Journal of Multivariate Analysis 100 (2009) 2044–2054

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On asymptotic theory for multivariate GARCH models

Christian M. Hafner [a,*], Arie Preminger [b]

[a] Institut de Statistique and CORE, Université Catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

[b] Department of Economics, University of Haifa, Mount Carmel, Haifa 31905, Israel

ARTICLE INFO

Article history:
Received 8 October 2007
Available online 31 March 2009

AMS subject classifications:
62M10
62F12
91B84

JEL classification:
C14
C22

Keywords:
Multivariate GARCH models
VEC
Geometric ergodicity
Consistency
Asymptotic normality

ABSTRACT

The paper investigates the asymptotic theory for a multivariate GARCH model in its general vector specification proposed by Bollerslev, Engle and Wooldridge (1988) [4], known as the VEC model. This model includes as important special cases the so-called BEKK model and many versions of factor GARCH models, which are often used in practice. We provide sufficient conditions for strict stationarity and geometric ergodicity. The strong consistency of the quasi-maximum likelihood estimator (QMLE) is proved under mild regularity conditions which allow the process to be integrated. In order to obtain asymptotic normality, the existence of sixth-order moments of the process is assumed.

1. Introduction

The theory of univariate GARCH models and its various extensions can be considered as well established, due to the enormous literature that has evolved since the introduction of the model by Engle [1] and Bollerslev [2]. In recent years, an increasing interest has focused on multivariate GARCH models, for several reasons. One is the importance of multivariate models in economic and finance theory such as portfolio selection or asset pricing. Second, computing power has risen to a point where the practical implementation of these models becomes feasible even for many assets and long time series. Third, many alternative specifications have been proposed to deal with the problem of balancing large numbers of parameters and flexibility. Fourth, statistical theory of multivariate GARCH models is not a trivial extension of the theory of univariate models and is not established under weak conditions. Indeed, asymptotic theory for multivariate GARCH models is as yet available only for particular specifications and under relatively strong conditions. The complex model structure makes the study of the likelihood function difficult, and the method of the proofs in the univariate case cannot be used directly. For a recent review of multivariate GARCH models we refer to [3].

This paper derives results for the general multivariate GARCH model, known as the VEC model, which was proposed by Bollerslev, Engle and Wooldridge [4]. The model contains as important special cases the BEKK model of Engle and Kroner [5], the factor GARCH models of Diebold and Nerlove [6], and Engle, Ng and Rothschild [7], the orthogonal GARCH model of Alexander [8], the generalized orthogonal GARCH model of van der Weide [9] and the full factor GARCH model of Vrontos et al. [10].

* Corresponding author.
E-mail address: christian.hafner@uclouvain.be (C.M. Hafner).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.03.011

Dealing with a general multivariate conditionally heteroskedastic model, the early paper by Bollerslev and Wooldridge [11] provides high level assumptions under which the QMLE is consistent and asymptotically normal, even if the true data generating process is not conditionally normal. Strong consistency of the QMLE has been shown by Jeantheau [12] for the model of Bollerslev [13] having constant conditional correlations (CCC), for which Ling and McAleer [14] show asymptotic normality. For the BEKK model, Comte and Lieberman [15] establish consistency and asymptotic normality under the existence of second-order and eighth-order moments of the data, respectively, which rules out the possibility of integrated GARCH. Recently, Hafner and Preminger [16] proved asymptotic properties of the QMLE for an integrated factor GARCH model. Their results are obtained under the finiteness of the fourth-order moment of the innovations.

The existence of a stationary and ergodic solution for the VEC model is important in establishing our asymptotic results. Bougerol and Picard [17] give necessary and sufficient conditions for strict stationarity and ergodicity of a univariate GARCH(p, q) model in terms of the top Lyapounov exponent. Their results extend the results of Nelson [18] for the GARCH(1, 1) model. As for the multivariate GARCH models, Dennis, Hansen and Rahbek [19] give sufficient conditions for geometric ergodicity of the so-called BEKK representation of the multivariate ARCH(q) model. Ling and McAleer [14] establish conditions under which the CCC model of Bollerslev [13] has a strictly stationary solution. Recently, Kristensen [20] provided sufficient conditions for geometric ergodicity for a variety of multivariate GARCH models which includes the VEC model (see also [21]. However, these results were established under the assumption of covariance stationarity and do not include the important case of integrated processes.

The main part of this paper establishes asymptotic results for the VEC model. Strong consistency is obtained under mild regularity conditions which allow for an integrated process and require the existence of moments of order two of the innovations. Asymptotic normality of the QMLE is shown under the assumption that the moment of order six of the process is finite. No conditions on the shape of the innovation distribution are required other than the existence of moment conditions. Further, we provide sufficient conditions for geometric ergodicity of the VEC model. For simplicity, we assume that the innovations are independently and identically distributed. However, it would be possible to replace this assumption by one that assumes a strictly stationary and ergodic martingale difference process.

Throughout the paper, we denote by $\rho(A)$ the spectral radius of any square matrix A, i.e., $\rho(A) = \max\{|\lambda_i| : \lambda_i$ is an eigenvalue of $A\}$. We use $\|\cdot\|$ as a matrix operator norm induced by some vector norm. Since we use the Euclidean vector norm, $\|\cdot\|$ denotes the spectral norm, i.e. $\|A\| = \rho^{1/2}(A'A)$. $O(1)$ (or $o(1)$ denotes a series of nonstochastic variables that are bounded (or converge to zero); $O_P(1)$ (or $o_P(1)$) denotes a series of random variables that are bounded (or converge to zero) in probability. The symbol $\to_{a.s.}(\to_D)$ denotes convergence almost surely (or in distribution).

2. Geometric ergodicity

A general multivariate GARCH specification has been proposed by Bollerslev, Engle and Wooldridge [4], usually called the VEC model. It can be written as

$$y_t = H_t^{1/2}\xi_t \tag{1}$$

where ξ_t is an i.i.d. centered random variables. The conditional covariance matrix is given by $\mathrm{E}(y_t y_t'|\mathcal{F}_{t-1}) = H_t$, where $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \ldots)$. In practice, the most common model order is VEC(1, 1), which is given by

$$h_t = \omega + A\eta_{t-1} + Bh_{t-1} \tag{2}$$

where $\eta_t = \mathrm{vec\,h}(y_t y_t')$ and $h_t = \mathrm{vec\,h}(H_t) \in \mathbb{R}^d$, $d = N(N+1)/2$. Let $\mathcal{M}_+(N)$ be the space of real symmetric positive definite $N \times N$ matrices. We assume that A and B are such that $H_t \in \mathcal{M}_+(N)$ for any given choice of $y_{t-1} \in \mathbb{R}^N$ and $H_{t-1} \in \mathcal{M}_+(N)$.

