and Markov-switching models.

Second, although Drost and Nijman (1993) mention estimation issues, no theoretically sound procedure for estimation is available in the weak GARCH context. It is the purpose of this paper to derive a large sample theory of inference for the two-stage representation. We use a LS procedure, which amounts to minimizing the linear prediction errors in both equations. Potential alternative approaches are generalized method of moments (GMM) procedures, quasi–maximum likelihood (QML) methods, and simulation-based methods. As for GMM, which can be seen as an extension of LS, the idea would be to exploit the infinite set of moment conditions implied by the innovations in each ARMA equation, along the lines of Hansen and Singleton (1996). However, to our knowledge, the existing theory on the GMM cannot be straightforwardly applied to our setting. Although a stochastic volatility structure could be used to compute a QMLE, this would require a complete specification of the first two conditional moments. In contrast, our estimation procedure does not require specification of any functional form other than the two ARMA equations. Clearly, our estimator will be strictly inefficient relative to full-information

MLE, or even QMLE and GMM. Two references on these techniques applied to volatility models are Ruiz (1994) and Andersen and Sørensen (1996). Actually, weak GARCH representations have little interest when a strong model is available. In practice, this is rarely the case, and a misspecified model is likely to be selected. In such situations, QML-based inference can lead to very poor forecasts, as we shall see. Finally, the simulation-based methods (see, e.g., Gouriéroux, Monfort, and Renault, 1993; Gallant and Tauchen, 1996; Broze, Scaillet, and Zakoïan, 1998) are inappropriate in our context. Because we do not specify the distribution of the innovations, and because they are not independent, we are unable to simulate the model. The examples presented subsequently show that a given representation is compatible with many DGP's.

The paper proceeds as follows. Section 2 provides the relevant definitions and some important illustrations of the concept of weak GARCH. Section 3 presents the estimation method and the asymptotic results. Apart from some moment conditions, along with some standard assumptions on the lag polynomials, strong mixing and strict stationarity of the *observable process* are sufficient to derive the results. Section 4 reports the results of various simulation experiments. Section 5 is devoted to forecasting issues. Section 6 concludes the paper and summarizes its main results. All derivations and proofs are collected in the Appendix.

2. WEAK GARCH REPRESENTATIONS: DEFINITION AND EXAMPLES

Consider any strictly stationary, purely nondeterministic process $(X_t)_{t\in\mathbb{Z}}$, admitting moments up to order four. From the Wold theorem, (X_t) admits an infinite moving-average (MA) representation. Let us assume that this MA(∞) can be inverted to obtain a finite order ARMA representation of the form

$$X_t + \sum_{i=1}^{P} \phi_i X_{t-i} = \epsilon_t + \sum_{i=1}^{Q} \psi_i \epsilon_{t-i}, \tag{1}$$

where (ϵ_t) is a sequence of centered, uncorrelated random variables with common variance $\sigma^2 > 0$ and where the polynomials $\Phi(z) = 1 + \phi_1 z + \cdots + \phi_P z^P$ and $\Psi(z) = 1 + \psi_1 z + \cdots + \psi_Q z^Q$ have all their zeros outside the unit disk and have no common zero. Without loss of generality, assume that ϕ_P and ψ_Q are both not equal to zero (by convention $\phi_0 = \psi_0 = 1$). With these assumptions, process (ϵ_t) can be interpreted as the linear innovation of (X_t), i.e., $\epsilon_t = X_t - E(X_t | H_X(t-1))$, where $H_X(t-1)$ denotes the closed span of $(X_s; s < t)$. The $(\epsilon_t^2)_{t\in\mathbb{Z}}$ process is clearly second-order stationary and purely nondeterministic. Therefore, it admits a Wold decomposition. Again, we assume that it can be inverted to obtain an ARMA equation of the form

$$\epsilon_t^2 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 = \omega + u_t + \sum_{i=1}^{q} \beta_i u_{t-i}, \tag{2}$$

where (u_t) is a white noise (with variance $\zeta^2 > 0$). We make similar standard regularity assumptions on the two polynomials $\phi(z) = 1 + \alpha_1 z + \cdots + \alpha_p z^p$ and $\psi(z) = 1 + \beta_1 z + \cdots + \beta_q z^q$ as we did on $\Phi(z)$ and $\Psi(z)$. Therefore (u_t) is the linear innovation of (ϵ_t^2). Denote the lag operator by L. Defining $\pi(L) = 1 - \sum_{i=1}^{\infty} \pi_i L^i = \psi(L)^{-1}\phi(L)$ we set $h_t = E(\epsilon_t^2 | H_{\epsilon^2}(t-1)) = E(\epsilon_t^2) + \sum_{i=1}^{\infty} \pi_i (\epsilon_{t-i}^2 - E(\epsilon_t^2))$.

Remarks.

(a) It is well known that any stationary strong (or semistrong) GARCH(p',q') process with a finite fourth-order moment admits a representation of the form (2) with $p = \max\{p',q'\}$ and $q = p'$. In this strong GARCH setting, h_t is the conditional variance of ϵ_t; moreover, the constraints $\omega > 0$, $\pi_i \geq 0$ $(\forall i)$, which guarantee the positivity of h_t, are usually imposed (see Nelson and Cao, 1992).

(b) Our definition is more general than that of weak GARCH proposed by Drost and Nijman (1993). From both definitions, u_t in (2) is the linear innovation of ϵ_t^2. In the Drost and Nijman approach, u_t is also orthogonal to all past values of ϵ_t. This additional constraint ensures the stability of the class under temporal aggregation. Because temporal aggregation is not the focus of the present paper, we derive our asymptotic results under weaker assumptions. To avoid the introduction of a new label in the GARCH literature, and because it is consistent with the concept of weak ARMA, we refer to the two-stage representation (1)–(2) as weak ARMA-GARCH or weak GARCH.

We now consider some interesting particular cases of processes admitting weak ARMA-GARCH representations. The first four have already been introduced in the GARCH literature and provide nice interpretations. Examples 5–7 illustrate the possibility of getting weak GARCH representations of some (strong) nonlinear processes that, a priori, do not seem to be related to the GARCH framework. Throughout the section, we assume that the unconditional moments are finite, as required in the definition. The σ-field generated by $(\epsilon_s; s < t)$ is denoted by $\underline{\epsilon_{t-1}}$.

Example 1 (temporal aggregation of a strong GARCH) (Drost and Nijman, 1993)

Let (X_t) be generated by an ARMA(P',Q') model with semistrong GARCH innovation, i.e.,

$$E(\epsilon_t | \underline{\epsilon_{t-1}}) = 0 \quad \text{and} \quad E(\epsilon_t^2 | \underline{\epsilon_{t-1}}) = \sigma_t^2 = c + \sum_{i=1}^{q'} a_i \epsilon_{t-i}^2 + \sum_{i=1}^{p'} b_i \sigma_{t-i}^2.$$

We assume in addition that the marginal distribution of X_t is symmetric. Thus, for any integer m, the process $(X_{mt})_{t\in\mathbb{Z}}$ follows an ARMA(P,Q) process with weak GARCH(p,q) errors, where $P = P'$, $Q = P' + [(Q' - P')/m]$, and $p = q = \max\{p',q'\} + \frac{1}{2}Q(Q+1)$. Obviously, the linear innovation of (X_{mt}) is a

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

martingale difference, but one can show in particular cases that the corresponding representation (2) is not strong, in the sense that the white noise is not a martingale difference. In particular, this result is useful for statistical purposes when only low frequency data from a high frequency (semi-)strong GARCH are available.

Example 2 (quadratic GARCH)

We here consider a modification of the basic GARCH given by

$$E(\epsilon_t|\underline{\epsilon_{t-1}}) = 0 \quad \text{and} \quad E(\epsilon_t^2|\underline{\epsilon_{t-1}}) = \sigma_t^2 = \left(c + \sum_{i=1}^{q'} a_i \epsilon_{t-i}\right)^2 + \sum_{i=1}^{p'} b_i \sigma_{t-i}^2,$$

where the b_i's are nonnegative. This model is a particular case of the quadratic ARCH model introduced by Sentana (1995). Put $\eta_t = \epsilon_t^2 - \sigma_t^2$ and observe that the η_t's are uncorrelated and that they are uncorrelated with any variable belonging to the future (by the martingale difference assumption) and the past of ϵ_t (by the conditional variance assumption). Rewriting the equation determining σ_t^2, we end up with

$$\epsilon_t^2 = c^2 + \sum_{i=1}^{\max\{p',q'\}} (a_i^2 + b_i)\epsilon_{t-i}^2 + v_t,$$

where $v_t = 2c\sum_{i=1}^{q'} a_i \epsilon_{t-i} + \sum_{i\neq j} a_i a_j \epsilon_{t-i}\epsilon_{t-j} + \eta_t - \sum_{i=1}^{p'} b_i \eta_{t-i}$. It is now easy to check that $E(v_t) = 0$ and $E(v_t v_{t-k}) = 0$, $\forall k > \max\{p',q'\}$. Hence (v_t) is a $\mathrm{MA}(\max\{p',q'\})$ process, from which we deduce that (ϵ_t) is a weak $\mathrm{GARCH}(\max\{p',q'\}, \max\{p',q'\})$.

Example 3 (unobserved GARCH)

A number of recent papers have focused on GARCH models observed with errors, examples being Harvey, Ruiz, and Sentana (1992), Gouriéroux et al. (1993), and King, Sentana, and Wadhwani (1994). These models take the form

$$\epsilon_t = e_t + W_t, \qquad e_t = \sigma_t Z_t, \qquad \sigma_t^2 = c + \sum_{i=1}^{q'} a_i e_{t-i}^2 + \sum_{i=1}^{p'} b_i \sigma_{t-i}^2, \tag{3}$$

where (Z_t) and (W_t) are mutually independent and are independent and identically distributed (i.i.d.) centered sequences. Their variances are 1 and σ_W^2, respectively. Unlike the other GARCH-type models, unobserved GARCH are not easy to estimate, for it is not possible to deduce analytically the density function of ϵ_t conditional on its past values. Other approaches for estimating model (3) are the Kalman filter or simulation-based methods (see the references outlined earlier).

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

Simple algebra shows that $\mathrm{Cov}(\epsilon_t^2, \epsilon_{t-h}^2) = \mathrm{Cov}(e_t^2, e_{t-h}^2)$, $\forall h > 0$. Because (e_t) is a strong GARCH(p', q') process, the autocovariance structure of its square is determined by

$$\mathrm{Cov}(e_t^2, e_{t-h}^2) = \sum_{i=1}^{\max\{p',q'\}} (a_i + b_i)\mathrm{Cov}(e_t^2, e_{t-h+i}^2), \qquad h > p'.$$

The same relation holds for (ϵ_t^2) except that it requires $h > \max\{p', q'\}$ (because the last term in the sum is $\mathrm{Cov}(e_t^2, e_{t-h+\max\{p',q'\}}^2)$, so it cannot be replaced by $\mathrm{Cov}(\epsilon_t^2, \epsilon_{t-h+\max\{p',q'\}}^2)$ unless $h > \max\{p', q'\}$). Finally, (ϵ_t) is a weak GARCH process of the form

$$\epsilon_t^2 - \sum_{i=1}^{\max\{p',q'\}} (a_i + b_i)\epsilon_{t-i}^2 = c + \left(1 - \sum_{i=1}^{\max\{p',q'\}} a_i + b_i\right)\sigma_W^2 + u_t + \sum_{i=1}^{\max\{p',q'\}} \beta_i u_{t-i},$$

where the β_i's are different from the $-b_i$'s (unless $\sigma_W = 0$). Note that the AR part in this representation is not affected by the presence of the disturbance W_t.

Example 4 (asymmetric GARCH)

In El Babsiri and Zakoïan (2000), the following model is considered:

$$\epsilon_t = \sigma_{t,+} Z_t^+ + \sigma_{t,-} Z_t^-,$$

where $Z_t^+ = \max(Z_t, 0)$, $Z_t^- = \min(Z_t, 0)$ with (Z_t) a symmetrically distributed i.i.d.(0,1) process, and $\sigma_{t,+}$ and $\sigma_{t,-}$ are two positive processes, measurable with respect to $\underline{\epsilon_{t-1}}$. The main interest of the model is to allow for different volatility processes: one for the positive part of ϵ_t, namely, $V(\epsilon_t^+ | \underline{\epsilon_{t-1}}) = \sigma_{t,+}^2 V(Z_t^+)$, and one for the negative part $V(\epsilon_t^- | \underline{\epsilon_{t-1}}) = \sigma_{t,-}^2 V(Z_t^-)$. In addition, assume that each volatility process reacts symmetrically to past innovations as

$$(I - A_+(L))\sigma_{t,+}^2 = \omega_+ + B_+(L)\epsilon_t^2 \quad \text{and} \quad (I - A_-(L))\sigma_{t,-}^2 = \omega_- + B_-(L)\epsilon_t^2,$$

where $A_+(L), B_+(L), A_-(L), B_-(L)$ are some lag polynomials, $(I - A_+(L))$ and $(I - A_-(L))$ are invertible, and $\omega_+ > 0, \omega_- > 0$. The (ϵ_t) process is not a martingale difference in general, because $E(\epsilon_t | \underline{\epsilon_{t-1}}) = (\sigma_{t,+} - \sigma_{t,-})E(Z_t^+)$. However, for some appropriate parameterizations of $\sigma_{t,+}$ and $\sigma_{t,-}$, (ϵ_t) is shown to be a white noise. Then we have

$$C(L)\epsilon_t^2 = \omega + (I - A_+(L))(I - A_-(L))u_t,$$

where $C(L) = [(I - A_+(L))(I - A_-(L)) - 0.5((I - A_-(L))B_+(L) + (I - A_-(L))B_-(L))]$ and $u_t = \epsilon_t^2 - E(\epsilon_t^2 | \underline{\epsilon_{t-1}})$, ω is a constant. Therefore, if the regularity assumptions on the polynomials are satisfied, (ϵ_t) is a weak GARCH.

