Les Cahiers de la Chaire / N°13

GARCH(1,1) models with exogenously-driven volatility : structure and estimation

Nazim Regnard, Jean-Michel Zakoian

CHAIRE
Finance & Développement Durable

GARCH(1,1) models with exogenously-driven volatility : structure and estimation

Nazim Regnard [*]and Jean-Michel Zakoïan[†]

Abstract: This paper considers GARCH(1,1) models in which the time-varying coefficients are functions of the realizations of an exogenous stochastic process. Time series generated by this model are in general nonstationary. Necessary and sufficient conditions are given for the existence of non-explosive solutions, and for the existence of moments of these solutions. The asymptotic properties of the quasi-maximum likelihood estimator are derived under mild assumptions and its finite sample properties are investigated by simulations.

*Keywords:*Asymptotic normality, Existence of nonexplosive solutions, GARCH, Nonstationary processes, Quasi-maximum likelihood estimation, Strong consistency, Time-varying models.

1 Introduction

The autoregressive conditionally heteroskedastic model (ARCH) introduced by Engle (1982), and the so-called generalized ARCH of Bollerslev (1986), were the first time series models in which the dynamics was driven by volatility, that is the conditional variance. Volatility being considered a measure of risk, these classes of models have gained an enormous popularity in the financial econometrics literature. Fitted GARCH models often lead to the conclusion that the underlying process is (close to be) non covariance stationary. In fact, as put forward e.g. by Lamoureux and Lastrapes (1990), Gray (1996), and more recently Mikosch and Starica (2004), the apparent high persistence might be spuriously caused by

[*]EDF, Laboratoire FIME

[†]GREMARS Université Lille 3, Laboratoire FIME and CREST. Address: CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr. Recherche réalisée avec le soutien de la chaire "Finance et développement durable".

parameters varying through the sample, corresponding to different volatility regimes. To cope with this problem, numerous Markov-switching (MS) GARCH models have been proposed, in the spirit of the class developed by Hamilton (1989) for ARMA models; see e.g. Hamilton and Susmel (1994), Dueker (1997), Haas, Mittnik and Paolella (2004).

Contrary to other classes of time series models (in particular the ARMA), MS - as well as standard - GARCH models, are only designed for *strictly stationary* time series. The strict stationarity conditions have been established by Nelson (1990) for the GARCH(1,1), and by Bougerol and Picard (1992) for the GARCH(p,q). They have been extended to the MS-GARCH(p,q) by Francq, Roussignol and Zakoïan (2001). When the stricti stationarity conditions do not hold, the standard GARCH model only admits a degenerate solution (the square of the process converges in probability to infinity for any initial value). The same remark holds for the various extensions of the standard GARCH, in particular those allowing for asymmetric responses to shocks (see Teräsvirta (2007) and the references therein) or long memory GARCH models (see Giraitis, Leipus and Surgailis (2007)).

The model of this paper combines changes in regimes and nonstationarity. The changes of regimes occur at known dates and are driven by an observed process. More precisely, we consider GARCH(1,1) models in which the coefficients are functions of the *realizations* s_t of an exogenous stochastic process (S_t) taking its values in a finite set, $\{1,\dots,d\}$ say. The model will then be conditional on (S_t). Contrary to Markov-switching models, the switching-regime process is observed. The model is capable of generating solutions which are not stationary without being explosive.

Models with time-varying coefficients have been considered by many authors in the time series literature; see e.g. Dalhaus (1997), Azrak and Mélard (2006). Models in which the coefficients are periodic functions have been proposed by Basawa and Lund (2001), among others, and in the GARCH framework by Bollerslev and Ghysels (1996). However, periodic models are not suitable when the change of dynamics occur at irregular dates. GARCH models with time varying coefficients have been considered by Polzehl and Spokoiny (2006). Time series models in which the coefficients are subordinated to an exogenous process have been recently proposed and analyzed for the conditional mean by Bibi and Francq (2003), Francq and Gautier (2004a, 2004b). They consider ARMA models with time-varying coefficients driven by an observed process and prove asymptotic properties for least-squares and generalized least-squares estimators. This paper follows the same approach for the modeling of conditional variances and studies the properties of the quasi-maximum likelihood estimator (QMLE).

The paper is organized as follows. In Section 2 we introduce the time-varying GARCH model and give the main interpretations. The existence of non explosive solutions is discussed in Section 3. We also derive conditions for the existence of moments. Section 4 is devoted to the statistical inference. The properties of the QMLE, namely the strong consistency and asymptotic normality, are established. Proofs and technical lemmas are relegated to an appendix.

2 Model specification

We consider the time-varying coefficients GARCH(1,1) model

$$\epsilon_t = \sqrt{h_t}\eta_t, \quad h_t = \omega(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)h_{t-1}, \quad t \in \mathbb{Z} \tag{2.1}$$

where (η_t) is a sequence of independent and identically distributed (iid), centered variables with unit variance; (s_t) is a sequence of coefficients with values in a finite set $E = \{e_1, \ldots, e_d\}$; the functions $\omega(\cdot), \alpha(\cdot), \beta(\cdot)$ are defined on E with values in $\mathbb{R}^+$ with $\omega(\cdot) > 0$.

In each "regime" (i.e. for each value of s_t) the volatility is that of a standard GARCH(p, q) model. Of course, the existence of non trivial solutions to this model require conditions on the coefficients and the sequence (s_t). In the standard GARCH framework, the conditions derived by Bougerol and Picard (1992) ensure the existence of a *strictly stationary* solution. For Model (2.1), we will derive conditions ensuring the existence of a solution that is *not stationary*, in any sense. The probability properties will be investigated in the next section.

When the sequence (s_t) is purely periodic (with $s_{t+d} = s_t$ for any t), we retrieve the periodic-GARCH model introduced by Bollerslev and Ghysels (1996). However, it is often too restrictive to assume that the coefficients change in a purely periodic way. The model of this paper can accommodate situations where the volatility coefficients fluctuate in time but with a certain regularity in the fluctuations.

An potential application of Model (2.1) is the modeling of electricity prices. Appropriate modeling of electricity prices is crucial for the management and trading in electricity markets. Several different approaches have been proposed in the literature (see Haldrup and Nielsen (2006) for a recent reference in which a regime switching model is developed). Electricity prices present similarities with financial data but also important differences such as strong seasonal effects. In particular, the strong dependence between electricity demand and the weather conditions leads to periodic behaviors in the volatility of electricity prices (see e.g. Carnero, Koopman and Ooms (2003)). In the framework of Model (2.1), a modeling of the volatility of electricity prices can be based on a sequence (s_t) related to the temperature (or any other indicator of the weather conditions).

We now turn to the probability properties of the model.

3 Probability structure

Without further restrictions, there is no guarantee about the existence of nonexplosive solutions to Model (2.1). In this section we are interested in *nonanticipative solutions* (ϵ_t), *i.e.* such that ϵ_t is a function of the variables $\eta_{t-i}, i \geq 0$, for a given sequence (s_t).

3.1 Existence of non explosive solutions

Letting $a(x, y) = \alpha(x)y^2 + \beta(x)$ we have

$$h_t = \omega(s_t) + a(s_t, \eta_{t-1})h_{t-1}. \tag{3.1}$$

The first result gives a stability condition, based on the assumption that (s_t) visits the different states with a certain regularity. Let by convention $\log(0) = -\infty$.

Theorem 3.1 *For $j = 1, \dots d$ let $\mathcal{T}(t, j, n) = \{\tau \in \{0, \dots, n\} \mid s_{t-\tau} = e_j\}$. Assume that*

$$\forall t, \quad \frac{|\mathcal{T}(t, j, n)|}{n} \to \pi_j, \quad \text{when } n \to \infty \tag{3.2}$$

for some $\pi_j \geq 0$ with $\sum_{j=1}^{d} \pi_j = 1$. Then, if

$$\gamma_0 := \sum_{j=1}^{d} \pi_j E\{\log a(e_j, \eta_0)\} < 0, \tag{3.3}$$

Model (2.1) admits a nonanticipative solution (ϵ_t) given by

$$\epsilon_t = \left\{ \omega(s_t) + \sum_{n=1}^{+\infty} a(s_t, \eta_{t-1}) \dots a(s_{t-n+1}, \eta_{t-n})\omega(s_{t-n}) \right\}^{1/2} \eta_t. \tag{3.4}$$

If $\gamma_0 > 0$, for any starting value h_0 we have $h_t \to +\infty$, almost surely as $t \to \infty$. If, in addition, $E|\log \eta_0^2| < \infty$ then $\epsilon_t^2 \to +\infty$, almost surely as $t \to \infty$.

Remarks:

1. Assumption (3.2) is very general. Realizations of stationary and ergodic processes on E obviously satisfy this condition, with probability one. For example, with $d = 2$, the sequence $(s_t) = (\dots, 1, 0, 1, 0, \dots)$ yields $\pi_1 = \pi_2 = 0.5$, whereas the sequence $(s_t) = (\dots, 1, 0, 0, 1, 0, 0, \dots)$ yields $\pi_1 = 2/3$ and $\pi_2 = 1/3$ (with $e_1 = 0, e_2 = 1$). [1]

2. It should be noted that the solution obtained in (3.4) is nonstationary. Indeed, the random coefficients in the infinite sum depend on the values of the s_{t-n}. For instance if $\alpha(\cdot) = \beta(\cdot) = 0$, the solution process is defined by $\epsilon_t = \sqrt{\omega(s_t)}\eta_t$ and is non identically distributed when $\omega(\cdot)$ is not constant.

3. The *local* stationarity condition, i.e. the condition for stationarity in each regime, $E\{\log a(e_j, \eta_0)\} < 0$ for $j = 1, \dots, d$, implies the existence of a solution by (3.3). The converse is not true. Condition (3.3) does not entail local stationarity of all regimes.

4. Since $a(e_j, \eta_0) \geq \beta(e_j)$, a simple necessary condition for the existence of a solution to model (2.1) is given by
$$\prod_{j=1}^{d} \beta^{\pi_j}(e_j) < 1. \tag{3.5}$$

5. In the ARCH(1) case, a more explicit condition can be given. Since $\log a(e_j, \eta_0) = \log \alpha(e_j) + \log \eta_0^2$ we find
$$\prod_{j=1}^{d} \alpha^{\pi_j}(e_j) < e^{-E \log \eta_0^2}. \tag{3.6}$$
In particular this condition holds when the distribution of η_0 has a mass at zero. It can also be noted that when $\gamma_0 > 0$, the condition $E|\log \eta_0^2| < \infty$ is automatically satisfied in the ARCH case, and thus Theorem 3.1 shows that $\epsilon_t^2 \to +\infty$, a.s. Indeed, since $E \log^+ \eta_1^2 \leq E\eta_1^2 = 1$, where $\log^+(x) = \max\{\log(x), 0\}$ for $x > 0$, we have $E|\log \eta_0^2| = \infty$ if and only if $E \log \eta_0^2 = -\infty$. Hence, if $\prod_{j=1}^{d} \alpha^{\pi_j}(e_j) > e^{-E \log \eta_0^2}$ we must have $E \log \eta_0^2 > -\infty$.

[1] An example of a sequence (s_t) which *does not* satisfy this condition can be constructed as follows: let $s_0 = 0$ and, for $t \geq 1$, let $s_t = 1$ until $\frac{1}{t}\sum_{i=0}^{t} s_i > 3/4$. Then, let $s_t = 0$ until $\frac{1}{t}\sum_{i=0}^{t} s_i < 1/4$. Then let $s_t = 1$ until $\frac{1}{t}\sum_{i=0}^{t} s_i > 3/4$ and continue this way. The first ten elements of this sequence are $0, 1, 0, 0, 0, 0, 1, 1, 1, 1$. To complete the sequence let $s_{-t} = s_t$ for any $t \geq 0$. It follows that (3.2) is not satisfied because for $j = 1, 2$ and the sequence $(|\mathcal{T}(0, j, n)|/n)$ has two accumulation points 1/4 and 3/4.

A consequence of Theorem 3.1, without further assumption, is the existence of a finite moment, of small order, for the process (ϵ_t).

Theorem 3.2 *Under the assumptions of Theorem 3.1, we have*

$$E\{\epsilon_t^2\}^r < \infty, \quad \textit{for some } r > 0$$

where (ϵ_t) *is defined in (3.4).*

3.2 Predictions of the squares

For standard GARCH(1,1) models, the predictions of ϵ_t^2 are obtained from the ARMA(1,1) representation for the squares. Similarly, for Model (2.1), letting $u_t = \epsilon_t^2 - h_t = (\eta_t^2 - 1)h_t$ we have

$$\epsilon_t^2 = \omega(s_t) + (\alpha + \beta)(s_t)\epsilon_{t-1}^2 + u_t - \beta(s_t)u_{t-1}.$$

Letting $\delta_t = \epsilon_t^2 - \omega(s_t) - (\alpha + \beta)(s_t)\epsilon_{t-1}^2$, we thus have under the assumptions of Theorem 3.1,

$$\epsilon_t^2 \;\; = \;\; \omega(s_t) + (\alpha + \beta)(s_t)\epsilon_{t-1}^2 - \sum_{k \geq 0} \beta(s_t) \dots \beta(s_{t-k})\delta_{t-k-1} + u_t.$$

This representation is valid because (3.3) implies

$$\sum_{j=1}^{d} \pi_j \log \beta(e_j) \leq \sum_{j=1}^{d} \pi_j E\{\log a(e_j, \eta_0)\} < 0,$$

from which the existence of the infinite sum is deduced, by the arguments used in the proof of Theorem 3.1. Note that the expectation of u_t conditional on ϵ_t^2 past values is zero. The optimal predictor $\hat{\epsilon}_t^2$ of ϵ_t^2 is then

$$\hat{\epsilon}_t^2 = \omega(s_t) + (\alpha + \beta)(s_t)\epsilon_{t-1}^2 - \sum_{k \geq 0} \beta(s_t) \dots \beta(s_{t-k})\delta_{t-k-1}.$$

Predictions at higher horizons can be derived similarly. Contrary to standard GARCH models, predictions formulas are time dependent trough the coefficients s_t.