Let $x_t = (y_t', h_t')'$ denote the joint process which can be realized as a homogenous Markov chain with a state space $\mathcal{X}$, a subset of the Euclidean space. In what follows, we provide a sufficient condition for the existence of a unique stationary solution. Further, we establish the V-geometric ergodicity of the process. To introduce this notion of ergodicity, suppose that there exists a real (Borel) measurable function $V : \mathcal{X} \to [1, \infty)$ and a probability measure π on the Borel sets of $\mathcal{X}$ and constants $M_x < \infty$ and $\rho \in (0, 1)$ such that

$$\sup_{v:|v|\le V}\left|\mathrm{E}(v(x_t|x_0 = x)) - \int_{\mathcal{X}} v(u)\pi(\mathrm{d}u)\right| = M_x\rho^t \tag{3}$$

for all $x \in \mathcal{X}$, $t \ge 1$. The definition also assumes that $\mathrm{E}_\pi(V(x)) < \infty$, where the expectation is taken with respect to π. If $V(\cdot) \equiv 1$, the chain refers to the usual notion of geometric ergodicity. Thus, geometric ergodicity entails that the t-step transition probability measure $P^t(x, \cdot)$ defined by $P^t(x, A) = P(x_t \in A|x_0 = x)$ converges at a geometric rate for all $x \in \mathcal{X}$ to the probability measure $\pi(\cdot)$ with respect to the total variation norm. This probability measure is often referred to as the stationary probability measure of the process. If the process $\{x_t\}$ is initialized from the stationary distribution, V-geometric ergodicity implies that $\mathrm{E}(v(x_t)) < \infty$ for all $|v(\cdot)| \le V$. Furthermore, the conditional expectation $\mathrm{E}(v(x_t)|x_0 = x)$ converges at a geometric rate to the corresponding expectation taken with respect to the stationary distribution. Before establishing this type of asymptotic stability, we need some further notation. Consider the sequence of $N \times N$ matrices $\{\Delta_t\}$ given by

$$\Delta_t = \Delta(h_t, \xi_t) = \frac{\partial h_{t+1}}{\partial h_t'} = AD_N^+(\tilde{\Delta}_t \otimes I_N)D_N + B \tag{4}$$

where $\tilde{\Delta}_t = H_t^{1/2}\xi_t\xi_t' H_t^{-1/2}$. For some integer $m \geq 1$ and $t > m$, let

$$\gamma_m(\Delta) = \frac{1}{m}\mathrm{E}\log\left(\sup_{\hbar^m}\left\|\prod_{k=1}^{m}\Delta(h_{m-k+1}, \xi_{m-k+1})\right\|\right) \tag{5}$$

where $\hbar^m = \{(h_1', \ldots, h_m')' \in \mathbb{R}^{md} : \mathrm{vec}\,\mathrm{h}^{-1}(h_t) \in \mathcal{M}_+(N), \|h_t\| = 1, t = 1, \ldots, m\}$.

A sufficient condition for $\gamma_m(\Delta) < 0$ is $\mathrm{E}\log\left(\sup_{\hbar^1}\|\Delta(h_1, \xi_1)\|\right) < 0$. In the case of the univariate GARCH(1, 1) model, our measure coincides with the stability condition of Nelson [18]. We can show that, for any induced matrix norm, the term inside the expectation can be bounded a.s. by an i.i.d. term[1] (see Lemma 1 in the Appendix); hence $\gamma_m(\Delta)$ is well defined if $\mathrm{E}\|\xi_t\|^q < \infty$ for some $q > 0$. The computation of $\gamma_m(\Delta)$ involves Monte Carlo simulations and can be quite difficult. Next, we give sufficient conditions for the process to be geometrically ergodic.

Assumption 2.1. The centered random vectors $\{\xi_t\}$ have a positive lower semi-continuous density w.r.t. the Lebesgue measure on the set $\{\xi_t \in \mathbb{R}^N : \|\xi_t\| \leq \eta\}$, for some $\eta > 0$. The initial condition x_0 is independent of $\{\xi_t\}$.

Assumption 2.2. $\det(A) \neq 0$, $\rho(B) < 1$ and $\mathrm{E}\|\xi_t\|^{4r} < \infty$ for some $r > 0$.

Assumption 2.3. $\gamma_m(\Delta) < 0$ for some integer $m \geq 1$.

The first assumption is satisfied for a wide range of distributions for the innovations, such as the multivariate Gaussian and Student distributions. The second assumption is necessary to prove the irreducibility and aperiodicity of the process, by ensuring that the model is forward accessible and attains a globally attracting state; see e.g. [22,20] for details. The last assumption allows for integrated processes. To see this, consider for example the simple case $N = 2$, $B = 0$ and $A = I_2$. For $m = 4$, we find by simulations[2] that $\gamma_4(\Delta) = -0.1411 < 0$. Our first result is based on Markov chain theory; see e.g. [23].

Theorem 1. *Assume that Assumptions 2.1–2.3 hold. Then the Markov chain $\{x_t\}$ is V-geometrically ergodic. Further,* $\mathrm{E}(\|x_t\|^s) < \infty$ *for some $s \in (0, r]$, where the expectations are taken under the stationary distribution.*

Geometric ergodicity implies that, if the VEC model is initiated from its stationary probability measure, the process is stationary and β-mixing with exponential decay. Furthermore, an important consequence of Theorem 1 is that limit theorems can be applied to any measurable function of the data, for any given starting values, given the existence of suitable moments; see [23, Ch. 17] for details. As for sufficient conditions for the existence of second-order and fourth-order moments we refer the reader to [5,24], respectively (see also [20]). Under similar conditions, we can also cover the case where the innovations are a strictly stationary martingale difference sequence, showing the existence of a strictly stationary solution and some fractional moment for the process. The proof is available from the authors on request, and for simplicity, in what follows, the i.i.d. assumption is maintained.

3. Consistency and asymptotic normality

The estimation of the MV–GARCH model is usually done by maximum likelihood assuming that the innovation distribution is normal. Let us decompose the parameter vector as $\theta = (\omega', \mathrm{vec}(A)', \mathrm{vec}(B)')'$ and assume that $\theta \in \Theta \subset \mathbb{R}^p$, $p = m + 2m^2$. Furthermore, denote the true parameter vector by θ_0. The log likelihood, up to an additive constant, for a sample of n observations, takes the form

$$L_n(\theta) = -\frac{1}{2n}\sum_{t=1}^{n}\log\det(H_t(\theta)) + y_t' H_t^{-1} y_t = -\frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \tag{6}$$

where the starting value H_1 is a fixed matrix. Define the QMLE as $\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta)$. Let $\tilde{H}_t$ denote the covariance process where the starting values are drawn from their stationary distribution and let $\tilde{h}_t$, $\tilde{L}_n$ and $\tilde{l}_t$ be defined analogously. These terms will be used in the proofs. Note that in practice the use of these values is not possible. However, we will show that the choice of the initial values does not matter for the asymptotic properties of the QML estimator. We assume that the true parameters are such that H_t is positive definite. We allow that the process is a pure ARCH or even i.i.d. When $A = 0$ the conditional variance process is completely deterministic, and to identify the model we also set $B = 0$ in that case. To obtain strong consistency the following assumptions are made.

Assumption 3.1. The parameter space Θ is compact and $\rho(B) < 1$.

[1] Without loss of generality, we take the supremum over the unit ball in $\mathbb{R}^d$, since $\tilde{\Delta}_t$ is invariant to the rescaling of $\mathrm{vec}\,\mathrm{h}^{-1}(h_t) \in \mathcal{M}_+(N)$.

[2] The simulations used the MATLAB optimization toolbox to maximize $\gamma_4(\Delta)$.

Assumption 3.2. The observed sequence $\{y_t\}$ is strictly stationary and ergodic and $\mathrm{E}(\|y_t\|^s) < \infty$ for some $s > 0$.

Assumption 3.3. $\mathrm{E}\|\xi_t\|^2 < \infty$, $var(\xi_t) = I_N$.

Assumption 3.4. The VEC model is identifiable: If for any $\theta, \theta_0 \in \Theta$, $H_t(\theta) = H_t(\theta_0)$ a.s., then $\theta = \theta_0$.

We now have the following theorem.

Theorem 2. *Under Assumptions 3.1–3.4,* $\hat{\theta}_n \to_{a.s.} \theta_0$.