Example 5 (stochastic autoregressive volatility models)

The so-called stochastic volatility models have received increasing attention in finance and econometrics literature. In particular, the class of stochastic autoregressive volatility (SARV) models provides a direct generalization of GARCH models (see Andersen, 1994). The simplest model of this kind is

$$\epsilon_t = \sigma_t Z_t, \qquad \sigma_t^2 = c + d\sigma_{t-1}^2 + [a + b\sigma_{t-1}^2]v_t, \qquad c,d,b > 0, \qquad a \geq 0, \tag{4}$$

where (Z_t) and (v_t) are i.i.d.(0,1) processes, with Z_t independent of v_{t-j}, $j \geq 0$. Note that the GARCH(1,1) is obtained by specifying $v_t = Z_{t-1}^2 - 1$ and $a = 0$. Then some computations not reported here (see Francq and Zakoïan, 1997) show that the covariance structure of (ϵ_t^2) is characterized by $\mathrm{Cov}(\epsilon_t^2, \epsilon_{t-h}^2) = d\,\mathrm{Cov}(\epsilon_t^2, \epsilon_{t-h+1}^2)$, $\forall h > 1$. Therefore a weak GARCH(1,1) structure is obtained for (ϵ_t):

$$\epsilon_t^2 - d\epsilon_{t-1}^2 = c + u_t + \beta u_{t-1},$$

where (u_t) is a white noise and β can be computed explicitly.

Example 6 (bilinear processes)

Set $\epsilon_t = \nu_t \nu_{t-1}$, where (ν_t) is a centered i.i.d. sequence with unit variance and $E(\nu_t^8) < \infty$. Then it is easily seen that (ϵ_t) is centered and has zero autocorrelation but is not a martingale difference sequence. Moreover, computing the autocovariance function of the process (ϵ_t^2) reveals that it is a MA(1) of the form $\epsilon_t^2 = 1 + u_t - \theta u_{t-1}$, where (u_t) is a white noise and θ a parameter depending on the fourth moment of ν_t, which can be assumed inside the unit circle. Therefore (ϵ_t) is a weak ARMA(0,0)-GARCH(0,1) process.

We now aim to show that $(X_t) := (\epsilon_t^2 - 1)$ is itself a weak ARMA-GARCH: it is much less straightforward because we have to prove that the process (u_t^2) has an ARMA representation. We have

$$u_t^2 - \theta^2 u_{t-1}^2 = X_t^2 + 2X_t \sum_{i=1}^{\infty} \theta^i X_{t-i} := v_t. \tag{5}$$

Equation (5) determines the autoregressive part of the ARMA model for (u_t^2). To obtain the order of the moving average part, we show that v_t is a MA process. We have

$$\begin{aligned}\mathrm{Cov}(v_t, v_{t-k}) = {} & \mathrm{Cov}\left(X_t^2, X_{t-k}^2 + 2X_{t-k} \sum_{i=1}^{\infty} \theta^i X_{t-k-i}\right) \\ & + \mathrm{Cov}\left(2\theta X_t X_{t-1}, X_{t-k}^2 + 2X_{t-k} \sum_{i=1}^{\infty} \theta^i X_{t-k-i}\right) \\ & + \mathrm{Cov}\left(2X_t \sum_{i=2}^{\infty} \theta^i X_{t-i}, X_{t-k}^2 + 2X_{t-k} \sum_{i=1}^{\infty} \theta^i X_{t-k-i}\right).\end{aligned}$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

Because X_t is a function of (ν_t, ν_{t-1}), the first covariance on the right-hand side is equal to zero for all $k > 1$. Similarly $X_t X_{t-1}$ is a function of $(\nu_t, \nu_{t-1}, \nu_{t-2})$; hence the second term is null for all $k \geq 2$. Finally, from $E(X_t) = 0$ and the independence between X_t and $X_{t-k} (\forall k \geq 2)$, the last covariance is equal to zero. Hence we have proved that $(\epsilon_t^2 - 1)$ is a weak ARMA(0,1)-GARCH(1,2) process.

Example 7 (Markov-switching process)

In an interesting generalization of ARMA(p,q) models, Hamilton (1989) proposed a switching-regime Markov model that can accommodate complicated dynamics such as occasional shifts or asymmetric cycles. The parameters of the model are specified as functions of the state of an unobservable (or *hidden*) Markov chain. Pagan and Schwert (1990) considered a variant of it for modeling conditional variance in financial time series. In the following model, previously analyzed by Cai (1994) and Dueker (1997) (see also Hamilton and Susmel, 1994), the intercept in the conditional variance is subject to Markov switching. Let Δ_t denote an unobserved random variable that can take on the values $0,1,\dots,K-1$. Suppose that (Δ_t) can be described by a Markov chain with strictly positive transition probabilities $p_{ij} = P[\Delta_t = j | \Delta_{t-1} = i]$, for $i,j = 0,1,\dots,K-1$. We assume that the dynamics of a process (ϵ_t) take the form

$$\epsilon_t = \sigma_t Z_t, \qquad \sigma_t^2 = \mu(\Delta_t) + \sum_{i=1}^{q'} a_i \epsilon_{t-i}^2 + \sum_{i=1}^{p'} b_i \sigma_{t-i}^2 \tag{6}$$

with

$$\mu(\Delta_t) = \sum_{i=1}^{K} \mu_i \mathbb{1}_{\{\Delta_t = i-1\}}, \qquad 0 < \mu_1 < \mu_2 < \cdots < \mu_K, \tag{7}$$

where (Z_t) is an i.i.d.(0,1) process admitting a fourth moment, (Z_t) being independent of (Δ_t). Calculations reported in Francq and Zakoïan (1997) show that (ϵ_t) is a weak GARCH$(\max\{p',q'\} + K - 1, p' + K - 1)$ process of the form

$$\prod_{k=1}^{K-1} (1 - \lambda_k L)\left(I - \sum_{i=1}^{\max\{p',q'\}} (a_i + b_i) L^i\right)\epsilon_t^2 = \omega + \left(I + \sum_{i=1}^{p'+K-1} \beta_i L^i\right) u_t,$$

where $\lambda_1, \dots, \lambda_{K-1}$ are the eigenvalues different from 1 of $\mathbb{P} = (p_{ji})$. To be more specific, suppose $p' = q' = 0$ (i.e., $\sigma_t^2 = \mu(\Delta_t)$) and (Δ_t) is a two-state Markov chain with state space $\{0,1\}$ and $0 < p_{01} < 1$, $0 < p_{10} < 1$. Within this setup, the two states can be interpreted as high $(\Delta_t = 1)$ and low $(\Delta_t = 0)$ conditional variance regimes. Therefore (ϵ_t^2) admits a weak ARMA(1,1) representation of the form

$$\epsilon_t^2 - (1 - p_{01} - p_{10})\epsilon_{t-1}^2 = \omega + u_t + \beta u_{t-1}, \tag{8}$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

where (u_t) is a white noise and ω and β can be determined in terms of p_{01}, p_{10}, μ_1, μ_2, and $E(Z_t^4)$. In the case where $p_{01} + p_{10} = 1$, we have $\beta = 0$, and (ϵ_t^2) is (up to its mean) a white noise.

Other examples such as the β-ARCH process (Diebolt and Guégan, 1991) or a diffusion process can also be dealt with (for details, see Francq and Zakoïan, 1997). To conclude the section, it may be worth noting that one can easily construct examples of GARCH-type or stochastic volatility models that do not fit into the weak ARMA-GARCH notion proposed here. A first example is the class of fractionally integrated GARCH introduced by Baillie, Bollerslev, and Mikkelsen (1996). In these models, the conditional variance implies a slow hyperbolic rate of decay for the influence of lagged squared innovations, which precludes the existence of a finite order ARMA representation for the squared innovations. In the exponential GARCH (EGARCH) of Nelson (1991) the autocovariances of the squared innovations decay at an exponential rate but cannot be expressed as linear combinations of exponentials as for ARMA models.[5] Therefore, a weak GARCH representation is also precluded in this example.

3. ASYMPTOTIC RESULTS

In this section, we will consider the problem of estimating the ARMA-GARCH representation (1)–(2) via two-stage LS. The method involves two successive minimizations of sums of squared deviations about conditional *linear* expectations. It is worth noting that the standard asymptotic results existing in the time series literature (e.g., based on the martingale theory) are not applicable because we are working with weak representations. We follow the same approach as Francq and Zakoïan (1998) for weak ARMA representations. The symbols $\to_d$ and *a.s.* signify convergence in distribution and almost surely.

3.1. Consistency and Asymptotic Normality

Let $\theta_0^{(1)} = (\phi_1,\dots,\phi_P,\psi_1,\dots,\psi_Q)'$, $\theta_0^{(2)} = (\alpha_1,\dots,\alpha_p,\beta_1,\dots,\beta_q,\omega)'$, and $\theta_0 = (\theta_0^{(1)},\theta_0^{(2)})'$. For any $\theta = (\theta^{(1)},\theta^{(2)})' = (\theta_1^{(1)},\dots,\theta_{P+Q}^{(1)},\theta_1^{(2)},\dots,\theta_{p+q+1}^{(2)})$, we set $\Phi_{\theta^{(1)}}(z) = 1 + \theta_1^{(1)}z + \cdots + \theta_P^{(1)}z^P$, $\Psi_{\theta^{(1)}}(z) = 1 + \theta_{P+1}^{(1)}z + \cdots + \theta_{P+Q}^{(1)}z^Q$, $\phi_{\theta^{(2)}}(z) = 1 + \theta_1^{(2)}z + \cdots \theta_p^{(2)}z^p$, and $\psi_{\theta^{(2)}}(z) = 1 + \theta_{p+1}^{(2)}z + \cdots + \theta_{p+q}^{(2)}z^q$. For any positive constant δ, we define the parameter space $\Theta_\delta := \Theta_\delta^{(1)} \times \Theta_\delta^{(2)}$ as the compact set of all θ's such that the roots of the polynomials $\Phi_{\theta^{(1)}}$, $\Psi_{\theta^{(1)}}$, $\phi_{\theta^{(2)}}$, and $\psi_{\theta^{(2)}}$ have moduli $\geq 1 + \delta$. We choose δ small enough so that θ_0 belongs to Θ_δ. Now from invertibility of the lag polynomials and the stationarity assumption on (X_t^2), the equations

$$\Phi_{\theta^{(1)}}(L)X_t = \Psi_{\theta^{(1)}}(L)\epsilon_t(\theta^{(1)}), \qquad \forall t \in \mathbb{Z} \tag{9}$$

and

$$\phi_{\theta^{(2)}}(L)\epsilon_t^2(\theta^{(1)}) = \theta_{p+q+1}^{(2)} + \psi_{\theta^{(2)}}(L)u_t(\theta), \qquad \forall t \in \mathbb{Z} \tag{10}$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

define two second-order stationary sequences, $(\epsilon_t(\theta^{(1)}))$ and $(u_t(\theta^{(1)},\theta^{(2)}))$, for all $\theta \in \Theta_\delta$ (see, e.g., Brockwell and Davis, 1991).

Let $X_1, X_2, \ldots, X_n$ be a realization of length n of (X_t). For $0 < t \le n$, $\epsilon_t(\theta^{(1)})$ and $u_t(\theta)$ are approximated by $\tilde{\epsilon}_t(\theta^{(1)})$ and $\tilde{u}_t(\theta)$ obtained by replacing the unknown starting values by zero ($\tilde{\epsilon}_t(\theta^{(1)}) = 0$, $-Q+1 \le t \le 0$, and $\tilde{u}_t(\theta) = 0$, $-q+1 \le t \le 0$).

The random variable $\hat{\theta}_n := (\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})'$ is called a LS estimator if it satisfies, a.s.,

$$Q_n^{(1)}(\hat{\theta}_n^{(1)}) = \min_{\theta^{(1)} \in \Theta_\delta^{(1)}} Q_n^{(1)}(\theta^{(1)}), \qquad Q_n^{(2)}(\hat{\theta}_n) = \min_{\theta^{(2)} \in \Theta_\delta^{(2)}} Q_n^{(2)}(\hat{\theta}_n^{(1)}, \theta^{(2)}), \tag{11}$$

where

$$Q_n^{(1)}(\theta^{(1)}) = \frac{1}{n}\sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta^{(1)}), \qquad Q_n^{(2)}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{u}_t^2(\theta).$$

We have the following consistency theorem.

THEOREM 1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic process belonging to L^4 and satisfying* (1)–(2). *Let $(\hat{\theta}_n)$ be a sequence of* LS *estimators. Then*

$$\hat{\theta}_n \to \theta_0 \quad a.s.\ as\ n \to \infty.$$

To be more specific about the asymptotic behavior of $(\hat{\theta}_n)$, we need some additional assumptions on the observed process.

THEOREM 2. *Let $(X_t)_{t\in\mathbb{Z}}$ be a strictly stationary process satisfying* (1)–(2), *endowed with the sequence of strong mixing coefficients $(\alpha_X(k))_{k\in\mathbb{N}}$. Furthermore, assume that, for some $\nu > 0$, $E(X_t^{8+4\nu}) < \infty$ and $\sum_{k=0}^{\infty}[\alpha_X(k)]^{\nu/(2+\nu)} < \infty$. Then,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{pmatrix}\right), \tag{12}$$

where

$$V_{11} = J_{11}^{-1} I_{11} J_{11}^{-1}, \qquad V_{21} = V_{12}' = J_{22}^{-1}(I_{21} + J_{21}J_{11}^{-1}I_{11})J_{11}^{-1},$$

$$V_{22} = J_{22}^{-1}(I_{22} + J_{21}J_{11}^{-1}I_{11}J_{11}^{-1}J_{12} - I_{21}J_{11}^{-1}J_{12} - J_{21}J_{11}^{-1}I_{12})J_{22}^{-1}$$

and

$$I_{11} = \lim_{n\to\infty} \operatorname{Var}\left(\sqrt{n}\,\frac{\partial}{\partial\theta^{(1)}}\, Q_n^{(1)}(\theta_0^{(1)})\right),$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

$$I_{22} = \lim_{n\to\infty} \operatorname{Var}\left(\sqrt{n}\,\frac{\partial}{\partial\theta^{(2)}}\,Q_n^{(2)}(\theta_0)\right),$$

$$I_{12} = \lim_{n\to\infty} E\left(n\,\frac{\partial}{\partial\theta^{(1)}}\,Q_n^{(1)}(\theta_0^{(1)})\,\frac{\partial}{\partial\theta^{(2)\prime}}\,Q_n^{(2)}(\theta_0)\right), \qquad I_{21} = I_{12}',$$

$$J_{11} \overset{a.s.}{=} \lim_{n\to\infty}\left[\frac{\partial^2}{\partial\theta_i^{(1)}\partial\theta_j^{(1)}}\,Q_n^{(1)}(\theta_0^{(1)})\right], \qquad J_{22} \overset{a.s.}{=} \lim_{n\to\infty}\left[\frac{\partial^2}{\partial\theta_i^{(2)}\partial\theta_j^{(2)}}\,Q_n^{(2)}(\theta_0)\right],$$

$$J_{12} \overset{a.s.}{=} \lim_{n\to\infty}\left[\frac{\partial^2}{\partial\theta_i^{(1)}\partial\theta_j^{(2)}}\,Q_n^{(2)}(\theta_0)\right], \qquad J_{21} = J_{12}'.$$

To make the asymptotic normality result operational, it is crucial to be able to build a weakly consistent estimator of the asymptotic covariance matrix.