3.3 Existence of a non explosive solution with finite moments

We now consider the existence of second-order moments for the solution (3.4).

Theorem 3.3 *Let Assumption (3.2) hold. Then if*

$$\gamma_1 := \prod_{j=1}^{d} \{Ea(e_j, \eta_0)\}^{\pi_j} < 1, \tag{3.7}$$

Model (2.1) admits a non anticipative solution (ϵ_t) *with* $E\epsilon_t^2 < \infty$. *If* $\gamma_1 > 1$ *there is no nonanticipative solution* (ϵ_t) *such that* $E\epsilon_t^2 < \infty$.

The already given comment applies: the local second-order conditions, $E\{a(e_j, \eta_0)\} < 1$ for all j, are sufficient but non necessary for the global second-order condition.
The following is an extension to higher-order even moments.

Theorem 3.4 *Let* m *denote a strictly positive integer such that* $E\eta_t^{2m} < \infty$. *Under the assumptions of Theorem 3.1 and if*

$$\gamma_m := \prod_{j=1}^{d} \{Ea(e_j, \eta_0)^m\}^{\pi_j} < 1, \tag{3.8}$$

Model (2.1) admits a non anticipative solution (ϵ_t) *with* $E\epsilon_t^{2m} < \infty$. *If* $\gamma_m > 1$ *there is no nonanticipative solution* (ϵ_t) *such that* $E\epsilon_t^{2m} < \infty$.

Note that Condition (3.8) is stronger than (3.3) since, by the Jensen inequality,

$$\sum_{j=1}^{d} m\pi_j E\{\log a(e_j, \eta_0)\} \quad \leq \quad \sum_{j=1}^{d} \pi_j \log E\{a(e_j, \eta_0)\}^m = \log \prod_{j=1}^{d} \{Ea(e_j, \eta_0)^m\}^{\pi_j}.$$

3.4 Some special cases

(a) standard GARCH. The classical GARCH(1,1) model is a particular case of model (2.1), obtained when $d = 1$ (or equivalently when the functions ω, α and β are constant). In this case, (3.3) reduces to the strict stationarity condition established by Nelson (1990), and the solution in (3.4) is strictly stationary.

(b) periodic GARCH. The use of periodic models for characterizing seasonal time series has a long story. In particular, the class of periodic ARMA models has received a great attention (see e.g. Tiao and Gruppe (1980), Anderson and Vecchia (1983), Lund and Basawa (1990)). Periodic GARCH models where considered by Bollerslev and Ghysels (1996) and correspond to $s_{t+d} = s_t$ for any t, in our framework. For instance if $d = 5$, for daily series, the volatility at each day of the week has a different specification. The periodicity of s_t allows to simplify the stability conditions. Indeed we have $\pi_j = 1/d$, for $j = 1, \dots, d$ and Conditions (3.3) and (3.7) reduce to

$$E \log \prod_{j=1}^{d} a(e_j, \eta_0) < 0 \quad \text{and} \quad \prod_{j=1}^{d} E\left\{a(e_j, \eta_0)\right\} < 1.$$

It can be seen from (3.4) that in this case, the processes $(\epsilon_{dt+i})_t$, for $i = 1, \dots, d$, are strictly stationary.

(c) (s_t) **realization of a stationary process.** Model (2.1) can be viewed as being conditional on a realization (s_t) of a process (S_t), with values in E. Suppose that the process (S_t) is stationary, ergodic, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as (η_t), and independent of (η_t). By the ergodic theorem, (3.2) holds with $\pi_j = P(S_t = e_j)$. It follows that $\gamma_0 = E\{\log a(S_t, \eta_t)\}$.

It is interesting to compare Model (2.1), which is conditional on (S_t), with the unconditional model

$$\epsilon_{S,t} = \sqrt{h_{S,t}}\eta_t, \quad h_{S,t} = \omega(S_t) + \alpha(S_t)\epsilon_{S,t-1}^2 + \beta(S_t)h_{S,t-1}, \quad t \in \mathbb{Z}. \tag{3.9}$$

The proof of the following result is omitted for brevity but is available from the authors.

Theorem 3.5 *Suppose that (3.3) holds. Then Model (3.9) admits a unique strictly stationary solution* $(\epsilon_{S,t})$*, which is also ergodic and nonanticipative.*

The existence of (ϵ_t) in (3.4) is guaranteed under the assumptions of Theorem 3.5. However, it is important to distinguish the strictly stationary solution of Model (3.9),

$$\epsilon_{S,t} = \left\{\omega(S_t) + \sum_{n=1}^{+\infty} a(S_t, \eta_{t-1}) \dots a(S_{t-n+1}, \eta_{t-n})\omega(S_{t-n})\right\}^{1/2} \eta_t, \quad a.s. \tag{3.10}$$

from its realization, (ϵ_t), which is a *non stationary* solution of Model (2.1). The probabilistic properties of the two processes are not the same. For instance, we have seen that $E\epsilon_t^2 < \infty$ under (3.7). But this condition is neither necessary nor sufficient for $E\epsilon_{S,t}^2 < \infty$. In fact, the existence of the latter expectation requires more assumptions on the process (S_t) than those already made. For instance, in the case of a Markov chain (S_t), such conditions involve the transition probabilities of the chain (see Francq, Roussignol and Zakoïan, Theorem 2, 2001). By contrast, Condition (3.7) only involves the stationary probabilities of (S_t), not its dependence structure.

4 Quasi-maximum likelihood estimation

Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004) (hereafter FZ) have established the consistency and asymptotic normality of the QMLE for standard GARCH(p, q) models under mild conditions. See also Straumann (2005) for a comprehensive monograph on the estimation of conditionally heteroskedastic time series, in which the asymptotic properties of the QMLE are investigated when the volatility process is a general function of the form $\sigma_t^2 = g_\theta(X_{t-1}, \ldots, X_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2)$, where the observations are $X_1, \ldots, X_n$. However, these results cannot be used for the model of the present paper because the volatility is a *time dependent* function of the past variables.

In the previous section we have seen that, under appropriate conditions, Model (2.1) admits a well-defined solution. However this solution is in general nonstationary and, from a statistical point of view, this entails difficulties. In particular, the ergodic and central limit theorems for stationary sequences cannot be used in this framework.

For convenience we will index by zero the coefficients of the data-generating mechanism

$$\epsilon_t = \sqrt{h_t}\eta_t, \quad h_t = \omega_0(s_t) + \alpha_0(s_t)\epsilon_{t-1}^2 + \beta_0(s_t)h_{t-1}, \tag{4.1}$$

and denote by $\theta = (\omega(e_1), \ldots, \omega(e_d), \alpha(e_1), \ldots, \alpha(e_d), \beta(e_1), \ldots, \beta(e_d))' = (\theta_1, \ldots, \theta_{3d})'$ the vector of parameters, with true value θ_0. The parameter is assumed to belong to a parameter space $\Theta \subset]0, +\infty[^d \times [0, \infty[^{2d}$. The sequence (s_t) is observed, and the orders p, q and d are known a priori.

We introduce the following assumption.

A0: (s_t) is a realization of a process (S_t) which is stationary, ergodic, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as (η_t), and independent of (η_t). Moreover, (3.2) holds with $\pi_j = P(S_t = e_j),\ j = 1, \ldots, d$.

It can be noted that, by the ergodic theorem, the first part of **A0** implies (3.2), but only for almost all sequences (s_t). The second part of **A0** ensures that (s_t) is a sequence for which the convergence in (3.2) holds. Note also that **A0** is not in contradiction with the motivations of Model (2.1). Even if (s_t) is the realization of a stationary process, (ϵ_t) is in general non stationary.

Let $(\epsilon_1, \ldots, \epsilon_n)$ be a realization of length n of a nonanticipative solution (ϵ_t) to model

(4.1). Conditionally on initial values $\tilde{\epsilon}_0$ and $\tilde{\sigma}_0^2$ the gaussian quasi-likelihood is given by

$$L_n(\theta) = L_n(\theta; \epsilon_1, \dots, \epsilon_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2}\right),$$

where the $\tilde{\sigma}_t^2$ are defined recursively, for $t \geq 2$, by

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \omega(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\tilde{\sigma}_{t-1}^2.$$

with $\tilde{\sigma}_1^2 = \omega(s_1) + \alpha(s_1)\tilde{\epsilon}_0^2 + \beta(s_1)\tilde{\sigma}_0^2$. For instance, the initial values can be chosen as

$$\tilde{\epsilon}_0^2 = \tilde{\sigma}_0^2 = \epsilon_1^2 \quad \text{or} \quad \tilde{\epsilon}_0^2 = \tilde{\sigma}_0^2 = \gamma \tag{4.2}$$

for some constant γ which may depend on θ. A QMLE of θ is defined as any measurable solution $\hat{\theta}_n$ of

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\min_{\theta\in\Theta} \tilde{\mathbf{l}}_n(\theta), \tag{4.3}$$

where

$$\tilde{\mathbf{l}}_n(\theta) = n^{-1}\sum_{t=1}^{n} \tilde{\ell}_t, \qquad \text{and} \quad \tilde{\ell}_t = \tilde{\ell}_t(\theta) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \log\tilde{\sigma}_t^2.$$

Let $a_0(x, y) = \alpha_0(x)y^2 + \beta_0(x)$. We make the following assumptions.

A1: $\theta_0 \in \Theta$ and Θ is compact.

A2: For θ_0, we have $\sum_{j=1}^{d} \pi_j E\{\log a_0(e_j, \eta_0)\} < 0$ and $\forall\theta \in \Theta$, $\prod_{j=1}^{d} \beta^{\pi_j}(e_j) < 1$.

A3: There exist $r, \rho \in (0, 1)$ and $C > 0$ such that

$$\forall i > 0, \quad E\left\{a_0^r(S_t, \eta_{t-1}) \dots a_0^r(S_{t-i}, \eta_{t-i-1})\right\} < C\rho^{i+1}. \tag{4.4}$$

A4: η_t^2 has a nondegenerate distribution with $E\eta_t^2 = 1$.

A5: For all i, $\alpha_0(e_i) + \beta_0(e_i) \neq 0$ and there exist $\ell \in \{1, \dots, d\}$ and $k > 0$ such that $\alpha_0(e_\ell)\mathbb{P}(S_{t-k} = e_\ell, S_t = e_i) > 0$.

Assumption **A1** is standard. Note that the stability condition in **A2** is only imposed for the true value. For $\theta \neq \theta_0$ we impose the weaker condition in (3.5). Assumption **A3** is a technical assumption which will be discussed below. Assumption **A4** is required for the consistency of the QMLE of standard GARCH models (see the aforementioned references). Assumption **A5** is made for identifiability reasons. In particular it precludes, for the true parameter value, (i) regimes with constant volatility, and (ii) cancelation of *all* coefficients $\alpha_0(e_i)$. This assumption also implies $\mathbb{P}(S_t = e_i) > 0$ for all i. Notice that in the standard GARCH(1,1) case, with $d = 1$, **A5** reduces to $\alpha_{01} > 0$, which is the usual identifiability assumption. When $d > 1$, states with $\alpha_0(e_\ell) = 0$ are allowed (but then $\beta_0(e_\ell) > 0$).

4.1 Asymptotic results

Our first asymptotic result establishes the strong consistency of the QMLE.

Theorem 4.1 *Let $(\hat{\theta}_n)$ be a sequence of QMLE satisfying (4.3), with the initial conditions (4.2). Then, under* **A0**-**A5***, almost surely* $\hat{\theta}_n \to \theta_0$*, as* $n \to \infty$.

For the asymptotic normality, the following additional assumptions are made.

A6: θ_0 belongs to the interior of Θ.

A7: $\kappa_\eta = E\eta_t^4 < \infty$.

Assumption **A6** precludes the case where some GARCH coefficients are equal to zero. This assumption is necessary for the asymptotic normality because the estimator $\hat{\theta}_n$ is obtained under positivity constraints. See Francq and Zakoïan (2007) for the asymptotic distribution of the QMLE for standard GARCH with coefficients equal to zero. Assumption **A7** is required for the existence of the variance of the score vector. Without this assumption, usual estimators in the standard GARCH framework have non standard asymptotic distributions (see Hall and Yao (2003)).

In the theoretical developments considered in the sequel, for clarity we will sometimes index by S, η the expectation of a variable depending on both processes (S_t) and (η_t). Notice that the stability condition (3.3) takes the form

$$E_{S,\eta}\{\log a_0(S_0, \eta_0)\} < 0. \tag{4.5}$$

It will be convenient to introduce the model

$$\epsilon_{S,t} = \sqrt{h_{S,t}}\eta_t, \quad h_{S,t} = \omega_0(S_t) + \alpha_0(S_t)\epsilon_{S,t-1}^2 + \beta_0(S_t)h_{S,t-1}, \quad t \in \mathbb{Z}.$$

By Theorem 3.5, under (4.5), this model admits a strictly stationary solution $(\epsilon_{S,t})$ given (3.10). For any $\theta \in \Theta$ satisfying the second part of **A2**, denote by $\left(\sigma_{S,t}^2\right)_t = \left\{\sigma_{S,t}^2(\theta)\right\}_t$ the strictly stationary, ergodic and nonanticipative solution of

$$\sigma_{S,t}^2 = \omega(S_t) + \alpha(S_t)\epsilon_{t-1}^2 + \beta(S_t)\sigma_{S,t-1}^2. \tag{4.6}$$

Note that $\sigma_{S,t}^2(\theta_0) = h_{S,t}$.

Theorem 4.2 *Under Assumptions* **A0-A7**, $\sqrt{n}(\hat{\theta}_n - \theta)$ *is asymptotically* $\mathcal{N}(0, (\kappa_\eta - 1)J^{-1})$ *distributed, where*

$$J := E_{S,\eta}\left(\frac{1}{\sigma^4_{S,t}(\theta_0)} \frac{\partial \sigma^2_{S,t}(\theta_0)}{\partial \theta} \frac{\partial \sigma^2_{S,t}(\theta_0)}{\partial \theta'}\right) \tag{4.7}$$

is a positive-definite matrix.