Theorem 2 shows the strong consistency of the QMLE. Assumption 3.2 is satisfied if for the true VEC model Assumptions 2.1–2.3 hold and the process is initiated from its stationary distribution or in the infinite past. These assumptions allow for integrated GARCH processes, while, for the BEKK model, [15] require the existence of a second-order moment of the data to obtain consistency. We assume that one observes the stationary solution. One could weaken this assumption and only assume that there exists a stationary solution; see e.g. [25]. The requirement in Assumption 3.3 that the variance of ξ_t equals the identity matrix is made to ensure identifiability and is not restrictive. Assumption 3.4 is a high level assumption. Jeantheau [12] gives primitive conditions for identification in an extended version of the constant conditional correlation model of Bollerslev [13]. For the BEKK model we can apply the identification results in [5], while, for the factor GARCH models, identification conditions are given by Fiorentini and Sentana [27] and Doz and Renault [28]. As the squares and cross-products of a multivariate GARCH process have a VARMA representation, one could apply the identifiability conditions such as the final form or echelon form given by Lütkepohl ([26], Ch. 7).

We define the following matrices,

$$V = \mathrm{E}\left(\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta}\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta'}\right), \qquad J = -\mathrm{E}\left(\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta'}\right).$$

To establish asymptotic normality the following additional assumptions are made.

Assumption 3.5. The parameter θ_0 is an interior point of Θ.

Assumption 3.6. $\mathrm{E}\|y_t\|^6 < \infty$.

We can now state the asymptotic distribution of QML estimators.

Theorem 3. *Under Assumptions 3.1–3.6,* $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathrm{D}} N(0, J^{-1}VJ^{-1})$.

As in [14], to prove the asymptotic normality of the QMLE, we only require the second derivatives of the likelihood. Hence, the proof is simplified and we can reduce the requirement for higher-order moments. The asymptotic normality result for the BEKK model which is nested in the VEC specification was established by Comte and Lieberman [15]. They required the existence of the eighth-order moments.

Assumption 3.5 is needed to establish the asymptotic normality. Otherwise, when the parameters are on the boundary, other methods should be used. For example, consider the BEKK model $A = D_N^+(A^* \otimes A^*)D_N$ and $B = D_N^+(B^* \otimes B^*)D_N$, where A^* and B^* are $(N \times N)$ parameter matrices. To obtain identifiability, the upper left element of A_0^* is usually restricted to be non-negative; see [5]. Thus, if $\mathrm{vec}(A_0^*) = 0$, the distribution of $\sqrt{n}\mathrm{vec}(\hat{A}_n^*)$ cannot be normal. Andrews [29] and Francq and Zakoian [30] study in detail the distribution of the QMLE in that case. However, this issue is beyond the scope of this paper.

An alternative specification of the model is $h_t(\theta) = \omega(\theta) + A(\theta)\eta_{t-1} + B(\theta)h_{t-1}$ for some smooth functions $\omega(\theta)$, $A(\theta)$ and $B(\theta)$ that depend on some underlying parameter θ such that $H_t(\theta)$ is positive definite for all $\theta \in \Theta$.[3] Note that our results would still be valid for this formulation of the model as well. However, the notation would become slightly more involved and we decided to keep the current formulation of the model.

In the case where the innovations are indeed multi-normal, maximum likelihood estimation provides the most efficient estimator and $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathrm{D}} N(0, J^{-1})$; see [15, Lemma 1]. However, in the presence of non-Gaussian innovations the QMLE is consistent but not efficient. Given the results of Theorems 2 and 3 and mild regularity conditions on the innovation terms which appear in [31], one can construct semi-parametric estimators which are asymptotically more efficient than the QMLE.

4. Conclusions

For the VEC multivariate GARCH model, we have shown consistency of the QMLE under weak conditions that allow for the empirically relevant case of integrated GARCH. To obtain asymptotic normality, we require finite sixth moments. An issue that seems very difficult to address is the asymptotic distribution under even weaker assumptions, possibly including integrated GARCH. Furthermore, asymptotic theory for the dynamic conditional correlation (DCC) model of Engle [38] under weak conditions is another topic for future research.

[3] We are grateful to Dennis Kristensen for pointing out this possibility.

Acknowledgments

The authors thank Luc Bauwens, Jürgen Franke, Dennis Kristensen, Oliver Linton and Jiazhu Pan for helpful comments which helped to improve the paper. Hafner acknowledges financial support by the *Fonds Spéciaux de Recherche* (FSR 05) of the Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Appendix

Proof of Theorem 1. Since the $\operatorname{vec}\mathrm{h}(\cdot)$ operator is invertible on $\operatorname{vec}\mathrm{h}(\mathcal{M}_+(N))$ such that there is a one to one relation between h_t and H_t, we have that $y_t = (\operatorname{vec}\mathrm{h}^{-1}h_t)^{1/2}\xi_t$ and (2) can be written in its state space representation

$$h_t = \omega + G(h_{t-1}, \xi_{t-1}) = F(h_{t-1}, \varepsilon_t) \tag{7}$$

where $\varepsilon_t = \xi_{t-1}$. This process forms a homogenous Markov chain with state space$(\mathcal{D}, \mathcal{B})$, where $\mathcal{D} = \{h \in \mathbb{R}^{N(N+1)/2} \mid \operatorname{vec}\mathrm{h}^{-1}(h) \in \mathcal{M}_+(N)\}$ and $\mathcal{B}$ is a Borel σ-algebra on $\mathcal{D}$. Following [23, Ch. 7], we can define inductively a sequence of functions F_t, $t \geq 1$, by $F_{t+1}(x, \varepsilon_1, \dots, \varepsilon_{t+1}) = F(F_t(x, \varepsilon_1, \dots, \varepsilon_t), \varepsilon_{t+1})$, so that for any initial value h_0 we can solve (7) recursively for $t \geq 1$ to obtain $h_t = F_t(h_0, \varepsilon_1, \dots, \varepsilon_t)$, the associated deterministic control model for the nonlinear state space above; $\{\varepsilon_t\}$ is called the control sequence.

From Assumption 2.2 and the continuity of the spectral radius it follows that there exists an ε^* sufficiently small such that $\sup_{\hbar^1} \rho(\Delta(\cdot, \varepsilon^*)) < 1$ (recall that $\rho(B) < 1$). Since the function $F(h_{t-1}, \varepsilon_t)$ is Lipschitz, by an application of the contraction mapping theorem, $h_t \to h^*$ for all $h_0 \in \mathcal{D}$ and control sequence $\{\varepsilon_t = \varepsilon^*\}$. This result implies that the control model associated with h_t attains a globally attracting state, h^* (for a definition see [23, pp. 163–164]). Assumptions 2.1 and 2.2 imply that conditions C.1 and C.2 of Theorem 16 of Kristensen [20] hold, which implies that the model is forward accessible. Hence, it is a T-chain by Proposition 7.1.5 of Meyn and Tweedie [23].

