Electronic Journal of Statistics
Vol. 3 (2009) 956–992
ISSN: 1935-7524
DOI: 10.1214/09-EJS452

Statistical inference for non-stationary GARCH(p, q) models*

Ngai Hang Chan

Department of Statistics
Chinese University of Hong Kong, Shatin, NT, Hong Kong
e-mail: nhchan@sta.cuhk.edu.hk

and

Chi Tim Ng

Department of Applied Mathematics
Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong
e-mail: machitim@inet.polyu.edu.hk

Abstract: This paper studies the quasi-maximum likelihood estimator (QMLE) of non-stationary GARCH(p, q) models. By expressing GARCH models in matrix form, the log-likelihood function is written in terms of the product of random matrices. Oseledec's multiplicative ergodic theorem is then used to establish the asymptotic properties of the log-likelihood function and thereby, showing the weak consistency and the asymptotic normality of the QMLE for non-stationary GARCH(p, q) models.

AMS 2000 subject classifications: Primary 62G30; secondary 62M10.
Keywords and phrases: Asymptotic normality, consistency, non-stationary GARCH model, Oseledec's multiplicative ergodic theorem, product of random matrices, quasi-maximum likelihood estimator.

Received July 2009.

1. Introduction

Quasi-maximum likelihood estimator (QMLE) is commonly used in practice to estimate the parameters of ARCH-type models. Literature on statistical inference for the GARCH(p, q) models is considerable. Recent studies on the properties of the QMLE can be found in Berkes et al. [3], Berkes and Horváth [2], Straumann [15], and Robinson and Zaffaroni [13], among others. These papers establish the strong consistency and asymptotic normality of the QMLE by assuming that within a parameter space Θ, the GARCH(p, q) equation admits a strictly stationary solution for all $\theta \in \Theta$. In the contrary, Jensen and Rahbek ([7], [8]) relax the stationarity conditions and establish the asymptotic behavior of non-stationary GARCH$(1, 1)$ and ARCH(1) models. That is, the QMLEs of both stationary and non-stationary GARCH$(1, 1)$ model are asymptotically

*Research supported in part by grants from HKSAR-RGC-GRF 400306 and 400408. Helpful comments from an anonymous referee are gratefully acknowledged.

normal and consistent in certain senses. The purpose of this paper is to extend Jensen and Rahbek's result to general non-stationary GARCH(p,q) situations.

Consider the GARCH(p,q) model defined by

$$\begin{aligned} X_t^2 &= \sigma_t^2 \epsilon_t^2 , \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 , \end{aligned}$$

where ω, $\alpha = (\alpha_1, \ldots, \alpha_q)$, and $\beta = (\beta_1, \ldots, \beta_p)$ are strictly positive real constants while $\{\epsilon_t\}_{t \in Z}$ are independent and identically-distributed random variables of zero mean and unit variance. We assume that the polynomials $\alpha(z) = \sum_{j=1}^{q} \alpha_j z^j$ and $1 - \beta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$ are co-prime.

The GARCH model can be expressed in vector-matrix form $Y_t = A_t Y_{t-1} + b$, for $j \in Z$, where $Y_t = (\sigma_{t+1}^2, \ldots, \sigma_{t-p+2}^2, X_t^2, \ldots, X_{t-q+2}^2)^T$, $b = (\omega, 0, \ldots, 0)^{\mathrm{T}}$ and

$$A_t = \begin{pmatrix} \alpha_1 \epsilon_t^2 + \beta_1 & \beta_2 & \ldots & \beta_{p-1} & \beta_p & \alpha_2 & \alpha_3 & \ldots & \alpha_q \\ 1 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 \\ \epsilon_t^2 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} .$$

It is shown in Bougerol and Picard [4] that a GARCH(p,q) model admits a strict stationary solution if and only if the top Lyapunov exponent

$$\rho = \inf_{\ell \in N} \frac{1}{\ell + 1} \left\{ \mathrm{E} \log ||A_0 A_{-1} \cdots A_{-\ell}|| \right\}$$

is strictly negative.

In this paper, we are interested in the case of $\rho > 0$. Under this situation, the GARCH model does not admit any strictly stationary solution. However, a stochastic process $\{X_t\}_{0 \le t \le n}$ can nevertheless be defined by specifying the initial probability distribution of the vector Y_{-1}.

Throughout this paper, we assume that the observed data $\{X_t\}_{0 \le t \le n}$ are generated by the GARCH(p,q) model with parameters ω^0, $\alpha^0 = (\alpha_1^0, \ldots, \alpha_q^0)^T$, and $\beta^0 = (\beta_1^0, \ldots, \beta_p^0)^T$. The initial values of the variances $\{\sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{-p+1}^2\}$ and the returns $\{X_{-1}^2, X_{-2}^2, \ldots, X_{-q+1}^2\}$ are assumed to be fixed.

In what follows, a QMLE is constructed for estimating $\theta^0 = (\alpha^0, \beta^0)$ from the observed data $\{X_t\}_{-(q-1) \le t \le n}$. Let the parameter space of $\theta = (\alpha, \beta)$ be $\Theta \subset R^{p+q}$. The QMLE is constructed as if the innovation terms $\{\epsilon_t\}_{t \in Z}$ are standard normal random variables. The unobservable values ω^0 and $\{\sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{-p+1}^2\}$ are replaced by a given positive real number ω and a given sequence $H_0 =$

$(h_0, h_{-1}, \dots, h_{-p+1})$ respectively. For all $t = 1, 2, \dots$ and $\theta \in \Theta$, define the stochastic process

$$h_t(\theta) = \omega_0 + \sum_{i=1}^{p} \beta_i h_{t-i}(\theta) + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 .$$

It should be noted that if $\omega = \omega^0$, $H_0 = (\sigma_0^2, \sigma_{-1}^2, \dots, \sigma_{-p+1}^2)$ and $\theta = \theta^0$, then, $h_t(\theta^0) = \sigma_t^2$. The quasi log-likelihood function is defined as

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left[\frac{X_t^2}{h_t(\theta)} + \log(h_t(\theta)) \right] .$$

Following the approach of Jensen and Rahbek [7], weak consistency and asymptotic normality of the QMLE are established. This paper is organized as follows. The main theorem is presented in Section 2. The proof of the main theorem is outlined in Section 3. Concepts and results related to the Lyapunov exponents and the products of random matrices used in Section 3 are introduced in Sections 4 and 5 respectively. A Detail proof of the main theorem is given in the appendix.

2. Main results

Before stating the assumptions, an alternative vector-matrix representations of the GARCH model is introduced. Let $Y_t' = (Y_t, 1)^T$ and

$$A_t' = \begin{pmatrix} A_t & b \\ 0 & 1 \end{pmatrix} .$$

The GARCH model can be rewritten as $Y_t' = A_t' Y_{t-1}'$.

We assume the following conditions throughout this paper.

A1: $\mathrm{E}\epsilon_t^4 < \infty$, and $\mathrm{E}|\epsilon_t|^{-2\delta} < \infty$ for some $\delta > 0$.
A2: The top Lyapunov exponent of A for the data generating process is strictly positive.
A3: The top Lyapunov exponent of A' for the data generating process is strictly positive and simple (cf. Theorem 4.1 for the definition of simplicity).

Remark 2.1. Details about the concepts of the Lyapunov exponents related to the discussion in this paper are given in Appendix A.1. According to Oseledec's multiplicative ergodic theorem, $p + q - 1$ real numbers ($-\infty$ is allowed), called Lyapunov exponents, can be associated to a sequence of random matrices $A_1, A_2, \dots,$ to characterize the asymptotic behavior of the product $A_n A_{n-1} \dots A_1$. These Lyapunov exponents may have multiplicities greater than one. A Lyapunov exponent is called simple if its multiplicity is one. The greatest Lyapunov exponent is the top Lyapunov exponent defined in Section 1.

Remark 2.2. Assumptions A2 and A3 hold simultaneously can be illustrated via the following example. Consider the case that $p = q = 2$, $\alpha = (0.15, 0.1)$, and $\beta = (0.55, 0.35)$. The Lyapunov exponents are $(0.08, -0.92, -36.97)$ for A and $(0.08, 0.00, -0.92, -36.97)$ for A'. In this case, the top Lyapunov exponents of A and A' are simple and positive.

When the top Lyapunov exponent of A is strictly positive, we have the volatilities diverging to infinity, exhibiting the explosive behavior.

Lemma 2.1. *Let ρ be the top Lyapunov exponent of A and suppose that $\rho > 0$. Then, we have $\lim_{t\to\infty} \sigma_t^2 = +\infty$ a.s.*

The main results on consistency and asymptotic normality of the QMLE are given as follows.

Theorem 2.1. *Assume conditions A1–A3. Let $H_0 = (h_0, h_{-1}, \ldots, h_{-p+1})$ and ω be arbitrarily chosen fixed values. Here, all the elements in H_0 are non-negative but not all elements equal to zero. Then, there exist a positive-definite matrix Ω and a fixed open neighborhood $M(\theta^0)$ of θ^0, independent of n, such that*

(I) with probability tending to one as $n \to \infty$, the likelihood function $L_n(\theta)$ is uniquely minimized in $M(\theta^0)$,

(II) for $\theta_n = \arg\min_{M(\theta^0)} L_n(\theta)$ we have

$$\theta_n \to^p \theta^0$$

and

$$\sqrt{n}(\theta_n - \theta^0) \to^d \mathrm{N}(0, \mathrm{E}(1-\epsilon_t^2)^2\Omega^{-1}).$$

Remark 2.3. Theorem 2.1 guarantees the existence of a consistent local QMLE in an open neighborhood $M(\theta^0)$ of θ^0. As θ^0 is unknown, in practice, we search for the stationary points of $L_n(\theta)$ within

$$R_+^{p+q} = \{x > 0 : \quad x \in R^{p+q}\}$$

instead. Denote the set of such stationary points by T. Then, θ_n constructed in Theorem 2.1 belongs to T. That means, if n is sufficiently large, T contains a vector that is close enough to the true parameter θ^0. If T is a singleton, then the only element in T must equal to $\theta_n = \arg\min_{M(\theta^0)} L_n(\theta)$. Although the uniqueness of the stationary point is not guaranteed, based on simulations, Gauss-Newton type methods usually give a solution close to the true value θ^0 in most practical situations.

3. Proofs

This section provides proof of Lemma 2.1 and Theorem 2.1. Lemma 2.1 is shown in subsection 3.1. An outline of the proof of Theorem 2.1 will be given in subsection 3.2 while the technical details are given in the subsequent sections and the appendix. The following conventions are used throughout the paper.

Convention 3.1. *For k integers $1 \le i_1, i_2, \ldots, i_k \le p+q-1$, define*

$$\partial^{i_1 \cdots i_k} h_t(\theta) = \frac{\partial^k h_t(\theta)}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}}$$

and

$$h_{kt}^{i_1 \cdots i_k}(\theta) = \frac{\partial^{i_1 \cdots i_k} h_t(\theta)}{h_t(\theta)}.$$

Convention 3.2. *The notation e_i refers to a unit vector with the i-th component equaling one and other components equaling zero. When there is no confusion, the dimension of e_i is not specified.*

Convention 3.3. *Let x and y be two vectors with the same dimension. $x \gg y$ means that $x \ge y$ componentwise.*

Convention 3.4. *For any matrix M, M^{T} refers to its transpose and M_{ij} refers to the elements of the i-th row and j-th column.*

Convention 3.5. *Two matrix norms are used throughout this section. They are $\|\cdot\|_1$, the largest row sum of the matrix and the operator norm $\|\cdot\|$, i.e., $\|M\| = \sup_{|x|=1} |Mx|$.*

Convention 3.6. *Let Ω be the sample space. It can be chosen as the set that contains all sample paths of $\{\epsilon_t\}_{t \in Z}$. Let L be a shift operator on Ω.*

3.1. Proof of Lemma 2.1

Applying the recursive relationship $Y_t = A_t Y_{t-1} + b$ repeatedly, we have

$$\begin{aligned} \sigma_t^2 &= e_1^T \left\{ \left(\prod_{j=1}^{t} A_{t-j} \right) Y_{-1} + b + \sum_{j=1}^{t-1} \left(\prod_{i=1}^{j-1} A_{t-i} \right) b \right\} \\ &\ge e_1^T \left(\prod_{j=1}^{t} A_{t-j} \right) Y_{-1}. \end{aligned}$$

By Proposition 4.1, for all $Y_{-1} \gg 0$, $Y_{-1} \ne 0$ and $0 < \delta < \rho$, we have $e_1^T \left(\prod_{j=0}^{t} A_{t-j} \right) Y_{-1} > e^{n(\rho-\delta)}$ for sufficiently large t. Consequently, σ_t^2 diverges almost surely.

3.2. Outline of the Proofs of Theorem 2.1

Note that the process

$$\ell_t(\theta) = \frac{X_t^2}{h_t(\theta)} + \log h_t(\theta)$$

is not stationary and therefore, the ergodic theorem and central limit theorem are not directly applicable to establish the asymptotics of $L_n(\theta)$. In the case of

GARCH$(1,1)$, Jensen and Rahbek [7] suggest that the asymptotic properties of θ_n can be obtained without using the convergence and asymptotic normality of $L_n(\theta)$ if the derivatives of $\ell_t(\theta)$ up to order three can be approximated by some stationary processes. To generalize the results of GARCH$(1,1)$ to GARCH(p,q), the most difficult part of the proof is to show that the quantity $h_{t-j}(\theta^0)/h_t(\theta^0)$, which appears in the derivatives of $\ell_t(\theta)$, has the following two properties:

(1) For any fixed positive integer j, $h_{t-j}(\theta^0)/h_t(\theta^0)$ has limiting distribution when $t\to\infty$.

(2) For any $\theta\in\Theta$, the moments of $h_{t-j}(\theta)/h_t(\theta)$ decays exponentially as $j\to\infty$.

