Econometric Theory, **31**, 2015, 362–393.
doi:10.1017/S0266466614000334

ECONOMETRIC ANALYSIS OF VOLATILITY COMPONENT MODELS

FANGFANG WANG
University of Illinois at Chicago

ERIC GHYSELS
University of North Carolina at Chapel Hill

Volatility component models have received considerable attention recently, not only because of their ability to capture complex dynamics via a parsimonious parameter structure, but also because it is believed that they can handle well structural breaks or nonstationarities in asset price volatility. This paper revisits component volatility models from a statistical perspective and attempts to explore the stationarity of the underlying processes. There is a clear need for such an analysis, since any discussion about nonstationarity presumes we know when component models are stationary. As it turns out, this is not the case and the purpose of the paper is to rectify this. We also look into the sampling behavior of the maximum likelihood estimates of recently proposed volatility component models and establish their consistency and asymptotic normality.

1. INTRODUCTION

Asset price volatility is persistent and several models have been proposed to capture this salient stylized fact. ARCH-type models originated by Engle (1982) are the most popular. Yet, empirical evidence suggests that volatility dynamics is better described by component models. Engle and Lee (1999) introduced a volatility component model with additive long- and short-run components. Several others have proposed related two-factor volatility models, see e.g., Ding and Granger (1996), Alizadeh, Brandt, and Diebold (2002), Chernov, Gallant, Ghysels, and Tauchen (2003), Adrian and Rosenberg (2008), Engle, Ghysels, and Sohn (2013), and Duan (1997) among others.

The appeal of component models is their ability to capture complex dynamics via a parsimonious parameter structure. Yet, there is also another reason why component models are becoming more popular, and this is again motivated by empirical evidence. Several studies have reported evidence of so called structural breaks in asset price volatility, see for example, Andreou and Ghysels (2002), Berkes, Gombay, Horváth, and Kokoszka (2004), Chen and Gupta (1997),

The authors would like to thank the Editor, Associate Editor, and two referees for valuable comments which helped us improve the paper. The second author also acknowledges support of a Marie Curie FP7-PEOPLE-2010-IIF grant.
Address correspondence to Fangfang Wang, Department of Information and Decision Sciences, University of Illinois at Chicago, 601 South Morgan Street, Chicago, IL 60607-7124; e-mail: ffwang@uic.edu.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Horvath, Kokoszka, and Teyssière (2001), Horvath, Kokoszka, and Zhang (2006), Inclan and Tiao (1994), Kokoszka and Leipus (2000), and Kulperger and Yu (2005), among others. To address the nonstationarity in the data, it has been suggested that such breaks should be captured by the long-run component. Alternatively, locally stable GARCH models have been considered to handle nonstationarity - see e.g., Dahlhaus and Rao (2006).

The component model of Engle and Lee (1999) consists of two additive GARCH(1,1) components. One is identified as short-run (transitory) component, while the other is identified as long-run (trend) component. The component models that have been suggested recently are not of the additive ARCH-type, but instead have a multiplicative structure. The first to suggest a multiplicative component structure that accommodates nonstationarity is Engle and Rangel (2008), later extended by Engle, Ghysels, and Sohn (2013). These component models, also known as Spline-GARCH and GARCH-MIDAS, respectively, feature a multiplicative decomposition of conditional variance into a short-run (high-frequency) and long-run (low-frequency) components. The high-frequency volatility component in both models is driven by a so called unit variance GARCH(1,1) process since it mean-reverts to an unconditional variance equal to one. The low-frequency component picks up the nonstationarity. The difference between the two models is the specification of the low-frequency volatility. The Spline-GARCH model formulates the low-frequency volatility in a nonparametric manner so that the long-run variance is time varying. This makes the model much more flexible but at the cost of losing its mean-reverting property. Moreover, the long-run component of the Spline-GARCH is de facto a deterministic function of time. The GARCH-MIDAS model of Engle, Ghysels, and Sohn (2013) modified the dynamics of low-frequency volatility to be stochastic "by smoothing realized volatility in the spirit of MIDAS (mixed data sampling, see e.g., Ghysels, Santa-Clara, and Valkanov, 2004) filtering" so that it can incorporate directly data sampled at lower frequency (say, monthly or quarterly) than the asset returns (sampled at a daily basis).[1]

The economic implications of the aforementioned component models and their empirical application have been studied extensively by Engle and Lee (1999), Engle and Rangel (2008), and Engle, Ghysels, and Sohn (2013). However, the literature has not well covered the conditions that characterize nonstationarity issues of the components. This paper revisits component models from a probabilistic and statistical perspective and attempts to explore the stationarity of the underlying processes. There is a clear need for such an analysis, since any discussion about nonstationarity presumes we know when component models are stationary. As it turns out, this is not the case and the purpose of the paper is to rectify this.

The rest of the paper is organized as follows. Section 2 gives a brief overview of volatility component models. Section 3 explores the stationarity of the various component models. Model estimation and asymptotic analysis of quasi-maximum-likelihood estimators are discussed in Section 4. Section 5 provides concluding remarks. Proofs are collected in an Appendix.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

2. AN OVERVIEW OF COMPONENT MODELS

In this section, we will give a brief overview of volatility component models. Although most of our focus is on multiplicative component models, we start with filling a gap in the literature pertaining to additive component model, i.e., the Engle and Lee model. Denote by r_t the return on, say day t. Namely, let

$$r_t = \sqrt{h_t}\varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ is an independent and identically distributed (i.i.d.) sequence with 0 mean and unit variance, denoted by $IID(0,1)$. Engle and Lee (1999) extend the classic GARCH model by modeling the conditional volatility h_t as the sum of a so-called trend component and a transitory component. To be specific,

$$\begin{aligned} h_t &= \tau_t + g_t, \\ g_t &= \alpha(r_{t-1}^2 - \tau_{t-1}) + \beta g_{t-1}, \\ \tau_t &= \omega + \rho\tau_{t-1} + \phi(r_{t-1}^2 - h_{t-1}), \end{aligned} \tag{2}$$

where the parameters satisfy the following conditions

Assumption 2.1. The parameters α, β, ω, ρ, and ϕ in model (2) are positive, and $\alpha + \beta < \rho < 1$, $\phi < \beta$.

Assumption 2.1 guarantees that h_t is nonnegative, and the mean reversion rate of g_t—which is $\alpha + \beta$—is slower than that of τ_t. Therefore, τ_t is referred to as the trend component and g_t as the transitory component.

Instead of modeling conditional volatility as the sum of two components, the spline-GARCH model of Engle and Rangel (2008) and the GARCH-MIDAS model of Engle, Ghysels, and Sohn (2013) structure conditional volatility as the product of long- and short-run components. The analysis in this paper focuses on the latter specification. For r_t defined in (1), the conditional variance is characterized as:

$$\begin{aligned} h_t &= g_t\tau_t, \\ g_t &= (1-\alpha-\beta) + \alpha(r_{t-1}^2/\tau_{t-1}) + \beta g_{t-1}, \\ \tau_t &= m + \theta\textstyle\sum_{k=1}^{K}\varphi_k(\omega)RV_{t-k}, RV_t = \sum_{j=0}^{N-1} r_{t-j}^2, \end{aligned} \tag{3}$$

where the short-run component g_t is a unit variance GARCH(1,1) process. The long-run component τ_t is stochastic and driven by past realized volatilities (henceforth RV). Restrictions are imposed on the parameters $(\alpha, \beta, \theta, m, \omega)$.[2] In particular, they satisfy the following conditions:

Assumption 2.2. The parameters α, β, θ and m are positive, with $\alpha + \beta < 1$, and the weights $\varphi_k(\omega)$ are a nonnegative function of ω for all k with $\sum_{k=1}^{K}\varphi_k(\omega) = 1$.

The GARCH-MIDAS model of (3) is also referred to as GARCH-MIDAS model with rolling window realized volatility (RV) in Engle, Ghysels, and

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Sohn (2013). The weights $\{\varphi_k(\omega), k = 1, \ldots, K\}$ are predetermined in model (3), and ω could be a scalar or a vector. At first we will work in Section 3 with weighting schemes involving potentially several parameters. In Section 4, dealing with estimation issues, we will specialize the analysis to a scalar parameter ω. This observation prompts a discussion about the weighting schemes used in (3).

The most commonly used parameterizations for φ_k, as suggested in Engle, Ghysels, and Sohn (2013), are

$$\varphi_k(\omega) = \frac{e^{\omega_1 k + \omega_2 k^2}}{\sum_{j=1}^{K} e^{\omega_1 j + \omega_2 k^2}} \quad \text{Exponential Almon weight,} \tag{4}$$

$$\varphi_k(\omega) = \frac{(k/K)^{\omega_1 - 1}(1 - k/K)^{\omega_2 - 1}}{\sum_{j=1}^{K}(j/K)^{\omega_1 - 1}(1 - j/K)^{\omega_2 - 1}} \quad \text{Beta weight,} \tag{5}$$

$$\varphi_k(\omega) = \frac{\omega^k}{\sum_{j=1}^{K} \omega^j} \quad \text{Exponential weight,} \tag{6}$$

where the first two polynomial specifications involve two parameters $\omega \equiv (\omega_1, \omega_2)$, whereas the third and last is a function of a scalar parameter. However, both the Exponential Almon and Beta weight specifications can be restricted to a single-parameter settings. For the Exponential Almon this means ω_2 is zero and for the Beta weights it means that we set $\omega_1 = 1$. Such single-parameter constrained cases are often used in practice—see Ghysels, Sinko, and Valkanov (2006) and Ghysels (2013). This means that we still cover a fairly flexible class of weighting schemes when we turn our attention to asymptotic properties of estimators in Section 4. Obviously, specifications involving more than two parameters can be considered as well. Our analysis pertains to any weighting scheme which is a known function of a finite set of parameters.

The specification of the τ_t component for GARCH-MIDAS builds on a long tradition, going back to Merton (1980), Schwert (1989), and others, of measuring long-run volatility by realized volatility over a monthly or quarterly horizon. In a GARCH-MIDAS model, however, one does not view the realized volatility of a single quarter or month as the measure of interest. Instead, the τ_t component is specified via smoothing historical realized volatilities in the spirit of MIDAS regression and MIDAS filtering (as in respectively Ghysels, Santa-Clara, and Valkanov, 2006 and Ghysels, Santa-Clara, and Valkanov, 2005).

In this paper, we will revisit the component models from a probabilistic and statistical perspective. In particular, we attempt to explore stationarity of the underlying time series by examining their top Lyapunov exponents. To that end we close this section with some notation. Let $x^+ = \max(x, 0)$ while $x^- = \max(-x, 0)$. Denote by I_N a $N \times N$ identify matrix. We say the matrix $A = (a_{ij})_{n\times n}$ is nonnegative (or $A \geq 0$) if $a_{ij} \geq 0$ for any i, j, and A is positive (or $A > 0$) if $a_{ij} > 0$ for any i, j. For matrices A and B, $A \leq B$ if $B - A \geq 0$. The spectral radius of matrix A is denoted as $\rho(A)$ and $\rho(A) = \max_i(|\lambda_i|)$ where $\lambda_i \in \mathbb{C}$ is the eigenvalue of A. We consider the following norm in this paper:

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

$\|V\| = \max_{i=1,2,\ldots,n} |v_j|$ for $V = (v_1, \ldots, v_n)' \in \mathbb{R}^n$, and the induced matrix norm is $\|M\| = \max_{i=1,2,\ldots,n} \sum_{j=1}^{n} |m_{ij}|$ for $M = (m_{ij})_{n\times n} \in \mathbb{R}^{n\times n}$. Given a set U, $\overline{U}$ denotes its closure and U^0 its interior.

3. STATIONARITY

This section explores stationarity of the component model of Engle and Lee and the GARCH-MIDAS models with rolling-window RV. A standard approach to study stationarity of ARCH-type processes, following Bougerol and Picard (1992a), is to characterize the process via a stochastic difference equation with a Markovian representation. In particular, suppose that Y_t is R^d-valued random vector ($d \geq 1$) which satisfies the following stochastic difference equation

$$Y_t = A_t Y_{t-1} + B_t, \quad t \in \mathbb{Z}, \tag{7}$$

where A_t is a $R^{d\times d}$-valued random matrix and B_t is R^d-valued random vector, and $\{(A_t, B_t)\}$ are strictly stationary and ergodic.[3] If the model in (7) is stable, it will converge to its stationary solution. The stability of model (7) closely relates to $\gamma(A)$, the top Lyapunov exponent associated with the sequence $\{A_t\}$. We can define the top Lyapunov exponent associated with $\{A_t\}$ as $\gamma(A) = \inf_{t\in\mathbb{N}} E(\frac{1}{t}\log\|A_t A_{t-1}\ldots A_1\|)$ provided that $E\log^+\|A_0\| < \infty$. A more tractable expression for $\gamma(A)$, due to Furstenberg and Kesten (1960) and Kingman (1973), is

$$\gamma(A) = \lim_{t\to\infty} \frac{1}{t} E\log\|A_t A_{t-1}\ldots A_1\| \overset{a.s.}{=} \lim_{t\to\infty} \frac{1}{t}\log\|A_t A_{t-1}\ldots A_1\|. \tag{8}$$

Suppose further that $E\log^+\|B_0\| < \infty$. Bougerol and Picard (1992b) show that $\gamma(A) < 0$ is a *necessary and sufficient* (N&S) condition under which (7) has a unique stationary ergodic solution *provided that the model is irreducible and* $\{(A_t, B_t)\}$ *are i.i.d.*[4] For more general settings where $\{(A_t, B_t)\}$ are strictly stationary and ergodic, however, $\gamma(A) < 0$ is only a *sufficient* condition, as shown by Glasserman and Yao (1995).

Models (2) and (3) are similar in spirit to (7), as is typically the case for ARCH-type models. Yet, the structure of the matrices involved is different. We therefore need to elaborate.

The calculation of $\gamma(A)$ requires an analysis of the limiting behavior of the product of random matrices. Their study goes back to Furstenberg and Kesten (1960) and Furstenberg (1963) (see Goldsheid, 1991 for a comprehensive review). Kesten and Spitzer (1984) studied the convergence of the product of i.i.d. nonnegative matrices. Cohen and Newman (1984) considered $A_t A_{t-1}\ldots A_1$ when $\{A_t\}$ are i.i.d. and the entries of A_t are symmetric stable random variables, whereas Newman (1986) extended it to the case where the entries of A_t are normally distributed. Peres (1992) examined a product of positive matrices with Markovian dependence (see also Yao, 2001). An explicit formula for $\gamma(A)$, however, is not available in general. Therefore, most of the discussion on chaotic behavior using

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Lyapunov exponents relies on simulation, see for instance, Lu and Smith (1997), Whang and Linton (1999), and Vanneste (2010), among others.

In the first subsection it will be shown that the $\{A_t\}$ associated with model (2) are i.i.d. but have negative entries, while for model (3), $\{A_t\}$ are strictly stationary ergodic nonnegative matrices, as explained in the next subsection. We will examine the associated top Lyapunov exponents in this section and give explicit necessary and sufficient conditions for $\gamma(A) < 0$. Moreover, we will show that $\gamma(A) < 0$ is also a necessary condition for the stationarity of Model (3).