Furthermore, the existence of a globally attracting state and forward accessibility imply by Proposition 7.2.5 and Theorem 7.2.6 of Meyn and Tweedie [23] that the chain is irreducible. This also implies that all compact sets in $\mathcal{D}$ are small and can be used as test sets. Aperiodicity follows from the fact that any cycle of the associate control model must contain the globally attracting state h^*; see the proof of Proposition 7.4.1 in [23]. From a mean-value approximation around $h^* \in C$, some compact set in $\mathcal{D}$, and after solving (7) recursively, we get

$$\begin{aligned} h_t &= \varpi(\bar{h}^*_{t-1}, \xi_{t-1}) + \Delta(\bar{h}^*_{t-1}, \xi_{t-1})h_{t-1} \\ &= \varpi(\bar{h}^*_{t-1}, \xi_{t-1}) + \sum_{j=1}^{m-1}\prod_{k=1}^{j} \Delta(\bar{h}^*_{t-k}, \xi_{t-k})\varpi(\bar{h}^*_{t-j-1}, \xi_{t-j-1}) + \prod_{k=1}^{m} \Delta(\bar{h}^*_{t-k}, \xi_{t-k})h_{t-m} \end{aligned} \tag{8}$$

where $\varpi(h, \xi) = \omega + G(h^*, \xi) - \Delta(h, \xi)h^*$ and $\bar{h}^*_{t-k}$ is on the chord between h^* and h_{t-k}, $k = 1, \dots, m$. Note that $\bar{h}^*_{t-k} \in \mathcal{D}$ since, for all $\alpha \in [0, 1]$, $\operatorname{vec}\mathrm{h}^{-1}(\alpha h^* + (1-\alpha)h_{t-k}) \in \mathcal{D}$. By Assumption 2.3 and the same arguments used in [32, Proof of Theorem 2.1], we have that

$$\lambda = \mathrm{E}\left(\sup_{\hbar^m}\left\|\prod_{k=1}^{m}\Delta(h_{m-k+1}, \xi_{m-k+1})\right\|\right)^s < 1, \quad s \in (0, r). \tag{9}$$

Next consider the drift function $V(h) = 1 + \|h\|^s$. Using (8) and (9), the usual properties of the matrix norm and the c_r inequality, we observe that for all $h \in \mathcal{D}$

$$\begin{aligned} \mathrm{E}[V(h_t)|h_{t-m} = h_0] &\leq (1-\lambda) + \mathrm{E}\sup_{\hbar^1}\|\varpi(\cdot, \xi_{t-1})\|^s \\ &\quad + \sum_{j=1}^{m-1}\mathrm{E}\sup_{\hbar^j}\left\|\prod_{k=1}^{j}\Delta(\cdot, \xi_{t-k})\right\|^s \mathrm{E}\sup_{\hbar^1}\|\varpi(\cdot, \xi_{t-j-1})\|^s + \lambda\left(1 + \|h_0\|^s\right) \\ &= \lambda V(h_0) + b. \end{aligned}$$

Assumption 2.2 and Lemma 1 imply $\mathrm{E}(\sup_{\hbar^1}\|\Delta(\cdot, \xi_t)\|^r) < \infty$ and $\mathrm{E}(\sup_{\hbar^1}\|\varpi(\cdot, \xi_t)\|^r) < \infty$; hence $b < \infty$. Since the drift function is continuous and bounded on the compact set C, we have that $\mathrm{E}(V(h_t)|h_{t-m} = h_0) \leq \lambda V(h_0) + b \cdot 1_C$; hence the drift criterion is satisfied. Therefore, by combining the results of Meyn and Tweedie [23, Theorem 16.0.1] and Tjostheim [33], we have that $\{h_t\}$ is V-geometric ergodic and that $\mathrm{E}(\|h_t\|^s) < \infty$. The extension of this result to the Markov chain $x_t = (y_t', h_t')'$ follows from Proposition 1 of Meitz and Saikkonen [34]. □

Proof of Theorem 2. By Assumption 3.4 and similar arguments as in [11], we can show that θ_0 is *uniquely identifiable*, thus $\mathrm{E}[l_t(\theta_0)] < \mathrm{E}[l_t(\theta)]$ for all $\theta \neq \theta_0$. It remains to show that this result implies strong consistency. Suppose that $\hat{\theta}_n$ does not converge to θ_0 a.s., so for an arbitrary $\varepsilon > 0$ the event $\Omega = \{\limsup_{n\to\infty}\|\hat{\theta}_n - \theta_0\| \geq \varepsilon\}$ has a positive probability. Hence, there exists a compact set $\Lambda \subseteq \Theta \setminus \{\|\hat{\theta}_n - \theta_0\| \leq \varepsilon\}$ and a non-null set $\Omega' \subseteq \Omega$ such that $\hat{\theta}_n(\omega) \to \theta(\omega) \in \Lambda$ for all $\omega \in \Omega'$.

Next, we note that the sequence $\{\inf_{\theta\in\Lambda} \tilde{l}_t(\theta)\}_t$ results from a measurable transformation of $\{y_t, y_{t-1}, \ldots\}$, so it is stationary and ergodic. Lemma 2 implies that $\mathrm{E}\inf_{\theta\in\Lambda} \tilde{l}_t(\theta) \in \mathbb{R}\cup\{+\infty\}$, so by the ergodic theorem (see [35])

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \inf_{\theta\in\Lambda} \tilde{l}_t(\theta) = \mathrm{E}\inf_{\theta\in\Lambda} \tilde{l}_t(\theta). \tag{10}$$

Therefore, we have with positive probability

$$\begin{aligned}\mathrm{E}\tilde{l}_t(\theta_0) &\geq \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} l_t(\theta_0) \geq \liminf_{n\to\infty} \inf_{\theta\in\Lambda} \frac{1}{n}\sum_{t=1}^{n} l_t(\theta)\\ &= \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} l_t(\hat{\theta}_n(\omega))\\ &\geq \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \inf_{\theta\in\Lambda} \tilde{l}_t(\theta) + o(1) = \mathrm{E}\inf_{\theta\in\Lambda} \tilde{l}_t(\theta). \end{aligned} \tag{11}$$

The first inequality results from the ergodic theorem and Lemma 2. The first equality is obtained by the definition of the QML estimator. The last inequality results from Lemma 2(ii). Eq. (11) contradicts that θ_0 is uniquely identifiable, and, since $\varepsilon > 0$ can be arbitrarily small, the desired result follows. □

Proof of Theorem 3. By the strong consistency and Assumption 3.5, we have that, for sufficiently large n, $\hat{\theta}_n$ is contained a.s. in an arbitrarily small neighborhood of θ_0. Hence, the mean-value expansion of the score vector around θ_0 gives

$$0 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\partial l_t(\hat{\theta}_n)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial\theta} + \left[\left(\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_t(\bar{\theta}_n)}{\partial\theta\partial\theta'} + J\right) - J\right]\sqrt{n}(\hat{\theta}_n - \theta_0) \tag{12}$$

where $\bar{\theta}_n$ is between $\hat{\theta}_n$ and θ_0.

In Lemma 4, we show that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial\theta}$ obeys a CLT and that the first term inside the square brackets in (12) converges a.s. to zero. In addition, we have that the ij-th element of the expectation of the Hessian is given by $\mathrm{E}(\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta_i\partial\theta_j}|\mathcal{F}_{t-1}) = \mathrm{Tr}(\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1})$. Comte and Lieberman [15] show that the expectation of the Hessian is positive definite a.s., as otherwise the model is not identifiable. Hence, by solving (12) and applying the Slutsky theorem, the desired result follows. □

Lemma 1. *Under Assumptions 2.1 and 2.2,* $\mathrm{E}(\sup_{\hbar^1} \|\Delta(\cdot, \xi_t)\|^r) < \infty$ *for some* $r > 0$.