3.2. Covariance Matrix Estimation

Recently, several authors have proposed methods to estimate covariance matrices in various situations (see, e.g., Newey and West, 1987; Andrews, 1991; Hansen, 1992). In particular, Francq and Zakoïan (2000) consider the case of weak ARMA and models.[6] Using a similar approach, we now consider the estimation of the covariance matrix $V := (V_{ij})$ in Theorem 2. Define

$$\hat{J}_{11}(\theta^{(1)}) = \frac{2}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta^{(1)}}\,\tilde{\epsilon}_t(\theta^{(1)})\left(\frac{\partial}{\partial\theta^{(1)}}\,\tilde{\epsilon}_t(\theta^{(1)})\right)',$$

$$\hat{J}_{22}(\theta) = \frac{2}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta^{(2)}}\,\tilde{u}_t(\theta)\left(\frac{\partial}{\partial\theta^{(2)}}\,\tilde{u}_t(\theta)\right)',$$

$$\hat{J}_{12}(\theta) = \frac{2}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta^{(1)}}\,\tilde{u}_t(\theta)\left(\frac{\partial}{\partial\theta^{(2)}}\,\tilde{u}_t(\theta)\right)' + \frac{2}{n}\sum_{t=1}^{n}\tilde{u}_t(\theta)\,\frac{\partial^2}{\partial\theta^{(1)}\partial\theta^{(2)\prime}}\,\tilde{u}_t(\theta).$$

An empirical estimator of J_{11} (resp. J_{22}, resp. J_{21}) is then given by $\hat{J}_{11}(\hat{\theta}_n^{(1)})$ (resp. $\hat{J}_{22}(\hat{\theta}_n)$, resp. $\hat{J}_{21}(\hat{\theta}_n)$).

From the proof of Theorem 2, we have $I_{22} = \sum_{h=-\infty}^{+\infty}\Delta_h(\theta_0)$ where

$$\Delta_h(\theta) = 4E\left(u_t(\theta)\,\frac{\partial}{\partial\theta^{(2)}}\,u_t(\theta)\right)\left(u_{t+h}(\theta)\,\frac{\partial}{\partial\theta^{(2)}}\,u_{t+h}(\theta)\right)'.$$

To estimate I_{22}, we consider a sequence of real numbers (b_n) going to zero and a real-valued weight function $\kappa(\cdot)$. The latter is a bounded, even, and nonnegative definite function with compact support $[-a,a]$ and continuous at the origin with $\kappa(0) = 1$. Let, for $0 \le h \le n$,

$$\hat{\Delta}_h(\theta) = \frac{4}{n}\sum_{t=1}^{n-h}\tilde{u}_t(\theta)\,\frac{\partial}{\partial\theta^{(2)}}\,\tilde{u}_t(\theta)\left(\tilde{u}_{t+h}(\theta)\,\frac{\partial}{\partial\theta^{(2)}}\,\tilde{u}_{t+h}(\theta)\right)' := \hat{\Delta}_{-h}(\theta)'$$

and

$$\hat{I}_{22}(\theta) = \sum_{h=-T_n}^{+T_n} \kappa(hb_n)\hat{\Delta}_h(\theta),$$

where T_n is the integer part of a/b_n. Then $\hat{I}_{22}(\hat{\theta}_n)$ can be used as an estimator of I_{22}. The estimators of I_{11} and I_{12} are constructed similarly.

Finally, we estimate V by plugging the estimates of the I_{ij}'s and J_{ij}'s into the expressions for the V_{ij}'s. Note that the derivatives in the preceding expressions can be recovered recursively using (9)–(10). It is worth noting that, when the standard assumptions hold (i.e., when (u_t) is a martingale difference), all the $\Delta_h(\theta_0)$'s are equal to zero except for $h = 0$. In the general case, neglecting those terms would entail inconsistency in the estimation of the matrix I_{22}. Another point to be noticed is the presence of second-order derivatives in the matrix $\hat{J}_{12}(\theta)$. Indeed, the last sum does not vanish when n goes to infinity and $\theta = \theta_0$ because the derivatives cannot be written as linear functions of the $\epsilon_{t-i}^2(\theta^{(1)})$'s; therefore u_t and its second derivative are not orthogonal.

As for all asymptotic results, the validity of our theorems for approximating the distribution of the estimator in small samples can be legitimately questioned. In the next section we propose some Monte Carlo experiments illustrating the performance of the estimator in finite samples.

4. NUMERICAL ILLUSTRATION

To gauge the proposed estimation procedure, this section presents a Monte Carlo study of the finite sample properties of the two-stage LS estimator in several situations. There are two experiments conducted in the study. We first consider the Markov-switching process given by

$$X_t = 0.2X_{t-1} + (\sigma_1 + \sigma_2 \Delta_t)Z_t, \tag{13}$$

where (Z_t) is an i.i.d. $\mathcal{N}(0,1)$ process; (Δ_t) is a Markov chain with state space $\{0,1\}$ and transition probabilities $P(\Delta_t = 1/\Delta_{t-1} = 0) = p_{01}$ and $P(\Delta_t = 0/\Delta_{t-1} = 1) = p_{10}$, $0 < p_{10} < 1$, $0 < p_{01} < 1$. In addition (Z_t) and (Δ_t) are independent. Note that the innovation of X_t admits moments up to any order. In light of Section 2, we estimate the weak AR(1)-GARCH(1,1) representation

$$X_t = \phi X_{t-1} + \epsilon_t, \qquad \epsilon_t^2 + \alpha\epsilon_{t-1}^2 = \omega + u_t + \beta u_{t-1}. \tag{14}$$

In Table 1, we report the simulation results for 1,000 replications. The sample size is $n = 1{,}000$, $n = 2{,}000$, or $n = 5{,}000$. Designs 1 and 2 (resp. 1 and 3) are concerned with $\sigma_1 = 0.5$ and $\sigma_2 = 1$ (resp. $p_{10} = p_{01} = 0.1$), designs 3 and 4 (resp. 2 and 4) with $\sigma_1 = 0.5$ and $\sigma_2 = 5$ (resp. $p_{10} = p_{01} = 0.8$). Based on the correlation structure of (ϵ_t^2) (see (8)), the true values of the parameters in (14) are computed as a function of σ_1, σ_2, p_{10}, and p_{01}. For each design we show the mean, the standard deviation, the minimum, the maximum, and the median among the 1,000 estimator values.

Although the estimation of parameter ϕ appears quite accurate, it is seen that for the sample size $n = 1{,}000$, the estimators of the GARCH parameters are biased with high variability. However, it should be noted that, for all designs, the true values of α and β in (14) are very close, which complicates the estimation (because the identifiability assumption is that the lag polynomials in the ARMA equations have no common root). As expected, increasing the sample size improves the results substantially in terms of bias reduction and accuracy. For $n = 2{,}000$ the bias of the estimators is generally on the order of 5% of the parameter value. The results obtained for $n = 5{,}000$ confirm the consistency theorem of Section 3 in this particular case (bias on the order of 1 to 3% in most cases). A significant increase in the precision is also noticeable. Besides, an increase in the difference between the two regimes (i.e., $\sigma_2 = 5$ instead of 1) makes the estimation easier. Moreover, the case of low transitions between the regimes ($p_{10} = p_{01} = 0.1$), generally provides better performances whatever the sample size and the value of σ_2. Finally, note that the median of the estimators is typically closer to the true value than the mean. A possible explanation for this comes from the robustness of the median to outliers. The near nonidentification of the ARMA representations could cause such outliers.

The second experiment consists in simulating the unobserved GARCH model

$$X_t = 0.2X_{t-1} + W_t + \sigma_t Z_t, \qquad \sigma_t^2 = 1 + a(\sigma_{t-1}Z_{t-1})^2 + b\sigma_{t-1}^2, \tag{15}$$

where (Z_t) is i.i.d. $\mathcal{N}(0,1)$, (W_t) is i.i.d. $\mathcal{N}(0,\sigma_W^2)$, and the two processes are independent. Eighteen experiments of 1,000 replications each are conducted for $a = 0.1$, $b = 0.3$ or 0.6, and $\sigma_W^2 = 0$, 0.5, or 1, with the same sizes as in the first example. In this example, the existence of moments implies constraints on a and b. However, for the values taken in the example, we have $E(X_t^8) < \infty$.[7]

Once again, from the results of Section 2, we estimate the weak AR(1)-GARCH(1,1) model (14). To save space, we only report (see Table 2) the outcome concerning the coefficients of the GARCH equation. Note first that the case where $\sigma_W^2 = 0$ corresponds to a strong GARCH(1,1). The estimation results appear quite satisfactory although a large sample size may be needed to achieve great accuracy. This is particularly true with design 5 when the variance of the i.i.d. component is large with respect to that of the strong GARCH process $\sigma_t Z_t$ (which increases with the sum $a + b$). Other experiments not reported here reveal that, when $\sigma_W = 1$, unbiasedness can be observed for larger sample sizes (e.g., $n = 10{,}000$). Again the median is closer to the true values than the mean.

5. COMPARISON OF STRONG AND WEAK GARCH FORECASTS

Because GARCH models remain the most widely used by practitioners, a question of great interest is whether they can actually provide valuable forecasts

Table 1. Switching-regime Markov process

Size	Parameter	Value	Mean	Std. dev.	Min.	Max.	Median
Design 1: $p_{01} = 0.1$, $p_{10} = 0.1$, $\sigma_1 = 0.5$, $\sigma_2 = 1$							
$n = 1{,}000$	ϕ	0.2	0.20140	0.03483	0.08288	0.31910	0.20142
	α	−0.8	−0.63894	0.30127	−0.97913	0.87407	−0.73446
	β	−0.71928	−0.55114	0.30414	−0.95592	0.91657	−0.64536
	ω	0.125	0.22480	0.18971	0.01091	1.19988	0.16628
$n = 2{,}000$	ϕ	0.2	0.19945	0.02525	0.12838	0.28280	0.19965
	α	−0.8	−0.75169	0.13415	−0.95309	0.55086	−0.78447
	β	−0.71928	−0.66717	0.14571	−0.91582	0.63901	−0.69841
	ω	0.125	0.15528	0.08621	0.03050	1.10643	0.13356
$n = 5{,}000$	ϕ	0.2	0.20087	0.01565	0.14620	0.25137	0.20050
	α	−0.8	−0.78678	0.05840	−0.90336	−0.40598	−0.79576
	β	−0.71928	−0.70534	0.06747	−0.84331	−0.28257	−0.71330
	ω	0.125	0.13314	0.03638	0.06281	0.37760	0.12773
Design 2: $p_{01} = 0.8$, $p_{10} = 0.8$, $\sigma_1 = 0.5$, $\sigma_2 = 1$							
$n = 1{,}000$	ϕ	0.2	0.20051	0.02693	0.10933	0.29884	0.20115
	α	0.6	0.43226	0.34500	−0.97318	0.98026	0.52354
	β	0.53353	0.36266	0.34532	−1.00018	0.97331	0.45214
	ω	1	0.89540	0.22394	0.01765	1.31857	0.94186
$n = 2{,}000$	ϕ	0.2	0.19949	0.01884	0.12661	0.25558	0.19918
	α	0.6	0.49131	0.29158	−1.40789	2.06772	0.56295
	β	0.53353	0.42083	0.29871	−1.52476	2.07826	0.49224
	ω	1	0.93257	0.18175	−0.03187	2.06059	0.97036
$n = 5{,}000$	ϕ	0.2	0.20010	0.01255	0.15888	0.23864	0.20024
	α	0.6	0.56800	0.16766	−1.39390	1.21072	0.59809
	β	0.53353	0.50024	0.17335	−1.55701	1.20660	0.53003
	ω	1	0.97909	0.10842	−0.32117	1.22342	0.99721

		Design 3: $p_{01} = 0.1$, $p_{10} = 0.1$, $\sigma_1 = 0.5$, $\sigma_2 = 5$					
$n = 1{,}000$	ϕ	0.2	0.20107	0.04237	0.05434	0.32892	0.20263
	α	−0.8	−0.76594	0.12004	−0.95486	0.18970	−0.79039
	β	−0.67187	−0.63474	0.13751	−0.90483	0.39538	−0.66134
	ω	2.525	2.96807	1.65776	0.62611	15.4942	2.67123
$n = 2{,}000$	ϕ	0.2	0.19900	0.02799	0.08591	0.30190	0.19889
	α	−0.8	−0.79019	0.06144	−0.91539	−0.43149	−0.79763
	β	−0.67187	−0.66077	0.07739	−0.84420	−0.23790	−0.66677
	ω	2.525	2.64981	0.79962	1.06834	7.19805	2.54588
$n = 5{,}000$	ϕ	0.2	0.19939	0.01864	0.13950	0.27344	0.19871
	α	−0.8	−0.79498	0.03573	−0.87814	−0.56793	−0.79753
	β	−0.67187	−0.66594	0.04584	−0.78017	−0.40583	−0.66919
	ω	2.525	2.58872	0.46696	1.44402	5.24673	2.55710
		Design 4: $p_{01} = 0.8$, $p_{10} = 0.8$, $\sigma_1 = 0.5$, $\sigma_2 = 5$					
$n = 1{,}000$	ϕ	0.2	0.19919	0.02045	0.13710	0.26316	0.19922
	α	0.6	0.49779	0.22753	−0.43232	0.94742	0.55678
	β	0.49089	0.38614	0.22888	−0.55747	0.90882	0.43720
	ω	20.2	18.8066	3.08961	7.65502	25.3745	19.4038
$n = 2{,}000$	ϕ	0.2	0.19968	0.01520	0.14222	0.25413	0.19984
	α	0.6	0.55075	0.16228	−0.00848	0.87571	0.58355
	β	0.49089	0.44137	0.16333	−0.10422	0.83397	0.47047
	ω	20.2	19.4588	2.21605	11.8456	24.5062	19.7856
$n = 5{,}000$	ϕ	0.2	0.20014	0.00917	0.17590	0.22902	0.19994
	α	0.6	0.58638	0.12344	−1.25683	2.19215	0.59787
	β	0.49089	0.47618	0.13724	−1.89863	2.16598	0.48755
	ω	20.2	20.0003	1.56992	−2.21781	34.9662	20.1046