It is worthnoting that for the standard GARCH(1,1) (obtained for $d = 1$), Theorems 4.1 and 4.2 reduce to the results established by FZ. It can also be noted that the asymptotic covariance matrix not only depends on the stationary probabilities $(\pi(1), \dots, \pi(d))$ but depends on its whole dependence structure of the process (S_t).

4.2 Discussion of Assumption A3

Assumption **A3** can be verified in particular cases. If (S_t) is an independent process then, by **A0**, (S_t, η_t) is an independent process and the left-hand side of the inequality reduces to $\{E a_0^r(S_t, \eta_t)\}^{i+1}$. By the first part of **A2** we have $E a_0^r(S_t, \eta_t) < 1$ for some $r > 0$ (see Berkes, Horváth and Kokoszka (2003)). Thus **A3** holds true, under **A2**, when (S_t) is an independent process.

Another case of interest is when (S_t) is a Markov chain. Let $\rho(A)$ denote the spectral radius of any square matrix A.

Theorem 4.3 *Suppose that* (S_t) *is a homogeneous, stationary, irreducible, and aperiodic Markov chain and denote by* $p(k, \ell) = P(S_t = \ell \mid S_{t-1} = k)$ *the transition probabilities. Let* $g_r(x) = E\{a_0^r(x, \eta_t)\}$, *for* $r \in (0, 1)$, *and let*

$$\mathbb{P}_r = \begin{pmatrix} p(1,1)g_r(1) & \cdots & p(d,1)g_r(1) \\ \vdots & & \vdots \\ p(1,d)g_r(d) & \cdots & p(d,d)g_r(d) \end{pmatrix}.$$

Then (4.4) is verified if $\rho(\mathbb{P}_r) < 1$.

Finally, without any further assumption on (S_t), a sufficient condition for **A3** to hold can be given. Let $\overline{\alpha}_0 = \max_i\{\alpha_0(e_i)\}$ and $\overline{\beta}_0 = \max_i\{\beta_0(e_i)\}$. Then if

A3': $\quad E \log\{\overline{\alpha}_0 \eta_t^2 + \overline{\beta}_0\} < 0$,

A3 holds true, by the argument used for the independent case. Note however that **A3'** is more restrictive than **A2-A3** because it precludes "nonstationary regimes" (i.e. regimes in which the standard strict stationarity condition for GARCH is violated).

Appendix: Lemmas and Proofs

Proof of Theorem 3.1. Iterating (3.1) yields

$$\begin{aligned} h_t &= \omega(s_t) + \sum_{n=1}^{N} a(s_t, \eta_{t-1}) \dots a(s_{t-n+1}, \eta_{t-n}) \omega(s_{t-n}) \\ &\quad + a(s_t, \eta_{t-1}) \dots a(s_{t-N}, \eta_{t-N-1}) h_{t-N-1}. \end{aligned} \tag{A.1}$$

Note that the terms involved in the sum are nonnegative. Therefore this sum converges almost surely (a.s.) in $\mathbb{R}^+ \cup \{+\infty\}$ and we can use the Cauchy rule to see if the limit is finite. Letting $u_{n,t} = a(s_t, \eta_{t-1}) \dots a(s_{t-n+1}, \eta_{t-n}) \omega(s_{t-n})$, for $n \geq 1$, we have

$$\begin{aligned} u_{n,t}^{1/n} &= \{a(s_t, \eta_{t-1}) \dots a(s_{t-n+1}, \eta_{t-n}) \omega(s_{t-n})\}^{1/n} \\ &= \exp \frac{1}{n} \sum_{i=1}^{n} \log a(s_{t-i+1}, \eta_{t-i}) + \frac{1}{n} \log \omega(s_{t-n}) \\ &= \exp \frac{1}{n} \sum_{j=1}^{d} \sum_{i \in \mathcal{T}(t,j,n-1)} \log a(e_j, \eta_{t-i-1}) + \frac{1}{n} \log \omega(s_{t-n}). \end{aligned}$$

Note that, for $j = 1, \dots, d$, the process $(a(e_j, \eta_t))_t$ is strictly stationary. It follows by the strong law of large numbers that

$$u_{n,t}^{1/n} \to \exp \left(\sum_{j=1}^{d} \pi_j E\{\log a(e_j, \eta_0)\} \right) = \exp(\gamma_0) < 1, \quad a.s. \quad \text{when } n \to \infty.$$

Therefore, if (3.3) holds, the series $\sum_n u_{n,t}$ converges. The process defined by (3.4) is a real-valued solution to model (2.1). This solution is nonanticipative, ϵ_t being a function of the $\eta_{t-i}, i > 0$.

Now we assume that $\gamma_0 > 0$. For any starting value h_0 we have, in view of (A.1), for $t \geq 2$,

$$h_t \geq \prod_{i=0}^{t-2} a(s_{t-i}, \eta_{t-i-1}) \omega(s_1),$$

thus

$$\frac{1}{t} \log h_t \geq \frac{1}{t} \sum_{i=0}^{t-2} \log a(s_{t-i}, \eta_{t-i-1}) + \frac{1}{t} \log \omega(s_1),$$

and

$$\underline{\lim}_{t \to \infty} \frac{1}{t} \log h_t \geq \gamma_0 > 0.$$

It follows that $\log h_t \to +\infty, a.s.$ and then $h_t \to +\infty, a.s.$ Similarly, in view of $\epsilon_t^2 = h_t \eta_t^2$,

$$\frac{1}{t} \log \epsilon_t^2 \geq \frac{1}{t} \sum_{i=0}^{t-2} \log a(s_{t-i}, \eta_{t-i-1}) + \frac{1}{t} \log \omega(s_1) + \frac{1}{t} \log \eta_t^2.$$

Under the assumption that $\log \eta_t^2$ has a finite expectation, the last term in the right-hand side converges to 0 as $t \to +\infty$ and the conclusion follows.

□

Proof of Theorem 3.2. We will use the following lemma, which is proved in Berkes, Horváth and Kokoszka (2003).

Lemma A.1 *Let X be a positive random variable such that $E \log X < 0$ and $EX^u < \infty$, for some $u > 0$. Then there exists a number $r > 0$ such that $EX^r < 1$.*

Note that (3.3) can be equivalently written as $E \log\{\prod_{j=1}^d a(e_j, \eta_j)^{\pi_j}\} < 0$. We will apply the lemma to the variable $\prod_{j=1}^d a(e_j, \eta_j)^{\pi_j}$. We can take $u = 1$ since

$$E \prod_{j=1}^d a(e_j, \eta_j)^{\pi_j} = \prod_{j=1}^d Ea(e_j, \eta_j)^{\pi_j} < \prod_{j=1}^d \{Ea(e_j, \eta_j) + 1\} < \prod_{j=1}^d \{\alpha(e_j) + \beta(e_j) + 1\} < \infty.$$

It follows that

$$\prod_{j=1}^d Ea(e_j, \eta_j)^{r_0 \pi_j} < 1, \qquad \text{for some } r_0 \in (0, 1). \tag{A.2}$$

Let $J = \{j \mid 1 \le j \le d, \pi_j > 0\} \neq \emptyset$ and let $r = r_0 \min_{j \in J} \pi_j \in (0, 1)$. Using the elementary inequality $(x + y)^r \le x^r + y^r$ for any $x, y \ge 0$, we have, using (3.4), for $K = \max_i\{\omega(e_i)\}$

$$\begin{aligned} E(\epsilon_t^2)^r &\le E\left\{\omega(s_t)^r + \sum_{n=1}^{+\infty} a(s_t, \eta_{t-1})^r \ldots a(s_{t-n+1}, \eta_{t-n})^r \omega(s_{t-n})^r\right\} E(\eta_t^2)^r \\ &\le K^r \left\{1 + \sum_{n=1}^{+\infty} \left\{\prod_{i=1}^n E\left[a(s_{t-i+1}, \eta_0)^r\right]\right\}\right\} E(\eta_t^2)^r \\ &= K^r \left\{1 + \sum_{n=1}^{+\infty} v_{n,t}\right\} E(\eta_t^2)^r \end{aligned} \tag{A.3}$$

where $v_{n,t} = \prod_{j=1}^d \{Ea(e_j, \eta_0)^r\}^{|\mathcal{T}(t,j,n-1)|}$. To apply the Cauchy rule we compute

$$v_{n,t}^{1/n} = \prod_{j=1}^d \{Ea(e_j, \eta_0)^r\}^{|\mathcal{T}(t,j,n-1)|/n} \sim \prod_{j=1}^d \{Ea(e_j, \eta_0)^r\}^{\pi_j} = \prod_{j \in J} \{Ea(e_j, \eta_0)^r\}^{\pi_j}.$$

Moreover, in view of the Jensen inequality $E(X^\nu) \le (EX)^\nu$ for any $\nu \in [0, 1]$ and any positive variable X, and since $r \le r_0 \pi_j$ for all $j \in J$, we have

$$\begin{aligned} \prod_{j \in J} \{Ea(e_j, \eta_0)^r\}^{\pi_j} &= \prod_{j \in J} \left[E\{a(e_j, \eta_0)^{r_0 \pi_j}\}^{\frac{r}{r_0 \pi_j}}\right]^{\pi_j} \\ &\le \prod_{j \in J} \{Ea(e_j, \eta_0)^{r_0 \pi_j}\}^{\frac{r}{r_0}} < 1, \end{aligned}$$

where the last inequality follows from (A.2). The convergence of the infinite sum in (A.3) follows from the Cauchy rule and the theorem is proved.

□

Proof of Theorem 3.3. First note that, because $E\eta_0^2 = 1$, the expectation of $\log a(e_j, \eta_0)$ is well-defined on $\mathbb{R} \cup \{-\infty\}$. By the the Jensen inequality we have

$$\gamma_0 = \sum_{j=1}^{d} \pi_j E\{\log a(e_j, \eta_0)\} \quad \leq \quad \sum_{j=1}^{d} \pi_j \log E\{a(e_j, \eta_0)\} = \log \prod_{j=1}^{d} \{Ea(e_j, \eta_0)\}^{\pi_j} = \log \gamma_1.$$

This proves that (3.7) implies (3.3). Thus we know that, under (3.7), the solution in (3.4) exists. Squaring it and taking the expectation yields

$$\begin{aligned} E\epsilon_t^2 &= E\left\{\omega(s_t) + \sum_{n=1}^{+\infty} a(s_t, \eta_{t-1}) \dots a(s_{t-n+1}, \eta_{t-n})\omega(s_{t-n})\right\} E\eta_t^2 \\ &= \omega(s_t) + \sum_{n=1}^{+\infty} E\left\{a(s_t, \eta_{t-1}) \dots a(s_{t-n+1}, \eta_{t-n})\right\} \omega(s_{t-n}) \\ &= \omega(s_t) + \sum_{n=1}^{+\infty} \prod_{i=1}^{n} E\left\{a(s_{t-i+1}, \eta_{t-i})\right\} \omega(s_{t-n}), \end{aligned} \tag{A.4}$$

where the second equality follows from Beppo Levi's monotone convergence theorem and the last equality follows from the independence of the variables η_{t-i}. Proceeding as in the proof of Theorem 3.1 we consider the general term of the previous sum, given by

$$v_n = \prod_{i=1}^{n} E\left\{a(s_{t-i+1}, \eta_0)\right\} \omega(s_{t-n}) = \prod_{j=1}^{d} \left\{Ea(e_j, \eta_0)\right\}^{|\mathcal{T}(t,j,n-1)|} \omega(s_{t-n}).$$

Thus

$$v_n^{1/n} = \prod_{j=1}^{d} \left\{Ea(e_j, \eta_0)\right\}^{|\mathcal{T}(t,j,n-1)|/n} \omega(s_{t-n})^{1/n} \to \prod_{j=1}^{d} \left\{Ea(e_j, \eta_0)\right\}^{\pi_j} = \gamma_1,$$

as $n \to \infty$. In view of (3.7), $\lim_{n\to\infty} v_n^{1/n} < 1$ and, by the Cauchy rule, the infinite sum in (A.4) converges. We have shown that $E\epsilon_t^2 < \infty$.

Now suppose that $\gamma_1 > 1$ and let (ϵ_t) a nonanticipative solution of model (2.1). Then by (A.1), we have

$$\begin{aligned} E\epsilon_t^2 &\geq \omega(s_t) + \sum_{n=1}^{N} \prod_{i=1}^{n} E\left\{a(s_{t-i+1}, \eta_0)\right\} \omega(s_{t-n}) \\ &= \omega(s_t) + \sum_{n=1}^{N} \prod_{j=1}^{d} \left\{Ea(e_j, \eta_0)\right\}^{|\mathcal{T}(t,j,n-1)|} \omega(s_{t-n}). \end{aligned}$$

By the Cauchy rule, the sum in the left-hand side goes to infinity as $N \to \infty$. It follows that $E\epsilon_t^2 = \infty$.