3.1. Component Model of Engle and Lee

The conditional variance h_t in Model (2) consists of two additive GARCH(1,1) components, and hence the additive component model has a structure of GARCH(2,2)

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}, \tag{9}$$

where $\alpha_0 = \omega(1-\alpha-\beta)$, $\alpha_1 = \phi + \alpha$, $\alpha_2 = -(\phi(\alpha+\beta)+\alpha\rho)$, $\beta_1 = \rho+\beta-\phi$, and $\beta_2 = \phi(\alpha+\beta) - \rho\beta$ (see Engle and Lee, 1999). It is worth noting that the coefficients are not nonnegative. Under Assumption 2.1, $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_2 < 0$, $\beta_1 > 0$, and $\beta_2 < 0$.[5] The stationarity of GARCH models with nonnegative parameters has been well studied. See for instance, Nelson (1990), Bougerol and Picard (1992a), Chen and An (1998), Carrasco and Chen (2002), Francq and Zakoian (2006, 2009), Meitz and Saikkonen (2008), Lindner (2009), and Kristensen (2009), among others. However, the standard results regarding the stationarity cannot carry over to Model (2) directly.

Engle and Lee (1999) pointed out that r_t defined in (2) is weakly stationary if $\rho < 1$ and $\alpha + \beta < 1$.[6] We will investigate the strict stationarity in this subsection. To do so, define $Y_t = (h_{t+1}, h_t, r_t^2)'$, $B = (\alpha_0, 0, 0)'$, and

$$A_t \equiv A(\varepsilon_t^2) = \begin{pmatrix} \beta_1 + \alpha_1\varepsilon_t^2 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ \varepsilon_t^2 & 0 & 0 \end{pmatrix}. \tag{10}$$

Therefore, (r_t, h_t) is strictly stationary ergodic if and only if $Y_t = A_t Y_{t-1} + B$ has a unique strictly stationary ergodic solution. Let $\Phi(Z) = 1 - \beta_1 Z - \beta_2 Z^2$ and $\Theta(Z) = \alpha_1 + \alpha_2 Z$. Note that Y_t is irreducible if ε_t has a continuous component at 0, and $\Phi(Z)$ and $\Theta(Z)$ have no common roots and all the roots to $\Phi(Z)$ lie outside the unit circle (see for instance, Kristensen, 2009, p. 130).[7] We therefore have the following

Assumption 3.1. ε_t has a continuous component at 0.

PROPOSITION 3.1. *Under Assumptions 2.1 and 3.1, Y_t is irreducible.*

Denote by $\gamma(A)$ the top Lyapunov exponent associated with $\{A(\varepsilon_t^2)\}$. Since the A_t's are i.i.d. Y_t is strictly stationary ergodic if and only if $\gamma(A) < 0$. The next

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

proposition provides a sufficient condition for $\gamma(A) < 0$. Note that if A_t's were nonnegative, then $\gamma(A) \leq \log \rho(EA_1)$ (see Kesten and Spitzer, 1984). Therefore, in Proposition 3.2, we try to find a 'mirror image' of A_t and then bound $\gamma(A)$ by the logarithm of the spectral radius of the 'image'. We also need to following technical Assumption:

Assumption 3.2. The distribution of ε_t has unbounded support and $P(\varepsilon_t = 0) = 0$.

PROPOSITION 3.2. *Suppose that Assumptions 2.1, 3.1, and 3.2 hold. Further, assume that* $(1+\rho)(1+\alpha+\beta) \leq 2$, *then* $\gamma(A) < 0$, *and hence* Y_t *(and* r_t *defined in (2) as well) is strictly stationary. Moreover, under the additional assumption that* ε_t *is absolutely continuous with strictly positive Lebesgue density in a neighborhood of zero,* r_t *is* β*-mixing.*

3.2. The Class of GARCH-MIDAS Models

We next consider the GARCH-MIDAS with rolling window RV, i.e., model (3). Note that τ_t can be written as $\tau_t = m + \theta \sum_{l=1}^{N+K-1} c_l r_{t-l}^2$, where c_l's are combinations of the weights $\varphi_k(\omega)$ and satisfy $\sum_{l=1}^{N+K-1} c_l = N \sum_{k=1}^{K} \varphi_k(\omega) = N$. Let $\epsilon_t = \sqrt{g_t}\varepsilon_t$. Then $r_t = \sqrt{\tau_t}\epsilon_t$ can be viewed as a semistrong ARCH$(N+K-1)$ with multiplicative GARCH error. The GARCH error is strictly stationary and ergodic under Assumption 2.2. This is because $g_t = (1-\alpha-\beta) + (\alpha\varepsilon_{t-1}^2 + \beta)g_{t-1}$ is of the form (7) with i.i.d. coefficients—i.e., $A_t = \alpha\varepsilon_{t-1}^2 + \beta$ and $B_t = (1-\alpha-\beta)$, and the top Lyapunov exponent associated with $\{A_t\}$ is $\gamma(A) = E\log|A_1| \leq \log(\alpha+\beta) < 0$ due to the Jensen's inequality.

Consider the following stochastic difference equation

$$Y_t = A_t(\vec{c})Y_{t-1} + B_t, \tag{11}$$

where $Y_t = (r_t^2, r_{t-1}^2, \ldots, r_{t-N-K+2}^2)'$, $B_t = (mg_t\varepsilon_t^2, 0, \ldots, 0)'$, $\vec{c} = (c_1, c_2, \ldots, c_{N+K-1})'$, and

$$A_t(\vec{c}) = \begin{pmatrix} \theta g_t\varepsilon_t^2 c_1 & \ldots & \theta g_t\varepsilon_t^2 c_{N+K-2} & \theta g_t\varepsilon_t^2 c_{N+K-1} \\ 1 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & 0 \\ \vdots & & & \\ 0 & \ldots & 1 & 0. \end{pmatrix}. \tag{12}$$

Stationarity of the process in (3) is therefore equivalent to the stability of dynamic system (11). Let $\gamma(A)$ be the top Lyapunov exponent defined on $\{A_t(\vec{c})\}$. Note that $\{A_t(\vec{c})\}$ is strictly stationary ergodic. Equation (11) has a unique stationary ergodic solution if $\gamma(A) < 0$. We will show next that $\gamma(A) < 0$ is also a *necessary*

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

condition. Define $S_{t,n} = A_t(\vec{c})A_{t-1}(\vec{c})\dots A_{t-n+1}(\vec{c})$ for $n > 0$ and $S_{t,0} = 1$. Then

$$Y_t = S_{t,k}Y_{t-k} + \sum_{n=0}^{k-1} S_{t,n}B_{t-n}, \tag{13}$$

for $k = 1, 2, \dots$. Further, define matrices $H_{N+K-1}, G_{N+K-1} \in \mathbb{R}^{(N+K-1)\times(N+K-1)}$ as

$$H_{N+K-1} = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad G_{N+K-1} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \tag{14}$$

Then $A_t(\vec{c}) = \theta g_t \varepsilon_t^2 H_{N+K-1} D(\vec{c}) + G_{N+K-1}$, where $D(\vec{c}) = Diag(c_1, \dots, c_{N+K-1})$.

PROPOSITION 3.3. *Suppose that Assumption 2.2 holds. Model (3) has a unique strictly stationary ergodic solution if and only if* $\gamma(A) < 0$.

Note that the GARCH-MIDAS model (3) is also equivalent to the following representation

$$X_t = \check{A}_t X_{t-1} + B, \tag{15}$$

where $X_t = (\tau_{t+1}, r_t^2, \dots, r_{t-N-K+3}^2)'$, $B = (m, 0, \dots, 0)'$, and

$$\check{A}_t = \begin{pmatrix} \theta c_1 g_t \varepsilon_t^2 & \theta c_2 & \dots & \theta c_{N+K-2} & \theta c_{N+K-1} \\ g_t \varepsilon_t^2 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0. \end{pmatrix}. \tag{16}$$

The results stated in Proposition 3.3 apply to (15) as well with top Lyapunov exponent defined on $\{\check{A}_t\}$—using a similar argument. However, since it is easier to examine the top Lyapunov exponent defined on (12), the representation (15) will not be considered here.

Next we look for tractable conditions - both sufficient and necessary - for $\gamma(A) < 0$. We consider three cases: (1) $K = 1, N = 1$, (2) $K = 1, N > 1$, and (3) $K > 1, N \geq 1$.

When $K = N = 1$, we have $\gamma(A) = E\log(\theta g_0 \varepsilon_0^2)$ and therefore,

PROPOSITION 3.4. *Suppose that Assumption 2.2 holds. For* $K = N = 1$, *model (3) has a unique strictly stationary ergodic solution if* $\theta \leq 1$.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

This is due to strict concavity of log function and the Jensen's inequality. When $K = 1$ and $N > 1$, the weights vanish and $A_t(\vec{c})$ is simply

$$A_t(\iota) = \begin{pmatrix} \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 & \dots & \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \tag{17}$$

where ι represents a vector of 1's. Let $M(a) = aH_N + G_N$ for $a > 0$. Hence $A_t(\iota) = M(\theta g_t \varepsilon_t^2)$. It is easy to verify that $M(a)$ is a primitive matrix and hence is irreducible and aperiodic. It follows from the Perron–Frobenius theory that $\rho(M(a))$ is the maximal positive root of $f_a(\lambda) = det(\lambda I_N - M(a))$, and it is simple. Since $A_t(\iota)$ is closely related to the top Lyapunov exponent, we first look at a few properties regarding the spectral radius of $M(a)$.

LEMMA 3.1. *Consider matrix $M(a)$ with $a > 0$.*

1. *For any $k > 0$, $\rho(M(a)) \le k$ if and only if $f_a(k) \ge 0$.*
2. *The map $a \mapsto \rho(M(a))$ is nondecreasing and concave.*

Next, we present a sufficient condition for $\gamma(A) < 0$ when $K = 1$ and $N > 1$. Let $\zeta = (1-\alpha-\beta)/(1-\beta)$.

PROPOSITION 3.5. *Suppose that Assumptions 2.2 and 3.2 hold. For $K = 1$ and $N > 1$, the top Lyapunov exponent associated with $A_t(\iota)$ is negative if $\theta \le \zeta^{N-1}/(1+\zeta+\cdots+\zeta^{N-1})$.*

Proposition 3.5 can be extended to a more general situation.

PROPOSITION 3.6. *Suppose that Assumptions 2.2 and 3.2 hold. For $K > 1$ and $N \ge 1$, the top Lyapunov exponent associated with $A_t(\vec{c})$ is negative if $\theta \le \zeta^{K+N-2}/(c_1\zeta^{K+N-2} + c_2\zeta^{K+N-3} + \cdots + c_{K+N-2}\zeta + c_{K+N-1})$.*

Therefore, combining Propositions 3.4, 3.5, and 3.6, we have the following

PROPOSITION 3.7. *Suppose that Assumptions 2.2 and 3.2 hold. Model (3) has a unique strictly stationary ergodic solution if*

$$\theta \le \zeta^{K+N-2}/(c_1\zeta^{K+N-2} + c_2\zeta^{K+N-3} + \cdots + c_{K+N-2}\zeta + c_{K+N-1}), \tag{18}$$

where $\zeta = (1-\alpha-\beta)/(1-\beta)$. The solution is nonanticipative (or causal). Therefore, both r_t and τ_t are strictly stationary ergodic.

If Assumption 3.2 is removed, the sufficient condition should be that θ is strictly less than the right-hand side of (18). Proposition 3.7 presents a sufficient condition for $\gamma(A) < 0$. We next give the necessary condition.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

PROPOSITION 3.8. *Suppose that Assumption 2.2 holds. Model (3) has a unique strictly stationary ergodic solution only if* $\theta < 1/(N\zeta)$.

Note that under the conditions stated in Proposition 3.7, the unique strictly stationary ergodic solution to equation (11) is

$$Y_t = \sum_{n=0}^{\infty} S_{t,n} B_{t-n}, \quad \text{with probability one } (w.p.1). \tag{19}$$

This solution, however, is not integrable, and hence r_t is not weakly stationary. This is formally stated in the following proposition.

PROPOSITION 3.9. *Consider* r_t *defined by model (3), and suppose that Assumptions 2.2 and 3.2 hold. If the system starts from the infinite past, then* $Er_t^2 = \infty$ *for any* t. *If the system starts from time 0, then there exists* $t_0 \geq 1$ *such that* $Er_t^2 = \infty$ *for* $t > t_0$.

It follows that the marginal distribution of the stationary version of r_t is heavy-tailed. This is a desirable property in modeling financial time series. Although the strictly stationary ergodic solution is not integrable, we will show next that it is 'fractionally' integrable. We first present a general result.

PROPOSITION 3.10. *Consider the stochastic difference equation (7), and* $\{(A_t, B_t)\}$ *are strictly stationary ergodic. Suppose that* $E\|A_0\|^{\delta} < \infty$, *and* $E\|B_0\|^{\delta} < \infty$ *for some* $\delta > 0$ *and the top Lyapunov exponent* $\gamma(A) < 0$. *Then there exists* $0 < \delta^* < \min(\delta, 1)$ *such that* $E(\|Y_t\|^{\delta^*}) < \infty$ *for any* t, *and hence* $E(\log\|Y_t\|) < \infty$.

Therefore, as a corollary we have

COROLLARY 3.1. *Consider* r_t *defined by model (3), and suppose that Assumptions 2.2 and 3.2 hold. There exists* $0 < \delta^* < 1$ *such that* $E(r_t^2)^{\delta^*} < \infty$ *and hence* $E(\log r_t^2) < \infty$.

Corollary 3.1 is of importance in statistical inference, especially in studying the asymptotic behavior of quasi-maximum-likelihood estimators. Most of the discussion in Section 4 will rely on Corollary 3.1.

4. QMLE OF THE GARCH-MIDAS MODEL

The quasi-maximum-likelihood estimation (QMLE) of conditionally heteroscedastic time series models has been discussed by Engle (1982), Bollerslev (1986), Weiss (1986), Bollerslev and Wooldridge (1992), Lee and Hansen (1994), Lumsdaine (1996), Berkes, Horváth, and Kokoszka (2003), Francq and Zakoian (2004), and Straumann and Mikosch (2006), among many others. Recall that the component model of Engle and Lee can be viewed as a

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

GARCH (2,2). The QMLE asymptotic analysis of Engle (1982) and Bollerslev (1986) therefore straightforwardly applies.[8]

Unfortunately, the existing asymptotic results do not directly carry over to the GARCH-MIDAS class of models. In this section, we cover the asymptotic analysis of the quasi-maximum-likelihood estimator for the GARCH-MIDAS model with rolling-window RV.