Provided that (1) and (2) hold, the remaining of the proof is analogous to that in Jensen and Rahbek [7]. The proof of (1) and (2) are less trivial than that of the GARCH$(1,1)$ case. For property (1), take $j=1$ as an example. In the GARCH$(1,1)$ case, $h_{t-1}(\theta^0)/h_t(\theta^0)$ can be approximated by

$$\frac{h_{t-1}(\theta^0)}{h_t(\theta^0)}=\frac{h_{t-1}(\theta^0)}{\omega+(\beta_1+\alpha_1\epsilon_t^2)h_{t-1}}\approx\frac{1}{\beta_1+\alpha_1\epsilon_t^2}\,.$$

Here, the right-hand side is stationary. In the GARCH(p,q) case, the quantity $(\beta_1+\alpha_1\epsilon_t^2)h_{t-1}$ has to be replaced by

$$e_1^T A_t(h_{t-1},\dots,h_{t-p},X_{t-2}^2,\dots,X_{t-p-1}^2)^T\,,$$

which involves not only h_{t-1}, but also $h_{t-2},\dots,h_{t-p},X_{t-2}^2,\dots,X_{t-p-1}^2$. This complicates the matter. To establish the convergence of $h_{t-j}(\theta^0)/h_t(\theta^0)$, techniques for product of random matrices are indispensable. Property (1) is established in the following lemma, which is a consequence of Proposition 4.1 and Lemma 4.1 given in Section 4.

Lemma 3.1. *There exists a stationary, ergodic, and adapted stochastic vector-valued process $\{\eta_t\}$ such that*

$$\frac{A_t'A_{t-1}'\cdots A_0'Y_{-1}'}{e_1^T A_t'A_{t-1}'\cdots A_0'Y_{-1}'}-\eta_t'\to 0\quad\textit{almost surely},$$

where, $\eta_t'=(\eta_t,0)$. Equivalently,

$$(Y_t,1)/h_{t+1}-\eta_t'\to 0\quad\textit{almost surely.}$$

To establish property (2), we bound $h_{t-j}(\theta)/h_t(\theta)$ by the matrix

$$Q^{t,j}(\theta)=\prod_{i=1}^{j}\left(B+\alpha_1\epsilon_{t-i}^2e_1e_1^T\right)\,,$$

where

$$B=\begin{pmatrix}\beta_1 & \beta_2 & \dots & \dots & \beta_p\\ 1 & & & & \\ & \dots & & & \\ & & \dots & & \\ & & & 1 & 0\end{pmatrix}.$$

To see this, let $F \gg 0$ be a $(p+q-1)$-dimensional vector. Since all elements in A_{t-j} are non-negative,

$$A_{t-j}F \gg \begin{pmatrix} B + \alpha_1 \epsilon_{t-j}^2 e_1 e_1^T & 0 \\ 0 & 0 \end{pmatrix} F\,.$$

Applying the above step repeatedly,

$$e_1^T A_{t-1} \dots A_{t-j} F \geq e_1^T \left\{ \prod_{i=1}^{j} \begin{pmatrix} B + \alpha_1 \epsilon_{t-j}^2 e_1 e_1^T & 0 \\ 0 & 0 \end{pmatrix} \right\} F\,.$$

If the first component of F is one, then

$$e_1^T \{A_{t-1} \dots A_{t-j}\} F \geq (Q^{t,j})_{11}\,.$$

This inequality is applicable when $F = Y_{t-j-1}/h_{t-j}$. We have the following lemma. The proof of the lemma is given in Section 5.

Lemma 3.2. *For any positive number r, there exist positive constants κ_1, $\kappa_2(r) < \lambda$, such that*

$$\frac{1}{(Q^{t,\ell})_{11}} \leq O(\kappa_1^\ell) \text{ a.s. and } \left\{ \mathrm{E} \left(\frac{1}{(Q^{t,\ell})_{11}} \right)^r \right\}^{1/r} \leq O(\kappa_2^\ell)\,.$$

4. Product of random matrices

This section is devoted to establishing some properties related to the product of random matrices $P_t' = A_t' A_{t-1}' \dots A_1'$ that was used in Section 3 to establish Lemma 3.1. Recall that the GARCH model can be written in vector matrix notation $Y_t' = A_t' Y_{t-1}'$ (see Section 2). The product P_t' arises from applying the above recursive relationship repeatedly. Oseledec's multiplicative ergodic theorem and the concepts of Lyapunov exponents are essential tools for our purpose. According to Oseledec's multiplicative ergodic theorem, $p+q-1$ Lyapunov exponents are associated to P_t' to characterize the asymptotic behavior of P_t'.

The results on

$$\lim_{t\to\infty} \frac{1}{t} \log(e_1^T P_t' Y_{-1}')$$

and

$$\frac{A_t' A_{t-1}' \cdots A_0' Y_{-1}'}{e_1^T A_t' A_{t-1}' \cdots A_0' Y_{-1}'}$$

are given in subsection 4.2 and 4.3 respectively. Subsection 4.1 provides an introduction to Oseledec's multiplicative ergodic theorem which is essential to understanding the materials presented in subsections 4.2 and 4.3.

4.1. Oseledec's multiplicative ergodic theorem

Results related to Oseledec's multiplicative ergodic theorem are introduced in this subsection. References on this topic can be found in Ledrappier [10], the collections of Cohen et al. [5] and Arnold, Crauel and Eckmann [1]. Section 1.5 of Krengal [9] also provides a short introduction to some of the results.

Oseledec's multiplicative ergodic theorem is stated in Theorem 4.1, in which, Lyapunov exponents and their multiplicities are defined. Ledrappier's version of multiplicative ergodic theorem and some related results are stated without proof in Theorem 4.2.

Theorem 4.1 (Oseledec's Multiplicative Ergodic Theorem, see Krengal [9] and Cohen, et al. [6])**.** *Let $\{M_t(\omega)\}_{t\in Z}$ be a stationary and ergodic stochastic process of $d \times d$ random matrices such that $\mathrm{E}\log^+\|M_1\| < \infty$. Define $P_t = M_t M_{t-1}\cdots M_1$. Then, there exists an L-invariant measurable set $\Omega' \subset \Omega$, i.e., $L\Omega' = \Omega'$, with $P(\Omega') = 1$ such that in Ω', the following holds.*

(I). The limit

$$\lim_{t\to\infty}\left\{P_t^T(\omega)P_t(\omega)\right\}^{1/2t} = B(\omega)$$

exists. Let $\rho_1 > \rho_2 > \cdots > \rho_s > -\infty$ be distinct log-eigenvalues of $B(\omega)$ with multiplicities $r_1, r_2, \ldots, r_s$. Then, $\rho_1, \ldots, \rho_s$ and $r_1, r_2, \ldots, r_s$ are constants. The eigenvalues are called the Lyapunov exponents and the largest one is called the top Lyapunov exponent. If a Lyapunov exponent has a multiplicity one, then it is called simple.

(II). For $1 \le k \le s$, the random set

$$V_k(\omega) = \left\{x \in R^d : \lim_{t\to\infty}\frac{1}{t}\log|P_t x| \le \rho_k\right\}$$

is a subspace with dimension $r_k + \cdots + r_s$.

(III). The subspaces can be arranged in an asending order

$$V_s \subset V_{s-1} \subset \cdots \subset V_1 = R^d.$$

When $x \in V_k(\omega) - V_{k-1}(\omega)$,

$$\lim_{t\to\infty}\frac{1}{t}\log|P_t x| = \rho_k,$$

and $V_k(L\omega) = M_1(\omega)V_k(\omega)$.

Theorem 4.2. *Let $M_1(\omega), M_2(\omega), \ldots$ be a stationary and ergodic stochastic process of invertible $d\times d$ matrices such that $\mathrm{E}\log^+\|M_1\| < \infty$ and $\mathrm{E}\log^+\|M_0^{-1}\| < \infty$. Define $P_t = M_t M_{t-1}\cdots M_1$. Then, there exists $\Omega' \subset \Omega$ with $P(\Omega') = 1$ such that for all $\omega \in \Omega'$, we have a direct sum decomposition*

$$R^d = W_1(\omega) \oplus W_2(\omega) \oplus \cdots \oplus W_s(\omega),$$

with the following properties.

1. *For $1 \le k \le s$, and $u \in W_k$. We have*

$$\frac{1}{n}\log|P_t u| \to \rho_k\,.$$

2. $W_k(L\omega) = M_1(\omega)W_k(\omega)$.
3. *The dimension of $W_1(\omega)$ is r_1, the multiplicity of the top Lyapunov exponent.*
4. *Let $\xi_k(\omega) \in W_k(\omega)$ be a random vector. Then,*

$$\frac{P_t\xi_k(\omega)}{|P_t\xi_k(\omega)|} \in W_k(L^t\omega).$$

5. *If the dimension of $W_k(\omega)$ is 1 and $\xi_k(\omega)$ is the unit vector in $W_k(\omega)$, we have the version of LeJan [11] for Oseledec's theorem*

$$\frac{P_t\xi_k(\omega)}{|P_t\xi_k(\omega)|} = \xi_k(L^t\omega).$$

   In addition, $\xi_k(\omega)$ depends only on the values of M_t for $-\infty < t \le 0$ and hence the process $\xi_k(L^t\omega)$ is stationary, ergodic, and adapted.

4.2. Top Lyapunov exponent of A'_t

The purpose of the subsection is to prove Proposition 4.1 below.

Proposition 4.1. *Let $x \gg 0$, $x \neq 0$ be a $(p+q-1)$-dimensional non-random vector and $x' = (x^T, 1)^T$.*

$$\lim_{t\to\infty}\frac{1}{t}\log(e_1^T P'_t x') = \lim_{t\to\infty}\frac{1}{t}\log|P'_t x'| = \lim_{t\to\infty}\frac{1}{t}\log\|P'_t\| = \rho'.$$

Proof of Proposition 4.1. We prove this statement in two steps.

Step 1: To show that $\frac{1}{t}|P'_t x'| \to \rho'_1$. Note that $\rho'_1 > 0$, $|P'_t e_{p+q}| = 1$, and P_t and x are all non-negative, it is enough to show that $\frac{1}{t}\log|P'_t e_1|, \ldots, \frac{1}{t}\log|P'_t e_{p+q-1}|$ converge to the same limit ρ'_1.

Theorem 4.1 guarantees that $\frac{1}{t}\log|P'_t e_1|$ converges. To obtain the convergence of $\frac{1}{t}\log|P'_t e_2|, \ldots, \frac{1}{t}\log|P'_t e_{p+q-1}|$, the following identities will be used. Similar identities can be found in the proof of Theorem 1.3 in Bougerol and Picard [4]. For simplicity, we use the notation $\left(\Pi_{j=2}^n A'_j\right) = A'_t A'_{t-1}\cdots A'_2$. Then

$$\begin{aligned}
P'_t e_p &= \beta_p\left(\Pi_{j=2}^t A'_j\right)e_1,\\
P'_t e_{p+q-1} &= \alpha_q\left(\Pi_{j=2}^t A'_j\right)e_1,\\
P'_t e_k &= \beta_k\left(\Pi_{j=2}^t A'_j\right)e_1 + \left(\Pi_{j=2}^t A'_j\right)e_{k+1},\ \text{for } 2 \le k \le p-1,\\
P'_t e_{q+k} &= \alpha_k\left(\Pi_{j=2}^t A'_j\right)e_1 + \left(\Pi_{j=2}^t A'_j\right)e_{p+k+1},\ \text{for } 1 \le k \le q-2.
\end{aligned}$$

It can be seen from these identities that $\frac{1}{t}\log|P'_t e_2|, \ldots, \frac{1}{t}\log|P'_t e_{p+q-1}|$, converge to $\lim_{t\to\infty}\frac{1}{t}\log|P'_t e_1|$. Take $\frac{1}{t}\log|P'_t e_{p-1}|$ as an example. We have

$$P'_t e_{p-1} = \beta_{p-1}\left(\Pi_{j=2}^t A'_j\right)e_1 + \left(\Pi_{j=2}^t A'_j\right)e_p = \beta_{p-1}\left(\Pi_{j=2}^t A'_j\right)e_1 + \beta_p\left(\Pi_{j=3}^t A'_j\right)e_1$$

The required result follows from the following fact.
For all positive sequences a_n, b_n and positive constants k_1, k_2, we have $\frac{1}{t}\log a_n \to a$ and $\frac{1}{t}\log b_n \to b$ implies $\frac{1}{t}\log(k_1 a_n + k_2 b_n) \to \max\{a, b\}$.

That $\lim_{t\to\infty}\frac{1}{t}\log|P'_t e_1| = \rho'_1$ is a consequence of the above fact and the following inequalities,

$$|P'_t e_1| \le \|P'_t\| \le |P'_t e_1| + |P'_t e_2| + \cdots + |P'_t e_{p+q-1}| + 1\,.$$

The second inequality is obtained by considering y such that $\sup_{|y|=1}|P'_t y|$ is attained. Since the absolute values of all components of y must be smaller than one, we have

$$\|P'_t\| = |P'_t y| \le |P'_t e_1| + |P'_t e_2| + \cdots + |P'_t e_{p+q-1}| + 1\,.$$

Step 2: To show that $\frac{1}{t}\log(e_1^T P'_t x') \to \rho_1$. Clearly, we have $e_1 P'_t x \le |P'_t x'|$. To obtain a lower bound for $e_1 P'_t x$, we use Lemma A.16 to get

$$|P'_t x'| \le e_1 P'_t x' \sqrt{1 + \frac{1}{\beta_1^2} + \cdots + \frac{1}{\beta_{p-1}^2} + \frac{1}{\alpha_1^2} + \cdots + \frac{1}{\alpha_{q-1}^2}}.$$

This yields the required result. □

4.3. Asymptotic behavior of $P'_t = A'_t A'_{t-1} \cdots A'_1$

The purpose of this subsection is to establish the following lemma.