Proof. It is sufficient to show that $\sup_{\hbar^1} \|\Delta(\cdot, \xi_t)\|$ is bounded a.s. by an i.i.d. term, where we use the spectral norm. With the definition of Δ_t in (4), we have

$$\|\Delta_t\| = \lambda_{\max}(\Delta_t\Delta_t') \leq C_1\mathrm{Tr}(\tilde{\Delta}_t\tilde{\Delta}_t') + C_2$$

where $C_1 = N\|A\|\cdot\|D_N^+\|\cdot\|D_N\|$ and $C_2 = \|B\|$. The inequality results from [36, p. 20, 11(b)]). Now, using inequality 6(b) in [36, p. 44], we find

$$\mathrm{Tr}(\tilde{\Delta}_t\tilde{\Delta}_t') \leq \frac{1}{4}[\mathrm{Tr}(\tilde{\Delta}_t) + \mathrm{Tr}(\tilde{\Delta}_t')]^2 = \mathrm{Tr}^2(\tilde{\Delta}_t) = (\xi_t'\xi_t)^2.$$

Thus, by Assumption 2.2, $\mathrm{E}\sup_{\hbar^1} \|\Delta(\cdot, \xi_t)\|^r \leq C_1\mathrm{E}\|\xi_t\|^{4r} + C_2 < \infty$, and the desired result follows. □

Lemma 2. *Under Assumptions 3.1–3.4*

(i) $\mathrm{E}|\tilde{l}_t(\theta)|$ *is well defined and* $\mathrm{E}|\tilde{l}_t(\theta_0)| < \infty$
(ii) $\limsup_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\tilde{l}_t(\theta) - \frac{1}{n}\sum_{t=1}^{n} l_t(\theta)\right| = 0.$

Proof. (i) Let $\{\lambda_{it}(\theta)\}_{i=1}^N$ be the eigenvalues of $\tilde{H}_t(\theta)$. Note that since this matrix is positive definite by definition, its eigenvalues are positive. The compactness of the parameter space and the Wielandt–Hoffman theorem imply that eigenvalues are continuous functions of the matrix elements. Hence, the compactness of the parameter set implies that there exists a real positive number such that $0 < \lambda \leq \inf_{\theta\in\Theta}\lambda_{it}(\theta)$, for all i, t; hence

$$\det(\tilde{H}_t(\theta)) = \prod_{i=1}^{N}\lambda_{it}(\theta) \geq \lambda^N > 0$$

which implies that $\mathrm{E}\tilde{l}_t(\theta)$ is well defined and that $-\infty < \mathrm{E}\inf_{\theta\in\Theta}\tilde{l}_t(\theta)$. By Assumptions 3.2 and 3.3, Jensen's inequality and the c_r inequality, for some $s > 0$,

$$\begin{aligned}\mathrm{E}\log\det(\tilde{H}_t(\theta_0)) &= \mathrm{E}\frac{1}{s}\log[\det(\tilde{H}_t(\theta_0))]^s \le \frac{1}{s}\log\mathrm{E}|\det(\tilde{H}_t(\theta_0))^s| \\ &\le \frac{1}{s}\log\mathrm{E}\prod_{i=1}^{N}\lambda_i^{s/N}(\theta_0) \le C_1\log\mathrm{E}\|\tilde{H}_t(\theta_0)\|^s \\ &\le C_2\mathrm{E}\log\|\tilde{h}_t(\theta_0)\|^s < \infty. \end{aligned} \tag{13}$$

Hence, $\mathrm{E}(|\tilde{l}_t(\theta_0)|) < \infty$.

(ii) By solving (2) recursively, we get

$$h_t = h_0B^t + \sum_{j=0}^{t-1}B^j(\omega + A\eta_{t-j-1}); \tag{14}$$

hence for $t \ge 1$

$$\|H_t - \tilde{H}_t\| \le \|H_t - \tilde{H}_t\|_2 \le C_1\|h_t - \tilde{h}_t\| \le C_1\|B^t\|\cdot\|h_0 - \tilde{h}_0\| \tag{15}$$

where $\|\cdot\|_2$ is the Frobenius norm. Assumption 3.1 implies that $\bar{\rho} = \sup_{\theta\in\Theta}\rho(B) < 1$. Hence

$$\sup_{\theta\in\Theta}\left\|B^t\right\| = O(\bar{\rho}^t). \tag{16}$$

These results, Assumptions 3.2 and 3.3 and the c_r inequality imply that

$$\mathrm{E}\sup_{\theta\in\Theta}\|H_t - \tilde{H}_t\|^{s/2} = O(\bar{\rho}^t) \tag{17}$$

for some $s > 0$.

Next we show that $\sup_{\theta\in\Theta}\|H_t^{-1}(\theta)\| < \infty$ and $\sup_{\theta\in\Theta}\|\tilde{H}_t^{-1}(\theta)\| < \infty$. By definition, $H_t = D_t^{1/2}R_tD_t^{1/2}$, where $D_t = \mathrm{diag}(h_{11,t}(\theta),\ldots,h_{NN,t}(\theta))$ and R_t is the positive definite correlation matrix with eigenvalues $\{\psi_{it}(\theta)\}_{i=1}^N$, so $H_t^{-1} = D_t^{-1/2}R_t^{-1}D_t^{-1/2}$. As in Lemma 2(i), we can show that there exist positive real numbers h and ψ such that $0 < h < \inf_{\theta\in\Theta}h_{ii,t}(\theta)$ and $0 < \psi < \inf_{\theta\in\Theta}\psi_{it}(\theta)$ for all t. Thus,

$$\sup_{\theta\in\Theta}\left\|H_t^{-1}(\theta)\right\| \le h^{-1}\psi^{-1} < \infty. \tag{18}$$

In the first inequality we use that if λ is an eigenvalue of a nonsingular matrix A, then λ^{-1} is an eigenvalue of A^{-1}; see [36, p. 64, 3(c)]. Similarly, we can show that $\sup_{\theta\in\Theta}\left\|\tilde{H}_t^{-1}(\theta)\right\| < \infty$. A mean-value expansion of the likelihood function around $\tilde{h}_t$ gives

$$l_t - \tilde{l}_t = \frac{\partial l_t(\bar{h}_t)}{\partial h_t'}(h_t - \tilde{h}_t) = \frac{\partial l_t(\bar{h}_t)}{\partial h_t'}D_N^+\mathrm{vec}(H_t - \tilde{H}_t) \tag{19}$$

where $\bar{h}_t$ is evaluated on the chord between h_t and $\tilde{h}_t$ and

$$\frac{\partial\tilde{l}_t}{\partial h_t'} = \mathrm{vec}\left(\bar{H}_t^{-1} - y_t'\bar{H}_t^{-1}\otimes\bar{H}_t^{-1}y_t\right)'D_N. \tag{20}$$

We use the fact that, for any $m\times n$ matrices A, B,

(a) $\|\mathrm{vec}(A)\| = \|A\|_2 \le \sqrt{\min(m,n)}\|A\|$

(b) $\|\mathrm{vec}(A\otimes B)\| = |(I_m\otimes K_{nm}\otimes I_m)(\mathrm{vec}(A)\otimes\mathrm{vec}(B))| \le C_1\|A\|\cdot\|B\|$

where K_{mn} is the commutation matrix. Assumption 3.2, (18)–(20) and the c_r inequality imply

$$\mathrm{E}\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)|^{s/2} \le C_1\mathrm{E}\left[\|y_t\|^s\sup_{\theta\in\Theta}\left\|\bar{H}_t^{-1}\right\| + \sup_{\theta\in\Theta}\left\|\bar{H}_t^{-1}\right\|^{s/2}\right]\cdot\sup_{\theta\in\Theta}\left\|H_t - \tilde{H}_t\right\|^{s/2} = O(\bar{\rho}^t). \tag{21}$$

This result and the Markov inequality imply that, for an arbitrary $\varepsilon > 0$, we have that $\left\{\mathrm{P}(\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)| \ge \varepsilon)\right\}_t$ is summable. Therefore, by the Borel–Cantelli Lemma $\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)| \to_{a.s.} 0$, and the desired result follows from Césaro's mean theorem. □

Lemma 3. *Under Assumptions 3.1–3.6,*

(i) $\mathrm{E}\sup_{\theta\in\Theta}\left\|\dot{\tilde{H}}_{t,i}\right\|^3<\infty$

(ii) $\mathrm{E}\sup_{\theta\in\Theta}\left\|\ddot{\tilde{H}}_{t,ij}\right\|^3<\infty$

where $\dot{\tilde{H}}_{t,i}=\frac{\partial\tilde{H}_t}{\partial\theta_i}$ *and* $\ddot{\tilde{H}}_{t,ij}=\frac{\partial^2\tilde{H}_t}{\partial\theta_i\partial\theta_j}$.