Table 2. Unobserved GARCH process

Size	Parameter	Value	Mean	Std. dev.	Min.	Max.	Median
			Design 1: $\sigma_W^2 = 0$, $a = 0.1$, $b = 0.3$				
$n = 1{,}000$	$-\alpha$	0.4	0.29823	0.30831	−0.82380	0.97424	0.33915
	$-\beta$	0.3	0.20059	0.31172	−0.86275	0.95736	0.22940
	ω	1	1.16535	0.51245	0.04507	3.12467	1.09320
$n = 2{,}000$	$-\alpha$	0.4	0.36023	0.23030	−0.72932	1.23820	0.38182
	$-\beta$	0.3	0.26212	0.23489	−0.79561	1.13403	0.28285
	ω	1	1.06369	0.38074	−0.30585	2.79366	1.02521
$n = 5{,}000$	$-\alpha$	0.4	0.37931	0.13107	−0.09891	0.77063	0.39011
	β	0.3	0.28029	0.13671	−0.18893	0.70836	0.28474
	$-\omega$	1	1.03383	0.21793	0.37901	1.79815	1.01642
			Design 2: $\sigma_W^2 = 0$, $a = 0.1$, $b = 0.6$				
$n = 1{,}000$	$-\alpha$	0.7	0.58350	0.25720	−0.64892	0.97403	0.65134
	$-\beta$	0.6	0.48383	0.26640	−0.69887	0.94476	0.54408
	ω	1	1.38063	0.84472	0.08846	5.31795	1.15729
$n = 2{,}000$	$-\alpha$	0.7	0.64211	0.17471	−0.89081	0.97791	0.67760
	$-\beta$	0.6	0.54356	0.18126	−0.79294	0.96005	0.581564
	ω	1	1.18766	0.57168	0.07760	5.97737	1.06450
$n = 5{,}000$	$-\alpha$	0.7	0.68257	0.08268	0.01680	0.89930	0.69587
	$-\beta$	0.6	0.58253	0.09200	−0.03021	0.81264	0.59501
	ω	1	1.05673	0.27575	−0.37969	2.94330	1.01655
			Design 3: $\sigma_W^2 = 0.5$, $a = 0.1$, $b = 0.3$				
$n = 1{,}000$	$-\alpha$	0.4	0.22126	0.40761	−1.00338	0.98778	0.26878
	$-\beta$	0.34061	0.16338	0.40471	−0.99814	1.00261	0.19917
	ω	1.3	1.68033	0.87136	0.02546	4.27782	1.57302
$n = 2{,}000$	$-\alpha$	0.4	0.27334	0.33074	−0.96289	0.99412	0.30965
	$-\beta$	0.34061	0.21324	0.33078	−0.980411	0.99150	0.24178
	ω	1.3	1.57187	0.71003	0.01321	4.24562	1.48998

n = 5,000	$-\alpha$	0.4	0.36322	0.21197	−0.74862	0.98524	0.38921
	$-\beta$	0.34061	0.30313	0.21509	−0.78349	0.90865	0.32943
	ω	1.3	1.37804	0.45795	0.02107	3.66004	1.32416
			Design 4: $\sigma_W^2 = 0.5$, $a = 0.1$, $b = 0.6$				
n = 1,000	$-\alpha$	0.7	0.48222	0.34517	−0.95114	0.98148	0.59537
	$-\beta$	0.62494	0.40325	0.34269	−0.94758	0.97175	0.49728
	ω	1.15	1.97426	1.30439	0.06629	7.38536	1.55361
n = 2,000	$-\alpha$	0.7	0.60748	0.22965	−0.46322	1.08944	0.67231
	$-\beta$	0.62494	0.52896	0.23410	−0.56885	1.02812	0.58945
	ω	1.15	1.50274	0.87125	−0.33612	5.86556	1.25644
n = 5,000	$-\alpha$	0.7	0.67487	0.12422	−0.62096	0.89940	0.69607
	$-\beta$	0.62494	0.59656	0.13059	−0.69260	0.86537	0.61573
	ω	1.15	1.24454	0.47133	0.37894	6.17465	1.15961
			Design 5: $\sigma_W^2 = 1$, $a = 0.1$, $b = 0.3$				
n = 1,000	$-\alpha$	0.4	0.15011	0.41069	−0.94447	0.98951	0.06928
	$-\beta$	0.36066	0.10955	0.40932	−0.95626	0.98473	0.05990
	ω	1.6	2.25747	1.08690	0.02929	5.44834	2.38113
n = 2,000	$-\alpha$	0.4	0.19028	0.34755	−0.92548	1.18878	0.06831
	$-\beta$	0.36066	0.14967	0.34501	−0.94998	1.08105	0.06340
	ω	1.6	2.15347	0.92492	−0.51249	5.51787	2.38144
n = 5,000	$-\alpha$	0.4	0.24965	0.26827	−0.72320	0.91170	0.18316
	$-\beta$	0.36066	0.20880	0.26669	−0.77159	0.88827	0.13137
	ω	1.6	2.00143	0.71485	0.23908	4.59234	2.21131
			Design 6: $\sigma_W^2 = 1$, $a = 0.1$, $b = 0.6$				
n = 1,000	$-\alpha$	0.7	0.40740	0.36954	−0.97243	0.99310	0.51196
	$-\beta$	0.64143	0.34322	0.36259	−0.97168	1.00208	0.43046
	ω	1.3	2.54335	1.57383	0.03315	8.09213	2.11881
n = 2,000	$-\alpha$	0.7	0.53421	0.28828	−0.85637	1.24764	0.63935
	$-\beta$	0.64143	0.46892	0.28615	−0.73379	1.19732	0.56459
	ω	1.3	2.01531	1.23452	−0.95381	7.53990	1.57493
n = 5,000	$-\alpha$	0.7	0.62312	0.20407	−0.23986	0.91198	0.67974
	$-\beta$	0.64143	0.55822	0.20638	−0.23503	0.87724	0.61164
	ω	1.3	1.63357	0.88070	0.38799	5.09043	1.39956

under possible misspecification (i.e., when the strong assumptions do not hold). Most empirical studies based on GARCH-type models work with estimators computed from the following conditional (on initial conditions) Gaussian quasi-loglikelihood:

$$\mathcal{L} = -\frac{1}{2}\sum_{t=1}^{n}\log\sigma_t^2 - \frac{1}{2}\sum_{t=1}^{n}\frac{\epsilon_t^2}{\sigma_t^2},$$

where the ϵ_t's are obtained from (1) and σ_t^2 is computed recursively from

$$\sigma_t^2 = c + \sum_{i=1}^{q'} a_i\epsilon_{t-i}^2 + \sum_{i=1}^{p'} b_i\sigma_{t-i}^2 \tag{16}$$

under the positivity constraints $c > 0$, $a_i \geq 0$, $b_i \geq 0$. Letting $v_t = \epsilon_t^2 - \sigma_t^2$ and $r = \max(p', q')$, we can write

$$\epsilon_t^2 = c + \sum_{i=1}^{r}(a_i + b_i)\epsilon_{t-i}^2 + v_t - \sum_{i=1}^{p'} b_i v_{t-i}. \tag{17}$$

In the case of correct specification of the conditional variance σ_t^2, (17) can be used to derive the optimal forecasts of ϵ_t^2 that enjoy the property of being linear functions of past value of ϵ_t^2. Now, if the DGP only admits a weak ARMA-GARCH representation (i.e., if the conditional variance given by (16) is misspecified) the QML procedure can serve as a device for its estimation: the parameters of (16) are estimated and then plugged into (17). An equivalent approach has been followed by Drost and Nijman (1993).

To achieve insight into the asymptotic properties of this QML-based procedure, let us consider the expectation of the quasi score, evaluated at the true parameter value θ_0 of the weak GARCH representation

$$\begin{cases} E_{\theta_0}\left[\frac{1}{n}\frac{\partial\mathcal{L}}{\partial\theta^{(1)}}(\theta_0)\right] = E_{\theta_0}\left[\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2}{\partial\theta^{(1)}}(\theta_0)u_t\right] - E_{\theta_0}\left[\frac{\epsilon_t}{\sigma_t^2(\theta_0)}\frac{\partial\epsilon_t}{\partial\theta^{(1)}}\right], \\ E_{\theta_0}\left[\frac{1}{n}\frac{\partial\mathcal{L}}{\partial\theta^{(2)}}(\theta_0)\right] = E_{\theta_0}\left[\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2}{\partial\theta^{(2)}}(\theta_0)u_t\right]. \end{cases} \tag{18}$$

Because u_t is only the *linear* innovation of ϵ_t^2, and because the term in brackets in the last equality is obviously a nonlinear function of past values of ϵ_t^2, the expectation in the second equality will not vanish in general. The same conclusion holds for the first equality. Therefore, the estimators computed from the QML equations will unfortunately be inconsistent in general.

Of course if, in (18), (u_t) is a martingale difference sequence, then under some regularity conditions, the parameters can be estimated consistently by QML. In addition, the two-stage estimator proposed in this paper is likely to be inefficient relative to the QML estimator. However it should be emphasized that, even in the case of a martingale difference for (u_t), equation (2) is not necessarily compatible with a strong GARCH. Indeed, the parameter space must be constrained to ensure that σ_t^2 is positive (e.g., by imposing the classical nonnegativity constraint on the a_i's and b_i's).

To illustrate these issues, we have performed four numerical experiments designed to compare the ϵ_t^2 forecast errors. In each experiment, we consider a DGP compatible with the weak GARCH(1,1) structure

$$\epsilon_t^2 + \alpha\epsilon_{t-1}^2 = \omega + u_t + \beta u_{t-1}, \tag{19}$$

where, for ease of notation throughout the section, $\theta = (\alpha, \beta, \omega)'$ denotes the generic parameter vector and $\theta_0 = (\alpha_0, \beta_0, \omega_0)'$ denotes the true value. The linear prediction of ϵ_t^2 is given by

$$\hat{\epsilon}_t^2 = \frac{\omega}{1+\beta} + (\beta - \alpha)\sum_{i=0}^{\infty}(-\beta)^i \epsilon_{t-i-1}^2,$$

from which we can compute the mean-squared prediction error,

$$\begin{aligned} \text{MSE}(\theta) &:= E_{\theta_0}(\epsilon_t^2 - \hat{\epsilon}_t^2)^2 \\ &= \left[\frac{\omega_0(1+\alpha) - \omega(1+\alpha_0)}{(1+\beta)(1+\alpha_0)}\right]^2 + \left(1 + \frac{(\beta-\alpha)^2}{1-\beta^2}\right)\gamma(0) \\ &\quad - \frac{2(\beta-\alpha)(1-\alpha\beta)}{(1-\alpha_0\beta)(1-\beta^2)}\gamma(1), \end{aligned}$$

where $\gamma(0)$ and $\gamma(1)$ are, respectively, the variance and first autocovariance of the (ϵ_t^2) process. Of course this function is minimal for $\theta = \theta_0$ and is equal to $E(u_t^2)$ at this point.

For different sample sizes ($n = 100, 200, \ldots, 1{,}000$), we simulate 100 trajectories of the DGP. Then, for each trajectory, we estimate the weak GARCH(1,1) model (19) by the following: (i) the method proposed in the paper; (ii) QML. In the latter case, we estimate a strong GARCH; then we derive the corresponding weak GARCH model (17). For each estimated model, the mean-squared error (MSE) is computed and averaged across the 100 replications. We use it to compare the prediction errors resulting from both methods.

In the first experiment, the data were simulated from the strong GARCH(1,1) given by

$$\epsilon_t = \sigma_t Z_t, \qquad \sigma_t^2 = 1 + 0.4\epsilon_{t-1}^2 + 0.4\sigma_{t-1}^2, \tag{20}$$

where Z_t is an i.i.d. $\mathcal{N}(0,1)$ sequence. From (19), the true parameter value of the weak GARCH representation is $\theta_0 = (-0.8, -0.4, 1)$, and we have $V(u_t) = 450$. The estimation results are displayed in Figure 1. It shows that, unsurprisingly, for this strong GARCH, the QML method is more efficient than our two-stage estimation method, whatever the sample size.

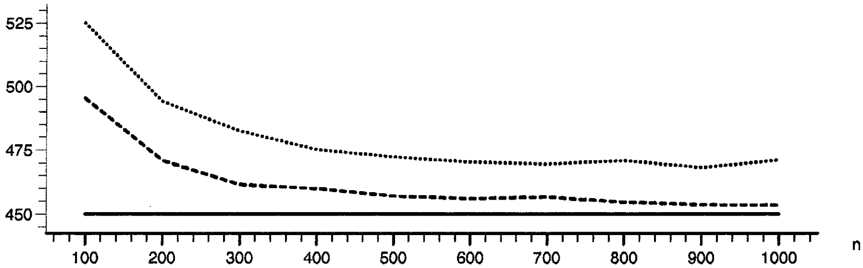

FIGURE 1. Comparison between the mean-squared error of prediction obtained with a strong GARCH model (dashed line) and a weak GARCH model (dotted line), as a function of the size n of a simulated trajectory of the strong GARCH(1,1) model (20). The full line corresponds to the minimal mean-squared error of prediction.

The second experiment deals with temporal aggregation. The DGP (20) has been maintained but we only use the low frequency observations ϵ_{2t}. The corresponding weak GARCH representation is given by

$$\epsilon_{2t}^2 + 0.64\epsilon_{2(t-1)}^2 = 1.8 + u_t - 0.466u_{t-1}, \qquad V(u_t) \simeq 618.24. \tag{21}$$

The results presented in Figure 2 confirm the observation of Drost and Nijman (1993, p. 922) that, for this weak GARCH model derived from a strong GARCH, "the asymptotic bias of the QMLE, if there is any, is very small." The next two examples will reveal that other weak GARCH models are not adequately estimated by QML.