Lemma 4.1. *Assume that* $\mathrm{E}\log^+\|A'_t\| < \infty$ *and the top Lyapunov exponent for* A'_t *is simple. Suppose that* $F' : \Omega \to R^{p+q}$ *is a random vector. Assume that*

$$\frac{1}{t}\log|A'_t A'_{t-1}\cdots A'_1 F'| \to \rho'_1$$

as $t \to \infty$, *then there exists a stationary, ergodic, and adapted stochastic* R^{p+q+1}*-valued process* $\{\eta'_t\}_{t\in Z}$ *such that as* $n \to \infty$,

$$\frac{A'_t A'_{t-1}\cdots A'_1 F'}{e_1^T A'_t A'_{t-1}\cdots A'_1 F'} - \eta'_t \to 0 \quad \textit{almost surely}$$

and for all $t \in Z$,

$$\eta'_{t+1} = \frac{A'_{t+1}\eta'_t}{e_1^T A'_{t+1}\eta'_t}.$$

Outline of the proof of Lemma 4.1:

To proof Lemma 4.1, we construct the stochastic process η_t' from Ledrappier's version of multiplicative ergodic theorem, which is stated in Theorem 4.2. This theorem associates a stationary and ergodic sequence of invertible matrices $\{M_t(\omega)\}_{t\in Z}$ with a random vector $\xi_1(\omega)$ such that

$$\frac{M_tM_{t-1}\dots M_1\xi_1(\omega)}{|M_tM_{t-1}\dots M_1\xi_1(\omega)|}=\xi_1(L^t\omega)\,,$$

provided that the top Lyapunov exponent of M_t is simple. It is natural to construct η_t' from $\xi_1(L^t\omega)$. However, since A_t' is not invertible, M_t cannot be chosen as A_t'. Here, we construct invertible matrix M_t and a linear transform $E_t : R^{\max(p,q)+1} \to R^{p+q}$ such that for any $x \in R^{\max(p,q)+1}$,

$$A_t'A_{t-1}'\dots A_1'E_0x = E_tM_tM_{t-1}\dots M_1x\,. \tag{4.1}$$

The proof is organized as follows. Firstly, the invertible random matrices M_t and the linear transform E_t are defined. Proposition 4.2 relates the matrix A_t' to M_t, which can be used to establish the identity (4.1). We show in Proposition 4.3 that A_t' and M_t share the same set of Lyapunov exponents except $-\infty$ that appears in A_t' only. These allow us to establish the asymptotic behavior of A_t' from those of M_t. Finally, we show that

$$\eta_t'=\frac{E_t\xi_1(L^t\omega)}{e_1^TE_t\xi_1(L^t\omega)}$$

can be served as an approximation to

$$\frac{A_t'A_{t-1}'\cdots A_1'F'}{e_1^TA_t'A_{t-1}'\cdots A_1'F'}\,.$$

Linking A_t' to the invertible matrix M_t:

M_t and the linear transform $E_t : R^{\max(p,q)+1} \to R^{p+q}$ that links A_t' and M_t, are constructed as follows.

Case $q > p$: Consider the vectors $Z_t = (\sigma_{t+1}^2,\dots,\sigma_{t-p+2}^2, X_{t-p+1}^2,\dots, X_{t-q+2}^2, 1)$. We have an alternative representation of the GARCH(p,q) model, $Z_t = M_tZ_{t-1}$, where

$$M_t=\left(\begin{array}{cccccccc|c}
\alpha_1\epsilon_t^2+\beta_1 & \dots & \dots & \alpha_p\epsilon_{t-p+1}^2+\beta_p & \alpha_{p+1} & \dots & \dots & \alpha_q & \omega \\
1 & & & & & & & & \\
 & \ddots & & & & & & & \\
 & & 1 & & & & & & \\
 & & & \epsilon_{t-p+1}^2 & & & & & \\
 & & & & 1 & & & & \\
 & & & & & \ddots & & & \\
 & & & & & & 1 & 0 & \\
\hline
 & & & & & & & & 1
\end{array}\right).$$

Define

$$E_t\xi = (\xi_1, \dots, \xi_p, \epsilon_t^2\xi_2, \dots, \epsilon_{t-p+2}^2\xi_p, \xi_{p+1}, \dots, \xi_{q+1})^{\mathrm{T}}$$

and

$$G = \begin{pmatrix} I_p & 0 & 0 & 0 \\ 0 & 0 & I_{q-p} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, we have $Y_t' = E_t Z_t$ and $Z_t = GY_t$.

Case $q \le p$**:** Consider the vectors $Z_t = (\sigma_{t+1}^2, \dots, \sigma_{t-p+2}^2, 1)$. Define

$$M_t = \left(\begin{array}{ccccccc|c} \alpha_1\epsilon_t^2 + \beta_1 & \dots & \alpha_q\epsilon_{t-q+1}^2 + \beta_q & \beta_{q+1} & \dots & & \beta_p & \omega \\ 1 & & & & & & & \\ & \ddots & & & & & & \\ & & \ddots & & & & & \\ & & & \ddots & & & & \\ & & & & & 1 & 0 & \\ \hline & & & & & & & 1 \end{array}\right),$$

$$E_t\xi = (\xi_1, \dots, \xi_p, \epsilon_t^2\xi_2, \dots, \epsilon_{t-q+2}^2\xi_q, \xi_{q+1})^{\mathrm{T}},$$

and

$$G = \begin{pmatrix} I_p & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We have $Z_t = M_t Z_{t-1}$, $Y_t' = E_t Z_t$ and $Z_t = GY_t$.

Remark 4.1. In order to apply Theorem 4.2, we need $\mathrm{E}\log^+ \|M_0^{-1}\| < \infty$. The choice of the norm here is immaterial as all matrix norms are equivalent. It is more convenient to work with the norm $\|\cdot\|_1$. For $\{M_t\}$ chosen in this subsection, the condition holds as $\mathrm{E}\log^+ |\epsilon_0|^{-2} < \infty$, or equivalently, $\mathrm{E}|\epsilon_0|^{-2\delta} < \infty$ for some $\delta > 0$.

Proposition 4.2. *For any* $\max(p, q) + 1$*-dimensional vector* x*, we have*

$$A_{t+1}' E_t x = E_{t+1} M_{t+1} x$$

and

$$E_0 G A_0' A_{-1}' \dots A_{-\min(p,q)+2}' = A_0' A_{-1}' \dots A_{-\min(p,q)+2}' .$$

Proof. Directly from the definition. ◻

Proposition 4.3. *Let* $\rho_s' < \rho_{s-1}' < \cdots < \rho_1'$ *be distinct Lyapunov exponents of* M_t *. Then,* $-\infty$ *and* $\rho_s', \dots, \rho_1'$ *are the Lyapunov exponents of* A_t'*. The multiplicities of a Lyapunov exponents* $\rho_s', \dots, \rho_1'$ *are the same for* A_t' *and* M_t *.*

Proof. First, we show that $-\infty$ is a Lyapunov exponent of A'_t. Let $J_r(\lambda)$ be the standard Jordan block of order r with diagonal elements equaling λ. Simple algebraic manipulations show that a non-random full-rank matrix P with $\min(p,q)-1$ columns can be found so that

$$A'_t P = P J_{\min(p,q)-1}(0)\,,$$

where the columns of P satisfy $A'_t P'_{\min(p,q)-1} = 0$ and $A'_t P'_i = P'_{i+1}$ for $1 \leq i \leq \min(p,q)-2$. As a result, for $n \geq \min(p,q)-1$,

$$(A'_t A'_{t-1} \cdots A'_1)^T (A'_t A'_{t-1} \cdots A'_1) P = 0,$$

showing that $-\infty$ is a Lyapunov exponent of A'_t with multiplicity at least equaling to $\min(p,q)-1$ (see Theorem 4.1).

By Theorem 4.1, we can find vector spaces

$$V_s(\omega) \subset V_{s-1}(\omega) \subset \cdots \subset V_1(\omega) = R^{\max(p,q)+1}$$

such that $\xi \in V_k(L^t\omega)$ if and only if

$$\lim_{t\to\infty} \frac{1}{t} \log |M_t M_{t-1} \cdots M_1 \xi| \leq \rho'_k.$$

Let V_P be the vector space spanned by the columns of P. Define a set of vector spaces

$$V_P \subset E_0 V_s \oplus V_P \subset \cdots \subset E_0 V_1 \oplus V_P = R^{p+q}.$$

What remains is to show that for $\eta \in E_0 V_k \oplus V_P$, we have

$$\lim_{t\to\infty} \frac{1}{t} \log |A'_t A'_{t-1} \cdots A'_1 \eta| \leq \rho'_k$$

and the dimension of $E_0 V_k + V_P$ is $\min(p,q) - 1 + r_k + \cdots + r_s$. Note that

$$(E_0 V_k \oplus V_P) - (E_0 V_{k-1} \oplus V_P) = E_0(V_k - V_{k-1}) \oplus V_P.$$

Consider $\eta = E_0\xi + \eta_P$ where $\xi \in V_k - V_{k-1}$ and $\eta_P \in V_P$. Then, by Proposition 4.2,

$$\lim_{t\to\infty} \frac{1}{t} \log |A'_t A'_{t-1} \cdots A'_1 \eta| = \lim_{t\to\infty} \frac{1}{t} \log |E_t M_t M_{t-1} \cdots M_1 \xi| = \rho'_k.$$

The linear transformation E_0 does not change the dimension of a vector space. In addition, for any $1 \leq k \leq s$, any elements in V_P , η_P, say, cannot be written as a linear combination of any basis of $E_0 V_k$, $\xi_1, \ldots, \xi_{r_k}$, say. To see this, assume on the contrary that $\eta_P = \sum_{i=1}^{r_k} c_i E_0 \xi_i$. Clearly, for $t \geq \min(p,q) - 1$, $A'_t \cdots A'_1 \eta_P = 0$. However,

$$A'_t \cdots A'_1 \sum_{i=1}^{r_k} c_i E_0 \xi_i = E_t \sum_{i=1}^{r_k} c_i M_t \cdots M_1 \xi_i\,.$$

By the invertibility of M , we have $M_t \cdots M_1 \xi_i$ are linearly independent and hence, $A'_t \cdots A'_1 \sum_{i=1}^{r_k} c_i E_0 \xi_i = 0$ if and only if all $c_i = 0$. As a result, the dimension of $E_0 V_k + V_P$ must be $\min(p,q) - 1 + r_k + \cdots + r_s$. ∎

Proof of Lemma 4.1. By Theorem 4.2, we have a direct sum decomposition $R^{\max(p,q)+1} = W_1(\omega) \oplus W_2(\omega) \oplus \cdots \oplus W_s(\omega)$ such that for any $\xi_k \in W_k(L^t\omega)$,

$$\lim_{t\to\infty} \frac{1}{t} \log |M_t M_{t-1} \cdots M_1 \xi_k| = \rho'_k$$

and

$$\frac{M_t M_{t-1} \cdots M_1 \xi_k}{|M_t M_{t-1} \cdots M_1 \xi_k|} \in W_k(\omega).$$

By Proposition 4.3, ρ'_1 for the matrices M_t is simple. Ignoring the sign, there is only one unit vector in $W_1(\omega)$ (see Theorem 4.2). Let this unit vector be $\xi_1(\omega)$. We now show that

$$\eta'_t = \frac{E_t \xi_1(L^t\omega)}{e_1^T E_t \xi_1(L^t\omega)}$$

meets our requirement. From Proposition 4.2

$$A'_t A'_{t-1} \cdots A'_{-(\min(p,q)-2)} = E_t M_t M_{t-1} \cdots M_1 G A'_0 \cdots A'_{-(\min(p,q)-2)}.$$

Decompose the vector

$$\xi_0 = G A'_0 \cdots A'_{-(\min(p,q)-2)} F'$$

into the components of $W_1(\omega), W_2(\omega), \ldots, W_s(\omega)$, then we have a random vector $(g_1, \ldots, g_s)(\omega)$ and unit vectors $\xi_i(\omega) \in W_i(\omega)$ such that for any integer n,

$$\xi_0 = g_1(\omega)\xi_1(\omega) + g_2(\omega)\xi_2(\omega) + \cdots + g_s(\omega)\xi_s(\omega).$$

For simplicity, define $P'_{t,k} = M_t M_{t-1} \cdots M_1 \xi_k(\omega)$. Note that $g_1(\omega) \neq 0$ almost surely. Otherwise, let $\Omega'' \subset \Omega$ be a measurable set such that $P(\Omega'') > 0$ and $g_1(\omega) = 0$ when $\omega \in \Omega''$. Without loss of generality, assume that $g_2(\omega) \neq 0$.

$$\begin{aligned} \frac{1}{t} \log |A'_t A'_{t-1} \cdots A'_{-(\min(p,q)-2)} F'| &= \frac{1}{t} \log |E_t M_t M_{t-1} \cdots M_1 \xi_0| \\ &= \frac{1}{t} \log |E_t \{g_2(\omega) P'_{t,2} + \cdots + g_s(\omega) P'_{t,s}\}| \,. \end{aligned}$$

Using Theorem 4.2,

$$\begin{aligned} & g_2(\omega) P'_{t,2} + \cdots + g_s(\omega) P'_{t,s} \\ = \quad & g_2(\omega) |P'_{t,2}| \xi_2(L^t\omega) + \cdots + g_s(\omega) |P'_{t,s}| \xi_s(L^t\omega)\,. \end{aligned}$$

When $t \to \infty$, the term $g_2(\omega)|P'_{t,2}|\xi_2(L^t\omega)$ dominates. In Ω'',

$$\frac{1}{t} \log |A'_t A'_{t-1} \cdots A'_{-(\min(p,q)-2)} F'| \to \rho'_2,$$

which contradicts the assumption.

Now, it can be seen that

$$\begin{aligned} & A'_t A'_{t-1} \cdots A'_{-(\min(p,q)-2)} F' \\ = \quad & g_1(\omega)|P'_{t,1}| \left\{ E_t \xi_1(L^t\omega) + \cdots + \frac{g_s(\omega)}{g_1(\omega)} \cdot \frac{|P'_{t,s}|}{|P'_{t,1}|} E_t \xi_s(L^t\omega) \right\} . \end{aligned}$$

Then

$$\frac{A'_t A'_{t-1} \cdots A'_{-(\min(p,q)-2)} F'}{e_1^T A'_t A'_{t-1} \cdots A'_{-(\min(p,q)-2)} F'} = \frac{E_t \xi_1(L^t\omega) + \cdots + \frac{g_s(\omega)}{g_1(\omega)} \cdot \frac{|P'_{t,s}|}{|P'_{t,1}|} E_t \xi_s(L^t\omega)}{e_1^T \left\{ E_t \xi_1(L^t\omega) + \cdots + \frac{g_s(\omega)}{g_1(\omega)} \cdot \frac{|P'_{t,s}|}{|P'_{t,1}|} E_t \xi_s(L^t\omega) \right\}} .$$

Since

$$\frac{E_t \xi_1(L^t\omega) + \cdots + \frac{g_s(\omega)}{g_1(\omega)} \cdot \frac{|P'_{t,s}|}{|P'_{t,1}|} E_t \xi_s(L^t\omega)}{e_1^T \left\{ E_t \xi_1(L^t\omega) + \cdots + \frac{g_s(\omega)}{g_1(\omega)} \cdot \frac{|P'_{t,s}|}{|P'_{t,1}|} E_t \xi_s(L^t\omega) \right\}} - \frac{E_t \xi_1(L^t\omega)}{|E_t \xi_1(L^t\omega)|} \to 0 \, ,$$

the process

$$\eta'_t = \frac{E_t \xi_1(L^t\omega)}{|E_t \xi_1(L^t\omega)|}$$

fulfills the requirement.