Let $\mathcal{U}$ be the parameter space which will be specified later. Define g_t and τ_t as

$$g_t(\Phi) = (1-\alpha-\beta) + \alpha \frac{r_{t-1}^2}{\tau_{t-1}(\Phi)} + \beta g_{t-1}(\Phi), \tag{20}$$

$$\tau_t(\Phi) = m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, \tag{21}$$

for $\Phi = (\alpha, \beta, m, \theta, \omega) \in \mathcal{U}$ and $t \in \mathbb{Z}$. Suppose that $\Phi_0 = (\alpha_0, \beta_0, m_0, \theta_0, \omega_0) \in \mathcal{U}$ is the true parameter such that

$$r_t = \sqrt{g_t(\Phi_0)\tau_t(\Phi_0)}\varepsilon_t, \quad t \in \mathbb{Z}, \tag{22}$$

where $\varepsilon_t \sim IID(0,1)$ and ε_t^2 has a nondegenerate distribution. Given a finite record of r_t: $\{r_t, 1 \le t \le T\}$ where $T \gg N+K$, define

$$\begin{aligned} \tilde{g}_t(\Phi) &= (1-\alpha-\beta) + \alpha \tfrac{r_{t-1}^2}{\tilde{\tau}_{t-1}} + \beta \tilde{g}_{t-1}(\Phi), \\ \tilde{\tau}_t(\Phi) &= m + \theta \textstyle\sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, \end{aligned} \tag{23}$$

for $t = N+K+1, \ldots, T$, and $\tilde{g}_{N+K}$ is an arbitrary number. The QMLE of Φ_0 is the minimizer of

$$\tilde{L}_T(\Phi) = \frac{1}{T-N-K} \sum_{t=N+K+1}^{T} \log \tilde{V}_t(\Phi) + \frac{r_t^2}{\tilde{V}_t(\Phi)}, \tag{24}$$

for $\Phi \in \mathcal{U}$, where $\tilde{V}_t(\Phi) = \tilde{g}_t(\Phi)\tilde{\tau}_t(\Phi)$. The estimator is denoted by $\tilde{\Phi}_T$. Consider a companion estimator, $\hat{\Phi}_T$, which minimizes

$$L_T(\Phi) = \frac{1}{T-(N+K)} \sum_{t=N+K+1}^{T} \log V_t(\Phi) + \frac{r_t^2}{V_t(\Phi)} \tag{25}$$

over $\Phi \in \mathcal{U}$ where $V_t(\Phi) = g_t(\Phi)\tau_t(\Phi)$. Clearly, $\tau_t(\Phi) = \tilde{\tau}_t(\Phi)$. Although $\hat{\Phi}_T$ is not feasible, it is theoretically tractable and $\hat{\Phi}_T - \tilde{\Phi}_T = o_p(1/\sqrt{T})$ (see the proof of Proposition 4.2).

The properties of $\tilde{\Phi}_T$ and $\hat{\Phi}_T$ are closely related to the choice of parameter space $\mathcal{U}$. Suppose that the parameter ω is of dimension d, i.e., $\omega = (\omega_1, \ldots, \omega_d)'$. The weights are parameterized by ω. We rule out the degenerate case that $\varphi_k \equiv 0$ for some $k \in \{1, 2, \ldots, K\}$. We also assume the weights $\varphi_k(\omega)$ satisfy the following conditions.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Assumption 4.1. There exists an open set $\Omega \subset \mathbb{R}^d$ such that for $k \in \{1, 2, \ldots, K-1\}$ and $\omega \in \Omega$, $\varphi_k(\omega) \geq 0$, $\varphi_1(\omega) + \cdots + \varphi_{K-1}(\omega) \leq 1$ and the second-order partial derivative $\partial^2 \varphi_k(\omega)/\partial \omega_j^2$ exists and is continuous on Ω, and the $(K-1)$-by-d matrix $\left(\partial \varphi_k(\omega)/\partial \omega_j\right)_{(K-1)\times d}$ has rank greater than or equal to d.

Remark 4.1. Because $\varphi_K = 1 - \varphi_1 - \cdots - \varphi_{K-1}$, under Assumption 4.1 we have $\varphi_k(\omega) = \varphi_k(\omega_0)$ if and only if $\omega = \omega_0$ for all k. For easy exposition we restrict our attention to the case where ω is a scalar for the rest of analysis, i.e., $d = 1$.

Note that Proposition 3.7 holds when $\beta = 0$ and/or $\theta = 0$, and $\theta = 0$ as it yields the regular GARCH(1,1). However, for estimation purpose we need to exclude such cases as it makes the weighting function unidentifiable. We will therefore require the true parameter $\theta_0 > 0$. Define

$$\mathcal{U}_1 = \Bigg\{ \Phi = (\alpha, \beta, m, \theta, \omega)' \in \mathbb{R}^5 : \alpha > 0, \beta \geq 0, \alpha + \beta < 1, m > 0, \\ 0 < \theta \leq \frac{\zeta^{K+N-2}}{c_1 \zeta^{K+N-2} + c_2 \zeta^{K+N-3} + \cdots + c_{K+N-2}\zeta + c_{K+N-1}}, \omega \in \Omega \Bigg\}, \tag{26}$$

where $\zeta = (1 - \alpha - \beta)/(1 - \beta)$. Moreover, define

$$\mathcal{U}_2 = \{\Phi = (\alpha, \beta, m, \theta, \omega)' \in \mathbb{R}^5 : \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1, m > 0, \theta > 0, \omega \in \Omega\}. \tag{27}$$

Clearly, $\mathcal{U}_2$ contains $\mathcal{U}_1$. The process r_t^2 is strictly stationary ergodic, if $\Phi_0 \in \mathcal{U}_1$ and Assumption 3.2 holds. The parameter space will satisfy the following assumption:

Assumption 4.2. The parameter space $\mathcal{U}$ is a compact subset of $\mathcal{U}_2$. The true parameter Φ_0 is in both $\mathcal{U}$ and $\mathcal{U}_1$, i.e., $\Phi_0 \in \mathcal{U} \cap \mathcal{U}_1$.

Consider $l_t(\Phi) = \log V_t(\Phi) + r_t^2 / V_t(\Phi)$.

LEMMA 4.1. *Under Assumptions 3.2, 4.2, and 4.1,* $E \sup_{\Phi \in \mathcal{U}} |\log(g_t(\Phi))|$ *and* $E \sup_{\Phi \in \mathcal{U}} |\log(\tau_t(\Phi))|$ *are finite, and hence* $E \sup_{\Phi \in \mathcal{U}} l_t^-(\Phi) < \infty$.[9]

It follows that under Assumption 4.2, $El_t(\Phi)$ is well defined for $\Phi \in \mathcal{U}$. Moreover, $E|l_t(\Phi_0)| < \infty$ but $E \sup_{\Phi \in \mathcal{U}} |l_t(\Phi)| = \infty$. Next we will show that Φ_0 is identifiably unique. In other words, $El_t(\Phi)$ is uniquely minimized at Φ_0.

LEMMA 4.2. *Under Assumptions 3.2, 4.2, and 4.1,* $El_t(\Phi) > El_t(\Phi_0)$ *for* $\Phi \in \mathcal{U}$ *and* $\Phi \neq \Phi_0$.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Because $El_t(\Phi)$ may not be finite for $\Phi \in \mathcal{U}$, the uniform Strong Law of Large Numbers (SLLN) does not apply. We need the following lemma for the proof of strong consistency of $\hat{\Phi}_T$ and $\tilde{\Phi}_T$.

LEMMA 4.3. *Suppose that Assumptions 3.2, 4.2, and 4.1 hold. For any compact set $C \subset \mathcal{U}$, then* $\liminf_{T\to\infty} \inf_{\Phi\in C} L_T(\Phi) \geq \inf_{\Phi\in C} El_t(\Phi)$ *almost surely (a.s.).*

The proof is skipped. The reader is referred to Lemma 3.11 of Pfanzagl (1969). The next lemma shows that the asymptotic property of $\hat{\Phi}_T$ is independent of the starting value.

LEMMA 4.4. *Under Assumptions 3.2 and 4.2,*

$$\lim_{T\to\infty} \sup_{\Phi\in\mathcal{U}} |L_T(\Phi) - \tilde{L}_T(\Phi)| = 0 \quad \textit{almost surely.} \tag{28}$$

Therefore, we have the following

PROPOSITION 4.1. *Under Assumptions 3.2, 4.2, and 4.1, $\hat{\Phi}_T$ and $\tilde{\Phi}_T$ converge to Φ_0 almost surely.*

Remark 4.2. Note that if $\varphi_k(\omega) > 0$ for $\omega \in \Omega$ and $k = 1, 2, \ldots, K$, then

$$\frac{\tau_t(\Phi_0)}{\tau_t} = \frac{m_0}{\tau_t} + \sum_{k=1}^{K} \frac{\theta_0\varphi_k(\omega_0) R V_{t-k}}{\tau_t} \leq \frac{m_0}{m} + \sum_{k=1}^{K} \frac{\theta_0\varphi_k(\omega_0)}{\theta\varphi_k(\omega)}. \tag{29}$$

We would have $E \sup_{\Phi\in\mathcal{U}} V_t^{-1} r_t^2 < \infty$ and $E \sup_{\Phi\in\mathcal{U}} |l_t(\Phi)| < \infty$. Therefore, the proof of Proposition 4.1 will be a direct application of the uniform SLLN (see, for instance, Theorem 3.3 of Gallant and White, 1988).

Further, we will show that $\tilde{\Phi}_T$ is also asymptotic normal, with an additional assumption. Let $\varphi = (\varphi_1, \ldots, \varphi_{K-1})$ and $\mathcal{V} = \{(x_1, \ldots, x_{K-1}) : x_1 > 0, \ldots, x_{K-1} > 0, x_1 + \cdots + x_{K-1} < 1\}$.

Assumption 4.3. $E(\varepsilon_t^4) < \infty$. The true parameter Φ_0 is an interior point of $\mathcal{U}$, i.e., $\Phi_0 \in \mathcal{U}^0$, and $\varphi(\omega_0) \in \mathcal{V}$.

By continuity of φ_k, there exists an open ball O_{ω_0} such that $\omega_0 \in O_{\omega_0} \subset \overline{O_{\omega_0}} \subset \Omega$ and $\varphi(O_{\omega_0}) \subset \mathcal{V}$. Therefore, one can always find a compact set $\mathcal{A}$ such that $\Phi_0 \in \mathcal{A} \subset \mathcal{U}^0$ and for $\Phi = (\alpha, \beta, m, \theta, \omega)' \in \mathcal{A}$, $\varphi(\omega) \in \mathcal{V}$. Define $\Sigma(\Phi) = E\left(V_t^{-2}(\Phi)\nabla V_t(\Phi)\nabla V_t(\Phi)'\right)$ for $\Phi \in \mathcal{A}$.[10]

LEMMA 4.5. *Suppose that Assumptions 3.2, 4.2, 4.1, and 4.3 hold. Then*

1. $\Sigma(\Phi)$ *exists and it is positive definite at* Φ_0.
2. $\sqrt{T}\nabla L_T(\Phi_0) \Longrightarrow N(0, (E\varepsilon_t^4 - 1)\Sigma(\Phi_0))$.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

3. *Let* $B(\Phi_0, 1/n) = \{\Phi \in \mathbb{R}^5 : \|\Phi - \Phi_0\| < 1/n\}$.

$$\limsup_{n\to\infty} \lim_{T\to\infty} \sup_{\Phi \in \overline{B(\Phi_0,1/n)} \cap \mathcal{A}} \|H(L_T)(\Phi) - \Sigma(\Phi_0)\| = 0 \quad a.s. \tag{30}$$

4. $\lim_{T\to\infty} \sqrt{T} \sup_{\Phi\in\mathcal{A}} \|\nabla L_T(\Phi) - \nabla \tilde{L}_T(\Phi)\| = 0$ *in probability.*

Lemma 4.5 allows us to establish the asymptotic normality of $\tilde{\Phi}_T$.

PROPOSITION 4.2. *Under Assumptions 3.2, 4.2, 4.1, and 4.3,*

$$\sqrt{T}(\tilde{\Phi}_T - \Phi_0) \Longrightarrow N(0, (E\varepsilon_t^4 - 1)\Sigma(\Phi_0)^{-1}).$$

Remark 4.3. Note that $g_t(\Phi_0) = (1 - \alpha_0 - \beta_0) + (\alpha_0 \varepsilon_{t-1}^2 + \beta_0) g_{t-1}(\Phi_0)$, and $\tau_t(\Phi_0) = m_0 + \theta_0 \sum_{k=1}^{K} \varphi_k(\omega_0) RV_{t-k}$. Denote by $\partial_\alpha, \partial_\beta, \partial_m, \partial_\theta$, and ∂_ω the partial derivatives with respect to α, β, m, θ, and ω, respectively. We have $\partial_m g_t(\Phi_0) = \partial_\theta g_t(\Phi_0) = \partial_\omega g_t(\Phi_0) = \partial_\alpha \tau_t(\Phi_0) = \partial_\beta \tau_t(\Phi_0) = 0$, and thus

$$\nabla g_t(\Phi_0) = (\partial_\alpha g_t(\Phi_0), \partial_\beta g_t(\Phi_0), 0, 0, 0)',$$
$$\nabla \tau_t(\Phi_0) = (0, 0, \partial_m \tau_t(\Phi_0), \partial_\theta \tau_t(\Phi_0), \partial_\omega \tau_t(\Phi_0))'.$$

Note also that

$$\begin{aligned} V_t^{-2}(\Phi_0) \nabla V_t(\Phi_0) \nabla V_t(\Phi_0)' &= \frac{\nabla g_t(\Phi_0) \nabla g_t(\Phi_0)'}{g_t(\Phi_0)^2} + \frac{\nabla \tau_t(\Phi_0) \nabla \tau_t(\Phi_0)'}{\tau_t(\Phi_0)^2} + \frac{\nabla \tau_t(\Phi_0) \nabla g_t(\Phi_0)'}{\tau_t(\Phi_0) g_t(\Phi_0)} \\ &\quad + \frac{\nabla g_t(\Phi_0) \nabla \tau_t(\Phi_0)'}{\tau_t(\Phi_0) g_t(\Phi_0)}. \end{aligned}$$

It follows that $\Sigma(\Phi_0) = E\left(V_t^{-2}(\Phi_0) \nabla V_t(\Phi_0) \nabla V_t(\Phi_0)'\right)$ is a block diagonal matrix, and $\Sigma(\Phi_0) = Diag(J_1, J_2)$ where $J_1 = E\left(g_t^{-2}(\Phi_0) dg_t dg_t'\right)$ and $J_2 = E\left(\tau_t^{-2}(\Phi_0) d\tau_t d\tau_t'\right)$, and $dg_t = (\partial_\alpha g_t(\Phi_0), \partial_\beta g_t(\Phi_0))'$ and $d\tau_t = (\partial_m \tau_t(\Phi_0), \partial_\theta \tau_t(\Phi_0), \partial_\omega \tau_t(\Phi_0))'$. The asymptotic variance becomes $(E\varepsilon_t^4 - 1)$ $Diag(J_1^{-1}, J_2^{-1})$. For the MLE $\tilde{\Phi}_T = (\tilde{\alpha}_T, \tilde{\beta}_T, \tilde{m}_T, \tilde{\theta}_T, \tilde{\omega}_T)'$, $(\tilde{\alpha}_T, \tilde{\beta}_T)'$, and $(\tilde{m}_T, \tilde{\theta}_T, \tilde{\omega}_T)'$ are asymptotically independent.

Remark 4.4. It should be noted that the results of QMLE can be easily extended to ω being multidimensional. The proofs of Lemmas 4.1–4.4 and Proposition 4.1 apply to a vector ω without modification. For Lemma 4.5 and Proposition 4.2, we just need to replace the (partial) derivative with respect to ω by the partial derivative with respect to each component of ω, and the discussion follows directly.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

5. CONCLUSION

This paper revisits the component models from a probabilistic and statistical perspective. Stationarity of two models is investigated: the component model of Engle and Lee (1999), and the GARCH-MIDAS models of Engle, Ghysels, and Sohn (2013). By examining the associated top Lyapunov exponents, we present explicit conditions on the space of parameters—both sufficient and necessary—for stationarity and ergodicity. This is necessary, because it allows one to easily track stationarity of the underlying process with given parameters and to manage model inference. A by-product regarding the analysis of Lyapunov exponents is that $\gamma(A) < 0$ is also a necessary condition for stationarity and ergodicity of Model (3).