$\square$

Proof of Theorem 3.4. By the argument used in the proof of Theorem 3.3 we have $m\gamma_0 \leq \log \gamma_m$, proving that when $\gamma_m < 1$ the solution in (3.4) exists. Define the usual L^m norm by $\|X\|_m = \{E|X|^m\}^{1/m}$ for $m \geq 1$. With the notation used in the proof of Theorem 3.1, letting $u_{0,t} = \omega(s_t)$ and $\mu_{2m} = E\eta_t^{2m}$, we have, in view of (3.4),

$$\begin{aligned}
\|\epsilon_t^2\|_m &= \mu_{2m}^{1/m} \left\{ E \lim_{N\to\infty} \uparrow \left(\sum_{n=0}^{N} u_{n,t} \right)^m \right\}^{1/m} \\
&= \mu_{2m}^{1/m} \left\{ \lim_{N\to\infty} \uparrow E \left(\sum_{n=0}^{N} u_{n,t} \right)^m \right\}^{1/m} \\
&= \mu_{2m}^{1/m} \lim_{N\to\infty} \left\| \sum_{n=0}^{N} u_{n,t} \right\|_m \leq \mu_{2m}^{1/m} \sum_{n=0}^{\infty} \|u_{n,t}\|_m ,
\end{aligned}$$

where the second equality follows from the monotone convergence theorem. Moreover,

$$\begin{aligned}
\|u_{n,t}\|_m &= \{E[a(s_t, \eta_{t-1}) \dots a(s_{t-n+1}, \eta_{t-n})\omega(s_{t-n})]^m\}^{1/m} \\
&= \left\{ \prod_{i=1}^{n} E\left[a(s_{t-i+1}, \eta_0)\right]^m \right\}^{1/m} \omega(s_{t-n}) \\
&= \left\{ \prod_{j=1}^{d} \{E[a(e_j, \eta_0)]^m\}^{|\mathcal{T}(t,j,n-1)|} \right\}^{1/m} \omega(s_{t-n}).
\end{aligned}$$

Thus $\|u_{n,t}\|_m^{1/n} \to \gamma_m^{1/m}$ which, by (A.5) and the Cauchy rule, shows that $\gamma_m < 1$ is a sufficient condition for the finiteness of $E(\epsilon_t^{2m})$. The proof of the necessary part uses similar arguments as that of Theorem 3.3.

The following lemma, which is an extension of Lemma 1 in Francq and Gautier (2004a), will be used in the proof of Theorems 4.1 and 4.2. To facilitate reading we sometimes denote by $\mathbb{P}_X(A)$ (resp. $E_X(Y)$) the probability of an event A (resp the expectation of a variable Y) which is measurable with respect to the $\sigma-$field generated by a process $X = (X_t)$. We denote the set of the infinite sequences valued in $\mathcal{E}$ by $\mathcal{E}^\infty$.

Lemma A.2 *Assume that (η_t) is a stationary and ergodic process and that* **A0** *holds. Let f be a measurable function, $f : \{e_1, \dots, e_d\}^\infty \times \mathbb{R}^\infty \to \mathbb{R}$, such that $Ef(S_t, S_{t-1}, \dots, \eta_t, \eta_{t-1}, \dots)$*

exists in $\mathbb{R}\cup\{-\infty,+\infty\}$*. Then, for* $\mathbb{P}_S$*-almost all sequence* (s_t)*,*

$$\frac{1}{n}\sum_{t=1}^{n} f(s_t,s_{t-1},\ldots,\eta_t,\eta_{t-1},\ldots)\to Ef(S_t,S_{t-1},\ldots,\eta_t,\eta_{t-1},\ldots),\quad \mathbb{P}_\eta - a.s.$$

More generally, if $X_n = X_n(S_n,S_{n-1},\ldots,\eta_n,\eta_{n-1},\ldots)$ *is a random variable which is measurable with respect to the* $\sigma-$*field* $\sigma(\{S_t,\eta_t,t\le n\}$ *and such that* $X_n\to X$ *a.s., where* X *is a random variable. Then, for* $\mathbb{P}_S$*-almost all sequence* (s_t)*,*

$$X_n(s_n,s_{n-1},\ldots,\eta_n,\eta_{n-1},\ldots)\to X,\quad \mathbb{P}_\eta - a.s.$$

Proof: The only difference between the first part of this lemma and Lemma 1 in Francq and Gautier (2004a) is that the expectation involved might not be finite. Their proof can be straightforwardly adapted using the following ergodic theorem: if (X_t) is a stationary and ergodic process such that $EX_1\in\mathbb{R}\cup\{+\infty\}$, then $n^{-1}\sum_{t=1}^n X_t$ converges $a.s.$ to EX_1 when $n\to\infty$ (see Billingsley (1995) p. 284 and 495).

The second part of the lemma is also proved following the lines of Francq and Gautier (2004a). We give the proof for completeness. The processes (S_t) and (η_t) being independent, we can assume, without loss of generality, that $\Omega=\Omega_1\times\Omega_2$ where $\Omega_1=\{e_1,\ldots,e_d\}^\infty$ and $\Omega_2=\mathbb{R}^\infty$, that $\mathcal{A}=\mathcal{A}_1\otimes\mathcal{A}_2$ where $\mathcal{A}_1$ and $\mathcal{A}_2$ are σ-fields defined on Ω_1 and Ω_2, and that $\mathbb{P}$ is a product probability $\mathbb{P}_1\otimes\mathbb{P}_2$, where $\mathbb{P}_1$ is defined on $\mathcal{A}_1$ and $\mathbb{P}_2$ is defined on $\mathcal{A}_2$. It follows that if $\omega=(\omega_1,\omega_2)$ with $\omega_1=(\omega_{1t}),\omega_2=(\omega_{2t})$, then $(S_t,\eta_t)(\omega)=(\omega_{1t},\omega_{2t})$. Define

$$\Omega^*=\{\omega\in\Omega : X_n(S_n(\omega),S_{n-1}(\omega),\ldots,\eta_n(\omega),\eta_{n-1}(\omega),\ldots)\nrightarrow X\}.$$

We have $\mathbb{P}(\Omega^*)=0$. For every $\omega_1\in\Omega_1$, define the set $\Omega^*_{\omega_1}=\{\omega_2\in\Omega_2 : (\omega_1,\omega_2)\in\Omega^*\}$. Since the integral of a positive function is equal to zero if and only if the function is zero almost everywhere,

$$\mathbb{P}(\Omega^*)=\int_{\Omega_1}\left\{\int_{\Omega_2}\mathbb{I}_{\Omega^*}(\omega_1,\omega_2)d\mathbb{P}_2(\omega_2)\right\}d\mathbb{P}_1(\omega_1)=0$$

implies that

$$\int_{\Omega_2}\mathbb{I}_{\Omega^*}(\omega_1,\omega_2)d\mathbb{P}_2(\omega_2)=\mathbb{P}_2(\Omega^*_{\omega_1})=\mathbb{P}(\Omega_1\times\Omega^*_{\omega_1})=0$$

when $\omega_1\notin\Omega_1^*$, for some measurable set Ω_1^* such that $\mathbb{P}_1(\Omega_1^*)=0$. If $\omega_1\notin\Omega_1^*$ and $\omega_2\notin\Omega^*_{\omega_1}$, then $(\omega_1,\omega_2)\notin\Omega^*$ which entails

$$X_n(s_n,s_{n-1},\ldots,\eta_n(\omega),\eta_{n-1}(\omega),\ldots)\to X,$$

when $(s_t)=\{S_t(\omega_1,\omega_2)\}$.

Proof of Theorem 4.1. The scheme of the proof is the same as that of the strong consistency for standard GARCH in FZ. See Pfanzagl (1969) for the consistency of minimum contrast estimators in a more general context. It will be convenient to approximate the sequence $(\tilde{\ell}_t(\theta))$ by a sequence $(\ell_t(\theta))$ which is independent of the initial values. Therefore, denote by $\left(\sigma_t^2\right)_t = \left\{\sigma_t^2(\theta)\right\}_t$ the solution of

$$\sigma_t^2 = \omega(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2$$

given by

$$\sigma_t^2 = c_t + \sum_{i=0}^{\infty} \beta(s_t)\ldots\beta(s_{t-i})c_{t-i-1}, \quad a.s. \tag{A.5}$$

where $c_t = \omega(s_t) + \alpha(s_t)\epsilon_{t-1}^2$. The almost sure convergence of the last sum is ensured by the second part of **A2** and arguments already given. Let

$$\mathbf{l}_n(\theta) = \mathbf{l}_n(\theta; \epsilon_n, \epsilon_{n-1}\ldots,) = n^{-1}\sum_{t=1}^{n} \ell_t, \quad \ell_t = \ell_t(\theta) = \frac{\epsilon_t^2}{\sigma_t^2} + \ln \sigma_t^2.$$

For any $\theta \in \Theta$ the strictly stationary, nonanticipative solution $(\sigma_{S,t}^2)$ to Model (4.6) is given by

$$\sigma_{S,t}^2 = \sigma_{S,t}^2(\theta) = c_{S,t}(\theta) + \sum_{i=0}^{\infty} \beta(S_t)\ldots\beta(S_{t-i})c_{S,t-i-1}(\theta), \quad a.s. \tag{A.6}$$

where $c_{S,t} = c_{S,t}(\theta) = \omega(S_t) + \alpha(S_t)\epsilon_{S,t-1}^2$. Note that the infinite sum converges because

$$E\log\beta(S_t) = \log\left(\prod_{j=1}^{d}\beta^{\pi_j}(e_j)\right) < 0 \quad \text{and } E|\log c_{S,t}(\theta)| < \infty$$

under **A2**. The latter inequality follows from *i)* below and the positivity of ω over Θ. Let $\ell_{S,t} = \ell_{S,t}(\theta) = \frac{\epsilon_{S,t}^2}{\sigma_{S,t}^2} + \ln \sigma_{S,t}^2$. We will establish the following intermediate results :

i) $E_{S,\eta}\{\epsilon_{S,t}^2\}^r < \infty$, where r is defined in **A3**.

ii) $\lim_{n\to\infty}\sup_{\theta\in\Theta}|\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| = 0, \quad a.s.$

iii) $\sigma_{S,t}(\theta) = \sigma_{S,t}(\theta_0), \quad \forall t,\ P_{\theta_0}\ a.s. \quad \Longrightarrow \quad \theta = \theta_0.$

iv) $E|\ell_{S,t}(\theta_0)| < \infty$, and if $\theta \neq \theta_0, \quad E\ell_{S,t}(\theta) > E\ell_{S,t}(\theta_0)$.

v) any $\theta \neq \theta_0$ has a neighborhood $V(\theta)$ such that

$$\liminf_{n\to\infty} \inf_{\theta^*\in V(\theta)} \tilde{\mathbf{l}}_n(\theta) > E\ell_{S,1}(\theta_0), \quad a.s.$$

The letter K is used for a positive generic constant, possibly random, whose value is allowed to change throughout the proof.

We start by proving *i)*. Using the elementary inequality $(x+y)^r \leq x^r + y^r$ for any $x, y \geq 0$, $r \in (0,1)$, by assumption **A3** and (3.10),

$$\begin{aligned} E_{S,\eta}(\epsilon_{S,t}^2)^r &\leq E\left\{\omega_0^r(S_t) + \sum_{n=1}^{+\infty} a_0^r(S_t,\eta_{t-1})\dots a_0^r(S_{t-n+1},\eta_{t-n})\omega^r(S_{t-n})\right\} E(\eta_t^2)^r \\ &\leq \{\max_i \omega_0^r(e_i)\}\left\{1 + \sum_{n=1}^{+\infty} Ea_0^r(S_t,\eta_{t-1})\dots a_0^r(S_{t-n+1},\eta_{t-n})\right\} E(\eta_t^2)^r \\ &\leq K\left\{1 + K\sum_{n=1}^{+\infty}\rho^n\right\} < \infty. \end{aligned}$$

Now we prove *ii)*. We have

$$\begin{aligned} \sup_{\theta\in\Theta} |\tilde{\sigma}_t^2 - \sigma_t^2| &= \sup_{\theta\in\Theta} \left|\beta(s_t)\dots\beta(s_2)\{\alpha(s_1)(\tilde{\epsilon}_0^2 - \epsilon_0^2) + \beta(s_1)(\tilde{\sigma}_0^2 - \sigma_0^2)\}\right| \\ &\leq K\beta(s_t)\dots\beta(s_2), \quad \forall t \geq 2 \end{aligned} \tag{A.7}$$

Thus, using $\log x \leq x - 1$, almost surely,

$$\begin{aligned} \sup_{\theta\in\Theta} |\tilde{\mathbf{l}}_n(\theta) - \mathbf{l}_n(\theta)| &\leq n^{-1}\sum_{t=1}^n \sup_{\theta\in\Theta}\left\{\left|\frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\tilde{\sigma}_t^2\sigma_t^2}\right|\epsilon_t^2 + \left|\log\left(1 + \frac{\sigma_t^2 - \tilde{\sigma}_t^2}{\tilde{\sigma}_t^2}\right)\right|\right\} \\ &\leq \left\{\sup_{\theta\in\Theta}\max_i \frac{1}{\omega^2(e_i)}\right\} Kn^{-1}\sum_{t=1}^n \beta(s_t)\dots\beta(s_2)\epsilon_t^2 \\ &\quad + \left\{\sup_{\theta\in\Theta}\max_i \frac{1}{\omega(e_i)}\right\} Kn^{-1}\sum_{t=1}^n \beta(s_t)\dots\beta(s_2). \end{aligned} \tag{A.8}$$

Now, using the second part of **A2**,

$$\{\beta(s_t)\dots\beta(s_2)\}^{1/t} = \prod_{j=1}^d \beta(e_j)^{|\mathcal{T}(t,j,t-2)|/t} \to \prod_{j=1}^d \beta^{\pi_j}(e_j) < 1$$

as $t \to \infty$. It follows that, for t large enough, $\beta(s_t)\dots\beta(s_2) < \beta_*^t$, for some $0 < \beta_* < 1$. Therefore the second term, in the right-hand side of (A.8) tends to 0 as $n \to \infty$, and the first one is bounded, for n large enough, by $Kn^{-1}\sum_{t=1}^n \beta_*^t\epsilon_t^2$. By Lemma A.2, it will be sufficient to show that $n^{-1}\sum_{t=1}^n \beta_*^t\epsilon_{S,t}^2$ almost surely converges to 0 as $n \to \infty$. This result follows from the Cesaro and Borel-Cantelli lemmas, and

$$\sum_{t=1}^{\infty}\mathbb{P}(\beta_*^t\epsilon_{S,t}^2 > \zeta) \leq \sum_{t=1}^{\infty}\frac{E(\beta_*^t\epsilon_{S,t}^2)^r}{\zeta^r} < \infty$$

for any $\zeta > 0$, where the last inequality follows from *i)*. Thus *ii)* is proved.