By Proposition 4.2 and the fact that

$$M_{t+1} \xi_1(L^n\omega) = \xi_1(L^{n+1}\omega),$$

we have $\eta'_{t+1} = \frac{A'_{t+1}\eta'_t}{e_1^T A'_{t+1}\eta'_t}$. □

5. Miscellaneous results on matrices

This appendix presents two results of the matrices B and $(Q^{t,j})_{11}$ introduced in Section 3. These two results are frequently used. In the following, $(B^j)_{ik}$ is the (i,k)-th element of the j-th power of B.

Proposition 5.1. *(I) $(B^j)_{11}$ satisfies the recursive relationship*

$$(B^j)_{11} = \sum_{i=1}^{\min(j,p)} \beta_i (B^{j-i})_{11}.$$

(II) For any $\delta > 0$, we have $(B^j)_{11} \le K(\rho(B) + \delta)^j$ for $j = 1, 2, \ldots$.
(III) If $\beta_1, \ldots, \beta_p > 0$, then the eigenvalue with the largest modulus λ_1 is real and positive, has a multiplicity of one, and is equal to $\rho(B)$. Also, for $1 \le i \le p$, the elements in the first row of the j-th power of B have order $(B^j)_{1i} = O(\lambda_1^j)$.

Proof. (I) The first conclusion is trivial.

(II) The characteristic equation of B is given by

$$\beta_1 + \frac{\beta_2}{\lambda} + \cdots + \frac{\beta_p}{\lambda^{p-1}} \quad = \quad \lambda\,. \tag{5.1}$$

Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues, then,

$$\begin{aligned} |1-\beta(z)| &= |(1-\lambda_1 z)\cdots(1-\lambda_p z)| \\ &\geq (1-|\lambda_1|\cdot|z|)\cdots(1-|\lambda_p|\cdot|z|) \\ &\geq (1-|\lambda_1|\cdot|z|)^p\,. \end{aligned}$$

Take $R=(|\lambda_1|+\delta)$. By Cauchy's estimation (see Theorem 10.26 in Rudin [14]), an upper bound is given by

$$(B^j)_{11} \leq \frac{1}{R^j}\cdot\frac{1}{(1-R|\lambda_1|)^p} = \frac{(|\lambda_1|+\delta)^{p+j}}{\delta^p}.$$

(III) Under the condition that $\beta_1, \ldots, \beta_p > 0$, the characteristic equation (5.1) has one and only one positive real root, which is also the root with the largest modulus.

Consider the Jordan decomposition $B = PJP^{-1}$. Normalizing the first component, the eigenvector corresponding to λ_1 is

$$\left(1, \frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_1^{p-1}}\right)^{\mathrm{T}},$$

and the corresponding row in P^{-1} with the first component normalized is

$$\Big(1, \lambda_1-\beta_1, \lambda_1^2-\beta_1\lambda_1-\beta_2, \ldots, \lambda_1^{p-1}-\beta_1\lambda_1^{p-2}-\cdots-\beta_{p-1}\Big)\,.$$

Note that all elements in this row vector must be greater than zero. If any one of them is negative, for example, the second component, then we have

$$\lambda_1^p - \beta_1\lambda_1^{p-1} - \cdots - \beta_p = (\lambda_1-\beta_1)\lambda_1^{p-1} - \beta_2\lambda_1^{p-2} - \cdots - \beta_p < 0.$$

The characteristic equation is no longer satisfied by λ_1. Since λ_1 only appears once in the Jordan matrix J and the coefficients of λ_1^j for $e_1^T B^j$ in the decomposition $B^j = PJ^jP^{-1}$ do not equal to zero, we have $(B^j)_{1i} = O(\lambda_1^j)$. □

Remark 5.1. A necessary and sufficient condition for B^j to decay exponentially is that $\rho(B) < 1$. This condition is equivalent to the condition that all roots of $1-\beta(z)=0$ lie outside the unit disc. The latter one will be used often.

Lemma 5.1. *For any positive number r, there exist positive constants κ_1, $\kappa_2(r) < \lambda$, such that*

$$\frac{1}{(Q^{t,\ell})_{11}} \leq O(\kappa_1^\ell) \ \textit{a.s. and} \ \left\{\mathrm{E}\left(\frac{1}{(Q^{t,\ell})_{11}}\right)^r\right\}^{1/r} \leq O(\kappa_2^\ell)\,.$$

Proof. Assume that the eigenvalue decomposition of B_{11} is $\lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_p u_p$. Here we set $\lambda_1 = \lambda$, i.e., the one with the largest modulus. Let $0 < \delta < \Re(u_1)$ be arbitrarily chosen, where $\Re(\cdot)$ refers to the real part of the number. Since

$$\lim_{i,j\to\infty} \frac{(B^i)_{11}(B^j)_{11}}{(B^{i+j})_{11}} = \Re(u_1)\,,$$

there exists an integer k_1 such that when $i, j > k_1$,

$$\frac{(B^i)_{11}(B^j)_{11}}{(B^{i+j})_{11}} > \Re(u_1) - \delta\,.$$

On the other hand, for any given i,

$$\lim_{j\to\infty} \frac{(B^i)_{11}(B^j)_{11}}{(B^{i+j})_{11}} = \frac{(B^i)_{11}}{\lambda^i}\,.$$

Choose $0 < \delta_i < \frac{(B^i)_{11}}{\lambda^i}$. Then, we can find $k_2 \geq k_1$ such that for all $0 \leq i \leq k_1,$ and $j > k_2$,

$$\frac{(B^i)_{11}(B^j)_{11}}{(B^{i+j})_{11}} > \frac{(B^i)_{11}}{\lambda^i} - \delta_i\,.$$

Take

$$\begin{aligned}\rho = \ \min\Bigg(&\frac{(B^l)_{11}(B^j)_{11}}{(B^{l+j})_{11}}, \frac{(B^i)_{11}}{\lambda^i} - \delta_i, \Re(u_1) - \delta | 0 \leq i \leq k_1, 0 \leq j \leq k_2,\\ &0 \leq l \leq k_2\Bigg) > 0\,,\end{aligned}$$

we have

$$0 \leq \rho \leq \frac{(B^i)_{11}(B^j)_{11}}{(B^{i+j})_{11}} \text{ for all } i, j \leq 1\,.$$

Here ρ depends on B and on the choices of δ and δ_is.

We turn to the multinomial expansion of $(Q^{t,\ell})_{11}$.

$$\begin{aligned}(Q^{t,\ell})_{11} &= e_1^T \left[\prod_{i=1}^{\ell}(B + \alpha_1 \epsilon_{t-i}^2 e_1 e_1^T)\right] e_1\\
&= (B^n)_{11} + \sum_{l=1}^{\ell} \sum_{1\leq j_1 \leq j_2 \leq \cdots \leq j_l \leq \ell} (B^{j_1-1})_{11} \alpha_1 \epsilon_{t-j_1}^2\\
&\quad \cdot (B^{j_2-j_1-1})_{11} \alpha_1 \epsilon_{t-j_2}^2 \cdots (B^{j_l-j_{l-1}-1})_{11} \alpha_1 \epsilon_{t-j_l}^2 \cdot (B^{\ell-j_l})_{11}\\
&\geq (B^{\ell})_{11} + \sum_{l=1}^{j} \sum_{1\leq j_1 \leq j_2 \leq \cdots \leq j_l \leq \ell} \rho \frac{(B^{j_1-1})_{11}}{(B^0)_{11}} \alpha_1 \epsilon_{t-j_1}^2\\
&\quad \cdot \rho \frac{(B^{j_2-1})_{11}}{(B^{j_1})_{11}} \alpha_1 \epsilon_{t-j_2}^2 \cdots \rho \frac{(B^{j_l-1})_{11}}{(B^{j_{l-1}})_{11}} \alpha_1 \epsilon_{t-j_l}^2 \cdot \rho \frac{(B^{\ell})_{11}}{(B^{j_l})_{11}}\\
&\geq \rho \prod_{i=1}^{\ell} \left(\frac{(B^i)_{11}}{(B^{i-1})_{11}} + \rho \alpha_1 \epsilon_{t-i}^2 \right)\,.\end{aligned}$$

Next, we show that

$$\mathrm{E}(\frac{(B^i)_{11}}{(B^{i-1})_{11}} + \rho\alpha_1\epsilon^2)^{-r} < K_1 < \frac{1}{\lambda}$$

for sufficiently large i and that $(Q^{t,\ell})_{11}$ decays exponentially almost surely. Note that $\frac{(B^i)_{11}}{(B^{i-1})_{11}} \to \lambda$, then for an arbitrarily chosen $\delta' > 0$, we have

$$\frac{(B^i)_{11}}{(B^{i-1})_{11}} > \lambda - \delta' \text{ for } j > N\,.$$

Let

$$f(x) = \mathrm{E}(\lambda - x + \rho\alpha_1\epsilon^2)^{-r} \text{ and } g(x) = \mathrm{E}\log(\lambda - x + \rho\alpha_1\epsilon^2)\,.$$

For non-degenerate ϵ^2, $f(0) < \lambda^{-r}$ and $g(0) > \log\lambda$. By the right-continuity of both $f(x)$ and $g(x)$ at $x = 0$, we can choose $\delta' > 0$ such that $K_1 = f(\delta') < \lambda^{-r}$ and $K_2 = g(\delta') > \log(\lambda)$ also. Almost sure exponential decay of $(Q^{t,\ell})_{11}$ follows from

$$\begin{aligned}
\frac{1}{\ell}\log(Q^{t,\ell})_{11} &\geq \frac{1}{\ell}\sum_{i=1}^{\ell}\log(\frac{(B^i)_{11}}{(B^{i-1})_{11}} + \rho\alpha_1\epsilon_{t-i}^2) \\
&\geq \frac{1}{\ell}\sum_{i=1}^{N}\log(\frac{(B^i)_{11}}{(B^{i-1})_{11}} + \rho\alpha_1\epsilon_{t-i}^2) + \frac{1}{\ell}\sum_{i=N+1}^{\ell}\log(\lambda - \delta' + \rho\alpha_1\epsilon_{t-i}^2) \\
&\to \mathrm{E}\log(\lambda - \delta' + \rho\alpha_1\epsilon^2) \;\text{ almost surely} \\
&= K_2 \\
&> \log\lambda\,.
\end{aligned}$$

□

Appendix A: A detail Proof of Theorem 2.1

We will closely follow the method of Jensen and Rahbek [7] to complete the proof of Theorem 2.1. It would be much easier to establish Theorem 2.1 with the additional assumptions that $\omega = \omega^0$ and $H_0 = (\sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{-p+1}^2)$, in which case $h_t(\theta^0) = \sigma_t^2$. Under such assumptions, we have the following Theorem A.1, which is proved in subsections A.1 to A.3. In subsection A.4, we prove Theorem 2.1 by showing that the difference

$$\sup_{M(\theta^0)} |L_n(\theta;\omega, h_0, h_{-1}, \ldots, h_{-p+1}) - L_n(\theta;\omega^0, \sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{-p+1}^2)|$$

and its derivatives up to order three converge in probability to zero.

Theorem A.1. *Suppose A1–A3 are satisfied. Let $H_0 = (\sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{-p+1}^2)$ and $\omega = \omega^0$. The conclusions in Theorem 2.1 hold.*

Some technical lemmae to be used are given in subsection A.5. In what follows, we give an outline of the proof of Theorem A.1. It suffices to construct positive-definite matrices Ω_1, Ω_2, and a neighbourhood of θ^0, $N(\theta^0)$ such that the following conditions C1–C3 hold. Then, Lemma 1 of Jensen and Rahbek [7] yields our results.

(C1)
$$\sqrt{n}\nabla L_n(\theta^0) \to^d \mathrm{N}(0, \Omega_1)\,,$$

(C2)
$$\nabla^2 L_n(\theta^0) \to^p \Omega_2\,,$$

and
(C3) the third-order derivatives are uniformly bounded by n-dependent random variables C_n,
$$\max_{i_1,i_2,i_3} \sup_{\theta\in N(\theta^0)} \left|\partial^{i_1,i_2,i_3} L_n(\theta)\right| \leq C_n\,,$$

where $C_n \to^p c$ for some $0 < c < \infty$.

The mean ergodic theorem and the martingale-array central limit theorem (see Pollard, [12]) can be be used to establish C1–C3 provided that we are able to construct stationary and ergodic stochastic processes which approximate $l_t(\theta^0)$ and its derivatives up to the second order. Similarly, to establish C3, we need stationary and ergodic stochastic processes $v_t^{i_1,i_2,i_3}$ such that $\mathrm{E}v_t^{i_1,i_2,i_3} < \infty$ and that $\sup_{\theta\in N(\theta^0)} \left|\partial^{i_1,i_2,i_3} l_t(\theta)\right| < v_t^{i_1,i_2,i_3}$.

The derivatives of $l_t(\theta)$ up to first three orders are given below (see equations 8-10 in Jensen and Rahbek [7]).

$$\partial^i l_t(\theta) = \left[1 - \frac{X_t^2}{h_t(\theta)}\right] h_{1t}^i(\theta)\,, \tag{A.1}$$

$$\partial^{i_1 i_2} l_t(\theta) = \left[1 - \frac{X_t^2}{h_t(\theta)}\right] h_{2t}^{i_1 i_2}(\theta) - \left[1 - 2\frac{X_t^2}{h_t(\theta)}\right] h_{1t}^{i_1}(\theta) h_{1t}^{i_2}(\theta)\,, \tag{A.2}$$

$$\begin{aligned}\partial^{i_1 i_2 i_3} l_t(\theta) = {} & \left[1 - \frac{X_t^2}{h_t(\theta)}\right] h_{3t}^{i_1 i_2 i_3}(\theta) - \left[1 - 2\frac{X_t^2}{h_t(\theta)}\right] \\ & \times \left[h_{2t}^{i_1 i_2}(\theta) h_{1t}^{i_3}(\theta) + h_{2t}^{i_1 i_3}(\theta) h_{1t}^{i_2}(\theta) + h_{2t}^{i_2 i_3}(\theta) h_{1t}^{i_1}(\theta)\right] \\ & + 2\left[1 - 3\frac{X_t^2}{h_t(\theta)}\right] h_{1t}^{i_1}(\theta) h_{1t}^{i_2}(\theta) h_{1t}^{i_3}(\theta)\,.\end{aligned} \tag{A.3}$$

Below, we consider the terms $X_t^2/h_t(\theta)$, $h_{1t}^i(\theta)$, $h_{2t}^{i_1 i_2}(\theta)$, and $h_{3t}^{i_1 i_2 i_3}(\theta)$ that appear in the above equations individually. Some useful identities are given.