We also show that the GARCH-MIDAS model has fat-tailed marginal distribution, a desirable property in modeling financial time series. We then study sampling behavior of the quasi-maximum-likelihood estimator of the model. The consistency and asymptotic normality of the QMLE are established.

NOTES

1. Applications of Spline-GARCH and GARCH-MIDAS include Baele, Bekaert, and Inghelbrecht (2010), Bauwens, Hafner, and Pierret (2013), Chauvet, Senyuz, and Yoldas (2010), Chen, Choudhry, and Wu (2013), Dorion (2010), Golosnoy, Gribisch, and Liesenfeld (2012), Hafner and Linton (2010), Asgharian, Hou, and Javed (2013), Maheu and McCurdy (2007), and Rangel (2011), among others.

2. We use generic parameter settings across all the models to avoid a proliferation of parameter notation. Therefore, it is important to note that the parameters in the models (2) and (3) are not related. For example, the ω in (2) is different from the ω appearing in (3).

3. The stationarity of Y_t of the form (7) has been studied extensively in the literature, see for instance, Pham (1985, 1986), Brandt (1986), Bougerol (1987), Meyn and Caines (1991), Bougerol and Picard (1992a, 1992b), and Glasserman and Yao (1995), among others.

4. Model (7) is said to be irreducible if no proper affine subspace $\mathcal{H} \in \mathbb{R}^d$ exists such that $\{A_t y + B_t : y \in \mathcal{H}\} \subseteq \mathcal{H}$. For details on irreducibility, the reader is referred to for example Meyn and Caines (1991), Bougerol and Picard (1992b), and Kristensen (2009).

5. It should be noted that the Jacobian determinant of the map from $(\omega, \alpha, \beta, \phi, \rho)$ to $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$ is $-(1-\alpha-\beta)(\alpha+\beta-\rho)^2$, and it is not zero under Assumption 2.1. Therefore, $(\omega, \alpha, \beta, \phi, \rho)$ can be identified from $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$.

6. See Engle and Lee (1999, p. 479).

7. A random variable X is said to have a continuous component at 0, if there exists a function $f > 0$ and a constant $\delta > 0$ such that $P(X \in \mathcal{A}) \geq \int_{\mathcal{A}\cap(-\delta,\delta)} f(x)dx$ for any Borel set $\mathcal{A} \subseteq \mathbb{R}$.

8. See the discussion in Engle and Lee (1999) for details.

9. Where $l_t^-(\Phi) = \max(-l_t(\Phi), 0)$.

10. We use ∇ to denote the vector differential operator so that ∇f is the gradient (column vector) of scalar function f, and $H(f)$ the Hessian matrix of f.

11. This notation is used only in the proof of Lemma 4.2. For a *set* U, U^0 represents the interior of U.

REFERENCES

Adrian, T. & J. Rosenberg (2008) Stock returns and volatility: Pricing the short-run and long-run components of market risk. *Journal of Finance* 63, 2997–3030.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Asgharian, H., A. J. Hou, & F. Javed (2013) The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS Approach, *Journal of Forecasting*, 32(7), 600–612.

Alizadeh, S., M. Brandt, & F. Diebold (2002) Range-based estimation of stochastic volatility models. *Journal of Finance* 57, 1047–1091.

Andreou, E. & E. Ghysels (2002) Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17, 579–600.

Baele, L., G. Bekaert, & K. Inghelbrecht (2010) The determinants of stock and bond return comovements. *Review of Financial Studies* 23, 2374–2428.

Basrak, B., R. Davis, & T. Mikosch (2001) Sample ACF of multivariate stochastic recurrence equations with application to GARCH. Preprint, Available at www. math. ku. dk/mikosch.

Bauwens, L., C. M. Hafner, & D. Pierret (2013) Multivariate volatility modeling of electricity futures, *Journal of Applied Econometrics*, 28(5), 743–761.

Berkes, I., E. Gombay, L. Horváth, & P. Kokoszka (2004) Sequential change-point detection in GARCH(p,q) models. *Econometric Theory* 20, 1140–1167.

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

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

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

Bougerol, P. (1987) Tightness of products of random matrices and stability of linear stochastic systems. *Annals of Probability* 15, 40–74.

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

Bougerol, P. & N. Picard (1992b) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1730.

Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18, 211–220.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chauvet, M., Z. Senyuz, & E. Yoldas (2010) What Does Realized Volatility Tell Us About Macroeconomic Fluctuations? Discussion paper, Working paper, University of California, Riverside.

Chen, M. & H. An (1998) A note on the stationarity and the existence of moments of the GARCH model. *Statistica Sinica* 8, 505–510.

Chen, W.-P., T. Choudhry, & C.-C. Wu (2013) The extreme value in crude oil and US dollar markets. *Journal of International Money and Finance* 36, 191–210.

Chen, J. & A. Gupta (1997) Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association* 92, 739–747.

Chernov, M., A. Gallant, E. Ghysels, & G. Tauchen (2003) Alternative models for stock price dynamics. *Journal of Econometrics* 116, 225–257.

Cohen, J. & C. Newman (1984) The stability of large random matrices and their products. *Annals of Probability* 12, 283–310.

Dahlhaus, R. & S. Rao (2006) Statistical inference for time-varying ARCH processes. *Annals of Statistics* 34, 1075–1114.

Ding, Z. & C. Granger (1996) Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics* 73, 185–215.

Dorion, C. (2010) Business Conditions, Market Volatility and Option Prices. Discussion paper, HEC Montreal Working paper.

Duan, J. (1997) Augmented GARCH (p,q) process and its diffusion limit. *Journal of Econometrics* 79, 97–127.

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

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Engle, R. & G. Lee (1999) A long-run and short-run component model of stock return volatility. In R. Engle & H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, pp. 475–497. Oxford University Press.

Engle, R. & J. Rangel (2008) The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21, 1187–1222.

Engle, R. F., E. Ghysels, & B. Sohn (2013) Stock market volatility and macroeconomic fundamentals, *Review of Economics and Statistics*, 95(3), 776–797.

Francq, C. & J. Zakoian (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. & J. Zakoian (2006) Mixing properties of a general class of GARCH (1, 1) models without moment assumptions on the observed process. *Econometric Theory* 22, 815–834.

Francq, C. & J. Zakoian (2009) A tour in the asymptotic theory of GARCH estimation. *Handbook of Financial Time Series*, pp. 85–111. Springer.

Francq, C. & J. Zakoian (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Furstenberg, H. (1963) Noncommuting random products. *Transactions of the American Mathematical Society* 108, 377–428.

Furstenberg, H. & H. Kesten (1960) Products of random matrices. *Annals of Mathematical Statistics* 31, 457–469.

Gallant, A. & H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell.

Ghysels, E. (2013) Matlab toolbox for MIDAS regressions. Available at `http://www.unc.edu/~eghysels`.

Ghysels, E., P. Santa-Clara, & R. Valkanov (2004) *The MIDAS Touch: Mixed Data Sampling Regression Models*. CIRANO.

Ghysels, E., P. Santa-Clara, & R. Valkanov (2005) There is a risk-return trade-off after all. *Journal of Financial Economics* 76, 509–548.

Ghysels, E., P. Santa-Clara, & R. Valkanov (2006) Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131, 59–95.

Ghysels, E., A. Sinko, & R. Valkanov (2006) MIDAS regressions: Further results and new directions. *Econometric Reviews* 26, 53–90.

Glasserman, P. & D. Yao (1995) Stochastic vector difference equations with stationary coefficients. *Journal of Applied Probability* 32, 851–866.

Goldsheid, I. (1991) Lyapunov exponents and asymptotic behaviour of the product of random matrices. *Lyapunov Exponents*, pp. 23–37. Springer.

Golosnoy, V., B. Gribisch, & R. Liesenfeld (2012) The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics* 167, 211–223.

Hafner, C. & O. Linton (2010) Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics* 159, 55–73.

He, C. & T. Terasvirta (1999) Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92, 173–192.

Horvath, L., P. Kokoszka, & G. Teyssière (2001) Empirical process of the squared residuals of an ARCH sequence. *Annals of Statistics* 29, 445–469.

Horvath, L., P. Kokoszka, & A. Zhang (2006) Monitoring constancy of variance in conditionally heteroskedastic time series. *Econometric Theory* 22, 373–402.

Inclan, C. & G. Tiao (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* 89, 913–923.

Kesten, H. & F. Spitzer (1984) Convergence in distribution of products of random matrices. *Probability Theory and Related Fields* 67, 363–386.

Kingman, J. (1973) Subadditive ergodic theory. *Annals of Probability* 1, 883–899.

Kokoszka, P. & R. Leipus (2000) Change-point estimation in ARCH models. *Bernoulli* 6, 513–540.

Kristensen, D. (2009) On stationarity and ergodicity of the bilinear model with applications to GARCH models. *Journal of Time Series Analysis* 30, 125–144.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Kulperger, R. & H. Yu (2005) High moment partial sum processes of residuals in GARCH models and their applications. *Annals of Statistics* 33, 2395–2422.

Lee, S. & B. Hansen (1994) Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–29.

Lindner, A. (2009) Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes. *Handbook of Financial Time Series*, pp. 43–69. Springer.

Lu, Z. & R. Smith (1997) Estimating local Lyapunov exponents. *Fields Institute Communications* 11, 135–151.

Lumsdaine, R. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1, 1) and covariance stationary GARCH (1, 1) models. *Econometrica* 64, 575–596.

Maheu, J. & T. McCurdy (2007) Components of market risk and return. *Journal of Financial Econometrics* 5, 560–590.

Meitz, M. & P. Saikkonen (2008) Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory* 24, 1291–1320.

Merton, R.C. (1980) On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323–361.

Meyn, S.P. & P.E. Caines (1991) Asymptotic behavior of stochastic systems possessing Markovian realizations. *SIAM Journal on Control and Optimization* 29, 535–561.

Mokkadem, A. (1990) Propriétés de mélange des processus autorégressifs polynomiaux. *Annales de l'Institut Henri Poincaré (B) Probabilités et statistiques* 26, 219–260.

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

Newman, C. (1986) The distribution of Lyapunov exponents: Exact results for random matrices. *Communications in Mathematical Physics* 103, 121–126.

Peres, Y. (1992) Domains of analytic continuation for the top Lyapunov exponent. *Annales de l'Institut Henri Poincaré (B) Probabilités et statistiques* 28, 131–148.

Pfanzagl, J. (1969) On the measurability and consistency of minimum contrast estimates. *Metrika* 14, 249–272.

Pham, D. (1985) Bilinear markovian representation and bilinear models. *Stochastic Processes and their Applications* 20, 295–306.

Pham, D. (1986) The mixing property of bilinear and generalised random coefficient autoregressive models. *Stochastic Processes and their Applications* 23, 291–300.

Rangel, J. (2011) Macroeconomic news, announcements, and stock market jump intensity dynamics. *Journal of Banking and Finance* 35, 1263–1276.

Schwert, G.W. (1989) Why does stock market volatility change over time? *Journal of Finance* 44, 1207–1239.

Seneta, E. (1981) *Non-negative Matrices and Markov Chains*. Springer Verlag.

Straumann, D. & T. Mikosch (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34, 2449–2495.

Vanneste, J. (2010) Estimating generalized Lyapunov exponents for products of random matrices. *Physical Review E* 81, 036701.

Weiss, A. (1986) Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2, 107–131.

Whang, Y. & O. Linton (1999) The asymptotic distribution of nonparametric estimates of the Lyapunov exponent for stochastic time series. *Journal of Econometrics* 91, 1–42.

Yao, J. (2001) On square-integrability of an AR process with Markov switching. *Statistics and Probability Letters* 52, 265–270.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

APPENDIX

Proof of Proposition 3.1. Note that $\Phi(Z)$ has roots outside the unit circle (see Engle and Lee, 1999). Moreover, since

$$\Phi(-\alpha_1/\alpha_2) = 1+\beta_1\alpha_1/\alpha_2-\beta_2\alpha_1^2/\alpha_2^2 = -\frac{1}{\alpha_2^2}(\rho-\alpha-\beta)^2\alpha\phi \neq 0,$$

if and only if $\rho-\alpha-\beta$ and α and ϕ are not 0, then $\Phi(Z)$ and $\Theta(Z)$ have no common roots, and hence Y_t is irreducible. ■

Proof of Proposition 3.2. We only need to show $\gamma(A) < 0$. Let

$$M_t = A(\varepsilon_t^2)A(\varepsilon_{t-1}^2)\dots A(\varepsilon_1^2), \qquad \tilde{M}_t = \tilde{A}(\varepsilon_t^2)\tilde{A}(\varepsilon_{t-1}^2)\dots\tilde{A}(\varepsilon_1^2),$$

where

$$\tilde{A}(x) = \begin{pmatrix} \beta_1+\alpha_1 x & -\beta_2 & -\alpha_2 \\ 1 & 0 & 0 \\ x & 0 & 0 \end{pmatrix}. \tag{A.1}$$

Then $E\tilde{A}(\varepsilon_t^2) = \tilde{A}(1)$. By induction $\|M_t\| \leq \|\tilde{M}_t\|$ for any t. Note that $\tilde{A}(\varepsilon_t^2)$ is not bounded almost surely and i.i.d. We then have $\lim_{t\to\infty}\frac{1}{t}\log\|\tilde{M}_t\| < \log(\rho(\tilde{A}(1)))$ (see Kesten and Spitzer, 1984, Thm. 2). Because

$$\rho(\tilde{A}(1)) = \frac{\alpha+\beta+\rho+\sqrt{(\alpha+\beta+\rho)^2+4(\alpha+\beta)\rho}}{2},$$

$\rho(\tilde{A}(1)) \leq 1$ if and only if $(1+\rho)(1+\alpha+\beta) \leq 2$.

Note that the characteristic function of $A(0)$ is $det(\lambda I_3 - A(0)) = \lambda(\lambda^2-\beta_1\lambda-\beta_2)$. The spectral radius of $A(0)$ is therefore less than 1. Define $V(y) = |y_1|+a|y_2|+a|y_3|$ for $y = (y_1, y_2, y_3)' \in \mathbb{R}^3$, where $a = (1-(\alpha_1+\beta_1))/4 > 0$. Let $\pi = (1+\alpha_1+\beta_1)/2 < 1$, and $B > 0$ is such that $(\alpha_0+1)/B < 1-\pi$. It is easy to verify that $E[V(Y_t)|Y_{t-1}] \leq \alpha_0+\pi V(Y_{t-1})$. Define $K = \{k \in \mathbb{R}^3 : V(k) \leq B\}$. $E[V(Y_t)|Y_{t-1} = y]$ is bounded for $y \in K$. On K^c, $E[V(Y_t)|Y_{t-1} = y] \leq \alpha_0+\pi V(y) \leq ((\alpha_0+1)/B+\pi)V(y)-1$. Under the additional assumption that ε_t is absolutely continuous with strictly positive Lebesgue density in a neighborhood of zero, Y_t is geometrically ergodic and hence r_t is β-mixing, which follows from Mokkadem (1990) (or Theorem 1 of Carrasco and Chen, 2002; Theorem 8 of Lindner, 2009). ■

Proof of Proposition 3.3. Note that $1 \leq \|A_0(\vec{c})\| \leq \theta g_0\varepsilon_0^2\|H_{N+K-1}D(\vec{c})\| + \|G_{N+K-1}\| \leq N\theta g_0\varepsilon_0^2+1$. We have $E(\log\|A_0(\vec{c})\|) \leq E\log(N\theta g_0\varepsilon_0^2+1) \leq E(N\theta g_0\varepsilon_0^2) < \infty$.