Proof. (i) The components of $\partial\tilde{h}_t/\partial\theta_i$ are given in the following.

$$\tilde{h}_t=\omega+A\eta_{t-1}+B\tilde{h}_{t-1}=\sum_{i=0}^{\infty}B^i(\omega+A\eta_{t-1-1})$$

$$\frac{\partial\tilde{h}_t}{\partial\omega'}_{(d\times d)}=(I_d-B)^{-1} \tag{22}$$

$$\begin{aligned}\frac{\partial\tilde{h}_t}{\partial a'}_{(d\times d^2)}&=(\eta'_{t-1}\otimes I_d)+\frac{\partial\tilde{h}_{t-1}}{\partial a'}(I_d\otimes B)\\&=\sum_{i=0}^{\infty}(I_d\otimes B)^i(\eta'_{t-i-1}\otimes I_d)=\sum_{i=0}^{\infty}(\eta'_{t-i-1}\otimes B^i)\end{aligned} \tag{23}$$

$$\begin{aligned}\frac{\partial\tilde{h}_t}{\partial b'}_{(d\times d^2)}&=\sum_{i=0}^{\infty}\left[(\omega+A\eta_{t-i-1})'\otimes I_d\right]\frac{\partial\mathrm{vec}(B^i)}{\partial\mathrm{vec}(B)'}\\&=\sum_{i=0}^{\infty}\left[(\omega+A\eta_{t-i-1})'\otimes I_d\right]\left[\sum_{j=0}^{i}(B')^{i-1-j}\otimes B^j\right]\end{aligned} \tag{24}$$

where $a=\mathrm{vec}(A)$ and $b=\mathrm{vec}(B)$. Let $\ell_{ij}=(i-1)d+j$ and $e^c_{\ell_{ij}}$ be a vector which contains a one at the ℓ_{ij} entry and zeros elsewhere, and whose dimension is the same as that of the vector c, where $c=a,b$ or ω. By definition,

$$\frac{\partial\tilde{H}_t}{\partial d_{ij}}=\mathrm{vec}^{-1}\left(\frac{\partial\tilde{h}_t}{\partial c'}e^c_{\ell_{ij}}\right) \tag{25}$$

where $d_{ij}=[A]_{ij}$,$[B]_{ij}$ or $[\omega]_{1j}$. Hence, to show that $\mathrm{E}\sup_{\theta\in\Theta}\left\|\dot{\tilde{H}}_{t,i}\right\|^3<\infty$ it suffices to prove that $\mathrm{E}\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{h}_t}{\partial c'}\right\|^3<\infty$. We see immediately that the derivatives with respect to the elements of ω are naturally bounded.

Note that, for any $m\times n$ matrices A, B,

$$\|A\otimes B\|=\lambda_{\max}[(A'\otimes B')(A\otimes B)]=\lambda_{\max}[(A'A\otimes B'B)]=\|A\|\cdot\|B\|, \tag{26}$$

and hence, from (16) and (26),

$$\sup_{\theta\in\Theta}\sum_{k=0}^{j}\left\|(B')^{j-1-k}\otimes B^k\right\|\le\sum_{k=0}^{j}\sup_{\theta\in\Theta}\|B^{j-1-k}\|\sup_{\theta\in\Theta}\|B^k\|\le C_1j\bar{\rho}^j. \tag{27}$$

Further, by using (23), (24), (26) and (27) and applying the Hölder and Minkowski inequalities, we get

$$\begin{aligned}\mathrm{E}\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{h}_t}{\partial b'}\right\|^3&\le\left\{\sum_{i=0}^{\infty}i\bar{\rho}^i\left[\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|(\omega+A\eta_{t-i-1})'\otimes I_m\right\|\right)^3\right]^{1/3}\right\}^3\\&\le\left\{\sum_{i=0}^{\infty}i\bar{\rho}^i\left[\mathrm{E}\left(\sup_{\theta\in\Theta}\|\omega+A\eta_{t-i-1}\|\right)^3\right]^{1/3}\right\}^3\\&\le\left\{\sum_{i=0}^{\infty}i\bar{\rho}^i(C_1+C_2\mathrm{E}^{1/6}\|y_t\|^6)\right\}^3<\infty\end{aligned} \tag{28}$$

and

$$\mathrm{E}\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{h}_t}{\partial a'}\right\|^3\le\left\{\sum_{i=0}^{\infty}\left[\mathrm{E}\left(\sup_{\theta\in\Theta}\|B\|^i\cdot\|\eta'_{t-1-i}\|\right)^3\right]^{1/3}\right\}^3$$

$$\leq \left(\sum_{i=0}^{\infty} \bar{\rho}^i \mathrm{E}^{1/6}|y_t|^6\right)^3 \leq C_1 \left(\sum_{i=0}^{\infty} \bar{\rho}^i\right)^3 < \infty \tag{29}$$

and the desired result follows.

(ii) The components of $\partial^2 \tilde{h}_t / \partial\theta_i \partial\theta_j$ are given in the following.

$$\begin{aligned}
\frac{\partial \mathrm{vec}}{\partial b'}\left(\frac{\partial \tilde{h}_t}{\partial a'}\right)_{(d^3\times d^2)} &= \left(I_{d^2} \otimes \frac{\partial \tilde{h}_{t-1}}{\partial a'}\right)_{(d^3\times d^4)} C_{(d^4\times d^2)} + \left[(I_d \otimes B') \otimes I_d\right]_{d^3\times d^3} \frac{\partial \mathrm{vec}}{\partial b'}\left(\frac{\partial \tilde{h}_{t-1}}{\partial a'}\right)_{d^3\times d^2} \\
&= \sum_{i=0}^{\infty} \left[(I_d \otimes B'^i) \otimes I_d\right] \left(I_{d^2} \otimes \frac{\partial \tilde{h}_{t-1-i}}{\partial a'}\right) \cdot C_{(d^4\times d^2)} \\
&= \sum_{i=0}^{\infty} \left[(I_d \otimes B'^i) \otimes I_d\right] \left[I_{d^2} \otimes \sum_{j=0}^{\infty} \eta'_{t-2-j-i} \otimes B^j\right] \cdot C_{(d^4\times d^2)}
\end{aligned} \tag{30}$$

where $C = (I_d \otimes K_{d1} \otimes I_d)$ and K_{d1} is the commutation matrix.

$$\begin{aligned}
\frac{\partial \mathrm{vec}}{\partial b'}\left(\frac{\partial \tilde{h}_t}{\partial b'}\right)_{(d^3\times d^2)} &= \sum_{i=0}^{\infty} \frac{\partial \mathrm{vec}}{\partial b'}\left(\tilde{h}'_{t-1-i} \otimes B^i\right) \\
&= C_1 \sum_{i=0}^{\infty} \left(\frac{\partial \tilde{h}_{t-1-i}}{\partial b'} \otimes \mathrm{vec}(B^i) + \tilde{h}_{t-1-i} \otimes \frac{\partial \mathrm{vec}(B^i)}{\partial b'}\right) \\
&= C_1 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left[(\omega + A\eta_{t-2-i-j})' \otimes I_d\right] \left[\sum_{k=0}^{j} (B')^{j-1-k} \otimes B^k\right] \otimes \mathrm{vec}(B^i) \\
&\quad + \tilde{h}_{t-1-i} \otimes \left[\sum_{k=0}^{i} (B')^{i-1-k} \otimes B^k\right]
\end{aligned} \tag{31}$$

where $\tilde{h}_{t-1-i} = \sum_{k=0}^{\infty} B^k(\omega + A\eta_{t-2-i-k})$.