First, when $\omega = \omega^0$ and $H_0 = (\sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{-p+1}^2)$, we have $h_t(\theta^0) = \sigma_t^2$ and

$$\frac{X_t^2}{h_t(\theta)} = \epsilon_t^2 \cdot \frac{h_t(\theta^0)}{h_t(\theta)}\,.$$

In particular, we have
$$\frac{X_t^2}{h_t(\theta^0)} = \epsilon_t^2\,.$$

The quantities $\partial^i h_t(\theta)$, $\partial^{i_1 i_2} h_t(\theta)$, and $\partial^{i_1 i_2 i_3} h_t(\theta)$ can be expressed in terms of $h_{t-j}(\theta)$, for $j = 1, 2, 3, \ldots$ Consider the following recursive relationship in vector-matrix form,

$$H_t(\theta) \quad = \quad BH_{t-1}(\theta) + \upsilon_{t-1} e_1 \, , \tag{A.4}$$

where $H_t(\theta) = (h_t, h_{t-1}, \ldots, h_{t-q+1})(\theta)$,

$$\upsilon_{t-1} = \omega + \sum_{i=1}^{q} \alpha_i X_{t-i} \, ,$$

and

$$B = \begin{pmatrix} \beta_1 & \beta_2 & \ldots & \ldots & \beta_p \\ 1 & & & & \\ & \ldots & & & \\ & & \ldots & & \\ & & & 1 & 0 \end{pmatrix} .$$

With (A.4), the recursive relationships for the derivatives of H_t up to order three can be obtained. For example, the first order derivatives are given by

$$\partial^i H_t(\theta) = (\partial^i B) H_{t-1}(\theta) + B(\partial^i H_{t-1}(\theta)) + (\partial^i \upsilon_{t-1}(\theta)) e_1 \, .$$

Applying the above recursive relationships repeatedly, we have

$$\begin{aligned}
H_t(\theta) \quad &= \quad B^t H_0 + \sum_{j=1}^{t} B^{j-1} \upsilon_{t-j} e_1 \, , && \text{(A.5)} \\
\partial^i H_t(\theta) \quad &= \quad \sum_{j=1}^{t} B^{j-1} (\partial^i B) H_{t-j}(\theta) \\
&\qquad + \sum_{j=1}^{t} B^{j-1} (\partial^i \upsilon_{t-j}(\theta)) e_1 \, , && \text{(A.6)} \\
\partial^{i_1 i_2} H_t(\theta) \quad &= \quad \sum_{j=1}^{t} B^{j-1} (\partial^{i_1} B)(\partial^{i_2} H_{t-j}(\theta)) \\
&\qquad + \sum_{j=1}^{t} B^{j-1} (\partial^{i_2} B)(\partial^{i_1} H_{t-j}(\theta)) \, , && \text{(A.7)} \\
\partial^{i_1 i_2 i_3} H_t(\theta) \quad &= \quad \sum_{j=1}^{t} B^{j-1} (\partial^{i_1} B)(\partial^{i_2 i_3} H_{t-j}(\theta)) \\
&\qquad + \sum_{j=1}^{t} B^{j-1} (\partial^{i_2} B)(\partial^{i_1 i_3} H_{t-j}(\theta)) \\
&\qquad + \sum_{j=1}^{t} B^{j-1} (\partial^{i_3} B)(\partial^{i_1 i_2} H_{t-j}(\theta)) \, . && \text{(A.8)}
\end{aligned}$$

By using these recursive relationships, stationary and ergodic stochastic processes $u_{kt}^{i_1 \ldots i_k}$ are constructed in subsection A.1 to approximate $h_{kt}^{i_1 \ldots i_k}$.

Conditions C1 and C2 are established in subsection A.2 using the results on $u_{kt}^{i_1 \ldots i_k}$ given in subsection A.1. Finally, in subsection A.3, a neighborhood $N(\theta^0)$ is constructed so that the items $h_t(\theta^0)/h_t(\theta)$, $h_{1t}^{i}(\theta)$, $h_{2t}^{i_1 i_2}(\theta)$, and $h_{3t}^{i_1 i_2 i_3}(\theta)$ that appear in Equations (A.1)–(A.3) are bounded by some stationary and ergodic stochastic processes within $N(\theta^0)$. Condition C3 is then a consequence of the mean ergodic theorem.

A.1. Approximating $h_{kt}^{i_1 \ldots i_k}$ by $u_{kt}^{i_1 \ldots i_k}$

This subsection is devoted to establishing the approximations to $h_{1t}^{i}(\theta^0)$ and $h_{2t}^{i_1 i_2}(\theta^0)$ by stationary and ergodic processes which are then used in subsection A.1 to guarantee C1 and C2. Since we are only interested in $\theta = \theta^0$ when establishing C1 and C2, we drop the term (θ^0) and write (α, β) instead of (α^0, β^0). Throughout this section, we assume that the conditions in Theorem A.1 hold.

Applying $Y_t = A_t Y_{t-1} + b$ repeatedly, h_t can be written as the sum of

$$e_1^T \left\{ \left(\prod_{i=1}^{j} A_{t-i} \right) Y_{t-j-1} + b + \sum_{i=1}^{j-1} \left(\prod_{k=1}^{i-1} A_{t-k} \right) b \right\}.$$

Note that the first term will be dominated when $h_t = \sigma_t^2 \to +\infty$, which is guaranteed by Lemma 2.1. In addition, we have

$$\frac{Y_{t-j-1}}{h_{t-j}} = \frac{A'_{t-j-1} A'_{t-j-2} \cdots A'_0 Y'_{-1}}{e_1^T A'_{t-j-1} A'_{t-j-2} \cdots A'_0 Y'_{-1}}.$$

It is shown in Lemma 3.1 that there exists a stationary, ergodic, and adapted stochastic vector-valued process $\{\eta_t\}$ such that when $t \to \infty$,

$$\frac{A'_t A'_{t-1} \cdots A'_0 Y'_{-1}}{e_1^T A'_t A'_{t-1} \cdots A'_0 Y'_{-1}} - (\eta_t, 0)^T \to 0 \ \text{ almost surely}.$$

The approximation to $h_{1t}^{i}(\theta^0)$ and $h_{2t}^{i_1 i_2}(\theta^0)$ are given by u_{1t}^{i} and $u_{2t}^{i_1 i_2}$. For $\theta_i = \beta_\mu$, where $\mu = 1, \ldots, p$, define

$$u_{1t}^{i} = \sum_{j=1}^{\infty} (B^{j-1})_{11} \left(e_1^T A_{t-1} \cdots A_{t-j-\mu+1} \eta_{t-j-\mu} \right)^{-1},$$

and for $\theta_i = \alpha_\mu$, where $\mu = 1, \ldots, q$, define

$$u_{1t}^{i} = \sum_{j=1}^{\infty} (B^{j-1})_{11} \epsilon_{t-j-\mu+1}^2 \left(e_1^T A_{t-1} \cdots A_{t-j-\mu+1} \eta_{t-j-\mu} \right)^{-1}.$$

The second order derivatives $h_{2t}^{i_1 i_2}$ are approximated by

$$\begin{aligned} u_{2t}^{i_1 i_2} &= \sum_{j=1}^{\infty} e_1^T B^{j-1} (\partial^{i_1} B) \left(e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1} \right)^{-1} u_{1,t-j}^{i_2} \\ &\quad + \sum_{j=1}^{\infty} e_1^T B^{j-1} (\partial^{i_2} B) \left(e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1} \right)^{-1} u_{1,t-j}^{i_1} . \end{aligned}$$

The moment condition for $\{u_{kt}^{i_1 \ldots i_k}\}$ is established in Lemma A.1. The relationships between $h_{kt}^{i_1 \ldots i_k}$ and $u_{kt}^{i_1 \ldots i_k}$ are given in Lemma A.2.

Lemma A.1. *For integer $k = 1, 2$, the processes $\{u_{kt}^{i_1 \ldots i_k}\}$ are stationary and ergodic with finite moments $E(u_{kt}^{i_1 \ldots i_k})^p < \infty$ for any integer $p > 0$.*

Proof. Here, the proof is given for the cases $\theta_i = \{\beta_1\}$ and $\{\theta_{i_1}, \theta_{i_2}\} = \{\beta_1, \beta_1\}$ only. Other situations can be handled in the same manner. Below, we show the existence of the moment $\mathrm{E}(u_{kt}^{i_1 \ldots i_k})^p$. Stationarity and ergodicity follow directly from the Lebesgue dominated convergence theorem.

When $\theta_i = \{\beta_1\}$, applying Minkowski's inequality and Lemma 3.2, we have

$$\begin{aligned} \left(\mathrm{E}(u^{\beta_1})^p\right)^{1/p} &\le \sum_{j=1}^{\infty} (B^{j-1})_{11} \left[\mathrm{E} \left(\frac{1}{e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1}} \right)^p \right]^{1/p} \\ &\le \sum_{j=1}^{\infty} (B^{j-1})_{11} \left[\mathrm{E} \left(\frac{1}{e_1^T A_{t-1} \cdots A_{t-j} e_1} \right)^p \right]^{1/p} \\ &\le \sum_{j=1}^{\infty} (B^{j-1})_{11} \left[\mathrm{E} \left(\frac{1}{(Q^{t,j})_{11}} \right)^p \right]^{1/p} \\ &< \infty . \end{aligned}$$

When $\{\theta_{i_1}, \theta_{i_2}\} = \{\beta_1, \beta_1\}$, $(Q^{t,j})_{11}$ and $u_{1,t-j}^{i_1}$ are independent. Using Minkowski's inequality and the preceding result of $\mathrm{E}(u_{1t}^{i})^p$, we have

$$\begin{aligned} &\left(\mathrm{E}(u_{2t}^{\beta_1 \beta_1})^p \right)^{1/p} \\ &\le \sum_{j=1}^{\infty} (B^{j-1})_{11} \left(\mathrm{E} \left(e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1} \right)^{-p} \right)^{1/p} \left[\mathrm{E} \left(u_{1,t-j}^{\beta_1} \right)^p \right]^{1/p} \\ &\quad + \sum_{j=1}^{\infty} (B^{j-1})_{11} \left(\mathrm{E} \left(e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1} \right)^{-p} \right)^{1/p} \left[\mathrm{E} \left(u_{1,t-j}^{\beta_1} \right)^p \right]^{1/p} \\ &\le \sum_{j=1}^{\infty} (B^{j-1})_{11} \left(\mathrm{E} \left((Q^{t,j})_{11} \right)^{-p} \right)^{1/p} \left[\mathrm{E} \left(u_{2,t-j}^{\beta_1} \right)^p \right]^{1/p} \\ &\quad + \sum_{j=1}^{\infty} (B^{j-1})_{11} \left(\mathrm{E} \left((Q^{t,j})_{11} \right)^{-p} \right)^{1/p} \left[\mathrm{E} \left(u_{2,t-j}^{\beta_1} \right)^p \right]^{1/p} \end{aligned}$$

$$
\begin{aligned}
&= \left[\mathrm{E}\left(u_{1t}^{\beta_1}\right)^p \right]^{1/p} \sum_{j=1}^{\infty} e_1^T B^{j-1} (\partial^{\beta_1} B) \left(\mathrm{E}\left((Q^{t,j})_{11}\right)^{-p} \right)^{1/p} \\
&\quad + \left[\mathrm{E}\left(u_{1t}^{\beta_1}\right)^p \right]^{1/p} \sum_{j=1}^{\infty} e_1^T B^{j-1} (\partial^{\beta_1} B) \left(\mathrm{E}\left((Q^{t,j})_{11}\right)^{-p} \right)^{1/p} \\
&< \infty .
\end{aligned}
$$

□

Lemma A.2. *If A3 is satisfied, we have for* $k = 1, 2$,

$$h_{kt}^{i_1 \ldots i_k} - u_{kt}^{i_1 \ldots i_k} \to^{L^p} 0,$$

$$\frac{1}{T} \sum_{j=1}^{T} [(h_{1t}^{i_1})(h_{1t}^{i_2}) - (u_{1t}^{i_1})(u_{1t}^{i_2})] \to^{L^p} 0,$$

and

$$\frac{1}{T} \sum_{j=1}^{T} (h_{2t}^{i_1 i_2} - u_{2t}^{i_1 i_2}) \to^{L^p} 0.$$

Proof. Step 1: First, we give upper and lower bounds for the differences $u_{kt}^{i_1 \ldots i_k} - h_{kt}^{i_1 \ldots i_k}$ and show that the upper and lower bounds converge to zero in L^p.

Here, we only consider the case $\{\theta_i\} = \{\beta_1\}$. In this case, we have

$$u_{1t}^{\beta_1} - h_{1t}^{\beta_1} = \sum_{j=1}^{t} (B^{j-1})_{11} \left(\frac{1}{e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1}} - \frac{h_{t-j}}{h_t} \right) .$$

It should be noted that for any integer j , the summand converges almost surely to zero as $t \to \infty$ according to Lemma 3.1 and it can be bounded by

$$
\begin{aligned}
& (B^{j-1})_{11} \left(\frac{1}{e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1}} - \frac{h_{t-j}}{h_t} \right) \\
\geq \quad & (B^{j-1})_{11} \left(\frac{1}{e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1}} - \frac{1}{e_1^T A_{t-1} \cdots A_{t-j} \frac{Y_{t-j-1}}{h_{t-j}}} \right) \\
= \quad & \frac{(B^{j-1})_{11}}{e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1}} \cdot \frac{e_1^T A_{t-1} \cdots A_{t-j} (\frac{Y_{t-j-1}}{h_{t-j}} - \eta_{t-j-1})}{e_1^T A_{t-1} \cdots A_{t-j} \frac{Y_{t-j-1}}{h_{t-j}}} \\
\geq \quad & -\frac{(B^{j-1})_{11}}{(Q^{t,j})_{11}} \max_{1 \leq k \leq p+q-1} \left| 1 - \frac{\eta_{t-j-1,k}}{Y_{t-j-1,k}/h_{t-j}} \right| .
\end{aligned}
$$

Here, the quantity

$$\max_{1 \leq k \leq p+q-1} \left| 1 - \frac{\eta_{t-1,k}}{Y_{t-1,k}/h_t} \right|$$

is bounded above by some random variable with finite moment according to Lemma A.16. In addition, it converges to zero almost surely as $t \to \infty$ by Lemma 3.1. Therefore, the moments of this quantity converge to zero by dominated convergence theorem. Also, we have

$$\mathrm{E}\left|(B^{j-1})_{11}\left(\frac{1}{e_1^T A_{t-1}\cdots A_{t-j}\eta_{t-j-1}} - \frac{h_{t-j}}{h_t}\right)\right|^p \to 0$$

by noting that the term

$$\mathrm{E}\left(\frac{(B^{j-1})_{11}}{(Q^{t,j})_{11}}\right)$$

vanishes as a result of Proposition 5.1 and Lemma 3.2.