Sufficiency follows from Theorem 3.1 of Glasserman and Yao (1995) (see also Bougerol and Picard, 1992b). The necessity is done through an argument similar to the proof of Theorem 1.3 of Bougerol and Picard (1992a).

Suppose that (11) has a unique strictly stationary ergodic solution. Let $t = 0$ in (13). Since $\sum_{n=0}^{k-1} S_{0,n}B_{-n} \leq Y_0$ almost surely for any k, then $\lim_{n\to\infty} S_{0,n}B_{-n} = 0$ almost surely. In other words,

$$\lim_{n\to\infty} \epsilon_{-n}^2 S_{0,n}e_1 = 0 \quad \text{almost surely},$$

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

where $\{e_1, e_2, \ldots, e_{N+K-1}\}$ is the canonical basis of R^{N+K-1}. Note that for $i = 1, 2, \ldots, N+K-2$,

$$\begin{aligned} S_{0,n}e_i &= S_{0,n-1}(\theta\epsilon_{-n+1}^2 H_{N+K-1}D(\vec{c}) + G_{N+K-1})e_i \\ &= \theta c_i \epsilon_{-n+1}^2 S_{0,n-1}e_1 + S_{0,n-1}e_{i+1}, \\ S_{0,n}e_{N+K-1} &= \theta c_{N+K-1}\epsilon_{-n+1}^2 S_{0,n-1}e_1. \end{aligned}$$

Therefore, $\lim_{n\to\infty} S_{0,n}e_i = 0$ almost surely for any i, and hence $\lim_{n\to\infty} S_{0,n} = 0$ almost surely. It follows from Lemma 3.4 of Bougerol and Picard (1992b) that $\gamma(A) < 0$. ■

Proof of Lemma 3.1. 1. Note that $f_a(\lambda) = \lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \cdots - a\lambda^2 - a\lambda - a$. For $|\lambda| > k$,

$$\begin{aligned} |f_a(\lambda)| &\geq |\lambda^N|(1 - \frac{a}{|\lambda|} - \frac{a}{|\lambda^2|} - \cdots - \frac{a}{|\lambda^N|}) \\ &> k^N(1 - \frac{a}{k} - \frac{a}{k^2} - \cdots - \frac{a}{k^N}) = f(k). \end{aligned}$$

Note that $\rho(M(a))$ is the largest positive root of $f_a(\lambda)$. Therefore, $\rho(M(a)) \leq k$ if and only if $f_a(k) \geq 0$.

2. Note that $\rho(M(a))$ is the maximal positive root of $f(\lambda) = det(\lambda I - M(a))$. It is simple and $\rho(M(a)) \geq |\lambda|$ for each root λ of $f(\lambda) = 0$. For easy exposition, we simply write $\rho(M(a))$ as λ. Since $f(\lambda) = \lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \cdots - a\lambda^2 - a\lambda - a = 0$,

$$a = \frac{\lambda^N}{\lambda^{N-1} + \lambda^{N-2} + \cdots + \lambda^2 + \lambda + 1} = \lambda - 1 + g(\lambda), \tag{A.2}$$

where $g(\lambda) = \frac{1}{h(\lambda)}$ and $h(\lambda) = \lambda^{N-1} + \lambda^{N-2} + \cdots + \lambda^2 + \lambda + 1$. Note that λ is a smooth function of a. To prove λ is a concave function of a is equivalent to show that $\frac{d^2\lambda(a)}{da^2} < 0$.

On one hand, taking derivative on both sides of (A.2) with respect to a, we have $1 = (1 + g')\lambda'$, where $g' = \frac{dg(\lambda)}{d\lambda}$ and $\lambda' = \frac{d\lambda(a)}{da}$. Furthermore, $0 = (1 + g')\lambda'' + g''(\lambda')^2$, where $g'' = \frac{d^2g(\lambda)}{d\lambda^2}$ and $\lambda'' = \frac{d^2\lambda(a)}{da^2}$. On the other hand, write $f(\lambda) = 0$ as $F(\lambda, a) = 0$. By implicit function theorem,

$$\lambda' = -\frac{F_a}{F_\lambda},$$

where $F_a = \frac{\partial F}{\partial a} = -h(\lambda) < 0$ and $F_\lambda > 0$ (since λ is the largest root of f and f goes to ∞ as λ goes to ∞ for fixed a). Hence $\lambda' > 0$ and $1 + g' > 0$.

To show $\lambda'' < 0$, it is sufficient to show that $g'' = \frac{2(h'(\lambda))^2 - h(\lambda)h''(\lambda)}{h^3(\lambda)} > 0$ or $\Delta = 2(h'(\lambda))^2 - h(\lambda)h''(\lambda) > 0$, where $h' = \frac{dh(\lambda)}{d\lambda}$ and $h'' = \frac{d^2h(\lambda)}{d\lambda^2}$. Note that

$$\begin{aligned} h(\lambda) &= \frac{\lambda^N - 1}{\lambda - 1}, \\ h'(\lambda) &= \frac{N\lambda^{N-1}}{\lambda - 1} - \frac{\lambda^N - 1}{(\lambda - 1)^2}, \\ h''(\lambda) &= \frac{N(N-1)\lambda^{N-2}}{\lambda - 1} - \frac{2N\lambda^{N-1}}{(\lambda - 1)^2} + \frac{2(\lambda^N - 1)}{(\lambda - 1)^3}. \end{aligned}$$

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Therefore,

$$\Delta = \frac{N\lambda^{N-2}[(N-1)\lambda^{N+1} - (N+1)\lambda^N + (N+1)\lambda - (N-1)]}{(\lambda-1)^3}. \tag{A.3}$$

Define

$$D(\lambda) = (N-1)\lambda^{N+1} - (N+1)\lambda^N + (N+1)\lambda - (N-1).$$

Then $D^{'}(\lambda) \doteq \frac{dD(\lambda)}{d\lambda} = (N-1)(N+1)\lambda^N - (N+1)N\lambda^{N-1} + (N+1)$ and $D^{''}(\lambda) \doteq \frac{d^2D(\lambda)}{d\lambda^2} = (N-1)N(N+1)\lambda^{N-2}(\lambda-1)$. Note that $D(1) = D^{'}(1) = D^{''}(1) = 0$ and $D^{''} < 0$ for $0 < \lambda < 1$, while on $\lambda > 1$, $D^{''} > 0$. It implies that $D^{'} > 0$ except $\lambda = 1$. Going one step further, we have $D > 0$ on $\lambda > 1$ and $D < 0$ on $0 < \lambda < 1$, which means $\Delta > 0$ on both $\lambda > 1$ and $0 < \lambda < 1$. By continuity, $\Delta > 0$ for $\lambda > 0$. It finishes the proof. ■

Proof of Proposition 3.5. Let $A_t = A_t(\iota)$. Note that $1 \le \|A_0\| \le \theta g_0 \varepsilon_0^2 \|H\| + \|G\| = \theta g_0 \varepsilon_0^2 + 1$, and hence $E(\log \|A_0\|) \le E\log(\theta g_0 \varepsilon_0^2 + 1) \le E\theta g_0 \varepsilon_0^2 < \infty$.

Note that $A_t = g_t(\theta\varepsilon_t^2 H + \frac{1}{g_t}G) \le g_t(\theta\varepsilon_t^2 H + \frac{1}{\zeta}G)$. Let $\tilde{A}_t = \theta\varepsilon_t^2 H + \frac{1}{\zeta}G$. $A_t, \tilde{A}_t$ are nonnegative. We then have

$$\|A_t A_{t-1} \dots A_0\| \le g_t g_{t-1} \dots g_0 \|\tilde{A}_t \tilde{A}_{t-1} \dots \tilde{A}_0\|.$$

It follows that

$$\gamma(A) \le E\log g_0 + \lim_t \frac{1}{1+t} E\log\|\tilde{A}_t \tilde{A}_{t-1} \dots \tilde{A}_0\| \le \tilde{\gamma},$$

where $\tilde{\gamma}$ is the top Lyapunov exponent associated with the sequence $\{\tilde{A}_t, t \in \mathbb{Z}\}$. Since $\tilde{A}_t$'s are i.i.d. and irreducible and ε has unbounded support, then $\tilde{\gamma} < 0$ if $\rho[E(\tilde{A}_0)] \le 1$ (see Kesten and Spitzer, 1984, Thm. 2). Note that $E(\tilde{A}_0) = \frac{1}{\zeta}M(\theta\zeta)$ and

$$\rho[E(\tilde{A}_0)] \le 1 \Longleftrightarrow \rho[M(\theta\zeta)] \le \zeta \Longleftrightarrow f_{\theta\zeta}(\zeta) = det(\zeta I - M(\theta\zeta)) \ge 0$$

by Lemma 3.1. $f_{\theta\zeta}(\zeta) \ge 0$ if $\theta \le \zeta^{N-1}/(1 + \zeta + \dots + \zeta^{N-1})$. The proof is complete. ■

Proof of Proposition 3.6. Note that $A_t(\vec{c}) = \theta g_t \varepsilon_t^2 H_{N+K-1} D(\vec{c}) + G_{N+K-1}$, where $D(\vec{c}) = Diag(c_1, \dots, c_{N+K-1})$. $1 \le \|A_0(\vec{c})\| \le \theta g_0 \varepsilon_0^2 \|H_{N+K-1} D(\vec{c})\| + \|G_{N+K-1}\| \le N\theta g_0 \varepsilon_0^2 + 1$. Hence $E(\log\|A_0(\vec{c})\|) \le E\log(N\theta g_0 \varepsilon_0^2 + 1) \le E(N\theta g_0 \varepsilon_0^2) < \infty$.

Note that $A_t(\vec{c}) \le g_t(\theta\varepsilon_t^2 H_{N+K-1} D(\vec{c}) + \frac{1}{\zeta} G_{N+K-1})$. Similar to the discussion in the proof of Proposition 3.5, we have $\gamma(A) < 0$ if $\rho(\theta\zeta H_{N+K-1} D(\vec{c}) + G_{N+K-1}) \le \zeta$. Let $f(\lambda) = \det(\lambda I_{N+K-1} - \theta\zeta H_{N+K-1} D(\vec{c}) - G_{N+K-1})$. Note that

$$f(\lambda) = \lambda^{N+K-1} - \theta\zeta c_1 \lambda^{N+K-2} - \theta\zeta c_2 \lambda^{N+K-3} - \dots - \theta\zeta c_{N+K-2}\lambda - \theta\zeta c_{N+K-1}.$$

The matrix $\theta\zeta H_{N+K-1} D(\vec{c}) + G_{N+K-1}$ is primitive. Similar to the discussion in the proof of Lemma 3.1(1), $\rho(\theta\zeta H_{N+K-1} D(\vec{c}) + G_{N+K-1}) \le \zeta$ if and only if $f(\zeta) \ge 0$, or equivalently, $\theta \le \zeta^{K+N-2}/(c_1\zeta^{K+N-2} + c_2\zeta^{K+N-3} + \dots + c_{K+N-2}\zeta + c_{K+N-1})$. ■

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Proof of Proposition 3.8. Note that the top Lyapunov exponent associated with $A_t(\vec{c})$, denoted by γ, should be strictly negative. Hence there exists $t_0 > 0$ such that $\|S_{t,t}\| < e^{t\gamma/2}$ for $t > t_0$ almost surely. It follows that, with probability one,

$$\sum_{i,j=1}^{N+K-1} ent_{ij} S_{t,t} < (N+K-1)e^{t\gamma/2},$$

where the notation $ent_{ij}A$ represents the entry in the i-th row and j-th column of a matrix A. Note also that $A_t(\vec{c}) = \theta g_t \varepsilon_t^2 H_{N+K-1} D(\vec{c}) + G_{N+K-1} \geq \theta\zeta\varepsilon_t^2 H_{N+K-1} D(\vec{c}) + G_{N+K-1} \geq 0$ almost surely. Let $L_t = \theta\zeta\varepsilon_t^2 H_{N+K-1} D(\vec{c}) + G_{N+K-1}$. L_t are i.i.d. Then we have

$$\sum_{i,j=1}^{N+K-1} ent_{ij}(EL_1)^t = E \sum_{i,j=1}^{N+K-1} ent_{ij}(L_t L_{t-1} \dots L_1) < (N+K-1)e^{t\gamma/2}. \tag{A.4}$$

Since EL_1 are nonnegative and primitive, it follows from the Perron–Frobenius theorem that $\sum_{i,j=1}^{N+K-1} ent_{ij}(EL_1)^t = O(\rho(EL_1))^t$ (see theorem 1.2 of Seneta, 1981; Kesten and Spitzer, 1984). Note that $\gamma < 0$. Together with (A.4), we have $\rho(EL_1) < 1$.

Let $f(\lambda) = \det(\lambda I_{N+K-1} - E(L_t))$. Note that $f(\lambda) = \lambda^{N+K-1} - \theta\zeta c_1 \lambda^{N+K-2} - \theta\zeta c_2 \lambda^{N+K-3} - \dots - \theta\zeta c_{N+K-2}\lambda - \theta\zeta c_{N+K-1}$. A similar argument to the proof of Lemma 3.1 yields that $f(1) > 0$, i.e., $\theta < 1/(N\zeta)$. ■

Proof of Proposition 3.9. Consider first $K = N = 1$. Suppose that the system starts from the infinite past. We have

$$r_t^2 = m \sum_{n=0}^{\infty} \theta^n \left(\prod_{j=0}^{n} g_{t-j} \varepsilon_{t-j}^2 \right) \text{ almost surely,} \tag{A.5}$$

and $Er_t^2 = m \sum_{n=0}^{\infty} \theta^n E\left(\prod_{j=0}^{n} g_{t-j}\varepsilon_{t-j}^2\right)$ due to Fubini's theorem. It is easy to verify that $\prod_{j=0}^{n} g_{t-j}$ can be expressed as a polynomial of g_{t-n} of degree $n+1$. In other words, $\prod_{j=0}^{n} g_{t-j} = \sum_{k=1}^{n+1} G_k^{(n)} g_{t-n}^k$, where $G_k^{(n)}$ is a function of $u_{t-1}^{k_1}, u_{t-2}^{k_2}, u_{t-3}^{k_3}, \dots, u_{t-n}^{k_n}$, and $u_t = \alpha\varepsilon_t^2 + \beta$, $0 \leq k_i \leq n$. As a result, we can rewrite Er_t^2 as

$$Er_t^2 = m \sum_{n=0}^{\infty} \theta^n \sum_{k=1}^{n+1} E\left(G_k^{(n)} \prod_{j=0}^{n} \varepsilon_{t-j}^2 \right) E(g_{t-n}^k).$$

Note that $\theta > 0$, and a necessary condition for $Er_t^2 < \infty$ is that $E(g_t^{n+1}) < \infty$ for $n \geq 0$. In other words, $Er_t^2 < \infty$ only if $E(u_0^{n+1}) < 1$ for any $n \geq 0$ (see for example, Bollerslev, 1986 or He and Terasvirta, 1999). Therefore, Er_t^2 is not finite when ε_t has unbounded support.