Let $M = \frac{\partial \mathrm{vec}}{\partial c'}\left(\frac{\partial \tilde{h}_t}{\partial c'}\right)$. Then,

$$\frac{\partial^2 \tilde{H}_t}{\partial d_{ij} \partial d_{gk}} = \mathrm{vec}^{-1}\left\{\mathrm{vec}^{-1}\left(M e^c_{\ell_{ij}}\right) e^c_{\ell_{gk}}\right\} \tag{32}$$

where d_{ij} and d_{gk} are defined above. Hence, to show that $\mathrm{E}\sup_{\theta\in\Theta} \left\|\ddot{\tilde{H}}_{t,ij}\right\|^3 < \infty$ it suffices to establish that $\mathrm{E}\sup_{\theta\in\Theta} \left\|\frac{\partial \mathrm{vec}}{\partial c'}\left(\frac{\partial \tilde{h}_t}{\partial c'}\right)\right\|^3$ is bounded.

We note that

$$\begin{aligned}
\mathrm{E}\sup_{\theta\in\Theta} \|\tilde{h}_{t-1-i}\|^3 &\leq \left\{\sum_{k=0}^{\infty} \sup_{\theta\in\Theta} \|B^k\| \left(\mathrm{E}\sup_{\theta\in\Theta} \|\omega + A\eta_{t-2-i-k}\|^3\right)^{1/3}\right\}^3 \\
&\leq \left\{\sum_{k=0}^{\infty} \bar{\rho}^k (C_1 + C_2 \mathrm{E}\|y_t\|^6)^{1/6}\right\}^3 < \infty
\end{aligned} \tag{33}$$

and recall that $\|\mathrm{vec}(B^i)\| = \|B^i\|_2 \leq d\|B^i\| \leq d\bar{\rho}^i$.

From (31) and (33), and by applying the Hölder and Minkowski inequalities,

$$\begin{aligned}
\left\{\mathrm{E}\sup_{\theta\in\Theta} \left\|\frac{\partial \mathrm{vec}}{\partial b'}\left(\frac{\partial \tilde{h}_t}{\partial b'}\right)\right\|^3\right\}^{1/3} &\leq C_1 \sum_{i=0}^{\infty} \bar{\rho}^i \sum_{j=0}^{\infty} j\bar{\rho}^j \left(\mathrm{E}\sup_{\theta\in\Theta} \left\|\omega + A\eta_{t-2-i-j}\right\|^3\right)^{1/3} \\
&\quad + C_2 \sum_{i=0}^{\infty} i\bar{\rho}^i \left(\mathrm{E}\sup_{\theta\in\Theta} \|\tilde{h}_{t-1-i}\|^3\right)^{1/3} < \infty.
\end{aligned}$$

Similarly, we can show that $\mathrm{E}\sup_{\theta\in\Theta} \left\|\frac{\partial \mathrm{vec}}{\partial b'}\left(\frac{\partial \tilde{h}_t}{\partial a'}\right)\right\|^3 < \infty$, which completes the proof. □

Lemma 4. *Under Assumptions 3.1–3.6,*

(i) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta}\to_D N(0,V)$

(ii) $-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\bar{\theta}_n)}{\partial\theta\partial\theta'}\to_{a.s.} J$*, a positive definite matrix, where* $\bar{\theta}_n$ *is between* $\hat{\theta}_n$ *and* θ_0*.*

Proof. (i) We first show that the normalized score function obeys the CLT. The components of the score function are given by

$$\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_i}=\mathrm{Tr}[(I_N-y_ty_t'\tilde{H}_t(\theta_0)^{-1})\dot{\tilde{H}}_{t,i}(\theta_0)\tilde{H}_t^{-1}(\theta_0)]. \tag{34}$$

From (1) and Assumption 3.2, it is clear that the score function is strictly stationary, ergodic and a martingale difference; that is, $\mathrm{E}\left[\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}|\mathcal{F}_{t-1}\right]=0$. If $\mathrm{E}\left\|\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}'\right\|<\infty$, we can apply the CLT of Scott [37] and the Cramér–Wold device to establish the asymptotic normality of the score function. We have

$$\begin{aligned}\mathrm{E}\left|\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_i}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_j}\right| &\le C_1\mathrm{E}\left(\left\|I_N-y_ty_t'\tilde{H}_t(\theta_0)^{-1}\right\|^2\left\|\dot{\tilde{H}}_{t,i}(\theta_0)\right\|\cdot\left\|\dot{\tilde{H}}_{t,j}(\theta_0)\right\|\right)\\ &\le 2C_1\mathrm{E}\left[\left(1+\left\|\tilde{H}_t^{1/2}(\theta_0)\xi_t\xi_t'\tilde{H}_t^{-1/2}(\theta_0)\right\|^2\right)\left\|\dot{\tilde{H}}_{t,i}(\theta_0)\right\|\cdot\left\|\dot{\tilde{H}}_{t,j}(\theta_0)\right\|\right]\\ &\le 2C_1\left(\mathrm{E}\|\xi_t\|^4+1\right)\mathrm{E}\left(\left\|\dot{\tilde{H}}_{t,i}(\theta_0)\right\|\cdot\left\|\dot{\tilde{H}}_{t,j}(\theta_0)\right\|\right)\\ &\le C_2\left[\mathrm{E}\left(\left\|\dot{\tilde{H}}_{t,i}(\theta_0)\right\|\right)^2\right]^{1/2}\left[\mathrm{E}\left(\left\|\dot{\tilde{H}}_{t,j}(\theta_0)\right\|\right)^2\right]^{1/2}<\infty.\end{aligned} \tag{35}$$

The first inequality uses (34) and the fact that, if A and B are $N\times N$ matrices, then $|\mathrm{Tr}(AB)|\le N\|A\|\cdot\|B\|$; see [36, p. 111, 6(a)]. The second inequality uses the c_r inequality. We can show that $\left\|\tilde{H}_t^{1/2}(\theta_0)\xi_t\xi_t'\tilde{H}_t^{-1/2}(\theta_0)\right\|^2\le C_1\left\|\xi_t\xi_t'\right\|^2\le C_2\|\xi_t\|^4$ a.s. as in Lemma 1. The third inequality is implied by the independence between ξ_t and $\tilde{H}_t$ and its derivatives. Further, from Assumption 3.6 and (18),

$$\mathrm{E}\|\xi_t\|^4=\mathrm{E}(\mathrm{Tr}(\xi_t'\xi_t))^2=\mathrm{E}(\mathrm{Tr}(y_t'H_t^{-1}y_t))^2\le C_1\mathrm{E}\|y_t\|^4<\infty.$$

This result and the Cauchy–Schwarz inequality imply the third inequality. The last inequality results from Lemma 3(i). Next, by Lemma 3 and using similar arguments as in [16, Lemma 4(i)], we can show that

$$\left|n^{-1/2}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta_i}-\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_i}\right|=o_{\mathrm{P}}(1). \tag{36}$$

Eq. (36) and the asymptotic equivalence lemma imply the desired result.