In the third experiment, the DGP is given by

$$\epsilon_t = \nu_t \nu_{t-1}, \tag{22}$$

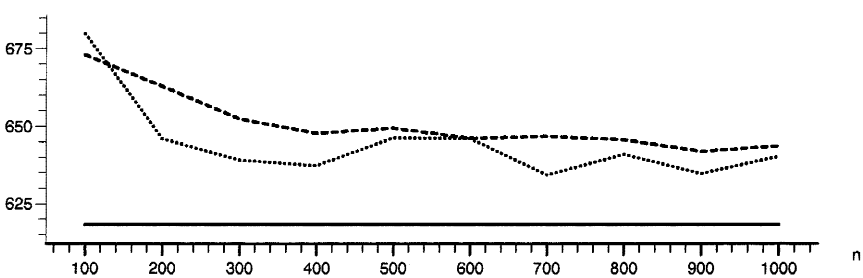

FIGURE 2. Comparison between the mean-squared error of prediction obtained with a strong GARCH model (dashed line) and a weak GARCH model (dotted line), as a function of the size n of a simulated trajectory of (ϵ_{2t}), where (ϵ_t) is the solution to model (20). The full line corresponds to the minimal mean-squared error of prediction.

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

where (ν_t) is an i.i.d. $\mathcal{N}(0,1)$ sequence. We know from Example 6 that (ϵ_t^2) is the MA(1):

$$\epsilon_t^2 = 1 + u_t + (2 - \sqrt{3})u_{t-1}, \tag{23}$$

where (u_t) is a white noise with variance equal to $2/(2 - \sqrt{3}) \simeq 7.464$. Figure 3 shows that the MSE's obtained by QML are very high, even for large samples. As a consequence, in this context, strong GARCH predictions are very poor. On the contrary, the MSE's deduced from the two-stage procedure appear satisfactory and seem to converge to the optimal value.

In the previous example, the best strong GARCH predictions were obtained for a model with a strictly negative coefficient (namely, $b = \sqrt{3} - 2$). In this sense, the weak GARCH equation was not compatible with a standard strong GARCH model, which may explain the failure of the QML procedure. In the next example, such a problem does not occur because the weak ARCH model is compatible with a standard strong ARCH model. Consider the stationary process (ξ_t^2) defined by

$$\xi_t^2 = 1 + u_t + 0.7\xi_{t-1}^2, \tag{24}$$

where (u_t) is the white noise defined by (23) and (22).[8] It is worth noting that the predictions of ξ_t^2 based on equation (24) coincide with those obtained from the strong ARCH(1) defined by $\sigma_t^2 = c + a\xi_{t-1}^2$, with $c = 1$, $a = 0.7$. However, because (u_t) is not a martingale difference sequence, the QMLE fails to converge to these optimal parameters. Figures 4 and 5 show that the strong ARCH predictions are much less accurate than those based on the weak ARCH representation, even for very large samples. For $n = 10{,}000, \ldots, n = 100{,}000$, the estimates of c and a are always close to 0.7 and 0.8, respectively.

Apart from considering the averaged MSE's, it is also of interest to compare the distributions of the 100 MSE's for the two estimation methods. In Table 3

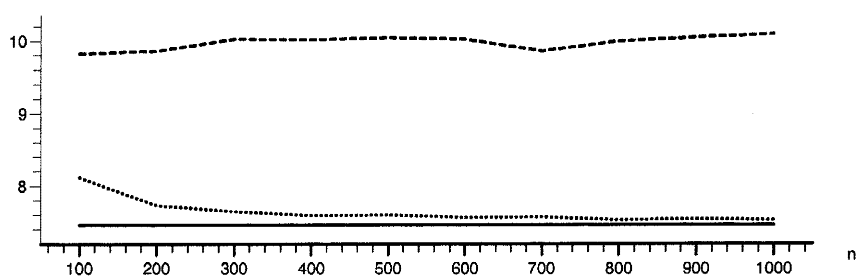

FIGURE 3. Comparison between the mean-squared error of prediction obtained with a strong GARCH model (dashed line) and a weak GARCH model (dotted line), as a function of the size n of a simulated trajectory of the weak GARCH(0,1) model (23). The full line corresponds to the minimal mean-squared error of prediction.

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

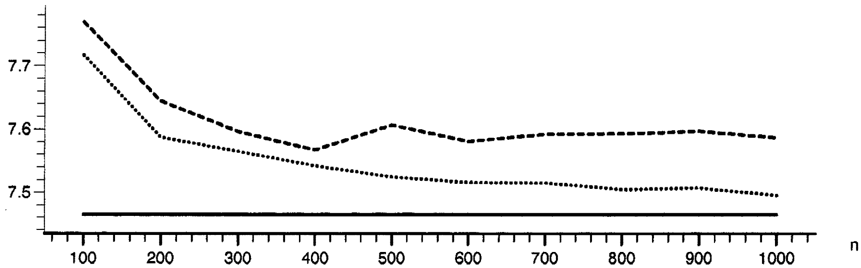

Figure 4. Comparison between the mean-squared error of prediction obtained with a strong GARCH model (dashed line) and a weak GARCH model (dotted line), as a function of the size n of a simulated trajectory of the weak ARCH(1) model (24). The full line corresponds to the minimal mean-squared error of prediction.

we have reported, for each model, the percentage of cases in which the MSE's obtained by QML are less than the MSE's obtained by our method. We have also tested the hypothesis that the median of the difference between the two MSE's is equal to zero against the alternative hypothesis that the median of the MSE obtained by QML is less than that obtained by the two-stage method. A Wilcoxon one-sample (matched pairs) signed ranks test was performed using the NAG routine G08AGF. The asymptotic p-values are given in parentheses. The results reported in Table 3 show that for the strong GARCH model (20), the superiority of the QML over the two-stage method is significant. Conversely, in the experiments where the DGP is not a strong GARCH, the hypothesis that the QML is not inferior to the two-stage method is rejected at any reasonable significance level (apart from the cases $n = 100$ and $n = 400$ in model (24)). It is worth noting that even in the second experiment (temporal

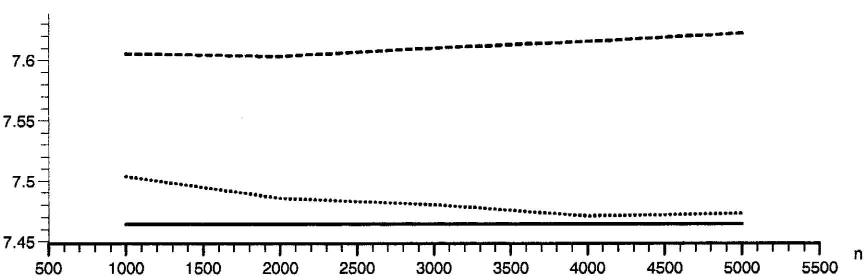

Figure 5. Comparison between the mean-squared error of prediction obtained with a strong GARCH model (dashed line) and a weak GARCH model (dotted line), as a function of the size n of a simulated trajectory of the weak ARCH(1) model (24), for sample sizes $n = 1{,}000, \ldots, n = 5{,}000$. The full line corresponds to the minimal mean-squared error of prediction.

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

TABLE 3. Percentage of cases in which the MSE's obtained by QML are less than the MSE's obtained by the two-stage method (p-values of the Wilcoxon rank test for the equality of the medians in parentheses)

	n				
Model	100	200	300	400	500
(20)	82 (1.000)	78 (1.000)	84 (1.000)	82 (1.000)	87 (1.000)
(21)	40 (.005)	35 (.000)	37 (.000)	32 (.000)	38 (.001)
(23)	22 (.000)	9 (.000)	3 (.000)	2 (.000)	3 (.000)
(24)	46 (.041)	44 (.015)	42 (.018)	47 (.157)	36 (.000)
	n				
Model	600	700	800	900	1,000
(20)	83 (1.000)	86 (1.000)	87 (1.000)	95 (1.000)	88 (1.000)
(21)	37 (.002)	30 (.000)	26 (.000)	25 (.000)	32 (.000)
(23)	0 (.000)	1 (.000)	2 (.000)	0 (.000)	0 (.000)
(24)	31 (.000)	31 (.000)	26 (.000)	22 (.000)	15 (.000)

aggregation), the differences between the two estimation methods are substantial, whatever the sample size.

Another view of the performance of the two estimation methods may be obtained through an out-of-sample forecasting experiment. The Monte Carlo experiment involves 10,000 replications of model (22)–(23). The sample sizes vary from 100 to 1,000. For each replication, we have estimated the weak GARCH representation by the two methods. Then we computed two one-step-ahead predictions based on the two methods. These two predictions were compared to one out-of-sample value. Averaging the squared prediction errors over the 10,000 replications leads to the results reported in Figure 6. They confirm that there can be notable gains in precision from using the two-stage estimation method instead of QML estimation. Similar out-of-sample experiments conducted with the other models led to the same findings as in the in-sample experiments. Therefore, they are not reported here.

In summary, several conclusions can be drawn from the experiments presented in this section. First, as we know from the very beginning, the QML is certainly more efficient than the two-stage method when the DGP is a strong GARCH, and, consequently, the predictions of the squares are more accurate. Second, weak GARCH representations can be incompatible with strong GARCH subject to positivity constraints. Imposing these constraints in QML estimation can therefore lead to very poor approximations of the DGP. Finally, even if no positivity problems occur, i.e., when the weak GARCH representation could be derived from a strong GARCH model, the QML may fail to estimate this opti-

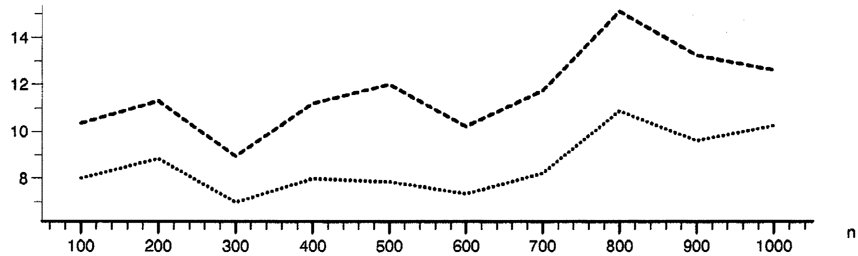

FIGURE 6. Out-of-sample one-step-ahead MSE's: comparison between a strong GARCH model (dashed line) and a weak GARCH model (dotted line), as function of the size n of a simulated trajectory of model (22)–(23).

mal (although misspecified) strong GARCH because the martingale difference assumptions do not hold. As a consequence, the predictive accuracy of the estimated strong GARCH can be very poor.

6. CONCLUSION

Of course, as are all statistical models, ARCH models are merely an approximation of the true DGP. In applied work, finding autocorrelation in the squared ordinary LS residuals is generally interpreted as evidence for the presence of ARCH. A misspecified model, however, will typically be selected because, as we have seen in this paper, such autocorrelation structure is compatible with severe misspecifications of strong GARCH.

This paper has proposed an asymptotic theory for weak ARMA-GARCH representations, using the principle of two-stage LS. The method is very simple to implement, and it is already used by practitioners. However, using confidence intervals based on strong assumptions can be misleading. In particular standard identification routines based on strong hypothesis on the innovation of the squared ordinary LS residuals can result in serious misspecifications. Most significantly, our approach can potentially serve as a basis for selecting and estimating some more specific classes of stochastic conditional variance models. We hope to report results on this topic in the near future.

NOTES

1. El Babsiri and Zakoïan (2000) have documented some asymmetry features of stock returns that are different from the so-called leverage effect. They typically found significant correlations between the current innovation and its past positive and/or negative parts. In addition they showed that conditional skewness and kurtosis can have huge fluctuations over time, just as conditional variances.

2. Recent papers that study asymptotic properties of a QMLE with different requirements on the finiteness of the unconditional variance or the assumed innovation density include Bollerslev

and Wooldridge (1992), Lee and Hansen (1994), Lumsdaine (1996), and Newey and Steigerwald (1997).

3. See however Leroux (1992), Rydén (1994), and Francq and Roussignol (1997) for conditions ensuring consistency of the maximum likelihood and the asymptotic normality of a sequence of pseudo-likelihood estimators.

4. It seems that our approach can be connected with that of Nelson (1992) and Nelson and Foster (1995). They have shown that, even when misspecified, a sequence of GARCH models can consistently estimate the underlying conditional variance of a (near-)diffusion (for increasingly higher sampling frequencies). The present paper also addresses misspecification of GARCH models but with a quite different perspective.

5. To see this, consider the EGARCH(1,1) model defined by $\epsilon_t = \sigma_t Z_t$, $\log \sigma_t^2 = c + b \log \sigma_{t-1}^2 + aZ_{t-1}$, where $|b| < 1$ and (Z_t) is an i.i.d. $N(0,1)$ process. Tedious computations show that $\forall h > 0$

$$\operatorname{Cov}(\epsilon_t^2, \epsilon_{t-h}^2) = \exp\left[\frac{2c}{1-b} + \frac{a^2}{(1-b)^2}\right]\left\{[1 + a^2 b^{2(h-1)}]\exp\left[\frac{a^2 b^h}{(1-b)^2}\right] - 1\right\}.$$

6. These results cannot be straightforwardly extended to the current context. However, we conjecture that a proof of consistency can be obtained along the same lines and leave it for future research.

7. The necessary and sufficient condition for existence of $E(X^8)$ is $105a^4 + b^4 + 18a^2b^2 + 4ab^3 + 60a^3b < 1$.

8. Easy computations show that

$$\xi_t^2 = \frac{1}{0.3} + \frac{\sqrt{3} - 2.7}{\sqrt{3} - 2} \sum_{h=0}^{\infty} \left[(\sqrt{3} - 2)^h - \frac{0.7^{h+1}}{\sqrt{3} - 2}\right](\epsilon_{t-h}^2 - 1) > 0.7044.$$

REFERENCES

Andersen, T.G. (1994) Stochastic autoregressive volatility: A framework for volatility modeling. *Mathematical Finance* 4, 103–119.

Andersen, T.G. & B.E. Sørensen (1996) GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 14, 328–352.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series.* New York: Wiley.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Baillie, R.T., T.P. Bollerslev, & H.O. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3–30.

Bollerslev, T.P. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 309–328.

Bollerslev, T.P., R.F. Engle, & D.B. Nelson (1994) ARCH Models. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2959–3038.

Bollerslev, T.P. & J.M. Wooldridge (1992) Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews* 11, 143–172.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods.* Springer-Verlag.

Broze, L., O. Scaillet, & J.M. Zakoïan (1998) Quasi-indirect inference for diffusion processes. *Econometric Theory* 14, 161–186.

Cai, J. (1994) A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics* 12, 309–316.

Davydov, Y.A. (1968) Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and Applications* 13, 691–696.