Now we turn to *iii)*. Let $\sigma_{S,t}(\theta) = \sigma_{S,t}(\theta_0), \quad \forall t,\ P_{\theta_0}\ a.s.$ Then we have

$$\begin{aligned} \omega_0(S_t) - \omega(S_t) &= \{\alpha(S_t) - \alpha_0(S_t)\}\epsilon_{S,t-1}^2 + \{\beta(S_t) - \beta_0(S_t)\}\sigma_{S,t-1}^2(\theta_0) \\ &= [\{\alpha(S_t) - \alpha_0(S_t)\}\eta_{t-1}^2 + \{\beta(S_t) - \beta_0(S_t)\}]\sigma_{S,t-1}^2(\theta_0). \end{aligned} \tag{A.9}$$

By **A5**, $\mathbb{P}(S_t = e_i) > 0$ for all i. Suppose that $\alpha(e_i) - \alpha_0(e_i) \neq 0$ for some i. Then, conditional on $(S_t = e_i)$ we have by (A.9),

$$\eta_{t-1}^2 = \{\alpha(e_i) - \alpha_0(e_i)\}^{-1}[\sigma_{S,t-1}^{-2}(\theta_0)\{\omega_0(e_i) - \omega(e_i)\} - \{\beta(e_i) - \beta_0(e_i)\}].$$

Because the right-hand side is independent from η_{t-1}^2, we deduce that η_{t-1}^2 is a.s. constant conditional on $(S_t = e_i)$. Thus, from the independence between η_{t-1}^2 and S_t, we find that η_{t-1}^2 is constant a.s. This is precluded by **A4**. Hence we have $\alpha(e_i) = \alpha_0(e_i)$ for all i.
In view of (A.9) we then have

$$\begin{aligned}\omega_0(S_t) - \omega(S_t) &= \{\beta(S_t) - \beta_0(S_t)\}\sigma_{S,t-1}^2(\theta_0) \\ &= \{\beta(S_t) - \beta_0(S_t)\}\{\omega_0(S_{t-1}) + \alpha_0(S_{t-1})\epsilon_{S,t-2}^2 + \beta_0(S_{t-1})\sigma_{S,t-2}^2\}.\end{aligned}$$

Let $i \in \{1, \ldots, d\}$ and let j_i the smallest integer $k > 0$ such that $\alpha_0(e_\ell)\mathbb{P}(S_{t-k} = e_\ell, S_t = e_i) > 0$ for some $\ell \in \{1, \ldots, d\}$. For ease of exposition we assume $j_i = 2$. Thus $\alpha_0(S_{t-1}) = 0$, a.s. conditional on $(S_t = e_i)$. We then have, conditional on $(S_t = e_i)$,

$$\begin{aligned}\omega_0(e_i) - \omega(e_i) &= \{\beta(e_i) - \beta_0(e_i)\}\{\omega_0(S_{t-1}) + \beta_0(S_{t-1})\sigma_{S,t-2}^2\} \\ &= \{\beta(e_i) - \beta_0(e_i)\}\{\omega_0(S_{t-1}) \\ &\quad +\beta_0(S_{t-1})[\omega_0(S_{t-2}) + \alpha_0(S_{t-2})\epsilon_{S,t-3}^2 + \beta_0(S_{t-2})\sigma_{S,t-3}^2]\}.\end{aligned}$$

It follows that conditional on $(S_t = e_i, S_{t-2} = e_\ell)$,

$$\begin{aligned}\omega_0(e_i) - \omega(e_i) &= \{\beta(e_i) - \beta_0(e_i)\}\{\omega_0(S_{t-1}) \\ &\quad +\beta_0(S_{t-1})[\omega_0(e_\ell) + (\alpha_0(e_\ell)\eta_{t-3}^2 + \beta_0(e_\ell))\sigma_{S,t-3}^2]\}.\end{aligned}$$

By independence arguments already given, the coefficient of η_{t-3}^2 must be equal to zero conditional on $(S_t = e_i, S_{t-2} = e_\ell)$:

$$\mathbb{P}[\{\beta(e_i) - \beta_0(e_i)\}\beta_0(S_{t-1})\alpha_0(e_\ell) = 0 \mid S_t = e_i, S_{t-2} = e_\ell] = 1.$$

Noting that $\beta_0(S_{t-1})\alpha_0(e_\ell) \neq 0$ a.s. conditional on $(S_t = e_i, S_{t-2} = e_\ell)$ (because by the first part of **A5**, $\alpha_0(S_{t-1}) = 0$ implies $\beta_0(S_{t-1}) \neq 0$), we can conclude that $\beta(e_i) = \beta_0(e_i)$. The argument clearly extends when $j_i > 2$. Thus we have proved that for any $i \in \{1, \ldots, d\}$, $\alpha(e_i) = \alpha_0(e_i)$ and $\beta(e_i) = \beta_0(e_i)$. It is straightforward from (A.9) that $\omega(e_i) = \omega_0(e_i)$ and *ii)* is proved.

Let us turn to *iv)*. First note that $E\ell_{S,t}(\theta)$ is well defined and belongs to $\mathbb{R}\cup\{\infty\}$, because $E\ell^-_{S,t}(\theta) \le E\log^- \sigma^2_{S,t}(\theta) \le \max\{0, -\log\min_i \omega(e_i)\} < \infty$. Remark that $E\ell_{S,t}(\theta) = \infty$, when, for instance, $\theta = (\omega, 0, \ldots, 0, \omega, 0, \ldots)$ and $E\epsilon^2_{S,t} = \infty$. However, we will show that $E|\ell_{S,t}(\theta_0)| < \infty$. It remains to show that $E\ell^+_{S,t}(\theta_0) < \infty$. By *i)*,

$$E\log\sigma^2_{S,t}(\theta_0) = E\frac{1}{r}\log\{\sigma^2_{S,t}(\theta_0)\}^r \le \frac{1}{r}\log E\{\sigma^2_{S,t}(\theta_0)\}^r < \infty.$$

Therefore

$$E\ell_{S,t}(\theta_0) = E\left\{\frac{\sigma^2_{S,t}(\theta_0)\eta_t^2}{\sigma^2_{S,t}(\theta_0)} + \log\sigma^2_{S,t}(\theta_0)\right\} = 1 + E\log\sigma^2_{S,t}(\theta_0) < \infty,$$

which proves the first part of *iv)*. Since $\log x \le x-1, \quad \forall x > 0$ and $\log x = x - 1$ if and only if $x = 1$, we have

$$\begin{aligned} E\ell_{S,t}(\theta) - E\ell_{S,t}(\theta_0) &= E\log\frac{\sigma^2_{S,t}(\theta)}{\sigma^2_{S,t}(\theta_0)} + E\frac{\sigma^2_{S,t}(\theta)}{\sigma^2_{S,t}(\theta_0)} - 1 \\ &\ge E\left\{\log\frac{\sigma^2_{S,t}(\theta)}{\sigma^2_{S,t}(\theta_0)} + \log\frac{\sigma^2_{S,t}(\theta_0)}{\sigma^2_{S,t}(\theta)}\right\} \ge 0 \end{aligned}$$

with equality if and only if $\frac{\sigma^2_{S,t}(\theta_0)}{\sigma^2_{S,t}(\theta)} = 1 \quad P_{\theta_0} a.s.$ In view of *iii)*, *iv)* then follows.

It remains to show *v)*. For any $\theta \in \Theta$ and any positive integer k, let $V_k(\theta)$ be the open ball with center θ and radius $1/k$. Following exactly the lines of proof in Francq and Zakoïan (2005), we have

$$\begin{aligned} \liminf_{n\to\infty} \inf_{\theta^*\in V_k(\theta)\cap\Theta} \tilde{\mathbf{l}}_n(\theta^*) &\ge \liminf_{n\to\infty} \inf_{\theta^*\in V_k(\theta)\cap\Theta} \mathbf{l}_n(\theta^*) - \limsup_{n\to\infty}\sup_{\theta\in\Theta}|\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| \\ &\ge \liminf_{n\to\infty} n^{-1}\sum_{t=1}^n \inf_{\theta^*\in V_k(\theta)\cap\Theta} \ell_t(\theta^*). \end{aligned}$$

Now we will apply Lemma A.2. We note that $\left\{\inf_{\theta^*\in V_k(\theta)\cap\Theta}\ell_t(\theta^*)\right\}_t$ is a measurable function of $(s_t, s_{t-1}, \ldots, \eta_t, \eta_{t-1}, \ldots)$, whose expectation exists in $\mathbb{R}\cup\{+\infty\}$. Thus we obtain

$$\liminf_{n\to\infty} n^{-1}\sum_{t=1}^n \inf_{\theta^*\in V_k(\theta)\cap\Theta}\ell_t(\theta^*) = E_S \inf_{\theta^*\in V_k(\theta)\cap\Theta}\ell_{S,1}(\theta^*).$$

By the Beppo-Levi theorem, when k increases to ∞, $E_S\inf_{\theta^*\in V_k(\theta)\cap\Theta}\ell_{S,1}(\theta^*)$ increases to $E_S\ell_{S,1}(\theta)$. Thus *v)* is proved. By a standard compactness argument the proof of Theorem 4.1 is completed.

□

Proof of Theorem 4.2. The proof follows the scheme of the asymptotic normality proof in FZ. Because $\hat{\theta}_n$ is strongly consistent and θ_0 belongs to the interior of Θ, the criterion derivative cancels at $\hat{\theta}_n$ and we have :

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\hat{\theta}_n) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0) + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\ell}_t(\theta_{ij}^*)\right)\sqrt{n}(\hat{\theta}_n-\theta_0), \end{aligned}$$

where the θ_{ij}^* are between $\hat{\theta}_n$ and θ_0. Let $\{\mathcal{F}_t\}$ denote the $\sigma-$algebra generated by the random variables $\eta_{t-i}, i \geq 0$. The norm of a matrix $A = (a_{ij})$ is defined by $\|A\| = \sum |a_{ij}|$. The proof of Theorem 4.2 consists of several intermediate steps.

i) $E_{S,\eta}\left\|\frac{\partial\ell_{S,t}(\theta_0)}{\partial\theta}\frac{\partial\ell_{S,t}(\theta_0)}{\partial\theta'}\right\| < \infty, \quad E_{S,\eta}\left\|\frac{\partial^2\ell_{S,t}(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty.$

ii) J is non singular and $\frac{1}{n}\sum_{t=1}^{n}\operatorname{Var}\left\{\frac{\partial\ell_t(\theta_0)}{\partial\theta} \mid \mathcal{F}_{t-1}\right\} \to (\kappa_\eta - 1)J, \quad a.s.$

iii) There exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that, for all $i,j,k \in \{1,\ldots,3d\}$,

$$E_{S,\eta}\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{\partial^3\ell_{S,t}(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| < \infty.$$

iv) $\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial\tilde{\ell}_t(\theta_0)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right\}\right\| \to 0$ and $\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2\tilde{\ell}_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right\}\right\| \to 0$ almost surely as $n \to \infty$.

v) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\ell_t(\theta_0)}{\partial\theta} \to \mathcal{N}(0,(\kappa_\eta-1)J).$

vi) $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\ell_t(\theta_{ij}^*)}{\partial\theta\partial\theta'} \to J(i,j), \quad a.s.$

The first and second derivatives of $\ell_{S,t} = \frac{\epsilon_{S,t}^2}{\sigma_{S,t}^2} + \log\sigma_{S,t}^2$ are given by :

$$\frac{\partial\ell_{S,t}}{\partial\theta} = \left\{1 - \frac{\epsilon_{S,t}^2}{\sigma_{S,t}^2}\right\}\left\{\frac{1}{\sigma_{S,t}^2}\frac{\partial\sigma_{S,t}^2}{\partial\theta}\right\}, \tag{A.10}$$

$$\frac{\partial^2\ell_{S,t}}{\partial\theta\partial\theta'} = \left\{1 - \frac{\epsilon_{S,t}^2}{\sigma_{S,t}^2}\right\}\left\{\frac{1}{\sigma_{S,t}^2}\frac{\partial^2\sigma_{S,t}^2}{\partial\theta\partial\theta'}\right\} + \left\{2\frac{\epsilon_{S,t}^2}{\sigma_{S,t}^2} - 1\right\}\left\{\frac{1}{\sigma_{S,t}^2}\frac{\partial\sigma_{S,t}^2}{\partial\theta}\right\}\left\{\frac{1}{\sigma_{S,t}^2}\frac{\partial\sigma_{S,t}^2}{\partial\theta'}\right\}. \tag{A.11}$$

For $\theta = \theta_0$, $\frac{\epsilon_{S,t}^2}{\sigma_{S,t}^2} = \eta_t^2$ is independent of the terms involving σ_t^2 and its derivatives. To prove *i)*, it will therefore be sufficient to show that:

$$E_{S,\eta}\left\|\frac{1}{\sigma_{S,t}^2}\frac{\partial\sigma_{S,t}^2}{\partial\theta}(\theta_0)\right\| < \infty, \quad E_{S,\eta}\left\|\frac{1}{\sigma_{S,t}^2}\frac{\partial^2\sigma_{S,t}^2}{\partial\theta\partial\theta'}(\theta_0)\right\| < \infty, \quad E_{S,\eta}\left\|\frac{1}{\sigma_{S,t}^4}\frac{\partial\sigma_{S,t}^2}{\partial\theta}\frac{\partial\sigma_{S,t}^2}{\partial\theta'}(\theta_0)\right\| < \infty. \tag{A.12}$$

By (A.6) we have, for $k \in \{1 \ldots, d\}$,

$$\frac{\partial \sigma^2_{S,t}}{\partial \omega(e_k)} = \mathbb{1}_{S_t=e_k} + \sum_{i=0}^{\infty} \beta(S_t)\ldots\beta(S_{t-i})\mathbb{1}_{S_{t-i-1}=e_k}, \tag{A.13}$$

$$\frac{\partial \sigma^2_{S,t}}{\partial \alpha(e_k)} = \epsilon^2_{S,t-1}\mathbb{1}_{S_t=e_k} + \sum_{i=0}^{\infty} \beta(S_t)\ldots\beta(S_{t-i})\epsilon^2_{S,t-i-2}\mathbb{1}_{S_{t-i-1}=e_k}, \tag{A.14}$$

$$\frac{\partial \sigma^2_{S,t}}{\partial \beta(e_k)} = \sum_{i=0}^{\infty} \frac{\beta(S_t)\ldots\beta(S_{t-i})}{\beta(e_k)} \left\{\sum_{j=0}^{i} \mathbb{1}_{S_{t-j}=e_k}\right\} c_{S,t-i-1}, \quad \text{if } \beta(e_k) \neq 0. \tag{A.15}$$