If the system starts from time 0, we have, by letting $k = t$ in equation (13),

$$r_t^2 = \theta^t \left(\prod_{j=0}^{t-1} g_{t-j} \varepsilon_{t-j}^2 \right) r_0^2 + m \sum_{n=0}^{t-1} \theta^n \left(\prod_{j=0}^{n} g_{t-j} \varepsilon_{t-j}^2 \right). \tag{A.6}$$

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

A similar argument yields

$$Er_t^2 = \theta^t \sum_{l=1}^{t} E\left(G_l^{(t-1)} \prod_{j=1}^{t-1} \varepsilon_{t-j}^2\right) E(g_1^l r_0^2) + m \sum_{n=0}^{t-1} \theta^n \sum_{l=1}^{n+1} E\left(G_l^{(n)} \prod_{j=1}^{n} \varepsilon_{t-j}^2\right) E(g_{t-n}^l),$$

and $Er_t^2 < \infty$ only if $E(g_1^l r_0^2) < \infty$ and $E(g_1^l) < \infty$ for $1 \le l \le t$. Therefore, there exists $t_0 \ge 1$ such that $Er_t^2 = \infty$ for $t > t_0$.

For K and N greater than 1, $Y_t = \sum_{n=0}^{\infty} S_{t,n} B_{t-n}$ w.p.1 if the system starts from the infinite past, and $EY_t = \sum_{n=0}^{\infty} E(S_{t,n} B_{t-n})$ due to Fubini's theorem. Note that $A_t(\vec{c}) = \theta g_t \varepsilon_t^2 H_{N+K-1} D(\vec{c}) + G_{N+K-1}$ and $B_t = m g_t \varepsilon_t^2 e_1$ where $e_1 = (1, 0, \ldots, 0)^T$. We obtained $S_{t,n} = A_t(\vec{c}) A_{t-1}(\vec{c}) \ldots A_{t-n+1}(\vec{c}) = \theta^n \left(\prod_{j=0}^{n-1} g_{t-j} \varepsilon_{t-j}^2\right) \left(H_{N+K-1} D(\vec{c})\right)^n + *$, where $*$ represents all the remaining terms, and

$$E(S_{t,n} B_{t-n}) = m\theta^n E\left(\prod_{j=0}^{n} g_{t-j} \varepsilon_{t-j}^2\right) \left(H_{N+K-1} D(\vec{c})\right)^n e_1 + *, \quad n \ge 0.$$

Using a similar argument one can show that Er_t^2 is not finite. If the system starts from time 0, then $Y_t = S_{t,t} Y_0 + \sum_{n=0}^{t-1} S_{t,n} B_{t-n}$, and $EY_t = E(S_{t,t} Y_0) + \sum_{n=0}^{t-1} E(S_{t,n} B_{t-n})$. A similar argument yields that there exists $t_0 \ge 1$ such that $Er_t^2 = \infty$ for $t > t_0$. ■

Proof of Proposition 3.10. Because $E\|A_0\|^\delta < \infty$ and $E\|B_0\|^\delta < \infty$ imply that $E \log^+ \|A_0\| < \infty$ and $E \log^+ \|B_0\| < \infty$, the equation $Y_t = A_t Y_{t-1} + B_t$ has a unique strictly stationary ergodic solution. The solution is $Y_t = \sum_{n=0}^{\infty} S_{t,n} B_{t-n}$, where $S_{t,n} = A_t A_{t-1} \ldots A_{t-n+1}$ for $n > 0$ and $S_{t,0} = 1$.

Consider $Y_0 = \sum_{n=0}^{\infty} S_{0,n} B_{-n}$. Note that $\lim_{n\to\infty} \frac{1}{n} E \log \|S_{0,n}\| = \gamma < 0$. There exists n_0 such that $E \log \|S_{0,n_0}\| < 0$. Note also that $E\|S_{0,n_0}\|^{\delta/n_0} \le E\|A_0\|^\delta < \infty$ due to the Hölder's inequality. One can find $0 < \delta_0 < \min(2, \delta)$ such that $E\|S_{0,n}\|^{\delta_0} < 1$. Moreover, there exist $0 < \tilde{c} < \infty$ and $0 < \tilde{\rho} < 1$ such that $E\|S_{0,k}\|^{\delta_0} \le \tilde{c}\tilde{\rho}^k$—see Remark 2.6 of Basrak, Davis, and Mikosch (2001), and Lemma 2.3 of Berkes, Horváth, and Kokoszka (2003). For $\delta_0/2 < 1$,

$$E\|Y_0\|^{\delta_0/2} \le \sum_{n=0}^{\infty} E\left(\|S_{0,n} B_{-n}\|^{\delta_0/2}\right) \le \sum_{n=0}^{\infty} (E\|S_{0,n}\|^{\delta_0})^{1/2} (E\|B_0\|^{\delta_0})^{1/2} < \infty.$$

It follows that $E(\log \|Y_t\|) \le \frac{1}{\delta_0/2} \log E(\|Y_t\|^{\delta_0/2}) < \infty$. ■

Proof of Corollary 3.1. Note that $E\|A_0(\vec{c})\| \le E(\theta g_0 \varepsilon_0^2)\|H_{N+K-1} D(\vec{c})\| + \|G_{N+K-1}\| = \theta \|H_{N+K-1} D(\vec{c})\| + \|G_{N+K-1}\|$, and $E\|B_0(\vec{c})\| = m$. The result follows from Proposition 3.10 immediately with $\delta = 1$. ■

The proof of Lemma 4.1 needs the following Lemma:

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

LEMMA A.1. *Suppose that* $\{a_i\}_{i=1}^n$ *are nonnegative. Let* $a = \sum_{i=1}^n a_i$ *and* $a > 0$. *Then*

$$\log\left(\sum_{i=1}^n a_i\right) \le \log n + \sum_{i=1}^n \frac{a_i}{a}\log a_i \le \log n + \sum_{i=1}^n \log^+ a_i, \tag{A.7}$$

$$\log^+\left(\prod_{i=1}^n a_i\right) \le \sum_{i=1}^n \log^+ a_i. \tag{A.8}$$

Since the results are elementary, its proof is skipped.

Proof of Lemma 4.1. Under Assumption 4.2, r_t^2 and τ_t are strictly stationary ergodic, and $E\log(r_t^2) < \infty$. Equation (20) and $0 \le \beta < 1$ imply that $g_t = (1-\alpha-\beta)/(1-\beta) + \alpha\sum_{k=0}^\infty \beta^k \frac{r_{t-1-k}^2}{\tau_{t-1-k}}$ almost surely. Thus, for δ^* defined in Corollary 3.1,

$$\begin{aligned}(g_t)^{\delta^*} &\le \frac{(1-\alpha-\beta)^{\delta^*}}{(1-\beta)^{\delta^*}} + \sum_{k=0}^\infty (\alpha\beta^k/m)^{\delta^*}(r_{t-1-k}^{2\delta^*}) \\ &\le K + K\sum_{k=0}^\infty (\bar\beta)^{k\delta^*}(r_{t-1-k}^{2\delta^*}),\end{aligned} \tag{A.9}$$

for some constants $K > 0$ and $0 < \bar\beta < 1$ due to the compactness of $\mathcal{U}$. Note also that $\inf_{\Phi\in\mathcal{U}} g_t(\Phi) > 0$ almost surely. Therefore, $E\sup_{\Phi\in\mathcal{U}}|\log(g_t)| = (\delta^*)^{-1}E\sup_{\Phi\in\mathcal{U}}|\log(g_t^{\delta^*})| < \infty$.

Rewrite τ_t as $\tau_t = m + \theta\sum_{l=1}^{N+K-1} c_l r_{t-l}^2$, where c_l's are combinations of the weights $\varphi_k(\omega)$ and satisfy $\sum_{l=1}^{N+K-1} c_l = N\sum_{k=1}^K \varphi_k(\omega) = N$. An application of Lemma A.1 yields

$$\begin{aligned}\log m \le \log\tau_t &\le \log(N+K) + \log^+(m) + (N+K-1)\log^+(\theta) \\ &+ \sum_{l=1}^{N+K-1} \log^+(c_l) + \log^+(r_{t-1}^2).\end{aligned}$$

We have $E\sup_{\Phi\in\mathcal{U}}|\log(\tau_t)| < \infty$, due to Corollary 3.1, Assumption 4.1, and the compactness of $\mathcal{U}$.

Since $l_t^-(\Phi) = (\log V_t(\Phi) + r_t^2/V_t(\Phi))^- \le (\log V_t(\Phi))^-$, we have $E\sup_{\Phi\in\mathcal{U}} l_t^-(\Phi) < \infty$. ■

Proof of Lemma 4.2. Note that $El_t(\Phi) - El_t(\Phi_0) = E(-\log(\frac{V_t(\Phi_0)}{V_t(\Phi)}) + \frac{V_t(\Phi_0)}{V_t(\Phi)} - 1) \ge 0$, and the equality holds if and only if $V_t(\Phi_0) = V_t(\Phi)$. Therefore, to complete the proof, one just needs to show that $V_t(\Phi_0) = V_t(\Phi)$ if and only if $\Phi_0 = \Phi$. The sufficiency is obvious, so we just need to prove that $V_t(\Phi_0) = V_t(\Phi)$ implies $\Phi_0 = \Phi$.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Note that

$$\begin{aligned}\tau_t g_t &= \left[m+\theta\sum_{l=2}^{K+N-1} c_l r_{t-l}^2+\theta c_1 r_{t-1}^2\right]\left[\eta+\eta(1+\sum_{n\geq 2}\prod_{j=2}^{n} u_{t-j})u_{t-1}\right]\\ &= \left[m+\theta\sum_{l=2}^{K+N-1} c_l r_{t-l}^2+\theta c_1 r_{t-1}^2\right]\\ &\quad\times\left[\eta+\eta\left(1+\sum_{n\geq 2}\prod_{j=2}^{n} u_{t-j}\right)\beta+\eta\left(1+\sum_{n\geq 2}\prod_{j=2}^{n} u_{t-j}\right)\alpha\varepsilon_{t-1}^2\right], \qquad \textbf{(A.10)}\end{aligned}$$

where $\eta = 1-\alpha-\beta > 0$ and $u_t = \alpha\varepsilon_t^2+\beta$. $V_t(\Phi_0) = V_t(\Phi)$ implies that ε_{t-1} is measurable with respect to $\sigma(\varepsilon_s, s < t-1)$, unless the coefficients in front of ε_{t-1} are 0. In particular, we have

$$\theta c_1\eta\left(1+\sum_{n\geq 2}\prod_{j=2}^{n} u_{t-j}\right)\alpha V_{t-1} = \theta_0 c_1^0\eta_0\left(1+\sum_{n\geq 2}\prod_{j=2}^{n} u_{t-j}^0\right)\alpha_0 V_{t-1}^0, \qquad \textbf{(A.11)}$$

where we use the superscript ^{0} to indicate that the *function* is evaluated at Φ_0.[11]

Since $\theta_0 \neq 0$, then $\theta \neq 0$. Let

$$\begin{aligned}F_{t-2} &= \theta c_1\eta\left(1+\sum_{n\geq 2}\prod_{j=2}^{n} u_{t-j}\right)\alpha\\ &= \theta c_1\eta\alpha\left(1+\beta\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}\right]+\alpha\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}\right]\varepsilon_{t-2}^2\right).\end{aligned}$$

In view of (A.11), we have $F_{t-2} = F_{t-2}^0$, and hence

$$\theta c_1\eta\alpha^2\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}\right] = \theta_0 c_1^0\eta_0\alpha_0^2\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}^0\right], \qquad \textbf{(A.12)}$$

$$\theta c_1\eta\alpha\left(1+\beta\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}\right]\right) = \theta_0 c_1^0\eta_0\alpha_0\left(1+\beta_0\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}^0\right]\right). \qquad \textbf{(A.13)}$$

Note that the right-hand side of (A.12) is not 0. Then $\theta c_1\eta\alpha^2$ is not 0 as well. It follows that

$$\left(\beta_0\frac{\theta_0 c_1^0\eta_0\alpha_0}{\theta c_1\eta\alpha}-\beta\frac{\theta_0 c_1^0\eta_0\alpha_0^2}{\theta c_1\eta\alpha^2}\right)\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}^0\right] = 1-\frac{\theta_0 c_1^0\eta_0\alpha_0}{\theta c_1\eta\alpha}. \qquad \textbf{(A.14)}$$

This is true only if

$$\left(\beta_0\frac{\theta_0 c_1^0\eta_0\alpha_0}{\theta c_1\eta\alpha}-\beta\frac{\theta_0 c_1^0\eta_0\alpha_0^2}{\theta c_1\eta\alpha^2}\right) = 0, \quad 1-\frac{\theta_0 c_1^0\eta_0\alpha_0}{\theta c_1\eta\alpha} = 0.$$

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Therefore, $\theta c_1 \eta\alpha = \theta_0 c_1^0 \eta_0 \alpha_0$, and $\beta/\alpha = \beta_0/\alpha_0$, and hence (A.12) becomes

$$\alpha\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}\right] = \alpha_0\left[1+\sum_{n\geq 3}\prod_{j=3}^{n} u_{t-j}^0\right]. \tag{A.15}$$

Note that the left-hand side of (A.15) is

$$\alpha^2\left[1+\sum_{n\geq 4}\prod_{j=4}^{n} u_{t-j}\right]\varepsilon_{t-3}^2 + \alpha + \alpha\beta\left[1+\sum_{n\geq 4}\prod_{j=4}^{n} u_{t-j}\right].$$

Similarly we have

$$\alpha^2\left[1+\sum_{n\geq 4}\prod_{j=4}^{n} u_{t-j}\right] = \alpha_0^2\left[1+\sum_{n\geq 4}\prod_{j=4}^{n} u_{t-j}^0\right], \tag{A.16}$$

$$\alpha+\alpha\beta\left[1+\sum_{n\geq 4}\prod_{j=4}^{n} u_{t-j}\right] = \alpha_0+\alpha_0\beta_0\left[1+\sum_{n\geq 4}\prod_{j=4}^{n} u_{t-j}^0\right]. \tag{A.17}$$

It follows that $\alpha = \alpha_0$, and hence $\beta = \beta_0$, $\theta c_1 = \theta_0 c_1^0$, and $m - m_0 + \sum_{l=2}^{N+K-1}(\theta c_l - \theta_0 c_l^0) r_{t-l}^2 = 0$. In a similar argument and using the fact that $\sum_{l=1}^{N+K-1} c_l = N$ and Assumption 4.1, we have $m = m_0$, $\theta = \theta_0$, and $\omega = \omega_0$. ■

Proof of Lemma 4.4. Note that $L_T - \tilde{L}_T = \frac{1}{T-(N+K)}\sum_{t=N+K+1}^{T}\log(\frac{V_t}{\tilde{V}_t}) + \left(\frac{r_t^2}{V_t} - \frac{r_t^2}{\tilde{V}_t}\right) = \frac{1}{T-(N+K)}\sum_{t=N+K+1}^{T}\log(\frac{g_t}{\tilde{g}_t}) + \frac{r_t^2}{\tau_t g_t}\frac{\tilde{g}_t - g_t}{\tilde{g}_t}$. By Cesaro's Lemma, it suffices to show $\sup_{\Phi\in\mathcal{U}}|\log(\frac{g_t}{\tilde{g}_t})|$ and $\sup_{\Phi\in\mathcal{U}}|\frac{r_t^2}{\tau_t g_t}\frac{\tilde{g}_t - g_t}{\tilde{g}_t}|$ converge to 0 almost surely.