The lower bound for $u_{kt}^i - h_{kt}^i$ can be given by

$$u_{1t}^i - h_{1t}^i \geq -\sum_{j=1}^{t} \frac{(B^{j-1})_{11}}{(Q^{t,j})_{11}} \max_{1\leq k\leq p+q-1}\left|1 - \frac{\eta_{t-j-1,k}}{Y_{t-j-1,k}/h_{t-j}}\right| .$$

That the p-th moment of the upper bound converges to zero can be shown using Minkoswki's inequality and the following fact.

Let a_t and b_t be two sequences. If a_t decays exponentially and $b_t \to 0$, then, the sequence

$$x_t = \sum_{j=1}^{t} a_j b_{t-j}$$

converges to zero.

To see this, let n be an integer such that for $j > n$, we have $|b_n| < \delta$, where $\delta > 0$ is an arbitrarily small real number. Suppose that $|a_t| \leq K\lambda^t$. Then, for $t > n$,

$$|x_t| \leq \delta \sum_{j=n+1}^{t} |a_{t-j}| + \sum_{j=1}^{n} |a_{t-j}|\cdot|b_j| \leq K(1-\lambda)^{-1}\delta + \sum_{j=1}^{n} |a_{t-j}|\cdot|b_j|.$$

The last term converges to zero as $t \to 0$, hence the required result.

Next, we construct an upper bound for $u_{1t}^{\beta_1} - h_{1t}^{\beta_1}$. Note that by Proposition 5.1 and Lemma 3.2, the sum

$$\sum_{j=1}^{\infty}\left\{\mathrm{E}\left(\frac{(B^{j-1})_{11}}{(Q^{t,j})_{11}}\right)^p\right\}^{1/p}$$

converges. Then, suppose that n is an integer so that the sum

$$\sum_{j=n+1}^{\infty}\left\{\mathrm{E}\left(\frac{(B^{j-1})_{11}}{(Q^{t,j})_{11}}\right)^p\right\}^{1/p}$$

is sufficiently small. Assume that $t > n$, then we have

$$\begin{aligned} u_{1t}^i - h_{1t}^i &\leq \sum_{j=1}^{n} (B^{j-1})_{11} \left(\frac{1}{e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1}} - \frac{h_{t-j}}{h_t} \right) \\ &\quad + \sum_{j=n+1}^{\infty} \frac{(B^{j-1})_{11}}{e_1^T A_{t-1} \cdots A_{t-j} \eta_{t-j-1}} . \end{aligned}$$

Then we have the required result $h_{1t}^{\beta_1} - u_{1t}^{\beta_1} \to^{L^p} 0$.

Step 2: Define $v_{1t} = \sum_{j=1}^{\infty} \frac{(B^{j-1})_{11}}{(Q^{t,j})_{11}}$. Then v_{1t} can be used to bound the difference,

$$u_{1j}^{i_1} u_{1j}^{i_2} - h_{1j}^{i_1} h_{1j}^{i_2} = \frac{1}{2} \left[(u_{1j}^{i_1} + h_{1j}^{i_1})(u_{1j}^{i_2} - h_{1j}^{i_2}) + (u_{1j}^{i_2} + h_{1j}^{i_2})(u_{1j}^{i_1} - h_{1j}^{i_1}) \right] .$$

Below, we only consider the case $\{\theta_{i_1}, \theta_{i_2}\} = \{\beta_1, \beta_1\}$. We have

$$\begin{aligned} & \left\{ \mathrm{E} \left\| \frac{1}{n} \sum_{j=1}^{n} (u_{1j}^{\beta_1} + h_{1j}^{\beta_1})(u_{1j}^{\beta_1} - h_{1j}^{\beta_1}) \right\|^p \right\}^{1/p} \\ \leq & \left\{ \mathrm{E} \left\| \frac{2}{n} \sum_{j=1}^{n} v_{1t} (u_{1j}^{\beta_1} - h_{1j}^{\beta_1}) \right\|^p \right\}^{1/p} \\ \leq & \frac{2}{n} \left[\mathrm{E}(v_{1t})^{2p} \right]^{1/2p} \sum_{j=1}^{n} \left\{ (u_{1j}^{\beta_1} - h_{1j}^{\beta_1})^{2p} \right\}^{1/2p} \\ \to & \ 0 \ \text{ almost surely.} \end{aligned}$$

Step 3: That $u_{2t}^{i_1 i_2} - h_{2t}^{i_1 i_2} \to^{L^p} 0$ can be shown in a similar manner as in Step 1 by means of Lemma A.1 and the recursive relationship (A.7). □

A.2. Conditions C1 and C2

With the stationary and ergodic stochastic processes u_{1t}^i and $u_{2t}^{i_1 i_2}$ constructed in the last subsection, conditions C1 and C2 are established in this subsection. Again, since we are only interested in $\theta = \theta^0$ when establishing C1 and C2, we drop the term (θ^0) and write (α, β) instead of (α^0, β^0).

Define $\Omega = \left(\mathrm{E}(u_{1t}^{i_1} u_{1t}^{i_2}) \right)_{1 \leq i_1, i_2 \leq p+q}$, where Ω_1 and Ω_2 in Lemma 2.1 are chosen to be $\mathrm{E}(1-\epsilon^2)^2 \Omega$ and Ω respectively. Lemma A.3 gives C1 while Lemma A.4 gives C2. Lemma A.5 establishes the positive-definiteness of Ω.

Lemma A.3.

$$\sqrt{n} \nabla L_n \to^d \mathrm{N}(0, E(1-\epsilon_t^2)^2 \Omega_1).$$

Proof. The convergence of the first order derivative of the quasi log-likelihood function,

$$\partial^i L_n = \frac{1}{n}\sum_{t=1}^{n}(1-\epsilon_t^2)h_{1t}^i$$

can be obtained by the martingale central limit theorem. Using Lemma A.2 and the mean ergodic theorem, the sum of conditional covariances is

$$\begin{aligned}\frac{1}{n}\mathrm{E}(1-\epsilon_t^2)^2\sum_{j=1}^{n}h_j^{i_1}h_j^{i_2} &= \frac{1}{n}\mathrm{E}(1-\epsilon_t^2)^2\sum_{j=1}^{n}\left[h_{1t}^{i_1}h_{1t}^{i_2}-u_{1t}^{i_1}u_{1t}^{i_2}+u_{1t}^{i_1}u_{1t}^{i_2}\right]\\ &= \mathrm{E}(1-\epsilon_t^2)^2\ \cdot\mathrm{E}\left\{u_{1t}^{i_1}u_{1t}^{i_2}\right\}+o(1)\,.\end{aligned}$$

To show that the Linderberg condition holds, we bound h_{1t}^i by a stationary and ergodic process. For $\theta^i=\beta_\mu$, consider

$$h_{1t}^{i_1}\le\sum_{i=1}^{t}B_{11}^{i-1}\frac{h_{t-i-\mu+1}}{h_t}\le\sum_{i=1}^{t}B_{11}^{i-1}\left(Q_{11}^{t,i+\mu-1}\right)^{-1}=v_{1t}^i$$

and for $\theta^i=\alpha_\mu$ consider

$$h_{1t}^{i_1}\le\sum_{i=1}^{t}B_{11}^{i-1}\epsilon_{t-i-\mu+1}^2\frac{h_{t-i-\mu+1}}{h_t}\le\sum_{i=1}^{t}B_{11}^{i-1}\epsilon_{t-i-\mu+1}^2\left(Q_{11}^{t,i+\mu-1}\right)^{-1}=v_{1t}^i.$$

Here v_{1t}^i is stationary and ergodic by Lemma 3.2. The Linderberg condition holds as $h_{1t}^i\le v_{1t}^i$ and

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left((1-\epsilon_t^2)^2(h_{1t}^i)^2)\,1\{|(1-\epsilon_t^2)h_{1t}^i)|>\delta\sqrt{n}\}\right)\\ \le\ &\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left((1-\epsilon_t^2)^2(v_{1t}^i)^2\,1\{|(1-\epsilon_t^2)v_{1t}^i|>\delta\sqrt{n}\}\right)\\ =\ &\mathrm{E}\left((1-\epsilon_t^2)^2(v_{1t}^i)^2\,1\{|(1-\epsilon_t^2)v_{1t}^i|>\delta\sqrt{n}\}\right)\\ \to\ &0\,.\end{aligned}$$

□

Lemma A.4.

$$\partial^{i_1i_2}L_n(\theta^0)\to^p E\left(u_{1t}^{i_1}\right)\left(u_{1t}^{i_2}\right).$$

Proof.

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{(1-\epsilon_t^2)h_{2t}^{i_1i_2}-(1-2\epsilon_t^2)h_{1t}^{i_1}h_{1t}^{i_2}\right\}\\ =\ &\frac{1}{n}\sum_{i=1}^{n}\left\{(1-\epsilon_t^2)(h_{2t}^{i_1i_2}-u_{2t}^{i_1i_2}+u_{2t}^{i_1i_2})\right\}\end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{n}\sum_{i=1}^{n}\left\{(1-2\epsilon_t^2)(h_{1t}^{i_1}h_{1t}^{i_2}-u_{1t}^{i_1}u_{1t}^{i_2}+u_{1t}^{i_1}u_{1t}^{i_2})\right\} \\
\sim \quad & \frac{1}{n}\sum_{i=1}^{n}\left\{(1-\epsilon_t^2)u_{2t}^{i_1i_2}-(1-2\epsilon_t^2)u_{1t}^{i_1}u_{1t}^{i_2}\right\} \quad \text{in } L^p.
\end{aligned}
$$

Since ϵ_t^2 and (u_{1t}, u_{2t}) are independent, we have $\mathrm{E}(1-\epsilon_t^2)u_{2t}=0$ and $\mathrm{E}(2\epsilon_t^2-1)u_{1t}^{i_1}u_{1t}^{i_2}=\mathrm{E}u_{1t}^{i_1}u_{1t}^{i_2}$. □

Lemma A.5. Ω_1 *is positive-definite.*

Proof. In the following, the notations $h_{1t}=(h_{1t}^1,\ldots,h_{1t}^{p+q})$ and $u_{1t}=(u_{1t}^1,\ldots,u_{1t}^{p+q})$ are used. Let λ be a $p+q$-dimensional non-random constant vector such that $\lambda u_{1t}=0$. We need to show that λ must be zero. Consider the recursive relationship for h_t and its derivatives,

$$h_t=\omega+\beta_1h_{t-1}+\cdots+\beta_ph_{t-p}+\alpha_1X_{t-1}^2+\cdots+\alpha_qX_{t-q}^2.$$

For $\theta^i=\beta_\mu$,

$$\partial^ih_t=\beta_1\partial^ih_{t-1}+\cdots+\beta_p\partial^ih_{t-p}+h_{t-\mu}.$$

For $\theta^i=\alpha_\mu$,

$$\partial^ih_t=\beta_1\partial^ih_{t-1}+\cdots+\beta_p\partial^ih_{t-p}+X_{t-\mu}^2.$$

Then,

$$
\begin{aligned}
\lambda h_{1t} \quad = \quad & \beta_1\lambda h_{1,t-1}+\cdots+\beta_p\lambda h_{1,t-p} \\
& +\frac{1}{h_t}\left\{\lambda_1h_{t-1}+\cdots+\lambda_ph_{t-p}+\lambda_{p+1}X_{t-1}^2+\cdots+\lambda_{p+q}X_{t-q}^2\right\}.
\end{aligned}
$$

Let $P_{t,k}=\prod_{j=1}^k\eta_{t-j+1,2}$. Applying Lemma A.16, 3.1, A.2, and using the assumption that $\lambda u_{1t}=0$, we have

$$\lambda_1P_{t-1,1}+\cdots+\lambda_pP_{t-1,p}+\lambda_{p+1}\epsilon_{t-1}^2P_{t-1,1}+\cdots+\lambda_{p+q}\epsilon_{t-q}^2P_{t-1,q}=0.$$

On the other hand, we have

$$\beta_1P_{t-1,1}+\cdots+\beta_pP_{t-1,p}+\alpha_1\epsilon_{t-1}^2P_{t-1,1}+\cdots+\alpha_q\epsilon_{t-q}^2P_{t-1,q}=1.$$

Let $\beta_1^*=\beta_1+\lambda_1,\ldots,\beta_p^*=\beta_p+\lambda_p$ and $\alpha_1^*=\alpha_1+\lambda_{p+1},\ldots,\alpha_q^*=\alpha_q+\lambda_{p+q}$. Define

$$\psi_k=\sum_{j=1}^{\min(q,t)}\alpha_jB_{11}^{t-j} \text{ and } \psi_k^*=\sum_{j=1}^{\min(q,t)}\alpha_j^*(B_{11}^*)^{t-j},$$

where B^* is the matrix B formed by the parameters (α^*,β^*). It can be shown that

$$\sum_{j=1}^{\infty}\psi_j\epsilon_{t-j}^2P_{t-1,j}=1 \text{ and } \sum_{j=1}^{\infty}\psi_j^*\epsilon_{t-j}^2P_{t-1,j}=1.$$

What remains are similar to the arguments in Berkes et al. [3]. If there is a positive integer m such that $\psi_m \neq \psi_m^*$ and for all $0 < i < m$, $\psi_i = \psi_i^*$, then

$$\left[\epsilon_{t-m}^2 - \frac{1}{\psi_m^* - \psi_m} \sum_{j=1}^{\infty} (\psi_{m+j} - \psi_{m+j}^*) P_{t-m-1,j}\right] P_{t-1,m} = 0.$$

Since the first and second terms in the square bracket are independent, the distribution of ϵ_{t-m}^2 must be degenerate, which is impossible under our assumption. Thus we must have $\psi_j = \psi_j^*$ for all $j = 1, 2, \ldots$. Within the radius of convergence,

$$\psi(z) = \frac{\alpha(z)}{1 - \beta(z)} = \frac{\alpha^*(z)}{1 - \beta^*(z)} = \psi^*(z),$$

by the assumption that $\alpha(z)$ and $1 - \beta(z)$ are co-prime, $\alpha(z) = \alpha^*(z)$ and $\beta(z) = \beta^*(z)$. That is, $\lambda = 0$. $\square$

A.3. Condition C3

This subsection is devoted to bounding the quantities $h_t(\theta^0)/h_t(\theta)$, $h_{1t}^i(\theta)$, $h_{2t}^{i_1 i_2}(\theta)$, and $h_{3t}^{i_1 i_2 i_3}(\theta)$ that appear in Equations (A.1)–(A.3). The results are given in Lemmas A.8 to A.11. It should be noted that the conditions $\omega = \omega^0$ and $H_0 = (\sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{-p+1}^2)$ are never used in this subsection, and so, the results given here are applicable to proving Theorem 2.1, too.