Notice that by the positivity of coefficients in (A.13)-(A.15), the derivatives of $\sigma^2_{S,t}$ are nonnegative. By (A.14) we have :

$$\alpha(e_k)\frac{\partial \sigma^2_{S,t}}{\partial \alpha(e_k)} \leq c_{S,t}\mathbb{1}_{S_t=e_k} + \sum_{i=0}^{\infty} \beta(S_t)\ldots\beta(S_{t-i})c_{S,t-i-1}\mathbb{1}_{S_{t-i-1}=e_k} \leq \sigma^2_{S,t},$$

from which we deduce, if $\alpha(e_k) \neq 0$,

$$\frac{1}{\sigma^2_{S,t}}\frac{\partial \sigma^2_{S,t}}{\partial \alpha(e_k)} \leq \frac{1}{\alpha(e_k)}. \tag{A.16}$$

We similarly obtain, noting that $\omega(e_k) > 0$,

$$\frac{1}{\sigma^2_{S,t}}\frac{\partial \sigma^2_{S,t}}{\partial \omega(e_k)} \leq \frac{1}{\omega(e_k)}. \tag{A.17}$$

Turning to the derivative with respect to $\beta(e_k)$ we get, in view of (A.15),

$$\beta(e_k)\frac{\partial \sigma^2_{S,t}}{\partial \beta(e_k)} \leq \sum_{i=0}^{\infty}(i+1)\beta(S_t)\ldots\beta(S_{t-i})c_{S,t-i-1}. \tag{A.18}$$

Thus, since $\beta_0(e_k) > 0$,

$$\begin{aligned}
E_{S,\eta}\frac{1}{\sigma^2_{S,t}}\frac{\partial \sigma^2_{S,t}}{\partial \beta_0(e_k)}(\theta_0) &\leq E_{S,\eta}\frac{1}{\beta_0(e_k)}\sum_{i=0}^{\infty}\frac{(i+1)\beta_0(S_t)\ldots\beta_0(S_{t-i})c_{S,t-i-1}(\theta_0)}{\min_j \omega_0(e_j) + \beta_0(S_t)\ldots\beta_0(S_{t-i})c_{S,t-i-1}(\theta_0)} \\
&\leq \frac{1}{\beta_0(e_k)}\sum_{i=0}^{\infty}(i+1)E_{S,\eta}\left\{\frac{\beta_0(S_t)\ldots\beta_0(S_{t-i})c_{S,t-i-1}(\theta_0)}{\min_j \omega_0(e_j)}\right\}^{r/2} \\
&\leq \frac{K}{\beta_0(e_k)}\sum_{i=0}^{\infty}(i+1)\left\{E_S\left\{\beta_0(S_t)\ldots\beta_0(S_{t-i})\right\}^r E_{S,\eta}\left\{c_{S,t-i-1}(\theta_0)\right\}^r\right\}^{1/2} \\
&\leq \frac{K}{\beta_0(e_k)}\left\{E_{S,\eta}c^r_{S,t}(\theta_0)\right\}^{1/2}\sum_{i=0}^{\infty}(i+1)\left\{E\left\{\beta_0(S_t)\ldots\beta_0(S_{t-i})\right\}^r\right\}^{1/2} \\
&\leq \frac{K}{\beta_0(e_k)}\sum_{i=0}^{\infty}(i+1)\rho^{i/2} < K. \qquad \text{(A.19)}
\end{aligned}$$

The first inequality directly follows from (A.18). The second inequality relies on the elementary inequality $\frac{x}{1+x} \leq x^s$ for any $x \geq 0$ and $s \in [0, 1]$. The third inequality is based on the Hölder inequality and the fact that the $\omega_0(e_j)$ are positive. The elementary inequality $(a+b)^s \le a^s + b^s$, for all $a, b \ge 0$, and $s \in [0,1]$, together with *i)* in the proof of Theorem 4.1, entails the existence of a moment of order r for $c_{S,t}(\theta_0) = \omega_0(S_t) + \alpha_0(S_t)\epsilon^2_{S,t-1}$. From this, and the strict stationarity of $(c_{S,t}(\theta_0))$, we deduced the fourth inequality. Finally, the last two inequalities are direct consequences of **A3**, $\rho < 1$ and the positivity of $\beta_0(e_k)$ under **A6**. By (A.16), (A.17), (A.19) we can conclude that the first expectation in (A.12) exists.

Let us now turn to the second-order derivatives of $\sigma^2_{S,t}$. It follows from (A.13) that for any $k, \ell \in \{1, \dots, d\}$, $\frac{\partial^2 \sigma^2_{S,t}}{\partial \omega(e_k)\partial \omega(e_\ell)} = \frac{\partial^2 \sigma^2_{S,t}}{\partial \omega(e_k)\partial \alpha(e_\ell)} = 0$ and, if $\beta(e_\ell) \neq 0$,

$$\frac{\partial^2 \sigma^2_{S,t}}{\partial \omega(e_k)\partial \beta(e_\ell)} = \sum_{i=0}^{\infty} \frac{\beta(S_t)\dots\beta(S_{t-i})}{\beta(e_\ell)} \left\{ \sum_{j=0}^{i} \mathbb{1}_{S_{t-j}=e_\ell} \right\} \mathbb{1}_{S_{t-i-1}=e_k}. \tag{A.20}$$

Thus, using **A3**

$$\begin{aligned} E_{S,\eta} \frac{\beta_0(e_\ell)}{\sigma^2_{S,t}(\theta_0)} \frac{\partial^2 \sigma^2_{S,t}}{\partial \omega(e_k)\partial \beta(e_\ell)}(\theta_0) &\le \frac{1}{\min_j \omega_0(e_j)} \sum_{i=0}^{\infty} (i+1) E\beta_0(S_t)\dots\beta_0(S_{t-i}) \\ &\le K \sum_{i=0}^{\infty} (i+1)\rho^{i+1} < \infty. \end{aligned} \tag{A.21}$$

By (A.14) we find $\frac{\partial^2 \sigma^2_{S,t}}{\partial \alpha(e_k)\partial \alpha(e_\ell)} = 0$ and

$$\begin{aligned} \alpha(e_k)\beta(e_\ell) \frac{\partial^2 \sigma^2_{S,t}}{\partial \alpha(e_k)\partial \beta(e_\ell)} &= \alpha(e_k) \sum_{i=0}^{\infty} \beta(S_t)\dots\beta(S_{t-i}) \left\{ \sum_{j=0}^{i} \mathbb{1}_{S_{t-j}=e_\ell} \right\} \epsilon^2_{t-i-2} \mathbb{1}_{S_{t-i-1}=e_k} \\ &\le \sum_{i=0}^{\infty} (i+1)\beta(S_t)\dots\beta(S_{t-i}) c_{S,t-i-1}. \end{aligned}$$

The arguments used to prove (A.19) give for $\theta = \theta_0$,

$$E_{S,\eta} \frac{1}{\sigma^2_{S,t}(\theta_0)} \frac{\partial^2 \sigma^2_{S,t}}{\partial \alpha(e_k)\partial \beta(e_\ell)}(\theta_0) \le \frac{K}{\alpha_0(e_k)\beta_0(e_\ell)} < K. \tag{A.22}$$

Moreover, for $\ell \neq k$

$$\beta(e_k)\beta(e_\ell) \frac{\partial^2 \sigma^2_{S,t}}{\partial \beta(e_\ell)\partial \beta(e_k)} = \sum_{i=0}^{\infty} \beta(S_t)\dots\beta(S_{t-i}) \left\{ \sum_{h,j=0}^{i} \mathbb{1}_{S_{t-h}=e_k} \mathbb{1}_{S_{t-j}=e_\ell} \right\} c_{S,t-i-1},$$

while

$$\beta^2(e_k) \frac{\partial^2 \sigma^2_{S,t}}{\partial \beta(e_k)^2} = \sum_{i=0}^{\infty} \beta(S_t)\dots\beta(S_{t-i}) \left\{ \sum_{h,j=0}^{i} \mathbb{1}_{S_{t-h}=e_k} \mathbb{1}_{S_{t-j}=e_k} - \sum_{j=0}^{i} \mathbb{1}_{S_{t-j}=e_k} \right\} c_{S,t-i-1}.$$

By arguments used to show (A.19), we can conclude that

$$E_{S,\eta}\frac{1}{\sigma^2_{S,t}(\theta_0)}\frac{\partial^2\sigma^2_{S,t}}{\partial\beta(e_k)\partial\beta(e_\ell)}(\theta_0)\leq\frac{K}{\beta_0(e_k)\beta_0(e_\ell)}.$$

This, together with (A.21) and (A.22), proves the existence of the second expectation in (A.12). Now, since by (A.16) $\frac{1}{\sigma^2_{S,t}}\frac{\partial\sigma^2_{S,t}}{\partial\omega(e_k)}$ is bounded, and, by (A.16), $\frac{1}{\sigma^2_{S,t}}\frac{\partial\sigma^2_{S,t}}{\partial\alpha(e_k)}$ is bounded at θ_0, it is clear that

$$E_S\left\|\frac{1}{\sigma^4_{S,t}(\theta_0)}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\omega(e_k)}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta}\right\|<\infty,\quad E_S\left\|\frac{1}{\sigma^4_{S,t}(\theta_0)}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\alpha(e_k)}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta}\right\|<\infty.$$

In view of (A.18), arguments already used to prove (A.19), in particular the inequality $x/(1+x)\leq x^{r/2}$, and the Minkowski inequality, we have

$$\begin{aligned}&\left\{E_{S,\eta}\left(\frac{1}{\sigma_{S,t}(\theta_0)^2}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\beta(e_k)}\right)^2\right\}^{1/2}\\ \leq\ &\sum_{i=0}^{\infty}\frac{i+1}{\beta_0(e_k)}\left\{E_S\left\{\frac{\beta_0(S_t)\ldots\beta_0(S_{t-i})c_{t-i-1}(\theta_0)}{\min_j\omega_0(e_j)}\right\}^r\right\}^{1/2}<\infty.\end{aligned}$$

Finally we conclude by the Cauchy-Schwarz inequality that the third expectation in (A.12) exists.

Now we prove *ii)*. Note that J exists by (A.12). By (A.10) and *i)* we have

$$\begin{aligned}E_{S,\eta}\left\{\frac{\partial\ell_{S,t}(\theta_0)}{\partial\theta}\right\}&=E(1-\eta_t^2)E_{S,\eta}\left\{\frac{1}{\sigma^2_{S,t}(\theta_0)}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta}\right\}=0,\\ \mathrm{Var}_{S,\eta}\left\{\frac{\partial\ell_{S,t}(\theta_0)}{\partial\theta}\right\}&=E_{S,\eta}\left\{\frac{\partial\ell_{S,t}(\theta_0)}{\partial\theta}\frac{\partial\ell_{S,t}(\theta_0)}{\partial\theta'}\right\}\\ &=E\{(1-\eta_t^2)^2\}E_{S,\eta}\left\{\frac{1}{\sigma^4_{S,t}(\theta_0)}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta}\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta'}\right\}\\ &=(\kappa_\eta-1)J.\end{aligned}$$

Now suppose

$$\lambda'J\lambda=E_{S,\eta}\left[\frac{1}{\sigma^4_{S,t}(\theta_0)}\left(\lambda'\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta}\right)^2\right]=0$$

for some $\lambda\in\mathbb{R}^{3d}$. Then almost surely, $\lambda'\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta}=0$. Using the recursive definition of $\sigma^2_{S,t}$

in (4.6), and the stationarity of $\left\{\frac{\partial\sigma^2_{S,t}(\theta_0)}{\partial\theta}\right\}_t$, we get, almost surely

$$0=\lambda'\begin{pmatrix}\mathbb{1}_{S_t=e_1}\\ \vdots\\ \mathbb{1}_{S_t=e_d}\\ \epsilon^2_{S,t-1}\mathbb{1}_{S_t=e_1}\\ \vdots\\ \epsilon^2_{S,t-1}\mathbb{1}_{S_t=e_d}\\ \sigma^2_{S,t-1}(\theta_0)\mathbb{1}_{S_t=e_1}\\ \vdots\\ \sigma^2_{S,t-1}(\theta_0)\mathbb{1}_{S_t=e_d}\end{pmatrix}+\beta(S_t)\lambda'\frac{\partial\sigma^2_{S,t-1}(\theta_0)}{\partial\theta}=\lambda'\begin{pmatrix}\mathbb{1}_{S_t=e_1}\\ \vdots\\ \mathbb{1}_{S_t=e_d}\\ \epsilon^2_{S,t-1}\mathbb{1}_{S_t=e_1}\\ \vdots\\ \epsilon^2_{S,t-1}\mathbb{1}_{S_t=e_d}\\ \sigma^2_{S,t-1}(\theta_0)\mathbb{1}_{S_t=e_1}\\ \vdots\\ \sigma^2_{S,t-1}(\theta_0)\mathbb{1}_{S_t=e_d}\end{pmatrix}$$

Writing $\lambda=(\lambda_{\omega_1},\ldots,\lambda_{\omega_d},\lambda_{\alpha_1},\ldots,\lambda_{\alpha_d},\lambda_{\beta_1},\ldots,\lambda_{\beta_d})$ we thus have

$$\sum_{i=1}^d\lambda_{\omega_i}\mathbb{1}_{S_t=e_i}+\left\{\sum_{i=1}^d\lambda_{\alpha_i}\mathbb{1}_{S_t=e_i}\right\}\epsilon^2_{S,t-1}+\left\{\sum_{i=1}^d\lambda_{\beta_i}\mathbb{1}_{S_t=e_i}\right\}\sigma^2_{S,t-1}(\theta_0)=0,\quad a.s.\quad\text{(A.23)}$$

Suppose that for some $i\in\{1,\ldots,d\}$, $\lambda_{\alpha_i}\neq 0$. Recall that $\mathbb{P}(S_t=e_i)>0$. Then, conditional on $(S_t=e_i)$, (A.23) leads to

$$\epsilon^2_{S,t-1}=\frac{-1}{\lambda_{\alpha_i}}\left\{\lambda_{\omega_i}+\lambda_{\beta_i}\sigma^2_{S,t-1}(\theta_0)\right\},$$

or equivalently, conditional on $(S_t=e_i)$,

$$\eta^2_{t-1}=\frac{-1}{\sigma^2_{S,t-1}\lambda_{\alpha_i}}\left\{\lambda_{\omega_i}+\lambda_{\beta_i}\sigma^2_{S,t-1}(\theta_0)\right\}.$$

The right-hand side being measurable with respect to the σ-field generated by $\{S_t,S_u,\eta_u;u<t-1\}$, and η_{t-1} being independent of this $\sigma-$field conditional on $(S_t=e_i)$, this equality implies that η^2_{t-1} is a.s. constant. By the non degeneracy assumption in **A4**, we can conclude that $\lambda_{\alpha_i}=0$ for all i. For the same reason, it can be shown that : $\lambda_{\beta_i}=0,\forall i\in\{1,\ldots,d\}$. Thus we get $\lambda_{\omega_i}=0,\forall i\in\{1,\ldots,d\}$. Therefore $\lambda'J\lambda=0$ implies $\lambda=0$ which shows that J is nonsingular.