For $t > K+N$, because $g_t = \tilde{g}_t + \beta^{t-K-N}(g_{K+N} - \tilde{g}_{K+N})$ and $\mathcal{U}$ is compact, we have

$$\log\left(\frac{g_t}{\tilde{g}_t}\right) \leq \frac{g_t}{\tilde{g}_t} - 1 \leq \frac{1}{1-\alpha-\beta}\beta^{t-K-N} g_{K+N} \leq C_* \bar{\beta}^t g_{K+N}, \tag{A.18}$$

$$\left|\frac{r_t^2}{\tau_t g_t}\frac{g_t - \tilde{g}_t}{\tilde{g}_t}\right| \leq \frac{1}{(1-\alpha-\beta)^2 m}\beta^{t-K-N} r_t^2 g_{K+N} \leq C_* \bar{\beta}^t r_t^2 g_{K+N}, \tag{A.19}$$

for some constant $C_* > 0$, and $\bar{\beta} \in (0,1)$. Therefore, we just need to show that $\sup_{\Phi\in\mathcal{U}}\bar{\beta}^t g_{K+N}$ and $\sup_{\Phi\in\mathcal{U}}\bar{\beta}^t r_t^2 g_{K+N}$ converge to 0 almost surely.

Note that for any $\epsilon > 0$,

$$P(\sup_{\phi\in\mathcal{U}}\bar{\beta}^t r_t^2 g_{K+N} > \epsilon) \leq \epsilon^{-\delta^*/2}\bar{\beta}^{t\delta^*/2}(E(r_t^2)^{\delta^*})^{1/2}(E\sup_{\phi\in\mathcal{U}} g_{K+N}^{\delta^*})^{1/2},$$

where δ^* is defined in Corollary 3.1. Because $E(r_t^2)^{\delta^*}$ is finite and $E\sup_{\phi\in\mathcal{U}} g_{K+N}^{\delta^*} < \infty$ (due to inequality (A.9)), we have

$$\sum_{t=K+N+1}^{\infty} P(\sup_{\phi\in\mathcal{U}}\bar{\beta}^t r_t^2 g_{K+N} > \epsilon) < \infty.$$

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

It follows from the Borel–Cantelli lemma that $\sup_{\Phi\in\mathcal{U}}\bar{\beta}^t r_t^2 g_{K+N}$ converges to 0 almost surely. Similarly, we have $\sup_{\Phi\in\mathcal{U}}\bar{\beta}^t g_{K+N}$ converges to 0 almost surely. ■

Proof of Proposition 4.1. The proof uses an argument similar to Francq and Zakoian (2010), p. 159 so the proof is omitted. ■

Proof of Lemma 4.5. In what follows, we use $\partial_\alpha, \partial_\beta, \partial_m, \partial_\theta$, and ∂_ω to denote the partial derivatives with respect to α, β, m, θ, and ω, respectively. For easy reference, the parameters $\{\alpha, \beta, m, \theta, \omega\}$ are also referred to as $\{\phi_i, 1 \le i \le 5\}$ when there is no confusion. Let $\partial_i = \partial_{\phi_i}$ the partial derivative with respect to ϕ_i, and $\partial_{ij} = \partial_{\phi_i}\partial_{\phi_j}$.

(1) Note that $V_t^{-2}\nabla V_t \nabla V_t' = g_t^{-2}\nabla g_t \nabla g_t' + \tau_t^{-2}\nabla\tau_t\nabla\tau_t' + g_t^{-1}\tau_t^{-1}(\nabla g_t \nabla\tau_t' + \nabla\tau_t\nabla g_t')$, and

$$|\tau_t^{-1}\partial_m\tau_t| \le 1/m,\ |\tau_t^{-1}\partial_\theta\tau_t| \le 1/\theta, \tag{A.20}$$

$$|\tau_t^{-1}\partial_\omega\tau_t| \le \tau_t^{-1}\theta\sum_{l=1}^{N+K-1}\left|\frac{dc_l(\omega)}{d\omega}\right| r_{t-l}^2 \le \sum_{l=1}^{N+K-1}\left|\frac{dc_l(\omega)}{d\omega}\right|/c_l(\omega), \tag{A.21}$$

$$|g_t^{-1}\partial_m g_t| \le g_t^{-1}\alpha\sum_{k=0}^{\infty}\beta^k(r_{t-1-k}^2/\tau_{t-1-k})|\tau_{t-1-k}^{-1}\partial_m\tau_{t-1-k}| \le 1/m. \tag{A.22}$$

Similarly, $|g_t^{-1}\partial_\theta g_t| \le 1/\theta$, and $|g_t^{-1}\partial_\omega g_t| \le \sum_{l=1}^{N+K-1}|\frac{dc_l(\omega)}{d\omega}|/c_l(\omega)$. Moreover,

$$\begin{aligned}
|g_t^{-1}\partial_\alpha g_t| &\le 1/(1-\beta)g_t^{-1} + g_t^{-1}\sum_{k=0}^{\infty}\beta^k(r_{t-1-k}^2/\tau_{t-1-k}) \le 1/(1-\alpha-\beta)+1/\alpha,\\
|g_t^{-1}\partial_\beta g_t| &\le \alpha/(1-\beta)^2 g_t^{-1} + g_t^{-1}\alpha\sum_{k=1}^{\infty}k\beta^{k-1}(r_{t-1-k}^2/\tau_{t-1-k})\\
&\le \frac{\alpha}{(1-\beta)^2(1-\alpha-\beta)} + \sum_{k=1}^{\infty}k\frac{\alpha\beta^{k-1}(r_{t-1-k}^2/\tau_{t-1-k})}{(1-\alpha-\beta)/(1-\beta)+\alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}\\
&\le \frac{\alpha}{(1-\beta)^2(1-\alpha-\beta)} + \beta^{-1}\sum_{k=1}^{\infty}k\left(\frac{\alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}{(1-\alpha-\beta)/(1-\beta)}\right)^{\delta},
\end{aligned} \tag{A.23}$$

where $\delta = \delta^*/2$ and δ^* is defined in Corollary 3.1. It follows from (A.23) that

$$(E|g_t^{-1}\partial_\beta g_t|^2)^{1/2} \le \frac{\alpha}{(1-\beta)^2(1-\alpha-\beta)} + \beta^{-1}\sum_{k=1}^{\infty}k\sqrt{E\left(\frac{\alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}{(1-\alpha-\beta)/(1-\beta)}\right)^{\delta^*}}, \tag{A.24}$$

and hence $E|g_t^{-1}\partial_\beta g_t|^2 < \infty$. The discussion implies that $E\left(V_t^{-2}\nabla V_t\nabla V_t'\right)$ exists.

Next we will show $\Sigma(\Phi_0)$ is positive definite. Otherwise, there exists $p \in \mathbb{R}^5$ such that $p'\Sigma(\Phi_0)p = 0$ and $\|p\| = 1$. In other words, $p'\nabla V_t(\Phi_0) = 0$ almost surely for any t.

Note that $\nabla V_t = \nabla((1-\alpha-\beta)\tau_t) + \nabla(\alpha\tau_t/\tau_{t-1})r_{t-1}^2 + \nabla(\beta\tau_t/\tau_{t-1})V_{t-1} + \beta\tau_t/\tau_{t-1}\nabla V_{t-1}$. Let $\psi_t = \nabla((1-\alpha-\beta)\tau_t) + \nabla(\alpha\tau_t/\tau_{t-1})r_{t-1}^2 + \nabla(\beta\tau_t/\tau_{t-1})V_{t-1}$.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Then $p'\nabla V_t(\Phi_0) = 0$ almost surely for any t implies that $p'\psi_t(\Phi_0) = 0$ almost surely for any t.

Suppose $\psi_t = (\psi_{t1}, \ldots, \psi_{t5})'$. Then $\psi_{t1} = \tau_t(r_{t-1}^2/\tau_{t-1} - 1)$, $\psi_{t2} = \tau_t(g_{t-1} - 1)$, and $\psi_{tk} = g_t\partial_k\tau_t - (\alpha r_{t-1}^2/\tau_{t-1} + \beta g_{t-1})\tau_t\partial_k\tau_{t-1}/\tau_{t-1}$ for $k = 3, 4, 5$. $p'\psi_t(\Phi_0) = 0$ almost surely indicates

$$p_1(g_{t-1}\varepsilon_{t-1}^2 - 1) + p_2(g_{t-1} - 1) + g_t\sum_{k=3}^{5} p_k\frac{\partial_k\tau_t}{\tau_t} - (\alpha g_{t-1}\varepsilon_{t-1}^2 + \beta g_{t-1})\sum_{k=3}^{5} p_k\frac{\partial_k\tau_{t-1}}{\tau_{t-1}} = 0, \tag{A.25}$$

almost surely at Φ_0. Rewrite (A.25) as

$$\varepsilon_{t-1}^2\underbrace{g_{t-1}\left(p_1 - \alpha\sum_{k=3}^{5} p_k\frac{\partial_k\tau_{t-1}}{\tau_{t-1}}\right)}_{F_{t-2}} + g_t\sum_{k=3}^{5} p_k\frac{\partial_k\tau_t}{\tau_t} = \underbrace{g_{t-1}\left(\beta\sum_{k=3}^{5} p_k\frac{\partial_k\tau_{t-1}}{\tau_{t-1}} - p_2\right) + p_1 + p_2}_{G_{t-2}}.$$

Let $H_{t,2} = m + \theta\sum_{l\geq 2} c_l r_{t-l}^2$. We obtained

$$(H_{t,2} + \theta c_1 r_{t-1}^2)F_{t-2}\varepsilon_{t-1}^2 + (\eta + \beta g_{t-1} + \alpha g_{t-1}\varepsilon_{t-1}^2)\sum_{k=3}^{5} p_k(\partial_k H_{t,2} + \partial_k(\theta c_1)r_{t-1}^2) = G_{t-2}(H_{t,2} + \theta c_1 r_{t-1}^2),$$

where $\eta = 1 - \alpha - \beta$. It follows that

$$\theta c_1 F_{t-2} + \alpha g_{t-1}\sum_k p_k\partial_k(\theta c_1) = 0. \tag{A.26}$$

In view of (A.26), we have $(\theta c_1 p_1 + \alpha p_4 c_1 + \alpha p_5\theta\frac{dc_1(\omega)}{d\omega})\tau_{t-1} = \theta c_1\alpha\sum_{k=3}^{5} p_k\partial_k\tau_{t-1}$. Therefore,

$$\left(\theta c_1 p_1 + \alpha p_4 c_1 + \alpha p_5\theta\frac{dc_1(\omega)}{d\omega}\right)m = \theta c_1\alpha p_3,$$
$$\left(\theta c_1 p_1 + \alpha p_4 c_1 + \alpha p_5\theta\frac{dc_1(\omega)}{d\omega}\right)\theta c_l = \theta c_1\alpha\left(p_4 c_l + p_5\theta\frac{dc_l(\omega)}{d\omega}\right),$$

for $l = 1, 2, \ldots, N + K - 1$. Note that $\sum_l c_l = N$ and $\sum_l \frac{dc_l(\omega)}{d\omega} = 0$. Consequently,

$$c_1 p_1 = 0, \qquad \theta p_3 = m p_4, \qquad p_5\frac{dc_l(\omega)}{d\omega} = 0 \quad \text{for any } l. \tag{A.27}$$

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Taking expectation on both sides of (A.25), we have $\sum_{k=3}^{5} p_k E(g_1 \partial_k \tau_1/\tau_1) = 0$. In view of (A.27), we obtained $p_4 E(g_t) = 0$ and hence $p_4 = 0$. Therefore, $p \equiv 0$ and this contradicts the assumption that $\|p\| = 1$.

(2) Note that

$$\nabla L_T = \frac{1}{T}\sum_{t=1}^{T} \nabla l_t = \frac{1}{T}\sum_{t=1}^{T}\left(1 - \frac{r_t^2}{V_t}\right)\frac{\nabla V_t}{V_t}.$$

We have $\partial_i l_t(\Phi_0) = (1-\varepsilon_t^2)(\partial_i \tau_t(\Phi_0)/\tau_t(\Phi_0) + \partial_i g_t(\Phi_0)/g_t(\Phi_0))$. Because $\partial_i l_t(\Phi_0)$ is a measurable function of $\{\varepsilon_s^2, s \le t\}$, it is strictly stationary ergodic. Note also that $E(\partial_i l_t(\Phi_0))^2$ is finite, and $E(\partial_i l_t(\Phi_0)|r_s, s < t) = 0$. We have $\{\partial_i l_t(\Phi_0), t \in \mathbb{Z}\}$ is a martingale difference sequence with finite second moment. The asymptotic normality follows from the martingale central limit theorem, and the Cramer–Wold device.

(3) Let $B_n = \overline{B(\Phi_0, 1/n)} \cap \mathcal{A}$. B_n is not empty. Note that $\partial_{ij} L_T = \frac{1}{T-(N+K)} \sum_{t=N+K+1}^{T} \partial_{ij} l_t$, and

$$\partial_{ij} l_t = \left(1 - \frac{r_t^2}{\tau_t g_t}\right)\left(\frac{\partial_{ij} g_t}{g_t} + \frac{\partial_{ij}\tau_t}{\tau_t} + \frac{\partial_i g_t}{g_t}\frac{\partial_j \tau_t}{\tau_t} + \frac{\partial_j g_t}{g_t}\frac{\partial_i \tau_t}{\tau_t}\right) + \left(\frac{2r_t^2}{\tau_t g_t} - 1\right)\left(\frac{\partial_i g_t}{g_t} + \frac{\partial_i \tau_t}{\tau_t}\right)\left(\frac{\partial_j g_t}{g_t} + \frac{\partial_j \tau_t}{\tau_t}\right).$$

To prove (30), we first need to show $E \sup_{\Phi \in B_n} |\partial_i \partial_j l_t(\Phi)| < \infty$. Note that $|\tau_t^{-1}\partial_i \tau_t|$, $|g_t^{-1}\partial_\alpha g_t|$, $|g_t^{-1}\partial_\theta g_t|$, $|g_t^{-1}\partial_\omega g_t|$, and $|g_t^{-1}\partial_m g_t|$ are bounded on $\mathcal{A}$ (see Lemma 4.5(1)). We have

$$\frac{\tau_t(\Phi_0)}{\tau_t} = \frac{m_0}{\tau_t} + \sum_{l=1}^{N+K-1} \frac{\theta_0 c_l(\omega_0) r_{t-l}^2}{\tau_t} \le \frac{m_0}{m} + \sum_{l=1}^{N+K-1} \frac{\theta_0 c_l(\omega_0)}{\theta c_l(\omega)}, \tag{A.28}$$

and hence $|\tau_t^{-1}\tau_t(\Phi_0)|$ is bounded as well. An argument similar to the proof of Lemma 4.5(1) shows that $|\tau_t^{-1}\partial_{ij}\tau_t|$ is bounded on $\mathcal{A}$ as well. Therefore, it is sufficient to show that $E \sup_{\Phi \in B_n} |g_t^{-1} g_t(\Phi_0)|^2$, $E \sup_{\Phi \in B_n} |g_t^{-1}\partial_\beta g_t|^4$, and $E \sup_{\Phi \in B_n} |g_t^{-1}\partial_{ij} g_t|^2$ are finite.