(ii) The second derivative is given by

$$\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j}=\mathrm{Tr}\left[\left(I-y_ty_t'\tilde{H}_t^{-1}\right)\left(\ddot{\tilde{H}}_{t,ij}\tilde{H}_t^{-1}-\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\right)+y_ty_t'\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\right] \tag{37}$$

using the notation $\ddot{\tilde{H}}_{t,ij}=\frac{\partial^2\tilde{H}_t}{\partial\theta_i\partial\theta_j}$. As in Lemma 4(i) above, using Lemma 3, the Cauchy–Schwarz inequality and Minkowski's inequality, we get

$$\begin{aligned}\mathrm{E}\sup_{\theta\in\Theta}\left|\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j}\right| &\le C_1\mathrm{E}\sup_{\theta\in\Theta}\left[\|y_ty_t'\|\left(\left\|\ddot{\tilde{H}}_{t,ij}\right\|+2\left\|\dot{\tilde{H}}_{t,i}\right\|\cdot\left\|\dot{\tilde{H}}_{t,j}\right\|\right)\right]+C_2\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\ddot{\tilde{H}}_{t,ij}\right\|\right)+C_3\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\dot{\tilde{H}}_{t,i}\right\|\cdot\left\|\dot{\tilde{H}}_{t,j}\right\|\right)\\ &\le C_1\left(\mathrm{E}\left(\|y_t\|\right)^4\right)^{1/4}\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\ddot{\tilde{H}}_{t,ij}\right\|^2\right)^{1/2}+2C_2\left[\mathrm{E}\left(\|y_t\|^6\right)\right]^{1/6}\left[\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\dot{\tilde{H}}_{t,i}\right\|^3\right)\right]^{1/3}\left[\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\dot{\tilde{H}}_{t,j}\right\|^3\right)\right]^{1/3}\\ &\quad+C_3\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\ddot{\tilde{H}}_{t,ij}\right\|\right)+C_4\left[\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\dot{\tilde{H}}_{t,j}\right\|^2\right)\right]^{1/2}\mathrm{E}\left(\sup_{\theta\in\Theta}\left\|\dot{\tilde{H}}_{t,i}\right\|^2\right)^{1/2}<\infty.\end{aligned}$$

From the ergodic theorem (see [35])

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j'}-\mathrm{E}\left(\frac{\partial\tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j'}\right)\right|\to_{a.s}0. \tag{38}$$

By Lemma 3, and using the same arguments as in [16, Lemma 4(ii)], we obtain

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta)}{\partial\theta_i\partial\theta_j'}-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j'}\right|\to_{a.s}0. \tag{39}$$

Hence, by (38) and (39), Theorem 2 and direct calculation, the desired result follows. □

References

[1] R.F. Engle, Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, Econometrica 50 (1982) 987–1008.
[2] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (1986) 307–327.
[3] L. Bauwens, S. Laurent, J.V.K. Rombouts, Multivariate GARCH models: A survey, Journal of Applied Econometrics 21 (2006) 79–109.
[4] T. Bollerslev, R. Engle, J. Wooldridge, A capital asset pricing mode1 with time varying covariances, Journal of Political Economy 96 (1988) 116–131.
[5] R.F. Engle, K.F. Kroner, Multivariate simultaneous generalized ARCH, Econometric Theory 11 (1995) 122–150.
[6] F.X. Diebold, M. Nerlove, The dynamics of exchange rate volatility: A multivariate latent factor ARCH model, Journal of Applied Econometrics 4 (1989) 1–21.
[7] R.F. Engle, V. Ng, M. Rothschild, Asset pricing with a factor-ARCH structure: Empirical estimates for treasury bills, Journal of Econometrics 45 (1990) 213–237.
[8] C.O. Alexander, in: C.O. Alexander (Ed.), Orthogonal GARCH, in Mastering Risk, Vol. II, Prentice Hall, 2001, pp. 21–38.
[9] R. van der Weide, GO-GARCH: A multivariate generalized orthogonal GARCH model, Journal of Applied Econometrics 17 (2002) 549–564.
[10] I.D. Vrontos, P. Dellaportas, D.N. Politis, A full-factor multivariate GARCH model, Econometrics Journal 6 (2003) 311–333.
[11] T. Bollerslev, J. Wooldridge, Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances, Econometric Reviews 11 (1992) 143–72.
[12] T. Jeantheau, Strong consistency for multivariate of ARCH models, Econometric Theory 14 (1998) 70–86.
[13] T. Bollerslev, Modeling the coherence on short-run nominal exchange rates: A multivariate generalized ARCH model, Review of Economics and Statistics 72 (1990) 498–505.
[14] S. Ling, M. McAleer, Asymptotic theory for a vector ARMA–GARCH model, Econometric Theory 19 (2003) 280–310.
[15] F. Comte, O. Lieberman, Asymptotic theory for multivariate GARCH processes, Journal of Multivariate Analysis 84 (2003) 61–84.
[16] C.M. Hafner, A. Preminger, Asymptotic theory for a factor GARCH model, Econometric Theory 25 (2009) 336–363.
[17] P. Bougerol, N. Picard, Strict stationarity of generalized autoregressive processes, The Annals of Probability 20 (1992) 1714–1730.
[18] D.B. Nelson, Stationarity and persistence in the GARCH(1,1) model, Econometric Theory 6 (1990) 318–334.
[19] J.G. Dennis, E. Hansen, A. Rahbek, ARCH Innovations and their impact on cointegration rank testing, Working paper no. 22, Centre for Analytical Finance, University of Aarhus, 2002.
[20] D. Kristensen, Uniform ergodicity of a class of Markov chains with applications to time series models, Working paper, Economics department, Columbia University, 2007.
[21] F. Boussama, Ergodicity, mixing and estimation in GARCH models. Ph.D. Dissertation, University Paris 7, 1998.
[22] D.B.H. Cline, D.P. Pu, Verifying irreducibility and continuity of a nonlinear time series model, Statistics and Probability Letters 40 (1998) 139–148.
[23] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, New York, 1993.
[24] C.M. Hafner, Fourth moment structure of multivariate GARCH models, Journal of Financial Econometrics 1 (2003) 26–54.
[25] D. Kristensen, A. Rahbek, Asymptotics of the QMLE for a class of ARCH(q) models, Econometric Theory 21 (2005) 946–961.
[26] H. Lütkepohl, Introduction to Multiple Time Series Analysis, 2nd ed., Springer Verlag, Berlin, New York, 1993.
[27] G. Fiorentini, E. Sentana, Identification, estimation and testing of conditionally heteroskedastic factor models, Journal of Econometrics 102 (2001) 143–164.
[28] C. Doz, E. Renault, Conditionally heteroskedastic factor models: Identification and instrumental variables estimation, CIRANO working paper, 2004.
[29] D.W.K. Andrews, Testing when a parameter is on the boundary of the maintained hypothesis, Econometrica 69 (2001) 683–734.
[30] C. Francq, J.-M. Zakoian, Quasi-likelihood inference in GARCH processes when some coefficients are equal to zero, Stochastic Processes and their Applications 117 (2007) 1265–1284.
[31] C.M. Hafner, J.V.K. Rombouts, Semiparametric multivariate volatility models, Econometric Theory 23 (2006) 251–280.
[32] B. Basrak, R.A. Davis, T. Mikosch, Regular variation of GARCH processes, Stochastic Processes and their Applications 99 (2002) 95–115.
[33] D. Tjøstheim, Non-linear time series and Markov chains, Advances in Applied Probability 22 (1990) 587–611.
[34] M. Meitz, P. Saikkonen, Ergodicity, mixing and existence of moments of a class of Markov models with applications, Econometric Theory 24 (2008) 1291–1320.
[35] P. Billingsley, Probability and Measure, John Wiley, New York, 1995.
[36] H. Lütkepohl, Handbook of Matrices, John Wiley and Sons, 1996.
[37] D.J. Scott, Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach, Advances in Applied Probability 5 (1973) 119–137.
[38] R.F. Engle, Dynamic conditional correlation — A simple class of multivariate GARCH models, Journal of Business and Economic Statistics 20 (2002) 339–350.