Finally, using again Lemma (A.2),

$$\frac{1}{n}\sum_{t=1}^n\operatorname{Var}\left\{\frac{\partial\ell_t(\theta_0)}{\partial\theta}\mid\mathcal{F}_{t-1}\right\}\to\operatorname{Var}_{S,\eta}\left\{\frac{\partial\ell_{S,t}(\theta_0)}{\partial\theta}\right\},\quad a.s.$$

which completes the proof of *ii)*.

To prove *iii)*, we differentiate (A.11):

$$\begin{aligned}
\frac{\partial^3 \ell_{S,t}(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} &= \left\{1-\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)}\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial^3\sigma^2_{S,t}(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\}\\
&+ \left\{2\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)}-1\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_i}\right\}\left\{\frac{1}{\sigma^2_{S,t}}\frac{\partial^2\sigma^2_{S,t}(\theta)}{\partial\theta_j\partial\theta_k}\right\}\\
&+ \left\{2\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)}-1\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_j}\right\}\left\{\frac{1}{\sigma^2_{S,t}}\frac{\partial^2\sigma^2_{S,t}(\theta)}{\partial\theta_i\partial\theta_k}\right\}\\
&+ \left\{2\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)}-1\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_k}\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial^2\sigma^2_{S,t}(\theta)}{\partial\theta_i\partial\theta_j}\right\}\\
&+ \left\{2-6\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)}\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_i}\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_j}\right\}\left\{\frac{1}{\sigma^2_{S,t}(\theta)}\frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_k}\right\}.
\end{aligned} \tag{A.24}$$

We first prove that $\left\{1-\epsilon^2_{S,t}/\sigma^2_{S,t}(\theta)\right\}$ is integrable. Note that $\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)}$ is not uniformly integrable over Θ because we did not assume the existence of a second-order moment for $\epsilon^2_{S,t}$ and, for instance, $\sigma^2_{S,t}(\theta)$ can be constant for some values of θ. However, we will show that $\left\{1-\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)}\right\}$ is uniformly integrable in a neighborhood of θ_0. Let Θ^* a compact set containing θ_0 and included in the interior of Θ. For all $\delta>0$, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0, with $\mathcal{V}(\theta_0)\subseteq\Theta^*$, such that $\forall\theta\in\mathcal{V}(\theta_0), \forall i\in\{1,\dots,3d\}, \quad (1-\delta)\theta_i<\theta_{0,i}<(1+\delta)\theta_i$. Note that

$$\frac{c_{S,t}(\theta_0)}{c_{S,t}(\theta)}\le\frac{\omega_0(S_t)}{\omega(S_t)}+\frac{\alpha_0(S_t)}{\alpha(S_t)}\le 2(1+\delta).$$

Then, in view of the elementary inequality $x/(y+z)\le(x/z)(z/y)^s$ for any $s\in[0,1]$, $x,y,z>0$,

$$\begin{aligned}
&\sup_{\theta\in\mathcal{V}(\theta_0)}\frac{\sigma^2_{S,t}(\theta_0)}{\sigma^2_{S,t}(\theta)}\\
\le\quad &\sup_{\theta\in\mathcal{V}(\theta_0)}\left\{\frac{c_{S,t}(\theta_0)+\sum_{k=0}^{\infty}\beta_0(S_t)\dots\beta_0(S_{t-i})c_{S,t-i-1}(\theta_0)}{c_{S,t}(\theta)+\sum_{k=0}^{\infty}\beta(S_t)\dots\beta(S_{t-i})c_{S,t-i-1}(\theta)}\right\}\\
\le\quad & 2(1+\delta)+\sum_{k=0}^{\infty}\frac{\beta_0(S_t)\dots\beta_0(S_{t-i})c_{S,t-i-1}(\theta_0)}{\omega(S_t)+\beta(S_t)\dots\beta(S_{t-i})c_{S,t-i-1}(\theta)}\\
\le\quad & 2(1+\delta)+\sum_{i=0}^{\infty}\frac{\beta_0(S_t)\dots\beta_0(S_{t-i})c_{S,t-i-1}(\theta_0)}{\beta(S_t)\dots\beta(S_{t-i})c_{S,t-i-1}(\theta)}\left(\frac{\beta(S_t)\dots\beta(S_{t-i})c_{S,t-i-1}(\theta)}{\omega(S_t)}\right)^s\\
\le\quad & K+K\sum_{i=0}^{\infty}\left(\frac{1+\delta}{(1-\delta)^s}\right)^{i+1}c^s_{S,t-i-1}(\theta_0)\left\{\beta_0(S_t)\dots\beta_0(S_{t-i})\right\}^s.
\end{aligned} \tag{A.25}$$

Choosing $s=r/2$ and δ small enough such that $\frac{(1+\delta)}{(1-\delta)^{r/2}}\rho^{1/2}<1$ and using **A3**, together with *i)* in the proof of Theorem 4.1 and the Cauchy-Schwarz inequality, we obtain

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)} = E \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\sigma^2_{S,t}(\theta_0)}{\sigma^2_{S,t}(\theta)} < \infty.$$

Now choosing $s = r/4$ and δ small enough so that $\frac{(1+\delta)}{(1-\delta)^{r/4}}\rho^{1/4} < 1$ we find from (A.25), the Cauchy-Schwarz inequality and arguments already used

$$\begin{aligned}
&\left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)} \right\|_2 = \kappa_\eta^{1/2} \left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\sigma^2_{S,t}(\theta_0)}{\sigma^2_{S,t}(\theta)} \right\|_2 \\
\leq\ & K + K \left\{ \sum_{i=0}^{\infty} \left(\frac{1+\delta}{(1-\delta)^{r/4}} \right)^{(i+1)} \left\| \beta_0^{r/4}(S_t)\dots\beta_0^{r/4}(S_{t-i}) c^{r/4}_{S,t-i-1}(\theta_0) \right\|_2 \right\} \\
\leq\ & K + K \left\{ \sum_{i=0}^{\infty} \left(\frac{1+\delta}{(1-\delta)^{r/4}} \right)^{(i+1)} \left\{ E\left\{\beta_0(S_t)\dots\beta_0(S_{t-i})\right\}^r E_{S,\eta} c^r_{S,t-i-1} \right\}^{1/4} \right\} \\
\leq\ & K + K \left\{ \sum_{i=0}^{\infty} \left(\frac{1+\delta}{(1-\delta)^{r/4}} \right)^{(i+1)} \rho^{(i+1)/4} \left\{ E_{S,\eta} c^r_{S,t} \right\}^{1/4} \right\} < \infty. \qquad \text{(A.26)}
\end{aligned}$$

Now we show that the second term into brackets in (A.24) is integrable. Note that, since for any $k, \ell \in \{1, \dots, d\}$, $\frac{\partial^2 \sigma^2_{S,t}}{\partial\omega(e_k)\partial\omega(e_\ell)} = \frac{\partial^2 \sigma^2_{S,t}}{\partial\omega(e_k)\partial\alpha(e_\ell)} = 0$, it suffices to consider the cases where at least two indexes among i, j, k belong to $\{2d+1, \dots, 3d\}$ in $\frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial^3 \sigma^2_{S,t}(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}$. We will only consider derivatives with respect to coefficients $\beta(e_k)$, the other cases being treated similarly. Following the lines of proofs of (A.18) and (A.19), we obtain

$$\beta(e_\ell)\beta(e_j)\beta(e_k) \frac{\partial^3 \sigma^2_{S,t}}{\partial\beta(e_\ell)\partial\beta(e_j)\partial\beta(e_k)} \leq \sum_{i=0}^{\infty} (i+1)^3 \beta(S_t)\dots\beta(S_{t-i}) c_{S,t-i-1}(\theta)$$

Now, as in the proof of (A.25):

$$\begin{aligned}
&\sup_{\theta\in\mathcal{V}(\theta_0)} \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial^3 \sigma^2_{S,t}(\theta)}{\partial\beta(e_\ell)\partial\beta(e_j)\partial\beta(e_k)} \\
\leq\ & K \sum_{i=0}^{\infty} (i+1)^3 \frac{1}{(1-\delta)^{r(i+1)}} \beta_0^r(S_t)\dots\beta_0^r(S_{t-i}) \left\{ \sup_{\theta\in\mathcal{V}(\theta_0)} c_{S,t-i-1}(\theta) \right\}^r.
\end{aligned}$$

Since $E\left\{\sup_{\theta\in\mathcal{V}(\theta_0)} c_{S,t-i}(\theta)\right\}^{2r} < \infty$ with r as in **A3**, choosing δ such that $\rho^{1/4} < (1-\delta)^r$, we then have following the proof of (A.26),

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial^3 \sigma^2_{S,t}(\theta)}{\partial\beta(e_\ell)\partial\beta(e_j)\partial\beta(e_k)} \right|^2 < \infty. \qquad \text{(A.27)}$$

More generally, it can similarly be shown that for any integer m, and an appropriate choice of δ that will depend on m,

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial^3 \sigma^2_{S,t}(\theta)}{\partial\beta(e_\ell)\partial\beta(e_j)\partial\beta(e_k)} \right|^m < \infty. \qquad \text{(A.28)}$$

By the Cauchy-Schwarz inequality, (A.26) and (A.27), we get

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \left\{ 1 - \frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)} \right\} \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial^3 \sigma^2_{S,t}(\theta)}{\partial\beta(e_\ell)\partial\beta(e_j)\partial\beta(e_k)} \right| < \infty. \tag{A.29}$$

To deal with the others terms of the sum in (A.24) we show that, similarly to (A.28),

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial^2 \sigma^2_{S,t}(\theta)}{\partial\beta(e_\ell)\partial\beta(e_j)} \right|^m < \infty, \quad E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial \sigma^2_{S,t}(\theta)}{\partial\beta(e_\ell)} \right|^m < \infty$$

for any integer m and appropriate δ. This allows to obtain, using the Hölder inequality, the existence of $\mathcal{V}(\theta_0)$ such that

$$\begin{aligned} & E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \left\{ 2 - 6\frac{\epsilon^2_{S,t}}{\sigma^2_{S,t}(\theta)} \right\} \left\{ \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_i} \right\} \left\{ \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_j} \right\} \left\{ \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_k} \right\} \right| \\ \leq & \left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_i} \right| \right\|_2 \max_j \left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\sigma^2_{S,t}(\theta)} \frac{\partial\sigma^2_{S,t}(\theta)}{\partial\theta_j} \right| \right\|_6^3 < \infty. \end{aligned}$$

The others terms of the sum in (A.24) can be treated this way. Therefore *iii)* is established.