Diebolt, J. & D. Guégan (1991) Le modèle de série chronologique autorégressive β-ARCH. *Comptes Rendus de l'Academie des Sciences de Paris*, Série I, 625–630.
Drost, F.C. & T. Nijman (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.
Drost, F.C. & B.J.M. Werker (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics* 74, 31–57.
Dueker, M.J. (1997) Markov switching in GARCH processes and mean-reverting stock market volatility. *Journal of Business and Economic Statistics* 15, 26–34.
El Babsiri, M. & J.M. Zakoïan (2000) Contemporaneous asymmetry in GARCH processes. *Journal of Econometrics*, forthcoming.
Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.
Francq, C. & M. Roussignol (1997) On white noises driven by hidden Markov chains. *Journal of Time Series Analysis* 18, 553–578.
Francq, C. & J.M. Zakoïan (1997) Estimating weak GARCH representations. CREST Discussion paper 9740.
Francq, C. & J.M. Zakoïan (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–165.
Francq, C. & J.M. Zakoïan (2000) Covariance matrix estimation for estimators of mixing Wold's ARMA. *Journal of Statistical Planning and Inference* 83, 369–394.
Gallant, A.R. & G.E. Tauchen (1996) Which moments of match? *Econometric Theory* 12, 657–681.
Ghysels, E., A.C. Harvey, & E. Renault (1996) Stochastic volatility. In C.R. Rao & G.S. Maddala (eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.
Gouriéroux, C., A. Monfort, & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.
Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.
Hamilton, J.D. & R. Susmel (1994) Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307–333.
Hansen, B.E. (1992) Heteroskedasticity and autocorrelation consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.
Hansen, L.P. & K.J. Singleton (1996) Efficient estimation of linear asset-pricing models with moving average errors. *Journal of Business and Economic Statistics* 14, 53–68.
Harvey, A.C., E. Ruiz, & E. Sentana (1992) Unobserved component time series models with ARCH disturbances. *Journal of Econometrics* 52, 129–158.
Ibragimov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and Applications* 7, 349–382.
King, M., E. Sentana, & S. Wadhwani (1994) Volatility and links between national stock markets. *Econometrica* 62, 901–933.
Lee, S.W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–58.
Leroux, B. (1992) Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and Their Applications* 40, 127–143.
Lumsdaine, R. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.
Meddahi, N. & E. Renault (1996) Aggregations and Marginalizations of GARCH and Stochastic Volatility Models. Discussion paper, GRENAQ.
Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
Nelson, D.B. (1992) Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model. *Journal of Econometrics* 52, 61–90.

Nelson, D.B. & C.Q. Cao (1992) Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* 10, 229–235.
Nelson, D.B. & D. Foster (1995) Filtering and forecasting with misspecified ARCH models II: Making the right forecast with the wrong model. *Journal of Econometrics* 67, 303–335.
Newey, W.K. & D.G. Steigerwald (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65 (3), 587–599.
Newey, W.K. & K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
Nijman, T. & E. Sentana (1996) Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71, 71–80.
Pagan, A.R. & G.W. Schwert (1990) Alternative models for conditional stock volatility, *Journal of Econometrics* 45, 267–290.
Ruiz, E. (1994) Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63, 289–306.
Rydén, T. (1994) Consistent and asymptotically normal parameters estimates for hidden Markov models. *Annals of Statistics* 22, 1884–1895.
Sentana, E. (1995) Quadratic ARCH models. *Review of Economic Studies* 62, 639–661.
Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D.V. Hinkley, & O.E. Barndorff-Nielsen (eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman and Hall.
Weiss, A.A. (1984) ARMA models with ARCH errors. *Journal of Time Series Analysis* 5, 129–143.
Weiss, A.A. (1986) Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2, 107–131.

APPENDIX: PROOFS OF THEOREMS 1 AND 2

Proof of Theorem 1. The theorem relies on a set of intermediate results that we now present. It will be convenient to consider the functions ($\forall\theta \in \Theta$): $O_n^{(1)}(\theta^{(1)}) = (1/n)\sum_{t=1}^{n}\epsilon_t^2(\theta^{(1)})$ and $O_n^{(2)}(\theta) = (1/n)\sum_{t=1}^{n}u_t^2(\theta)$, where $(\epsilon_t(\theta^{(1)}), u_t(\theta))$ is given by (9)–(10). We use $\|\cdot\|$ to signify the Euclidean norm $\|A\| = \{\text{tr}(A'A)\}^{1/2}$ and $\|\cdot\|_p$ to signify the L_p-norm $\|A\|_p = (E\|A\|^p)^{1/p}$. The first lemma was established by Francq and Zakoïan (1998, Theorem 1).

LEMMA 1. $\hat{\theta}_n^{(1)} \to \theta_0^{(1)}$ *a.s. as* $n \to \infty$.

LEMMA 2. *For any* $\theta \in \Theta_\delta$ *and any* $t \in Z$,

$$\begin{cases}\epsilon_t(\theta^{(1)}) = \epsilon_t & a.s.\\ u_t(\theta) = u_t & a.s.\end{cases} \Rightarrow \theta = \theta_0. \tag{A.1}$$

Proof. The proof that $\epsilon_t(\theta^{(1)}) = \epsilon_t$ (a.s.) $\Rightarrow \theta^{(1)} = \theta_0^{(1)}$ was given in Francq and Zakoïan (1998). The rest of the proof is similarly based on the innovation property of u_t along with the fact that representation (2) is the canonical one. ■

LEMMA 3. *For any* $\theta^{(2)} \in \Theta_\delta^{(2)}$, *let* $O_\infty^{(2)}(\theta_0^{(1)},\theta^{(2)}) = E_{\theta_0}u_t^2(\theta_0^{(1)},\theta^{(2)})$. *Then for any* $\theta^{(2)} \neq \theta_0^{(2)}$, $\theta^{(2)} \in \Theta_\delta^{(2)}$, *we have*

$$\zeta^2 = O_\infty^{(2)}(\theta_0^{(1)},\theta_0^{(2)}) < O_\infty^{(2)}(\theta_0^{(1)},\theta^{(2)}).$$

Proof. The lemma is a straightforward consequence of the innovation property of u_t and of Lemma 2. ■

LEMMA 4. *For any $\theta \in \Theta_\delta$ there exist two sequences of absolutely summable constants $(c_i(\theta^{(1)}))$ and $(d_i(\theta^{(2)}))$ and a constant $w(\theta^{(2)}))$ such that a.s. $(\forall t \in \mathbb{Z})$:*

$$\epsilon_t(\theta^{(1)}) = X_t + \sum_{i=1}^{\infty} c_i(\theta^{(1)}) X_{t-i} \tag{A.2}$$

and

$$u_t(\theta) = w(\theta^{(2)}) + \epsilon_t^2(\theta^{(1)}) + \sum_{i=1}^{\infty} d_i(\theta^{(2)}) \epsilon_{t-i}^2(\theta^{(1)}). \tag{A.3}$$

In addition, $\epsilon_t(\cdot)$ and $(\forall \theta^{(1)} \in \Theta_\delta^{(1)}) u_t(\theta^{(1)}, \cdot)$ are continuously differentiable functions, and for any $\theta \in \Theta_\delta$, any $m_1 \in \{1,\dots,P+Q\}$, and any $m_2 \in \{1,\dots,p+q+1\}$, there exist a constant $w_{m_2}(\theta^{(2)})$ and two absolutely summable sequences $(c_{i,m_1}(\theta^{(1)}))_{i\geq 1}$ and $(d_{i,m_2}(\theta^{(2)}))_{i\geq 0}$ such that

$$\frac{\partial}{\partial \theta_{m_1}^{(1)}} \epsilon_t(\theta^{(1)}) = \sum_{i=1}^{\infty} c_{i,m_1}(\theta^{(1)}) X_{t-i} \quad \text{and}$$

$$\frac{\partial}{\partial \theta_{m_2}^{(2)}} u_t(\theta) = w_{m_2}(\theta^{(2)}) + \sum_{i=1}^{\infty} d_{i,m_2}(\theta^{(2)}) \epsilon_{t-i}^2(\theta^{(1)}). \tag{A.4}$$

Moreover there exist $\rho \in [0,1[$ and $K \in [0,\infty[$ such that, for all $i \geq 1$,

$$\sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} |c_i(\theta^{(1)})| \leq K\rho^i, \qquad \sup_{\theta^{(2)} \in \Theta_\delta^{(2)}} |d_i(\theta^{(2)})| \leq K\rho^i, \tag{A.5}$$

$$\sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} |c_{i,m_1}(\theta^{(1)})| \leq K\rho^i, \qquad \sup_{\theta^{(2)} \in \Theta_\delta^{(2)}} |d_{i,m_2}(\theta^{(2)})| \leq K\rho^i. \tag{A.6}$$

Proof. Again, this is a straightforward consequence of the invertibility assumptions on the MA polynomials in (1)–(2). Similar results have been established in Francq and Zakoïan (1998); therefore we do not detail the proof.

Now we show the following almost sure uniform convergence result.

LEMMA 5. *We have a.s.*

$$\lim_{n\to\infty} \sup_{\theta^{(2)} \in \Theta_\delta^{(2)}} |O_n^{(2)}(\hat{\theta}_n^{(1)}, \theta^{(2)}) - O_n^{(2)}(\theta_0^{(1)}, \theta^{(2)})| = 0.$$

Proof. From (A.3) we have

$$\begin{aligned}
&|u_t^2(\hat{\theta}_n^{(1)}, \theta^{(2)}) - u_t^2(\theta_0^{(1)}, \theta^{(2)})| \\
&\quad = \left| \epsilon_t^2(\hat{\theta}_n^{(1)}) - \epsilon_t^2 + \sum_{i=1}^{\infty} d_i(\theta^{(2)}) [\epsilon_{t-i}^2(\hat{\theta}_n^{(1)}) - \epsilon_{t-i}^2] \right| |u_t(\hat{\theta}_n^{(1)}, \theta^{(2)}) + u_t(\theta_0^{(1)}, \theta^{(2)})| \\
&\quad \leq \left[|\epsilon_t(\hat{\theta}_n^{(1)}) - \epsilon_t| |\epsilon_t(\hat{\theta}_n^{(1)}) + \epsilon_t| + \sum_{i=1}^{\infty} |d_i(\theta^{(2)})| |\epsilon_{t-i}(\hat{\theta}_n^{(1)}) - \epsilon_{t-i}| |\epsilon_{t-i}(\hat{\theta}_n^{(1)}) + \epsilon_{t-i}| \right] \\
&\qquad \times |u_t(\hat{\theta}_n^{(1)}, \theta^{(2)}) + u_t(\theta_0^{(1)}, \theta^{(2)})|.
\end{aligned} \tag{A.7}$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

Now using a Taylor expansion we can write

$$|\epsilon_t(\hat{\theta}_n^{(1)}) - \epsilon_t| \le \left\| \frac{\partial}{\partial \theta^{(1)}} (\epsilon_t(\tilde{\theta}_{t,n}^{(1)})) \right\| \times \|\hat{\theta}_n^{(1)} - \theta_0^{(1)}\|, \quad \textbf{(A.8)}$$

where $\tilde{\theta}_{t,n}^{(1)}$ is between $\hat{\theta}_n^{(1)}$ and $\theta_0^{(1)}$. We deduce from (A.7) and (A.8) that

$$\begin{aligned} &\sup_{\theta^{(2)} \in \Theta_\delta^{(2)}} |O_n^{(2)}(\hat{\theta}_n^{(1)}, \theta^{(2)}) - O_n^{(2)}(\theta_0^{(1)}, \theta^{(2)})| \\ &\quad \le \Bigg[\frac{2}{n} \sum_{t=1}^{n} \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} \left(\left\| \frac{\partial}{\partial \theta^{(1)}} (\epsilon_t(\theta^{(1)})) \right\| \right) \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} |\epsilon_t(\theta^{(1)})| W_t \\ &\qquad + \frac{2}{n} \sum_{t=1}^{n} \sum_{i=1}^{\infty} \sup_{\theta^{(2)} \in \Theta_\delta^{(2)}} |d_i(\theta^{(2)})| \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} \left(\left\| \frac{\partial}{\partial \theta^{(1)}} (\epsilon_{t-i}(\theta^{(1)})) \right\| \right) \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} |\epsilon_{t-i}(\theta^{(1)})| W_t \Bigg] \\ &\qquad \times \|\hat{\theta}_n^{(1)} - \theta_0^{(1)}\|, \end{aligned} \quad \textbf{(A.9)}$$

where $W_t = \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} \sup_{\theta^{(2)} \in \Theta_\delta^{(2)}} |u_t(\theta^{(1)}, \theta^{(2)}) + u_t(\theta_0^{(1)}, \theta^{(2)})|$. Using (A.2)–(A.6) and the Hölder inequality, it can be shown that

$$E_{\theta_0} \left\{ \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} \left(\left\| \frac{\partial}{\partial \theta^{(1)}} (\epsilon_t(\theta^{(1)})) \right\| \right) \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} |\epsilon_t(\theta^{(1)})| W_t \right\} < \infty \quad \textbf{(A.10)}$$

and

$$E_{\theta_0} \sum_{i=1}^{\infty} \sup_{\theta^{(2)} \in \Theta_\delta^{(2)}} |d_i(\theta^{(2)})| \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} \left(\left\| \frac{\partial}{\partial \theta^{(1)}} (\epsilon_{t-i}(\theta^{(1)})) \right\| \right) \sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} |\epsilon_{t-i}(\theta^{(1)})| W_t < \infty. \quad \textbf{(A.11)}$$

Finally, the sums in brackets on the right-hand side of (A.9) involve ergodic positive processes (as measurable functions of X_t and its past values). Therefore, the ergodic theorem can be applied, and, in light of (A.10) and (A.11), the limit is finite a.s.