The neighborhood $N(\theta^0)$ is chosen as a rectangular region $\theta^L \ll \theta \ll \theta^U$ such that all components in θ^L are strictly positive. The notations $\theta^L = (\beta^L, \alpha^L)$ and $\theta^U = (\beta^U, \alpha^U)$ are used. Using Proposition 5.1 and Lemma 3.2, together with the continuity of $\mathrm{E}\frac{1}{(Q^{t,j})_{11}(\beta, \alpha^0)}$ with respect to β, if β^L is chosen enough close to β^U, then

$$\sup_{N(\theta^0)} \sum_{j=1}^{\infty} \frac{(B^j)_{11}(\beta^U)}{(Q^{t,j})_{11}(\beta^L, \alpha^0)}$$

converges almost surely and has a finite expectation.

For this selected neighborhood $N(\theta^0)$, we have the following two useful lemmas.

Lemma A.6.

$$\frac{h_t(\beta, \alpha^0)}{h_t(\beta, \alpha)} < \kappa^U \ \ \textit{and} \ \ \frac{h_t(\beta, \alpha^0)}{h_t(\beta, \alpha)} > \kappa^L$$

for some positive constants κ^U and κ^L.

Lemma A.7. *For $(\beta^1, \alpha), (\beta^2, \alpha) \in N(\theta^0)$, with $\beta^1 \gg \beta^0 \gg \beta^2$, we have*

$$\mathrm{E}\left\{\frac{h_{t-j}(\beta^1, \alpha)}{h_t(\beta^2, \alpha)}\right\}^r \leq O\left(\mathrm{E}\left\{\frac{1}{(Q^{t,j})_{11}(\beta^2, \alpha^0)}\right\}^r\right).$$

Proof of Lemma A.6. The lemma holds by setting

$$\kappa^U = \sup_{N(\theta^0)} \left\{ 1, \frac{\omega^0}{\omega}, \frac{\alpha_1^0}{\alpha_1}, \frac{\alpha_2^0}{\alpha_2}, \ldots, \frac{\alpha_p^0}{\alpha_p} \right\} \tag{A.9}$$

and

$$\kappa^L = \inf_{N(\theta^0)} \left\{ 1, \frac{\omega^0}{\omega}, \frac{\alpha_1^0}{\alpha_1}, \frac{\alpha_2^0}{\alpha_2}, \ldots, \frac{\alpha_p^0}{\alpha_p} \right\}. \tag{A.10}$$

This can be directly checked by the expansion formula (A.5) of $h_t(\theta)$,

$$h_t(\theta) = e_1^T B^t H_0 + \sum_{j=1}^{t} (B^{j-1})_{11} \upsilon_{t-j}(\theta).$$

□

Proof of Lemma A.7. By Lemma A.6 and (A.12) of Lemma A.13,

$$\begin{aligned} \frac{h_{t-j}(\beta^1, \alpha)}{h_t(\beta^2, \alpha)} &= \frac{h_{t-j}(\beta^1, \alpha)}{h_{t-j}(\beta^1, \alpha^0)} \frac{h_t(\beta^2, \alpha^0)}{h_t(\beta^2, \alpha)} \frac{h_{t-j}(\beta^1, \alpha^0)}{h_{t-j}(\beta^2, \alpha^0)} \frac{h_{t-j}(\beta^2, \alpha^0)}{h_t(\beta^2, \alpha^0)} \\ &\leq \frac{\kappa^U}{\kappa^L} \frac{h_{t-j}(\beta^1, \alpha^0)}{h_{t-j}(\beta^2, \alpha^0)} \frac{1}{(Q^{t,j})_{11}(\beta^2, \alpha^0)}. \end{aligned}$$

Note that $\frac{h_{t-j}(\beta^1,\alpha^0)}{h_{t-j}(\beta^2,\alpha^0)}$ is independent of $(Q^{t,j})_{11}(\beta^2, \alpha^0)$. We need to prove that there is a stationary and ergodic process v_{0t} such that $\mathrm{E}\frac{h_t(\beta^1,\alpha^0)}{h_t(\beta^2,\alpha^0)} < \mathrm{E}v_{0t} < \infty$. In fact, if $\beta^2 \ll \beta^0$, by Lemma A.14, we have

$$\begin{aligned} \frac{h_t(\beta^1, \alpha^0)}{h_t(\beta^2, \alpha^0)} &\leq \frac{h_t(\beta^L, \alpha^0)}{h_t(\beta^U, \alpha^0)} \\ &= 1 + (\beta_1^U - \beta_1^L) \sum_{j=1}^{t} (B^{j-1})_{11}(\beta^U) \frac{h_{t-j}(\beta^L, \alpha^0)}{h_t(\beta^L, \alpha^0)} + \cdots \\ &\quad + (\beta_q^U - \beta_q^L) \sum_{j=1}^{t} (B^{j-1})_{11}(\beta^U) \frac{h_{t-j-q+1}(\beta^L, \alpha^0)}{h_t(\beta^L, \alpha^0)} \\ &\leq 1 + (\beta_1^U - \beta_1^L) \sum_{j=1}^{t} (B^{j-1})_{11}(\beta^U) \frac{1}{(Q^{t,j})_{11}(\beta^L, \alpha^0)} + \cdots \\ &\quad + (\beta_q^U - \beta_q^L) \sum_{j=1}^{t} (B^{j-1})_{11}(\beta^U) \frac{1}{(Q^{t,j+q-1})_{11}(\beta^L, \alpha^0)} \\ &\leq 1 + (\beta_1^U - \beta_1^L) \sum_{j=1}^{\infty} (B^{j-1})_{11}(\beta^U) \frac{1}{(Q^{t,j})_{11}(\beta^L, \alpha^0)} + \cdots \\ &\quad + (\beta_q^U - \beta_q^L) \sum_{j=1}^{\infty} (B^{j-1})_{11}(\beta^U) \frac{1}{(Q^{t,j+q-1})_{11}(\beta^L, \alpha^0)} \\ &= v_{0t}. \end{aligned}$$

□

In the remaining subsection, the bounds for $h_t(\theta^0)/h_t(\theta)$, $h^i_{1t}(\theta)$, $h^{i_1 i_2}_{2t}(\theta)$, and $h^{i_1 i_2 i_3}_{3t}(\theta)$ are given in Lemmas A.8 to A.11.

Lemma A.8. *There exists a stationary and ergodic process* $\{v_{0t}\}$ *such that*

$$\sup_{\theta \in N(\theta^0)} \left\{ \frac{h_t}{h_t(\theta)} \right\} \le v_{0t}$$

and the r*-th moment* $\mathrm{E} v^r_{0t} < \infty$ *for* $r = 1, 2, 3, 4$.

Proof. Let $\theta \in \Theta$ and partition the vector θ into β and α. Then

$$\frac{h_t}{h_t(\theta)} = \frac{h_t(\beta^0, \alpha^0)}{h_t(\beta, \alpha^0)} \frac{h_t(\beta, \alpha^0)}{h_t(\beta, \alpha)} .$$

We establish a bound for the right-hand side. By Lemma A.6,

$$\frac{h_t(\beta^0, \alpha^0)}{h_t(\beta, \alpha^0)} < \kappa^U ,$$

where κ^U is defined in (A.9), which is non-stochastic and does not depend on θ. Consider the quantity

$$\frac{h_t(\beta^0, \alpha^0)}{h_t(\beta, \alpha^0)}.$$

By (A.13) of Lemma A.13,

$$h_t(\beta, \alpha^0) \ge h_t(\beta^L, \alpha^0) .$$

Together with Lemma A.14 and (A.12) in Lemma A.13,

$$\begin{aligned}
\sup_\theta \frac{h_t}{h_t(\beta, \alpha^0)} \quad &\le \quad \frac{h_t}{h_t(\beta^L, \alpha^0)} \\
&= \quad 1 + (\beta_1 - \beta^L_1) \sum_{j=1}^{t} (B^{j-1})_{11} \frac{h_{t-j}(\beta^L, \alpha^0)}{h_t(\beta^L, \alpha^0)} \\
&\qquad + \cdots + (\beta_p - \beta^L_p) \sum_{j=1}^{t} (B^{j-1})_{11} \frac{h_{t-j-q+1}(\beta^L, \alpha^0)}{h_t(\beta^L, \alpha^0)} \\
&\le \quad 1 + (\beta_1 - \beta^L_1) \sum_{j=1}^{t} (B^{j-1})_{11} \frac{1}{(Q^{t,j})_{11}(\beta^L, \alpha^0)} \\
&\qquad + \cdots + (\beta_p - \beta^L_p) \sum_{j=1}^{t} (B^{j-1})_{11} \frac{1}{(Q^{t,j+q-1})_{11}(\beta^L, \alpha^0)} \\
&\le \quad 1 + (\beta_1 - \beta^L_1) \sum_{j=1}^{\infty} (B^{j-1})_{11} \frac{1}{(Q^{t,j})_{11}(\beta^L, \alpha^0)} \\
&\qquad + \cdots + (\beta_p - \beta^L_p) \sum_{j=1}^{\infty} (B^{j-1})_{11} \frac{1}{(Q^{t,j+q-1})_{11}(\beta^L, \alpha^0)} \\
&= \quad v_{0t} \, . \qquad\qquad \text{(A.11)}
\end{aligned}$$

The result for higher moments can be obtained by applying Minkoswki's Inequality to (A.11). □

Lemma A.9. *There exist a neighbourhood $N(\theta^0)$ and a stationary and ergodic processes $\{v_{1t}^i\}$ such that*

$$\sup_{\theta \in N(\theta^0)} h_{1t}^i(\theta) \leq v_{1t}^i,$$

and the r-th moment $\mathrm{E}v_{1t}^r < \infty$ *for* $r = 1, 2, 3, 4$.

Proof. Suppose that there exist θ^L and θ^U such that $\theta^L \ll \theta \ll \theta^U$ for all $\theta \in N(\theta^0)$. We consider the derivatives with respect to θ_i for $\theta_i = \alpha_\mu$ and $\theta_i = \beta_\mu$ respectively as follows.

Case (i): When $\theta^i = \beta_\mu$, the derivative becomes

$$\sum_{j=1}^{t} \frac{(B^{j-1})_{11} h_{t-j-\mu+1}(\beta, \alpha)}{h_t(\beta, \alpha)}.$$

By Lemma A.7,

$$\mathrm{E}\frac{h_{t-j-\mu+1}(\beta, \alpha)}{h_t(\beta, \alpha)} < \kappa \frac{1}{(Q^{t,j})_{11}(\beta, \alpha^0)},$$

thus, the derivative is bounded by

$$\kappa \sum_{j=1}^{\infty} \frac{(B^{j-1})_{11}(\beta^U)}{(Q^{t,j+\mu-1})_{11}(\beta^L, \alpha^0)} = v_{1t}^i,$$

which is almost surely convergent with finite expectation.

Case (ii): When $\theta^i = \alpha_\mu$, the derivative becomes

$$\sum_{j=1}^{t} \frac{(B^{j-1})_{11} \epsilon_{t-j-\mu+1}^2 h_{t-j-\mu+1}(\beta, \alpha)}{h_t(\beta, \alpha)},$$

which is bounded by

$$\kappa \sum_{j=1}^{\infty} \epsilon_{t-j-\mu+1}^2 \frac{(B^{j-1})_{11}(\beta^U)}{(Q^{t,j+\mu-1})_{11}(\beta^L, \alpha^0)} = v_{1t}^i.$$

□

Lemma A.10. *There exists a stationary and ergodic process $\{v_{2t}^{i_1 i_2}\}$ such that*

$$\sup_{\theta \in N(\theta^0)} h_{2t}^{i_1 i_2} \leq v_{2t}^{i_1 i_2},$$

and the r-th moment $\mathrm{E}(v_{2t}^{i_1 i_2})^r < \infty$ *for* $r = 1, 2, 3, 4$.

Proof. From equation (A.7) for the second derivatives, we only need to consider the term

$$\sum_{j=1}^{t} B^{j-1}(\partial^{i_1}B)(\partial^{i_2}H_{t-j}(\theta)).$$

Consider the case that $\theta_{i_1}=\beta_{\mu_1}$. We have

$$\begin{aligned}
& \sum_{j=1}^{t}(B^{j-1})_{11}\frac{\partial^{i_2}h_{t-j-\mu_1}(\theta)}{h_{t-j-\mu_1}(\theta)}\frac{h_{t-j-\mu_1}(\theta)}{h_t(\theta)} \\
\leq\ & \sum_{j=1}^{t}(B^{j-1})_{11}(\beta^U)\left\{\sup_{\theta\in N(\theta^0)}\frac{\partial^{i_2}h_{t-j-\mu_1}(\theta)}{h_{t-j-\mu_1}(\theta)}\right\}\left\{\frac{h_{t-j-\mu_1}(\beta^U,\alpha)}{h_t(\beta^L,\alpha)}\right\} \\
\leq\ & \sum_{j=1}^{\infty}(B^{j-1})_{11}(\beta^U)\frac{v^{i_2}_{1,t-j-\mu+1}}{(Q^{t,j+\mu-1})_{11}(\beta^L,\alpha^0)} \\
=\ & v^{i_1i_2}_{2t}\,.
\end{aligned}$$

□

A result similar to Lemma A.10 is stated below without proof.

Lemma A.11. *There exists a stationary and ergodic process* $\{v^{i_1i_2i_3}_{3t}\}$ *such that*

$$\sup_{\theta\in N(\theta^0)} h^{i_1i_2i_3}_{3t}\leq v^{i_1i_2i_3}_{3t},$$

and the r-th moment $\mathrm{E}(v^{i_1i_2i_3}_{3t})^r<\infty$ *for* $r=1,2,3,4$.

A.4. Proofs of Theorem 2.1

Consider the neighborhood $N(\theta^0)$ constructed in subsection 3.5. Theorem 2.1 can be shown using Theorem A.1. In Lemma A.12, results on the asymptotic properties of the differences

$$L_n(\theta,\omega,H_0)-L_n(\theta,\omega_0,H_0^0)\,,$$

where $H_0^0=(\sigma_0^2,\sigma_{-1}^2,\ldots,\sigma_{-p+1}^2)^T$, are given. Theorem 2.1 directly follows from Lemma A.12 and Theorem A.1.