Consider n such that $n > n_0 \equiv \lfloor 1 + \beta_0^{-1}(1-\beta_0^\delta)^{-1/(1-\delta)} \rfloor$, where $\delta = \delta^*/4$ and δ^* is defined in Corollary 3.1. For $\Phi \in B_n$,

$$\frac{g_t(\Phi_0)}{g_t} \le \frac{\eta_0}{\eta} + \sum_{k=0}^{\infty} \frac{\alpha_0 \beta_0^k r_{t-1-k}^2/\tau_{t-1-k}^0}{\eta + \alpha\beta^k r_{t-1-k}^2/\tau_{t-1-k}} \le \frac{\eta_0}{\eta} + \frac{\alpha_0}{\alpha^{1-\delta}\eta^\delta m^\delta} \sum_{k=0}^{\infty} \frac{\beta_0^k}{\beta^{k(1-\delta)}} \frac{\tau_{t-1-k}}{\tau_{t-1-k}^0} r_{t-1-k}^{2\delta}, \tag{A.29}$$

where $\eta = 1 - \alpha - \beta$, and the superscript ^{0} indicates that the quantity is evaluated at Φ_0. Note that $\frac{\eta_0}{\eta}$, $\frac{\alpha_0}{\alpha^{1-\delta}\eta^\delta m^\delta}$, and $\frac{\tau_{t-1-k}}{\tau_{t-1-k}^0}$ are bounded on $\mathcal{A}$. For $\Phi \in B_n$,

$$\frac{\beta_0}{\beta^{(1-\delta)}} \le \frac{\beta_0}{\beta_0^{(1-\delta)} - n^{-(1-\delta)}} \le \frac{\beta_0}{\beta_0^{(1-\delta)} - n_0^{-(1-\delta)}} < 1.$$

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

Let $\rho_0 = \frac{\beta_0}{\beta_0^{(1-\delta)} - n_0^{-(1-\delta)}}$. Using (A.29), we have $\sup_{\Phi\in B_n} \frac{g_t^0}{g_t} \le K + K\sum_{k=0}^{\infty} \rho_0^k r_{t-1-k}^{2\delta}$ for some constant $K > 0$. It follows from Corollary 3.1 that

$$E \sup_{\Phi\in B_n} |g_t^{-1} g_t(\Phi_0)|^2 < \infty. \tag{A.30}$$

Take δ appearing in (A.23) as $\delta^*/4$, then in a similar argument we have

$$E \sup_{\Phi\in B_n} |g_t^{-1} \partial_\beta g_t|^4 < \infty. \tag{A.31}$$

Moreover, note that $\partial_\alpha g_t = (g_t - 1)/\alpha$, and $\partial_{\alpha\alpha} g_t = 0$. Therefore, $\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\alpha\beta} g_t|$, $\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\alpha\theta} g_t|$, $\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\alpha m} g_t|$, $\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\alpha\omega} g_t|$ are square integrable. Note also that

$$\begin{aligned}
\left|\frac{\partial_{\beta\beta} g_t}{g_t}\right| &\le \frac{2\alpha}{(1-\beta)^3\eta} + \beta^{-2}\sum_{k=2}^{\infty} k(k-1)\frac{\alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}{\eta + \alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})} \\
&\le \frac{2\alpha}{(1-\beta)^3\eta} + \beta^{-2}\sum_{k=2}^{\infty} k(k-1)\left(\frac{\alpha\beta^k r_{t-1-k}^2}{m\eta}\right)^{\delta^*/2}, \qquad \text{(A.32)} \\
\left|\frac{\partial_{\beta m} g_t}{g_t}\right| &\le \sum_{k=1}^{\infty} k\frac{\alpha\beta^{k-1}(r_{t-1-k}^2/\tau_{t-1-k})}{\eta + \alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}\left|\frac{\partial_m \tau_{t-1-k}}{\tau_{t-1-k}}\right| \\
&\le \frac{1}{m\beta}\sum_{k=1}^{\infty} k\left(\frac{\alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}{\eta}\right)^{\delta^*/2}, \qquad \text{(A.33)} \\
\left|\frac{\partial_{mm} g_t}{g_t}\right| &\le \sum_{k=0}^{\infty}\frac{\alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}{\eta + \alpha\beta^k(r_{t-1-k}^2/\tau_{t-1-k})}\left(2\left|\frac{\partial_m \tau_{t-1-k}}{\tau_{t-1-k}}\right|^2 + \left|\frac{\partial_{mm}\tau_{t-1-k}}{\tau_{t-1-k}}\right|\right) \\
&\le \sum_{k=0}^{\infty}\left(\frac{\alpha\beta^k r_{t-1-k}^2}{\eta\tau_{t-1-k}}\right)^{\delta^*/2}\left(2\left|\frac{\partial_m \tau_{t-1-k}}{\tau_{t-1-k}}\right|^2 + \left|\frac{\partial_{mm}\tau_{t-1-k}}{\tau_{t-1-k}}\right|\right). \qquad \text{(A.34)}
\end{aligned}$$

We have $E\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\beta\beta} g_t| < \infty$ and $E\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\beta m} g_t| < \infty$ due to (A.20), Corollary 3.1 and $\mathcal{A}$ being compact. Similarly, we have $E\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\beta\theta} g_t|$, $E\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\beta\omega} g_t|$, $E\sup_{\Phi\in B_n} |g_t^{-1}\partial_{m\theta} g_t|$, $E\sup_{\Phi\in B_n} |g_t^{-1}\partial_{m\omega} g_t|$, and $E\sup_{\Phi\in B_n} |g_t^{-1}\partial_{\theta\omega} g_t|$ are finite.

Because $\partial_{ij} l_t$ is strictly stationary ergodic—it is a measurable function of $\{r_s^2, s \le t\}$, and

$$\begin{aligned}
\sup_{\Phi\in B_n} \|H(L_T)(\Phi) - \Sigma(\Phi_0)\| \le{}& \sup_{\Phi\in B_n} \|H(L_T)(\Phi) - EH(l_1)(\Phi)\| \\
&+ E\sup_{\Phi\in B_n} \|H(l_1)(\Phi) - H(l_1)(\Phi_0)\|,
\end{aligned}$$

and $E\sup_{\Phi\in B_n} \|H(l_t)(\Phi)\|$ is O(1) uniformly in t, (30) holds due to the dominated convergence theorem and the uniform SLLN.

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

(4) It suffices to show $\frac{1}{\sqrt{T}}\sum_{t=N+K+1}^{T}\sup_{\Phi\in\mathcal{A}}|\partial_i l_t(\Phi)-\partial_i\tilde{l}_t(\Phi)|$ converges to 0 in probability for each i. Note that

$$
\begin{aligned}
|\partial_i l_t-\partial_i\tilde{l}_t| &= \left|\left(1-\frac{r_t^2}{V_t}\right)\frac{\partial_i V_t}{V_t}-\left(1-\frac{r_t^2}{\tilde{V}_t}\right)\frac{\partial_i\tilde{V}_t}{\tilde{V}_t}\right| \\
&\le \left|\frac{g_t-\tilde{g}_t}{\tilde{g}_t}\frac{r_t^2}{V_t}\frac{\partial_i V_t}{V_t}\right|+\left|\left(\frac{\partial_i g_t}{g_t}-\frac{\partial_i\tilde{g}_t}{\tilde{g}_t}\right)\left(1-\frac{r_t^2}{\tilde{V}_t}\right)\right|, \\
\left|\frac{\partial_i g_t}{g_t}-\frac{\partial_i\tilde{g}_t}{\tilde{g}_t}\right| &\le |g_t^{-1}\partial_i g_t|\left|\frac{\tilde{g}_t-g_t}{\tilde{g}_t}\right|+\tilde{g}_t^{-1}|\partial_i g_t-\partial_i\tilde{g}_t|, \\
|\partial_i g_t-\partial_i\tilde{g}_t| &\le (t-K-N)\beta^{t-K-N-1}|g_{K+N}-\tilde{g}_{K+N}|+\beta^{t-K-N}|\partial_i g_{K+N}|.
\end{aligned}
$$

In view of (A.18) and (A.19), we have, for $\Phi\in\mathcal{A}$,

$$
\begin{aligned}
|\partial_i l_t-\partial_i\tilde{l}_t| &\le C_*\bar{\beta}^t r_t^2 g_{K+N}\left|\frac{\partial_i V_t}{V_t}\right| \\
&\quad + C_*\bar{\beta}^t(|\partial_i g_t|g_{K+N}+t|g_{K+N}-\tilde{g}_{K+N}|+|\partial_i g_{K+N}|)\left|1-\frac{r_t^2}{\tilde{V}_t}\right|,
\end{aligned}
$$

for some constant $C_*>0$ and $0<\bar{\beta}<1$. Therefore, for $0<\delta<1$,

$$
E(\sup_{\Phi\in\mathcal{A}}|\partial_i l_t(\Phi)-\partial_i\tilde{l}_t(\Phi)|)^{\delta}\le C_*^{\delta}t^{\delta}\bar{\beta}^{\delta t}E(t,\delta), \tag{A.35}
$$

where

$$
\begin{aligned}
E(t,\delta) &= t^{-\delta}E\left[\sup_{\Phi\in\mathcal{A}}r_t^2 g_{K+N}\left|\frac{\partial_i V_t}{V_t}\right|+(|\partial_i g_t|g_{K+N}\right. \\
&\quad \left.+t|g_{K+N}-\tilde{g}_{K+N}|+|\partial_i g_{K+N}|)|1-\frac{r_t^2}{\tilde{V}_t}|\right]^{\delta} \\
&\le E\left[\sup_{\Phi\in\mathcal{A}}r_t^2 g_{K+N}\left|\frac{\partial_i V_t}{V_t}\right|\right]^{\delta}+E\left[\sup_{\Phi\in\mathcal{A}}|\partial_i g_t|g_{K+N}\left|1-\frac{r_t^2}{\tilde{V}_t}\right|\right]^{\delta} \\
&\quad +E\left[\sup_{\Phi\in\mathcal{A}}|g_{K+N}-\tilde{g}_{K+N}|\left|1-\frac{r_t^2}{\tilde{V}_t}\right|\right]^{\delta}+E\left[\sup_{\Phi\in\mathcal{A}}|\partial_i g_{K+N}|\left|1-\frac{r_t^2}{\tilde{V}_t}\right|\right]^{\delta}.
\end{aligned}
$$

Note that $Er_t^{2\delta_*}$ and $E(\sup_{\Phi\in\mathcal{A}}g_t)^{\delta_*}$ are finite (see Corollary 3.1, (A.9)). In a similar argument and using the fact that $\tau_t^{-1}\partial_i\tau_t$ is bounded on $\mathcal{A}$, $E(\sup_{\Phi\in\mathcal{A}}\partial_i g_t)^{\delta_*}$ and $E(\sup_{\Phi\in\mathcal{A}}V_t^{-1}\partial_i V_t)^{\delta_*}$ are finite. Take $\delta=\delta_*/3$, then $E(t,\delta_*/3)$ is bounded, say by C_{**}. It follows that, $\forall\epsilon>0$,

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press

$$
\begin{aligned}
&P\left(\frac{1}{\sqrt{T}}\sum_{t=N+K+1}^{T}\sup_{\Phi\in\mathcal{A}}|\partial_i l_t(\Phi)-\partial_i\tilde{l}_t(\Phi)|>\epsilon\right)\\
&\quad\le(\sqrt{T}\epsilon)^{-\delta^*/3}E\left[\sum_{t=N+K+1}^{T}\sup_{\Phi\in\mathcal{A}}|\partial_i l_t(\Phi)-\partial_i\tilde{l}_t(\Phi)|\right]^{\delta^*/3}\\
&\quad\le(\sqrt{T}\epsilon)^{-\delta^*/3}\sum_{t=N+K+1}^{T}C_*^{\delta^*/3}t^{\delta^*/3}(\bar{\beta}^{\delta^*/3})^t C_{**}\\
&\quad\longrightarrow 0\quad\text{as } t\to\infty.
\end{aligned}
$$

Therefore, $\lim_{T\to\infty}\sqrt{T}\sup_{\Phi\in\mathcal{A}}\|\nabla L_T(\Phi)-\nabla\tilde{L}_T(\Phi)\|=0$ in probability. ■

Proof of Proposition 4.2. We first show that $\sqrt{T}(\hat{\Phi}_T-\Phi_0)\Longrightarrow N(0,(E\varepsilon_t^4-1)\Sigma(\Phi_0)^{-1})$. Note that $-\nabla L_T(\Phi_0)=H(L_T)(\bar{\Phi}_T)(\hat{\Phi}_T-\Phi_0)$ where $\bar{\Phi}_T$ is between Φ_0 and $\hat{\Phi}_T$. Since $\bar{\Phi}_T$ converges to Φ_0 a.s., it follows from Lemma 4.5(3) that $H(L_T)(\bar{\Phi}_T)$ converges to $\Sigma(\Phi_0)$ a.s. Note that $\Sigma(\Phi_0)$ is invertible. Therefore, $\sqrt{T}(\hat{\Phi}_T-\Phi_0)=-(\Sigma(\Phi_0))^{-1}(1+o_p(1))\sqrt{T}\nabla L_T(\Phi_0)$. The asymptotic normality follows from Lemma 4.5(2) and Slutsky's theorem.

Next we will show $\sqrt{T}(\tilde{\Phi}_T-\hat{\Phi}_T)=o_p(1)$. Note that $\nabla L_T(\tilde{\Phi}_T)-\nabla L_T(\hat{\Phi}_T)=H(L_T)(\breve{\Phi}_T)(\tilde{\Phi}_T-\hat{\Phi}_T)$ where $\breve{\Phi}_T$ is between $\tilde{\Phi}_T$ and $\hat{\Phi}_T$. On one hand, $\sqrt{T}(\nabla L_T(\tilde{\Phi}_T)-\nabla L_T(\hat{\Phi}_T))=\sqrt{T}(\nabla L_T(\tilde{\Phi}_T)-\nabla\tilde{L}_T(\tilde{\Phi}_T))$ converges to 0 in probability due to Lemma 4.5(4). On the other hand, $H(L_T)(\breve{\Phi}_T)$ converges to $\Sigma(\Phi_0)$ a.s. due to Lemma 4.5(3). Therefore, $\sqrt{T}(\tilde{\Phi}_T-\hat{\Phi}_T)=(\Sigma(\Phi_0))^{-1}(1+o_p(1))(\nabla L_T(\tilde{\Phi}_T)-\nabla\tilde{L}_T(\tilde{\Phi}_T))$ converges to 0 in probability.

An application of Slutsky's theorem yields that $\sqrt{T}(\tilde{\Phi}_T-\Phi_0)$ converges to $N(0,(E\varepsilon_t^4-1)\Sigma(\Phi_0)^{-1})$ in distribution. ■

https://doi.org/10.1017/S0266466614000334 Published online by Cambridge University Press