Because $\|\hat{\theta}_n^{(1)} - \theta_0^{(1)}\|$ converges to zero a.s., Lemma 5 is proved. ■

LEMMA 6. *We have a.s.*

$$\lim_{n \to \infty} \sup_{\theta \in \Theta_\delta} \sqrt{n} |Q_n^{(2)}(\theta) - O_n^{(2)}(\theta)| = 0 \quad \textbf{(A.12)}$$

and for $i \in \{1,2\}$,

$$\lim_{n \to \infty} \sup_{\theta \in \Theta_\delta} \left| \sqrt{n} \frac{\partial}{\partial \theta^{(i)}} (Q_n^{(i)}(\theta) - O_n^{(i)}(\theta)) \right| = 0. \quad \textbf{(A.13)}$$

Proof. In view of Lemma 4 we have $\tilde{\epsilon}_t(\theta^{(1)}) = X_t + \sum_{i=1}^{t-1} c_i(\theta^{(1)}) X_{t-i}$. Then from (A.5)

$$\sup_{\theta^{(1)} \in \Theta_\delta^{(1)}} |\epsilon_t(\theta^{(1)}) - \tilde{\epsilon}_t(\theta^{(1)})| \le \sum_{i \ge t} K \rho^i |X_{t-i}|, \quad \text{a.s.} \quad \textbf{(A.14)}$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

Using (A.3), (A.5), and (A.14) we have

$$\sup_{\theta\in\Theta_\delta}|\tilde{u}_t(\theta)-u_t(\theta)| \le \sum_{i=0}^{t-1} K\rho^i \sum_{j\ge t-i} K\rho^j |X_{t-i-j}| \sup_{\theta\in\Theta_\delta}|\epsilon_{t-i}(\theta^{(1)})+\tilde{\epsilon}_{t-i}(\theta^{(1)})|$$
$$+ K \sup_{\theta\in\Theta_\delta}\sum_{i=t}^{\infty}\rho^i\epsilon_{t-i}^2(\theta^{(1)}).$$

Because there exists a constant M_1, independent of t, such that $E\sup_{\theta\in\Theta_\delta}|\epsilon_t(\theta^{(1)}) + \tilde{\epsilon}_t(\theta^{(1)})| \le M_1$, the Markov inequality and the Borel–Cantelli lemma show that

$$K^2\rho_1^t \sum_{i=0}^{t-1}\sup_{\theta\in\Theta_\delta}|\epsilon_{t-i}(\theta^{(1)})+\tilde{\epsilon}_{t-i}(\theta^{(1)})| \le K_1\rho_2^t, \quad \text{a.s.}$$

for $K_1 > 0$ and $\rho_1 < \rho_2 < 1$. Similarly we show that $K\sum_{i=t}^{\infty}\rho_1^i \sup_{\theta\in\Theta_\delta}\epsilon_{t-i}^2(\theta^{(1)}) \le K_2\rho_2^t$, a.s. for some constant $K_2 > 0$. Therefore, we obtain

$$\sup_{\theta\in\Theta_\delta} n|Q_n^{(2)}(\theta)-O_n^{(2)}(\theta)| \le \sum_{t=1}^{n}\sup_{\theta\in\Theta_\delta}|\tilde{u}_t(\theta)-u_t(\theta)|\left(\sup_{\theta\in\Theta_\delta}|u_t(\theta)| + \sup_{\theta\in\Theta_\delta}|\tilde{u}_t(\theta)|\right)$$
$$\le \sum_{t=1}^{\infty}\rho_2^t\left(\sup_{\theta\in\Theta_\delta}|u_t(\theta)| + \sup_{\theta\in\Theta_\delta}|\tilde{u}_t(\theta)|\right)\left(K_1\sum_{k\ge0}\rho^k|X_{-k}| + K_2\right),$$

which is finite a.s. (because it is positive and its expectation is finite). We deduce (A.12). By the same arguments (using (A.6) instead of (A.5)) we obtain (A.13). ■

LEMMA 7. *For any $\theta_*^{(2)} \in \Theta_\delta^{(2)}$, $\theta_*^{(2)} \neq \theta_0^{(2)}$, there exists a neighborhood $V(\theta_*^{(2)})$ of $\theta_*^{(2)}$ such that $V(\theta_*^{(2)}) \subset \Theta_\delta^{(2)}$ and*

$$\liminf_{n\to\infty}\inf_{\theta^{(2)}\in V(\theta_*^{(2)})} Q_n^{(2)}(\hat{\theta}_n^{(1)},\theta^{(2)}) > \zeta^2, \quad a.s.$$

Proof. We have

$$\inf_{\theta^{(2)}\in V(\theta_*^{(2)})} Q_n^{(2)}(\hat{\theta}_n^{(1)},\theta^{(2)}) \ge \inf_{\theta^{(2)}\in V(\theta_*^{(2)})} O_n^{(2)}(\theta_0^{(1)},\theta^{(2)})$$
$$- \sup_{\theta^{(2)}\in\Theta_{\delta/2}^{(2)}}|O_n^{(2)}(\hat{\theta}_n^{(1)},\theta^{(2)}) - Q_n^{(2)}(\hat{\theta}_n^{(1)},\theta^{(2)})|$$
$$- \sup_{\theta^{(2)}\in\Theta_{\delta/2}^{(2)}}|O_n^{(2)}(\hat{\theta}_n^{(1)},\theta^{(2)}) - O_n^{(2)}(\theta_0^{(1)},\theta^{(2)})|. \quad \textbf{(A.15)}$$

Let $V_m(\theta_*^{(2)})$ be the open sphere with center $\theta_*^{(2)}$ and radius $1/m$. Let

$$S_m(t) = \inf_{\theta^{(2)}\in V_m(\theta_*^{(2)})\cap\Theta_\delta^{(2)}} u_t^2(\theta_0^{(1)},\theta^{(2)}).$$

The ergodic theorem shows that a.s.

$$\inf_{\theta^{(2)}\in V_m(\theta_*^{(2)})\cap\Theta_\delta^{(2)}} O_n^{(2)}(\theta_0^{(1)},\theta^{(2)}) \ge \frac{1}{n}\sum_{t=1}^{n} S_m(t) \to E_{\theta_0}S_m(t),$$

as n tends to infinity. Because $u_t^2(\theta)$ is a smooth function of θ, $S_m(t)$ increases to $u_t^2(\theta_0^{(1)},\theta_*^{(2)})$ as m goes to infinity. Therefore $\lim_{m\to\infty}E_{\theta_0}S_m(t) = E_{\theta_0}u_t^2(\theta_0^{(1)},\theta_*^{(2)})$. But

$E_{\theta_0} u_t^2(\theta_0^{(1)}, \theta_*^{(2)}) > E_{\theta_0} u_t^2(\theta_0^{(1)}, \theta_0^{(2)}) = \zeta^2$, because $u_t(\theta_0^{(1)}, \theta_0^{(2)})$ is the linear innovation of ϵ_t. Hence $\liminf_{m\to\infty} \liminf_{n\to\infty} \inf_{\theta^{(2)} \in V_m(\theta_*^{(2)})} O_n^{(2)}(\theta_0^{(1)}, \theta^{(2)}) > \zeta^2$. Because, from Lemmas 5 and 6, the two suprema in (A.15) converge to zero as n goes to infinity, the stated result follows. ■

The proof of Theorem 1 follows from Lemma 7 and a standard compactness argument. ■

Proof of Theorem 2. Again, the proof of the theorem consists of a sequence of lemmas.

LEMMA 8. *The random vector*

$$Z_n = \begin{pmatrix} \sqrt{n} \dfrac{\partial}{\partial \theta^{(1)}} O_n^{(1)}(\theta_0^{(1)}) \\ \sqrt{n} \dfrac{\partial}{\partial \theta^{(2)}} O_n^{(2)}(\theta_0) \end{pmatrix}$$

has a limiting normal distribution with zero mean and covariance matrix

$$I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}.$$

Proof. Using the fact that ϵ_t and $u_t(\theta_0)$ are, respectively, the linear innovations of X_t and ϵ_t^2, it is straightforward to show that Z_n is centered. Instead of Z_n, we can equivalently prove asymptotic normality of $v'Z_n$ where $v = (\lambda', \mu')'$, $\forall \lambda \in \mathbb{R}^{P+Q}$, $\mu \in \mathbb{R}^{p+q+1}$. We have

$$\lambda' \sqrt{n} \frac{\partial}{\partial \theta^{(1)}} O_n^{(1)}(\theta_0^{(1)}) + \mu' \sqrt{n} \frac{\partial}{\partial \theta^{(2)}} O_n^{(2)}(\theta_0) = \frac{2}{\sqrt{n}} \sum_{t=1}^{n} (\lambda' Y_t^{(1)} + \mu' Y_t^{(2)}),$$

where

$$Y_t^{(1)} = \epsilon_t \frac{\partial}{\partial \theta^{(1)}} \epsilon_t(\theta_0^{(1)}) = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} c_i(\theta_0^{(1)})[c_{j,m}(\theta_0^{(1)})] X_{t-i} X_{t-j}, \tag{A.16}$$

$$\begin{aligned} Y_t^{(2)} &= u_t(\theta_0) \frac{\partial}{\partial \theta^{(2)}} u_t(\theta_0) \\ &= \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} d_i(\theta_0^{(2)})[d_{j,m}(\theta_0^{(2)})] \epsilon_{t-i}^2 \epsilon_{t-j}^2 \\ &\quad + w(\theta_0^{(2)}) \sum_{j=1}^{\infty} [d_{j,m}(\theta_0^{(2)})] \epsilon_{t-j}^2 + [w_m(\theta_0^{(2)})] \sum_{i=0}^{\infty} d_i(\theta_0^{(2)}) \epsilon_{t-i}^2, \end{aligned} \tag{A.17}$$

with $c_0(\theta_0^{(1)}) = d_0(\theta_0^{(2)}) = 1$. To use the strong mixing property of (X_t) we truncate all the sums involved in the expansions of $Y_t^{(1)}$ and $Y_t^{(2)}$. For any positive integer r, let $\mathbf{Y}_{t,r} = \lambda \mathbf{Y}_{t,r}^{(1)} + \mu \mathbf{Y}_{t,r}^{(2)}$, where

$$\mathbf{Y}_{t,r}^{(1)} = \sum_{i=0}^{r}\sum_{j=1}^{r} c_i(\theta_0^{(1)})[c_{j,m}(\theta_0^{(1)})]X_{t-i}X_{t-j},$$

$$\mathbf{Y}_{t,r}^{(2)} = \sum_{i=0}^{r}\sum_{j=1}^{r} d_i(\theta_0^{(2)})[d_{j,m}(\theta_0^{(2)})]\left(\sum_{k_1=0}^{r} c_{k_1}(\theta_0^{(1)})X_{t-i-k_1}\right)^2\left(\sum_{k_2=0}^{r} c_{k_2}(\theta_0^{(1)})X_{t-j-k_2}\right)^2$$
$$+ w(\theta_0^{(2)})\sum_{j=1}^{r}[d_{j,m}(\theta_0^{(2)})]\left(\sum_{k_1=0}^{r} c_{k_1}(\theta_0^{(1)})X_{t-j-k_1}\right)^2$$
$$+ [w_m(\theta_0^{(2)})]\sum_{i=0}^{r} d_i(\theta_0^{(2)})\left(\sum_{k_1=0}^{r} c_{k_1}(\theta_0^{(1)})X_{t-i-k_1}\right)^2. \quad \textbf{(A.18)}$$

Because $\mathbf{Y}_{t,r}$ is a function of a finite number of values of the process (X_t), it is strongly mixing. Therefore, the central limit theorem for strongly mixing processes (Ibragimov, 1962) can be applied. It implies that $(2/\sqrt{n})\sum_{t=1}^{n}(\mathbf{Y}_{t,r} - E_{\theta_0}\mathbf{Y}_{t,r})$ has a limiting $\mathcal{N}(0,\tilde{I}_r)$ distribution. We will show (in point (ii), which follows) that matrix I exists. Standard calculations show that $\tilde{I}_r \to (\lambda,\mu)I(\lambda,\mu)'$ as $r\to\infty$.

(i) We will show that the asymptotic distribution of the untruncated random variable $n^{-1/2}\sum_{t=1}^{n} Y_t$ (where $Y_t = \lambda' Y_t^{(1)} + \mu' Y_t^{(2)}$) is equal to the limit (as $r\to\infty$) of the asymptotic distribution (as $n\to\infty$) of the truncated random variable $n^{-1/2}\sum_{t=1}^{n}\mathbf{Y}_{t,r}$. Let $\mathbf{Z}_{t,r} = \lambda'(\mathbf{Y}_t^{(1)} - \mathbf{Y}_{t,r}^{(1)}) + \mu'(\mathbf{Y}_t^{(2)} - \mathbf{Y}_{t,r}^{(2)})$. It suffices to prove that $E\big((1/\sqrt{n})\sum_{t=1}^{n}(\mathbf{Z}_{t,r} - E_{\theta_0}\mathbf{Z}_{t,r})\big)\big(1/\sqrt{n})\sum_{t=1}^{n}(\mathbf{Z}_{t,r} - E_{\theta_0}\mathbf{Z}_{t,r})\big)'$ converges to zero uniformly in n as $r\to\infty$. A straightforward adaptation of a result given by Anderson (1971, Corollary 7.7.1, p. 426) will provide the advanced result. Because the computations are very similar for all the sums involved in the definition of $\mathbf{Z}_{t,r}$, we will only detail one of them. For instance, let

$$\mathbf{U}_{t,r} = \sum_{i=r+1}^{\infty}\sum_{j=1}^{\infty} d_i(\theta_0^{(2)})[d_{j,m}(\theta_0^{(2)})]\epsilon_{t-i}^2\epsilon_{t-j}^2$$

and let us show that for all $m\in\{1,\ldots,p+q+1\}$

$$\sup_n V\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{U}_{t,r}(m)\right)\overset{r\to\infty}{\longrightarrow} 0. \quad \textbf{(A.19)}$$

By a classical argument, it is sufficient to show that

$$\sum_{h=-\infty}^{\infty}|\mathrm{Cov}(\mathbf{U}_{t,r}(m),\mathbf{U}_{t+h,r}(m))|\overset{r\to\infty}{\longrightarrow} 0.$$

Using Lemma 4 and the assumption that $EX_t^{8+4\nu}<\infty$ we show that $M := E\epsilon_t^{8+4\nu}<\infty$. By the Cauchy–Schwarz inequality, we deduce that

$$|\mathrm{Cov}(\epsilon_{t_1}^2\epsilon_{t_2}^2,\epsilon_{t_3}^2\epsilon_{t_4}^2)| \le M < \infty \quad \textbf{(A.20)}$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

and that there exist a positive constant C and a constant $\rho \in [0,1[$ such that

$$|\mathrm{Cov}(\mathbf{U}_{t,r}(m),\mathbf{U}_{t+h,r}(m))|$$

$$= \Bigg| \sum_{i,i'=r+1}^{\infty} \sum_{j,j'=1}^{\infty} d_i(\theta_0^{(2)}) d_{i'}(\theta_0^{(2)}) d_{j,m}(\theta_0^{(2)}) d_{j',m}(\theta_0^{(2)}) \times \mathrm{Cov}(\epsilon_{t-i}^2 \epsilon_{t-j}^2, \epsilon_{t+h-i'}^2 \epsilon_{t+h-j'}^2) \Bigg|$$

$$\leq M \sum_{i=r+1}^{\infty} |d_i(\theta_0^{(2)})| \sum_{j=1}^{\infty} |d_{j,m}(\theta_0^{(2)})| \sum_{i'=r+1}^{\infty} |d_{i'}(\theta_0^{(2)})| \sum_{j'=1}^{\infty} |d_{j',m}(\theta_0^{(2)})| \leq C\rho^{2r}.$$