Lemma A.12. *For all* $\omega>0$ *and* $H_0>0$, *the first order derivatives satisfy*

$$\sqrt{n}\left(\partial^i L_n(\theta^0,\omega,H_0)-\partial^i L_n(\theta^0,\omega^0,H_0^0)\right)\to^p 0\,,$$

the second order derivatives satisfy

$$\left(\partial^{i_1i_2}L_n(\theta^0,\omega,H_0)-\partial^{i_1i_2}L_n(\theta^0,\omega^0,H_0^0)\right)\to^p 0,$$

and the third order derivatives satisfy

$$\sup_{N(\theta^0)}\left(\partial^{i_1i_2i_3}L_n(\theta,\omega,H_0)-\partial^{i_1i_2i_3}L_n(\theta,\omega^0,H_0^0)\right)\to^p 0\,,$$

where θ_i are chosen from α and β.

Proof of Lemma A.12. Consider a $p+1$-dimensional close set U containing $\varphi^0 = (\omega^0, H_0^0)$ as an interior point and covering all $\varphi = (\omega, H_0)$ of interest. Below, the notations $h_t(\theta, \varphi)$ and $L_n(\theta, \varphi)$ are used. The lemma can be shown by using the mean-value theorem and the following convergence results,

$$\sup_{\varphi \in U} \sqrt{n} \partial^{i_0 i_1} L_n(\theta^0, \varphi) \to^p 0,$$

$$\sup_{\varphi \in U} |\partial^{i_0 i_1 i_2} L_n(\theta^0, \varphi)| \to^p 0,$$

$$\sup_{\varphi \in U, \theta \in N(\theta^0)} |\partial^{i_0 i_1 i_2 i_3} L_n(\theta^0, \varphi)| \to^p 0,$$

where the variable associated with the index i_0 is chosen from φ, while that associated with i_1, i_2, i_3 are chosen from (β, α).

The derivatives up to order three that appear in the above relations are given in Equations (A.1)–(A.3). The fourth order derivatives can be obtained by differentiating Equation (A.3). By using Lemma A.15, the convergence results hold if the following two conditions are satisfied.

1. $X_t^2 / h_t(\theta, \varphi)$ and the quantities $h_{kt}^{i_1 \ldots i_k}$ that relate to differentiations with respect to (α, β) only are bounded by some stationary and ergodic processes with finite unconditional moments.
2. The quantities $h_{kt}^{i_1 \ldots i_k}$ that involve differentiations with respect to φ decay almost surely at the rate $\leq O(t^k \mu^t)$ for some non-negative integer k and $0 < \mu < 1$.

The first condition is established in Lemmas A.8 to A.11. We now show that the second condition holds. By Equation (A.5), when $\theta^{i_0} = \omega$,

$$\partial^{i_0} H_t(\theta, \varphi) = \sum_{j=1}^{t} B^{t-j} e_1 = \begin{cases} O(\lambda^t) & \text{if } \lambda > 1, \\ O(1) & \text{if } \lambda < 1, \\ O(t) & \text{if } \lambda = 1. \end{cases}$$

where λ is the eigenvalue of B with the largest modulus. Similarly, when $\theta_{i_0} = h_{-\mu+1}$, we have

$$\partial^{i_0} H_t(\theta, \varphi) = B^{t-j} e_\mu = O(\lambda^t).$$

Note that

$$\frac{1}{h_t(\theta, \varphi)} = \frac{h_t(\theta^0, \varphi)}{h_t(\theta, \varphi)} \cdot \frac{1}{h_t(\theta^0, \varphi)}.$$

A bound for the first term on the right-hand side is given in Lemma A.8. For the second term, we have

$$\frac{1}{h_t(\theta^0, \varphi)} \leq \frac{1}{e_1^T A_{t-1} \cdots A_0 Y_{-1}}.$$

According to Proposition 4.1, for all $0 < \delta < \rho$,

$$h_t(\theta^0, \varphi) \geq O(e^{(\rho - \delta)t}).$$

If δ is chosen so that $O(e^{(\rho-\delta)t}) > O((Q^{t,t})_{11}(\theta^0))$, then,

$$\sup_{\varphi\in U,\theta\in N(\theta^0)} \frac{\partial^{i_0} h_t(\theta,\varphi)}{h_t(\theta,\varphi)} = O(\mu^t)$$

for some $0 < \mu < 1$. Using similar arguments as presented above, it is not difficult to show that we have in general,

$$\sup_{\varphi\in U,\theta\in N(\theta^0)} \frac{\partial^{i_0\dots i_k} h_t(\theta,\varphi)}{h_t(\theta,\varphi)} = O(t^{k'}\mu^t)$$

for some non-negative integer $k' \le k$ and $0 < \mu < 1$. $\square$

A.5. Technical Lemmae

Lemma A.13. *If $\beta \ll \beta^0$, for $t \ge 1$ and $j \le t$, the following inequality holds*

$$h_t(\beta,\alpha^0) \quad > \quad (Q^{t,j})_{11}(\beta,\alpha^0) h_{t-j}(\beta,\alpha^0) . \tag{A.12}$$

Let $\theta^1 = (\beta^1,\alpha)$ and $\theta^2 = (\beta^2,\alpha)$, here both vectors share the same parameters α and further assume that $\beta^1 \gg \beta^2$, then

$$h_t(\beta^1,\alpha) \quad \ge \quad h_t(\beta^2,\alpha) . \tag{A.13}$$

Proof. Inequality (A.13) follows from

$$\begin{aligned} h_t(\beta^1,\alpha) &= e_1^T B^t(\beta^1) H_0 + \sum_{j=1}^t (B^{j-1})_{11}(\beta^1) \upsilon_{t-j}(\alpha) \\ &\ge e_1^T B^t(\beta^2) H_0 + \sum_{j=1}^t (B^{j-1})_{11}(\beta^2) \upsilon_{t-j}(\alpha) \\ &= h_t(\beta^2,\alpha) . \end{aligned}$$

Using (A.4) and (A.13), we have

$$\begin{aligned} H_t(\beta,\alpha^0) &= BH_{t-1}(\beta,\alpha^0) + \upsilon_{t-1}(\alpha^0) e_1 \\ &\gg BH_{t-1}(\beta,\alpha^0) + \alpha_1^0 \epsilon_{t-1}^2 h_{t-1}(\beta^0,\alpha^0) e_1 \\ &\gg BH_{t-1}(\beta,\alpha^0) + \alpha_1^0 \epsilon_{t-1}^2 h_{t-1}(\beta,\alpha^0) e_1 \\ &= (B + \alpha_1^0 \epsilon_{t-1}^2 e_1 e_1^T) H_{t-1}(\beta,\alpha^0) \\ &\gg \cdots \\ &\gg Q^{t,j}(\beta,\alpha^0) H_{t-j}(\beta,\alpha^0) . \end{aligned}$$

By noting that all entries in the above inequality are positive,

$$h_t(\theta^L) > (Q^{t,j})_{11}(\theta^L) h_{t-j}(\beta,\alpha^0) .$$

$\square$

Lemma A.14. *Given two sets of parameters (β^1,α) and (β^2,α), the following expansion holds,*

$$H_t(\beta^1,\alpha)-H_t(\beta^2,\alpha)=\sum_{j=1}^{t}B^{j-1}(\beta^1)(B(\beta^1)-B(\beta^2))H_{t-j}(\beta^2,\alpha)\,.$$

In particular, the first element is given by

$$\begin{aligned}h_t(\beta^1,\alpha)-h_t(\beta^2,\alpha) &= (\beta^1_1-\beta^2_1)\sum_{j=1}^{t}B^{j-1}(\beta^1)h_{t-j}(\beta^2,\alpha)\\ &\quad +\cdots+(\beta^1_q-\beta^2_q)\sum_{j=1}^{t}B^{j-1}(\beta^1)h_{t-j-q+1}(\beta^2,\alpha)\,.\end{aligned}$$

Proof. The recursive relationships for $H_t(\beta^1,\alpha)$ and $H_t(\beta^2,\alpha)$ are given by

$$H_t(\beta^1,\alpha)=B(\beta^1)H_{t-1}(\beta^1,\alpha)+\upsilon_{t-1}(\alpha)$$

and

$$H_t(\beta^2,\alpha)=B(\beta^2)H_{t-1}(\beta^1,\alpha)+\upsilon_{t-1}(\alpha),$$

respectively. It follows that

$$\begin{aligned}&H_t(\beta^1,\alpha)-H_t(\beta^2,\alpha)\\ =\;& B(\beta^1)H_{t-1}(\beta^1,\alpha)-B(\beta^2)H_{t-1}(\beta^2,\alpha)\\ =\;& (B(\beta^1)-B(\beta^2))H_{t-1}(\beta^1,\alpha)+B(\beta^2)(H_{t-1}(\beta^1,\alpha)-H_{t-1}(\beta^2,\alpha))\\ =\;& \cdots\\ =\;& B^t(\beta^2)(H_0-H_0)+\sum_{j=1}^{t}B^{j-1}(\beta^1)(B(\beta^1)-B(\beta^2))H_{t-j}(\beta^2,\alpha)\\ =\;& \sum_{j=1}^{t}B^{j-1}(\beta^1)(B(\beta^1)-B(\beta^2))H_{t-j}(\beta^2,\alpha)\,.\end{aligned}$$

□

Lemma A.15. *If $\{a_n\}$ is a stationary ergodic process with finite unconditional expectation and $b_n\to 0$ almost surely, then*

$$\frac{1}{n}\sum_{t=1}^{n}a_tb_t\to 0\quad\textit{almost surely}.$$

Moreover, if $\sum_{t=1}^{n}tb_t$ converges almost surely, then

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}a_tb_t\to 0\quad\textit{almost surely.}$$

Proof. Using mean ergodic theorem, we have

$$\frac{1}{n}\sum_{t=1}^{n} a_t \to \mathrm{E}a_n \quad \text{almost surely}\,.$$

Simple mathematical analysis arguments yield the lemma. □

Lemma A.16. *Let η'_t be defined in Lemma 4.1 and $\eta''_t = P'_t x'/e_1^T P'_t x'$. For $p > 1$, define $\zeta'_t = e_2^T \eta'_t$ and $\zeta''_t = e_2^T \eta''_t$. For $p = 1$, define $\zeta'_t = \epsilon_t^{-2}\eta'_t$ and $\zeta''_t = \epsilon_t^{-2}\eta''_t$. We have*

(I) $A'_{t+1}\eta'_t = (\zeta'_{t+1})^{-1}\eta'_{t+1}$ and $A'_{t+1}\eta''_t = (\zeta''_{t+1})^{-1}\eta'_{t+1}$.

(II) $e_1^T\eta'_t = e_1^T\eta''_t = 1$.

(III) Define

$$K_t = 1 + \beta_1 + \frac{\beta_2}{\beta_1} + \cdots + \frac{\beta_p}{\beta_{p-1}} + \alpha_1\epsilon_t^2 + \frac{\alpha_2}{\alpha_1} + \cdots + \frac{\alpha_q}{\alpha_{q-1}}\,.$$

Then, $\zeta'_t \geq K_t^{-1}$ and $\zeta''_t \geq K_t^{-1}$.

(IV) For $2 \leq k \leq p$,

$$\begin{aligned} e_k^T\eta'_t &= \prod_{j=1}^{k-1}\zeta'_{t-j+1} \leq \beta_{k-1}^{-1}\,, \\ e_k^T\eta''_t &= \prod_{j=1}^{k-1}\zeta''_{t-j+1} \leq \beta_{k-1}^{-1}\,, \\ \frac{e_k^T\eta'_t}{e_k^T\eta''_t} &\leq \beta_1^{-(k-1)}\prod_{j=1}^{k-1}K_{t-j+1}\,. \end{aligned}$$

(IV) For $1 \leq k \leq q-1$,

$$\begin{aligned} e_{p+k}^T\eta'_t &= \epsilon_{t-k+1}^2\prod_{j=1}^{k}\zeta'_{t-j+1} \leq \alpha_k^{-1}\,, \\ e_{p+k}^T\eta''_t &= \epsilon_{t-k+1}^2\prod_{j=1}^{k}\zeta''_{t-j+1} \leq \alpha_k^{-1}\,, \\ \frac{e_k^T\eta'_t}{e_k^T\eta''_t} &\leq \alpha_1^{-k}\prod_{j=1}^{k}K_{t-j+1}\,. \end{aligned}$$

Proof. This is a consequence of Lemma 4.1. □

References

[1] Arnold, L., Crauel, H. and Eckmann, J.P. (1991). *Lyapunov exponents: Proceedings of a conference held in Oberwolfach, May 28-June 2, 1990. Lecture Notes in Mathematics* **1486**. Springer-Verlag, Berlin. MR1178942

[2] Berkes, L. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* **32**, 633–655. MR2060172

[3] Berkes, L., Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**, 201–228. MR1997027

[4] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52**, 115–127. MR1165646

[5] Cohen, J.E., Kesten, H. and Newman, C.M. (1986a). *Random Matrices and Their Applications.* American Mathematical Society, Rhode Island. MR0841077

[6] Cohen, J.E., Kesten, H. and Newman, C.M. (1986b). Oseledec's multiplicative ergodic theorem: a proof, *in Random matrices and their applications,* edited by Cohen, J.E., Kesten, H. and Newman, C.M., American Mathematical Society, 23–30.

[7] Jensen, S.T. and Rahbek, A. (2004a). Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**, 1203–1226. MR2101955

[8] Jensen, S. and Rahbek, A. (2004b). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* **72**, 641–646. MR2036735

[9] Krengal, U. (1985). *Ergodic Theorems.* W. de Gruyter, New York. MR0797411

[10] Ledrappier, F. (1984). Quelques proprietes des exposants caracteristiques, *in Ecole d'ete de probabilites de Saint-Flour XII - 1982,* edited by Hennequin, P.L., Springer, Berlin, 305–396. MR0876081

[11] LeJan, Y. (1991). A second order extension of Oseledets theorem, *in Lyapunov exponents: proceedings of a conference held in Oberwolfach, May 28-June 2, 1990. Lecture Notes in Mathematics* **1486**, Arnold, L., Crauel, H. and Eckmann, J.P. (eds.), 81–85. Springer-Verlag, Berlin. MR1178948

[12] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York. MR0762984

[13] Robinson, P.M. and Zaffaroni, P. (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models. *Annals of Statistics* **34,** 1049–1074. MR2278351

[14] Rudin, W. (1987). *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, New York. MR0924157

[15] Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models.* Springer, New York. MR2142271