To show *iv)*, we first note that the derivatives of σ_t^2 are deduced from (A.13)-(A.15), with (S_t) replaced by (s_t). We have

$$\tilde{\sigma}_t^2 = c_t + \sum_{i=0}^{t-3} \beta(s_t)\dots\beta(s_{t-i})c_{t-i-1} + \beta(s_t)\dots\beta(s_2)\tilde{c}_1 + \beta(s_t)\dots\beta(s_1)\tilde{\sigma}_0^2 \tag{A.30}$$

where $\tilde{c}_1 = \omega(s_1) + \alpha(s_1)\tilde{\epsilon_0}^2$. Analogously to (A.13)-(A.15), we have for $k \in \{1,\dots,d\}$

$$\begin{aligned} \frac{\partial\tilde{\sigma}_t^2(\theta)}{\partial\omega(e_k)} &= \mathbb{1}_{s_t=e_k} + \sum_{i=0}^{t-3} \beta(s_t)\dots\beta(s_{t-i})\mathbb{1}_{s_{t-i-1}=e_k} \\ &+ \beta(s_t)\dots\beta(s_2)\frac{\partial\tilde{c}_1}{\partial\omega(e_k)} + \beta(s_t)\dots\beta(s_1)\frac{\partial\tilde{\sigma}_0^2}{\partial\omega(e_k)}, \end{aligned} \tag{A.31}$$

$$\begin{aligned} \frac{\partial\tilde{\sigma}_t^2(\theta)}{\partial\alpha(e_k)} &= \epsilon_{t-1}^2\mathbb{1}_{s_t=e_k} + \sum_{i=0}^{t-3} \beta(s_t)\dots\beta(s_{t-i})\epsilon_{t-i-2}^2\mathbb{1}_{s_{t-i-1}=e_k} + \beta(s_t)\dots\beta(s_2)\frac{\partial\tilde{c}_1}{\partial\alpha(e_k)} \\ & +\beta(s_t)\dots\beta(s_1)\frac{\partial\tilde{\sigma}_0^2}{\partial\alpha(e_k)}, \end{aligned} \tag{A.32}$$

$$\begin{aligned} \frac{\partial\tilde{\sigma}_t^2(\theta)}{\partial\beta(e_k)} &= \sum_{i=0}^{t-3} \frac{\beta(s_t)\dots\beta(s_{t-i})}{\beta(e_k)} \left\{ \sum_{j=0}^{i} \mathbb{1}_{s_{t-j}=e_k} \right\} c_{t-i-1} \\ & +\frac{\beta(s_t)\dots\beta(s_2)}{\beta(e_k)} \left\{ \sum_{j=0}^{t-2} \mathbb{1}_{s_{t-j}=e_k} \right\} \tilde{c}_1 + \frac{\beta(s_t)\dots\beta(s_1)}{\beta(e_k)} \left\{ \sum_{j=0}^{t-1} \mathbb{1}_{s_{t-j}=e_k} \right\} \tilde{\sigma}_0^2, \end{aligned} \tag{A.33}$$

if $\beta(e_k) \neq 0$. Notice that $\frac{\partial\tilde{\sigma}_0^2}{\partial\theta} = 0$ when the initial conditions are given by the first alternative in (4.2). The second-order derivatives have similar expressions. We have seen in the proof of *ii)* in Theorem 4.1 that for t sufficiently large, $\beta(s_t)\dots\beta(s_2) < \beta_*^t$, for some $0 < \beta_* < 1$. It follows that, almost surely,

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\sigma_t^2}{\partial\theta} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta}\right\| < K\beta_*^t, \quad \sup_{\theta\in\Theta}\left\|\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta} - \frac{\partial^2\tilde{\sigma}_t^2}{\partial\theta\partial\theta'}\right\| < K\beta_*^t, \quad \forall t \tag{A.34}$$

In view of (A.7) we have,

$$\left|\frac{1}{\sigma_t^2(\theta)} - \frac{1}{\tilde{\sigma}_t^2(\theta)}\right| = \left|\frac{\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)\tilde{\sigma}_t^2(\theta)}\right| \le \frac{K\beta_*^t}{\sigma_t^2(\theta)}, \quad \frac{\sigma_t^2(\theta)}{\tilde{\sigma}_t^2(\theta)} \le 1 + K\beta_*^t. \tag{A.35}$$

Since

$$\frac{\partial\tilde{\ell}_t}{\partial\theta}(\theta) = \left\{1 - \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\theta)}\right\}\left\{\frac{1}{\tilde{\sigma}_t^2(\theta)}\frac{\partial\tilde{\sigma}_t^2(\theta)}{\partial\theta}\right\} \quad \text{and} \quad \frac{\partial\ell_t}{\partial\theta}(\theta) = \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2(\theta)}\right\}\left\{\frac{1}{\sigma_t^2(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\right\}$$

we have, using (A.35) and the first inequality in (A.34),

$$\begin{aligned}\left|\frac{\partial\tilde{\ell}_t}{\partial\theta_i}(\theta_0) - \frac{\partial\ell_t}{\partial\theta_i}(\theta_0)\right| &= \left|\left\{\frac{\epsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\} + \left\{1 - \frac{\epsilon_{S,t}^2}{\tilde{\sigma}_t^2}\right\}\left\{\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}\right.\\ &\quad \left. + \left\{1 - \frac{\epsilon_t^2}{\tilde{\sigma}_t^2}\right\}\left\{\frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial\sigma_t^2}{\partial\theta_i} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta_i}\right\}(\theta_0)\right|\\ &\le K\beta_*^t(1+\eta_t^2)\left|1 + \left\{\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_i}\right\}\right|.\end{aligned}$$

It follows that

$$\left|n^{-1/2}\sum_{t=1}^{n}\left\{\frac{\partial\ell_t(\theta_0)}{\partial\theta_i} - \frac{\partial\tilde{\ell}_t(\theta_0)}{\partial\theta_i}\right\}\right| \le Kn^{-1/2}\sum_{t=1}^{n}\beta_*^t(1+\eta_t^2)\left|1 + \left\{\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_i}\right\}\right| \tag{A.36}$$

To show that the right-hand side a.s. converges to 0, it suffices to prove that the sum a.s. converges as $n \to \infty$. The independence between η_t and $\sigma_{S,t}^2(\theta_0)$ and *i)* entail that,

$$\begin{aligned}&E\left(\sum_{t=1}^{\infty}\beta_*^t(1+\eta_t^2)\left|1 + \left\{\frac{1}{\sigma_{S,t}^2(\theta_0)}\frac{\partial\sigma_{S,t}^2(\theta_0)}{\partial\theta_i}\right\}\right|\right)\\ \le\ & 2\left(1 + E_S\left|\frac{1}{\sigma_{S,t}^2(\theta_0)}\frac{\partial\sigma_{S,t}^2(\theta_0)}{\partial\theta_i}\right|\right)\sum_{t=1}^{\infty}\beta_*^t < \infty\end{aligned}$$

which shows that $\sum_{t=1}^{n}\beta_*^t(1+\eta_t^2)\left|1 + \left\{\frac{1}{\sigma_{S,t}^2(\theta_0)}\frac{\partial\sigma_{S,t}^2(\theta_0)}{\partial\theta_i}\right\}\right|$ converges a.s. Thus, by a straightforward extension of Lemma A.2, $\sum_{t=1}^{n}\beta_*^t(1+\eta_t^2)\left|1 + \left\{\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_i}\right\}\right|$ converges a.s. Thus the right-hand side of (A.36) converges to zero, a.s., showing the first part of *iv)*. The second convergence of *iv)* follows by similar arguments.

To prove *v)* we will use the following Central Limit Theorem (CLT) for martingale difference (see Billingsley, 1995, p.476).

Lemma A.3 *Let $\{(X_t), \mathcal{F}_t\}$ a $\mathbb{R}^d$-martingale difference such that*

a) There exists a $d \times d$ matrix A such that, when $n \to \infty$,

$$\forall \lambda \in \mathbb{R}^d, \quad \frac{1}{n}\sum_{t=1}^{n} Var\{\lambda' X_t | \mathcal{F}_{t-1}\} \to \lambda' A \lambda, \quad a.s.$$

b)

$$\forall \lambda \in \mathbb{R}^d, \forall \epsilon > 0, \quad \sum_{t=1}^{n} E\left\{ \left(\frac{\lambda' X_t}{\sqrt{n}}\right)^2 \mathbb{I}_{\{|\frac{\lambda' X_t}{\sqrt{n}}| > \epsilon\}} \right\} \to 0.$$

Then $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \xrightarrow{d} \mathcal{N}(0, A)$.

We will apply Lemma A.3 with $X_t = \frac{\partial}{\partial\theta}\ell_t(\theta_0)$. We have already shown that (X_t) is a martingale difference and that condition a) is satisfied with $A = J$. In view of arguments used to prove i), it can be shown that :

$$E\left\{\lambda' \frac{\partial \ell_{S,t}}{\partial\theta}(\theta_0)\right\}^4 < \infty, \quad \forall \lambda \in \mathbb{R}^{3d}.$$

Using Lemma A.2, we get:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} E\left\{\lambda' \frac{\partial}{\partial\theta}\ell_t(\theta_0)\right\}^4 = E_{S,\eta}\left\{\lambda' \frac{\partial}{\partial\theta}\ell_{S,t}(\theta_0)\right\}^4 < \infty \quad \forall \lambda \in \mathbb{R}^{3d}, \quad a.s$$

This result entails, using the Cauchy-Schwarz and Markov inequalities

$$\forall \epsilon > 0, \qquad \sum_{t=1}^{n} E\left[\left\{\frac{\lambda' \frac{\partial}{\partial\theta}\ell_t(\theta_0)}{\sqrt{n}}\right\}^2 \mathbb{1}_{\{|\frac{\lambda' \frac{\partial}{\partial\theta}\ell_t(\theta_0)}{\sqrt{n}}| > \epsilon\}}\right]$$
$$\leq \frac{1}{\epsilon^2 n^2} \sum_{t=1}^{n} E\left\{\lambda' \frac{\partial}{\partial\theta}\ell_t(\theta_0)\right\}^4 \to 0.$$

We get now all the conditions needed to use Lemma A.3 and obtain the asymptotic normality results *v)*.

The proof of *vi)* is similar to that of Theorem 2.2-*v)* in FZ. For brevity it is omitted.

The proof of Theorem 4.2 is now completed.

Proof of Theorem 4.3. Let $\pi(k) = P(S_t = k)$ denote the stationary probabilities of the Markov chain (S_t) and let $\Pi_r = (\pi(1)g_r(1), \ldots, \pi(d)g_r(d))'$. Because the variables $\{a_0^r(S_t, \eta_{t-1}), \ldots a_0^r(S_{t-i}, \eta_{t-i-1})$ are independent conditional on $(S_t, \ldots, S_{t-i})$, and by the Markov property the expectation in (4.4) is equal to

$$\begin{aligned} E\left[E\left\{a_0^r(S_t, \eta_{t-1}) \ldots a_0^r(S_{t-i}, \eta_{t-i-1}) \mid S_t, \ldots, S_{t-i}\right\}\right] &= E\{g_r(S_t) \ldots g_r(S_{t-i})\} \\ &= (1, \ldots, 1)\mathbb{P}_r^i \Pi_r, \end{aligned}$$

where the last equality follows from Lemma 1 in Francq and Zakoïan (2005). Consider the matrix norm defined by $\|A\| = \sum_{i,j} |A(i,j)|$, where $A(i,j)$ denotes the generic element of some matrix A. The matrix norm being multiplicative we have

$$E\left\{a_0^r(S_t, \eta_{t-1}) \ldots a_0^r(S_{t-i}, \eta_{t-i-1})\right\} \leq \|(1, \ldots, 1)\| \|\mathbb{P}_r^i\| \|\Pi_r\|$$

It follows from the Jordan decomposition that $\|\mathbb{P}_r^i\| \leq K\{\rho(\mathbb{P}_r)\}^i$, for some constant K. The conclusion follows.

$\square$

References

Anderson, P.L. and A.V. Vecchia (1983) Asymptotic results for for periodic autoregressive moving-average processes. *Journal of Time Series Analysis*, 14, 1–18.

Azrak, R. and G. Mélard (2006) Asymptotic Properties of Quasi-Maximum Likelihood Estimators for ARMA Models with Time-dependent coefficients. *Statistical Inference for Stochastic Processes*, 3, 279-330.

Basawa, I.V. and R.B. Lund (2001) Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, 22, 651–663.

Berkes, I., Horváth, L. and P. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

Bibi, A. and C. Francq (2003) Consistent and asymptotically normal estimators for time-dependent linear models. *Annals of the Institute of Statistical Mathematics*, 55, 41–68.

Billingsley, P. (1995) *Probability and Measure.* John Wiley & Sons, New York.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Bollerslev, T. and E. Ghysels (1996) Periodic autoregressive conditional heteroskedasticity. *Journal of Business and Economic Statistics* 14, 139–151.

Bougerol, P. and N. Picard (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

Carnero, M.A., Koopman, S.J. and M. Ooms (2003) Periodic Heteroskedastic RegARFIMA Models for Daily Electricity Spot Prices. Discussion Paper No. TI 03-071/4, Tinbergen Institute.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1–37.

Dueker, M.J. (1997) Markov Switching in GARCH processes and Mean-Reverting Stock Market Volatility. *Journal of Business and Economic Statistics* 15, 26–34.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987-1007.

Francq, C. and A. Gautier (2004a) Large Sample Properties of Parameter Least Squares Estimates for Time-Varying ARMA Models. *Journal of Time Series Analysis* 25, 765-783.

Francq, C. and A. Gautier (2004b) Estimation of Time-Varying ARMA Models with Markovian Changes in Regime. *Statistics and Probability Letters* 70, 243-251.

Francq, C., Roussignol, M. and J-M. Zakoïan (2001) Conditional Heteroskedasticity driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22, 197-220.

Francq, C. and J-M. Zakoïan (2004) Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605-637.

Francq, C. and J-M. Zakoïan (2005) The L^2 structures of standard and switching-regime GARCH models. *Stochastic Processes and their Applications* 115, 1557-1582.

Francq, C. and J-M. Zakoïan (2007) Quasi-Maximum Likelihood Estimation in GARCH Processes when some coefficients are equal to zero. *Stochastic Processes and their Applications*, 117, 1265-1284.

Giraitis, L., Leipus, R., and D. Surgailis (2007) ARCH(1) models and long memory properties. Forthcoming in *Handbook of Financial Time Series*, ed. by T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch. New York: Springer.

Gray, S.F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.

Haas, M., Mittnik, S., and M. Paolella (2004) A new approach to markov-switching GARCH models. *Journal of Financial Econometrics* 2, 493–530.

Haldrup, N. and M.O. Nielsen (2006) A regime switching long memory model for electricity prices. *Journal of Econometrics* 135, 349–376.

Hall, P. and Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285-317

Hamilton, J. D. and R. Susmel (1994) Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307-333.

Lamoureux, C.G. and W.D. Lastrapes (1990) Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. *Journal of Finance*, 45, 221–229.

Mikosch, T. and C. Starica (2004) Non-stationarities in Financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* 86, 378–390.

Nelson, D. B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.

Pfanzagl, J. On the measurability and consistency of minimum contrast estimators. *Metrika* 14, 249–272.

Polzehl, J. and V. Spokoiny (2006) Varying coefficient GARCH versus local constant volatility modelling. Comparison of the predictive power. Unpublished document. Weierstrass Institut, Berlin.

Straumann, D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics, Springer, Berlin.

Tiao, G.C. and M.R. Grupe (1980) Hidden periodic autoregressive-moving average models in time series data. *Biometrika* 67, 365–373.

Teräsvirta, T. (2007) An introduction to univariate GARCH models. Forthcoming in *Handbook of Financial Time Series*, ed. by T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch. New York: Springer.