To deal with $|\mathrm{Cov}(\epsilon_{t-i}^2 \epsilon_{t-j}^2, \epsilon_{t+h-i'}^2 \epsilon_{t+h-j'}^2)|$, we introduce the following truncations

$$[\epsilon_t]_{(r)} = \sum_{k=0}^{r} c_k(\theta_0^{(1)}) X_{t-k}, \qquad [\epsilon_t]^{(r)} = \sum_{k=r+1}^{\infty} c_k(\theta_0^{(1)}) X_{t-k},$$

and

$$[\epsilon_t \epsilon_{t'}]^{(r)} = \sum_{k \vee k' \geq r+1} c_k(\theta_0^{(1)}) c_{k'}(\theta_0^{(1)}) X_{t-k} X_{t'-k'},$$

where r is a positive integer. Using a Taylor expansion we have, for positive integers i, j, i', j', and h.

$$\epsilon_{t+h-i'}^2 \epsilon_{t+h-j'}^2 = [\epsilon_{t+h-i'}]_{([h/4])}^2 [\epsilon_{t+h-j'}]_{([h/4])}^2 + 2C(t+h,i',j')[\epsilon_{t+h-i'} \epsilon_{t+h-j'}]^{([h/4])} \tag{A.21}$$

for some $C(t+h,i',j')$, $|C(t+h,i',j')| \leq \sum_{k_1,k_2=0}^{\infty} |c_{k_1}(\theta_0^{(1)}) c_{k_2}(\theta_0^{(1)}) \times X_{t+h-i'-k_1} X_{t+h-j'-k_2}|$. Hence we have

$$|\mathrm{Cov}(\epsilon_{t-i}^2 \epsilon_{t-j}^2, \epsilon_{t+h-i'}^2 \epsilon_{t+h-j'}^2)|$$

$$\leq |\mathrm{Cov}(\epsilon_{t-i}^2 \epsilon_{t-j}^2, [\epsilon_{t+h-i'}]_{([h/4])}^2 [\epsilon_{t+h-j'}]_{([h/4])}^2)| + 2|\mathrm{Cov}(\epsilon_{t-i}^2 \epsilon_{t-j}^2, C(t+h,i',j')[\epsilon_{t+h-i'} \epsilon_{t+h-j'}]^{([h/4])})|.$$

By the Davydov inequality (Davydov, 1968), the first term on the right is bounded by

$$C_2 \|\epsilon_{t-i}^2 \epsilon_{t-j}^2\|_{2+\nu} \|[\epsilon_{t+h-i'}]_{([h/4])}^2 [\epsilon_{t+h-j'}]^2{}_{([h/4])}\|_{2+\nu} \times \left(\alpha_X\left(\min\left\{h+i-i'-\left[\frac{h}{4}\right], h+j-i'-\left[\frac{h}{4}\right], h+i-j'-\left[\frac{h}{4}\right], h+j-j'-\left[\frac{h}{4}\right]\right\}\right)\right)^{\nu/(2+\nu)}$$

$$\leq C_2 M^{1/(2+\nu)} \left(\alpha_X\left(\left[\frac{h}{4}\right]\right)\right)^{0(2+0)} \tag{A.22}$$

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

for $i' \vee j' \leq h/2$, where C_2 is a positive constant. Moreover by the Hölder inequality we have

$$\begin{aligned} &|E\epsilon_{t-i}^2\epsilon_{t-j}^2 C(t+h,i',j')[\epsilon_{t+h-i'}\epsilon_{t+h-j'}]^{([h/4])}| \\ &\quad \leq \|\epsilon_{t-i}^2\epsilon_{t-j}^2\|_{2+\nu}\|C(t+h,i',j')\|_{(4+2\nu)/(1+\nu)} \\ &\qquad \times \|[\epsilon_{t+h-i'}\epsilon_{t+h-j'}]^{([h/4])}\|_{(4+2\nu)/(1+\nu)}. \end{aligned} \tag{A.23}$$

The third term in this product is dominated by

$$\sum_{k\vee k'>[h/4]} |c_k(\theta_0^{(1)})c_{k'}(\theta_0^{(1)})|\,\|X_t\|_{(4+2\nu)/(1+\nu)} \leq C_3\rho^{[h/4]},$$

where C_3 is a positive constant. Similarly we show that the second term in (A.23) is finite, whereas the first one is bounded by $M^{1/(2+\nu)}$. Therefore the left-hand side in (A.23) is bounded by a term of the order $\rho^{[h/4]}$. Therefore we have shown that

$$\begin{aligned} &|\mathrm{Cov}(\epsilon_{t-i}^2\epsilon_{t-j}^2, \epsilon_{t+h-i'}^2\epsilon_{t+h-j'}^2)| \leq C_4\left(\alpha_X\left(\left[\frac{h}{4}\right]\right)\right)^{\nu/(2+\nu)} + C_5\rho^{[h/4]}, \\ &\quad \text{for } i'\vee j' \leq \frac{h}{2} \quad \text{and} \quad h \geq 0, \end{aligned} \tag{A.24}$$

for some positive constants C_4 and C_5.

Using this inequality, we show that, for $h \geq 0$ and $r < [h/2]$,

$$\begin{aligned} &|\mathrm{Cov}(\mathbf{U}_{t,r}(m), \mathbf{U}_{t+h,r}(m))| \\ &\quad \leq \sum_{r<i'<[h/2]}\ \sum_{0<j'<[h/2]}\ \sum_{i,j} |d_i(\theta_0^{(2)})d_{i'}(\theta_0^{(2)})d_{j,m}(\theta_0^{(2)})d_{j',m}(\theta_0^{(2)}) \\ &\qquad\qquad \times \mathrm{Cov}(\epsilon_{t-i}^2\epsilon_{t-j}^2, \epsilon_{t+h-i'}^2\epsilon_{t+h-j'}^2)| \\ &\qquad + \sum_{i'\geq[h/2]}\ \sum_{i,j,j'} |d_i(\theta_0^{(2)})d_{i'}(\theta_0^{(2)})d_{j,m}(\theta_0^{(2)})d_{j',m}(\theta_0^{(2)}) \\ &\qquad\qquad \times \mathrm{Cov}(\epsilon_{t-i}^2\epsilon_{t-j}^2, \epsilon_{t+h-i'}^2\epsilon_{t+h-j'}^2)| \\ &\qquad + \sum_{j'\geq[h/2]}\ \sum_{i,j,i'} |d_i(\theta_0^{(2)})d_{i'}(\theta_0^{(2)})d_{j,m}(\theta_0^{(2)})d_{j',m}(\theta_0^{(2)}) \\ &\qquad\qquad \times \mathrm{Cov}(\epsilon_{t-i}^2\epsilon_{t-j}^2, \epsilon_{t+h-i'}^2\epsilon_{t+h-j'}^2)| \\ &\quad \leq C_8\rho^r\left(\alpha_X\left(\left[\frac{|h|}{4}\right]\right)\right)^{\nu/(2+\nu)} + C_9\rho^r\rho^{[|h|/4]} + C_{10}\rho^r\rho^{|h|/2}, \end{aligned}$$

where C_8, C_9, and C_{10} are positive constants. The same inequality holds for $h < 0$. Therefore there exists a constant K such that

$$\sum_{h=-\infty}^{\infty} |\mathrm{Cov}(\mathbf{U}_{t,r}(m), \mathbf{U}_{t+h,r}(m))| \leq Kr\rho^r + K\rho^r + K\rho^r\sum_h (\alpha_X(h))^{\nu/(2+\nu)} \overset{r\to\infty}{\longrightarrow} 0.$$

(ii) It remains to prove the existence of the components of matrix I. We have

$$V\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_t^{(2)}\right) = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} \mathrm{Cov}(Y_t^{(2)}, Y_s^{(2)}).$$

For $(l,m) \in \{0,\dots,p+q\}^2$ and $h \in \mathbb{Z}$, let

$$c(h) = \mathrm{Cov}\left(\sum_{\substack{i=0\\j=1}}^{\infty} d_i(\theta_0^{(2)}) d_{j,l}(\theta_0^{(2)}) \epsilon_{t-i}^2 \epsilon_{t-j}^2, \sum_{\substack{i=0\\j=1}}^{\infty} d_i(\theta_0^{(2)}) d_{j,m}(\theta_0^{(2)}) \epsilon_{t+h-i}^2 \epsilon_{t+h-j}^2\right).$$

First suppose that $h \geq 0$. We have

$$|c(h)| \leq S_1 + S_2, \tag{A.25}$$

where

$$S_1 = \sum_{i\vee j\vee i'\vee j' > h/2} |d_i(\theta_0^{(2)}) d_{j,l}(\theta_0^{(2)}) d_{i'}(\theta_0^{(2)}) d_{j',m}(\theta_0^{(2)})| \times |\mathrm{Cov}(\epsilon_{t-i}^2 \epsilon_{t-j}^2, \epsilon_{t+h-i'}^2 \epsilon_{t+h-j'}^2)|,$$

$$S_2 = \sum_{i\vee j\vee i'\vee j' \leq h/2} |d_i(\theta_0^{(2)}) d_{j,l}(\theta_0^{(2)}) d_{i'}(\theta_0^{(2)}) d_{j',m}(\theta_0^{(2)})| \times |\mathrm{Cov}(\epsilon_{t-i}^2 \epsilon_{t-j}^2, \epsilon_{t+h-i'}^2 \epsilon_{t+h-j'}^2)|.$$

From (A.5), (A.6), and (A.20) there exists a constant C_1 such that

$$S_1 \leq C_1 \rho^{h/2}. \tag{A.26}$$

Thus for $h \geq 0$, in view of (A.25), (A.26), and (A.24) we have

$$|c(h)| \leq C_1 \rho^{h/2} + C_6\left(\alpha_X\left(\left[\frac{h}{4}\right]\right)\right)^{\nu/(2+\nu)} + C_7 \rho^{[h/4]}$$

for some positive constants C_6 and C_7. A similar inequality holds for $h \leq 0$. Therefore, from the strong mixing assumption of Theorem 2, the sequence $(|c(h)|)$ is summable. All other terms involved in $\mathrm{Cov}(Y_t^{(2)}, Y_s^{(2)})$ may be treated in the same way. Finally, by a classical application of the dominated convergence theorem

$$V\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_t^{(2)}\right) \to \sum_{h=-\infty}^{\infty} \mathrm{Cov}(Y_t^{(2)}, Y_{t+h}^{(2)}), \quad \text{as } n \to \infty.$$

We similarly prove the existence of the other components of matrix I. ■

LEMMA 9. *Almost surely the matrices J_{11}, J_{12}, and J_{22} exist and are strictly positive definite.*

Proof. The proof is very similar to the one given in Francq and Zakoïan (1998). It is mainly based on the ergodic theorem, applied to sequences involving second derivatives of the ϵ_t's and the $u_t(\theta_0^{(1)}, \theta_0^{(2)})$'s. The fact that these sequences belong to L^2 is deduced from expansions similar to (A.4). ■

https://doi.org/10.1017/S0266466600165041 Published online by Cambridge University Press

To complete the proof of Theorem 2, we make Taylor expansions around the true parameter values. First, we have

$$0 = \sqrt{n}\,\frac{\partial}{\partial \theta^{(1)}}\,Q_n^{(1)}(\hat{\theta}_n^{(1)}) = \sqrt{n}\,\frac{\partial}{\partial \theta^{(1)}}\,Q_n^{(1)}(\theta_0^{(1)}) + \left[\frac{\partial^2}{\partial \theta_i^{(1)}\partial \theta_j^{(1)}}\,Q_n^{(1)}(\theta_{n,i,j}^{(1)*})\right]\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_0^{(1)}),$$

where the $\theta_{n,i,j}^{(1)*}$'s are between $\hat{\theta}_n^{(1)}$ and $\theta_0^{(1)}$. Doing again a Taylor expansion we obtain

$$\left|\frac{\partial^2}{\partial \theta_i^{(1)}\partial \theta_j^{(1)}}\,Q_n^{(1)}(\theta_{n,i,j}^{(1)*}) - \frac{\partial^2}{\partial \theta_i^{(1)}\partial \theta_j^{(1)}}\,Q_n^{(1)}(\theta_0^{(1)})\right|$$

$$\leq \sup_{\theta^{(1)}\in\Theta_\delta^{(1)}}\left\|\frac{\partial}{\partial \theta^{(1)}}\left(\frac{\partial^2}{\partial \theta_i^{(1)}\partial \theta_j^{(1)}}\,Q_n^{(1)}(\theta^{(1)})\right)\right\|\|\theta_{n,i,j}^{(1)*} - \theta_0^{(1)}\| \to 0$$

a.s. as $n \to \infty$. Similarly, two successive Taylor expansions lead to

$$0 = \sqrt{n}\,\frac{\partial}{\partial \theta^{(2)}}\,Q_n^{(2)}(\theta_0^{(1)},\theta_0^{(2)}) + \left[\frac{\partial^2}{\partial \theta_i^{(2)}\partial \theta_j^{(1)}}\,Q_n^{(2)}(\theta_{n,i,j}^{(1)**},\theta_0^{(2)})\right]\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_0^{(1)})$$

$$+ \left[\frac{\partial^2}{\partial \theta_i^{(2)}\partial \theta_j^{(2)}}\,Q_n^{(2)}(\hat{\theta}_n^{(1)},\theta_{n,i,j}^{(2)*})\right]\sqrt{n}(\hat{\theta}_n^{(2)} - \theta_0^{(2)}),$$

where the $\theta_{n,i,j}^{(1)**}$'s (resp. $\theta_{n,i,j}^{(2)*}$'s) are between $\hat{\theta}_n^{(1)}$ and $\theta_0^{(1)}$ (resp. $\hat{\theta}_n^{(2)}$ and $\theta_0^{(2)}$). The second-order derivatives of $Q_n^{(2)}(\cdot)$ can then be handled in the same way as for $Q_n^{(1)}(\cdot)$. Therefore, using Lemmas 6, 8, and 9, the proof of Theorem 2 is routinely completed. ■