Econometric Analysis for Volatility Component Models

Fangfang Wang* Eric Ghysels†

September 21, 2011

Abstract

The volatility component models have received much 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 non-stationarities in asset price volatility. This paper revisits the component 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 non-stationarity 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.

*Department of Information and Decision Sciences, University of Illinois at Chicago. Email: ffwang@uic.edu

†Department of Economics, and Department of Finance, University of North Carolina at Chapel Hill. Email: eghysels@unc.edu

1 Introduction

Asset price volatility is persistent and several models have been proposed to capture this salient stylized fact. The 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, Ronald Gallant, Ghysels, and Tauchen (2003) and Adrian and Rosenberg (2008) among many 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), Horvath, Kokoszka, and Teyssière (2001), Horvath, Kokoszka, and Zhang (2006), Inclan and Tiao (1994), Kokoszka and Leipus (2000), Kulperger and Yu (2005), among others. To address the non-stationarity 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 non-stationarity - 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 of a multiplicative structure. The first to suggest a multiplicative component structure that accommodates non-stationarity is Engle and Rangel (2008), later extended by Engle, Ghysels, and Sohn (2008). These component models, also known as Spline-GARCH and GARCH-MIDAS respectively, feature a multiplicative decomposition of the 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 GARCH(1,1) process which mean-reverts to one. The low-frequency component picks up the non-stationarity. 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 non-parametric manner so that the unconditional variance is time varying. This makes the model much more flexible but at the cost of losing the mean-reverting property.

The GARCH-MIDAS model of Engle, Ghysels, and Sohn (2008) modified the dynamics of low-frequency volatility as a stochastic component "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).

The economic implications of the aforementioned component models and their empirical application have been studied intensively in Engle and Lee (1999), Engle and Rangel (2008), Engle, Ghysels, and Sohn (2008). However, the literature has not well covered the conditions that characterize non-stationarity issues of the components. This paper revisits the 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 non-stationarity 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 paper is organized as follows. Section 2 gives a brief overview of the volatility component models. We then outline the main results of the paper in Section 3, and contribution of this work. Section 4 explores the stationarity of the various component models. Model estimation and asymptote of the estimators are discussed in section 5. Section 6 gives the concluding remarks. Proofs are collected in an appendix.

2 An overview of component models

In this section, we will give a brief overview of the component models. Although most of our focus is on the multiplicative component model, we start with filling a gap in the literature pertaining to additive component models, that is the original Engle and Lee model. Denote by r_t the return on, say day t, and $h_t = Var(r_t|r_s, s < t)$. Engle and Lee (1999) extends the classic GARCH model by modeling the conditional volatility h_t as sum of a so-called trend component and a transitory component. To be specific,

$$\begin{aligned} h_t &= \tau_t + g_t \\ g_t &= \alpha(r_{t-1}^2 - \tau_{t-1}) + \beta g_{t-1} \\ \tau_t &= \omega + \rho\tau_{t-1} + \phi(r_{t-1}^2 - h_{t-1}) \end{aligned} \tag{2.1}$$

In addition, $r_t/\sqrt{h_t}$, denoted by ε_t, is assumed to be iid $N(0,1)$. The parameters appearing in (2.1) are positive, and satisfy $\alpha + \beta < \rho < 1$, $\phi < \beta$. This 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, $\omega/(1-\rho)$. Therefore, τ_t is referred to as trend component and g_t is transitory component.

Instead of modeling conditional volatility as a sum of two components, the spline-GARCH model of Engle and Rangel (2008) and the GARCH-MIDAS model of Engle, Ghysels, and Sohn (2008) structure the volatility as a product of long-run and short-run components. The spline-GARCH model is defined as

$$\begin{aligned} h_t &= g_t \tau_t \\ g_t &= (1-\alpha-\beta) + \alpha \tfrac{r_{t-1}^2}{\tau_{t-1}} + \beta g_{t-1} \\ \tau_t &= c \exp(w_0 t + \textstyle\sum_{i=1}^{k} w_i (t - t_{i-1})^2 1_{\{t > t_{i-1}\}}) \end{aligned} \tag{2.2}$$

where $\{0 = t_0 < t_1 < t_2 < \ldots < t_k = T\}$ is a partition of the time horizon T into k equally spaced intervals. The high-frequency component g is a unit GARCH(1,1) process. The low-frequency component τ is deterministic, and hence it equals the unconditional variance, ie $Var(r_t) = \tau_t$.

The GARCH-MIDAS model has a similar structure – g stays the same, while τ is modified to be stochastic:

$$\begin{aligned} h_t &= g_t \tau_t \\ g_t &= (1-\alpha-\beta) + \alpha \tfrac{r_{t-1}^2}{\tau_{t-1}} + \beta g_{t-1} \\ \tau_t &= m + \theta \textstyle\sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, RV_t = \textstyle\sum_{j=0}^{N-1} r_{t-j}^2 \end{aligned} \tag{2.3}$$

where $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$. Alternatively, Engle, Ghysels, and Sohn (2008) proposed a companion model which is able to handle seasonal/periodic time series. Suppose that $r_{t,i}$ is the return on day i and period t, and there are a fixed amount of days in a period.[1] The model goes as

$$\begin{aligned} r_{t,i} &= \sqrt{\tau_t g_{t,i}} \varepsilon_{t,i}, \quad \varepsilon_{t,i} \overset{iid}{\sim} N(0,1), \quad 1 \le i \le N, t \in \mathbb{Z} \\ g_{t,i} &= (1-\alpha-\beta) + \alpha \tfrac{r_{t,i-1}^2}{\tau_t} + \beta g_{t,i-1} \\ \tau_t &= m + \theta \textstyle\sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, \quad RV_t = \textstyle\sum_{i=1}^{N} r_{t,i}^2 \end{aligned} \tag{2.4}$$

where N is the number of days in a period, and $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ $m > 0$. $\varphi_k(\omega)$ are nonnegative functions of ω such that $\sum_{k=1}^{K} \varphi_k(\omega) = 1$. Since the long-run component varies from period to period but stay the same within a period, the GARCH-MIDAS model

[1] This assumption is not imposed in Engle, Ghysels, and Sohn (2008). We add this for theoretical derivation.

of (2.4) is also referred to as GARCH-MIDAS model with fixed time span realized volatility (RV) in Engle, Ghysels, and Sohn (2008). Therefore, the model featured in (2.3) is referred to as the GARCH-MIDAS model with rolling window RV. Note that the short-run component has short memory, Engle, Ghysels, and Sohn (2008) specifically assumes that the short-run component in (2.4) starts from the unconditional level at the beginning of each period, i.e., $E(g_{t,1}|\mathcal{F}_{t-1}) = 1$.

3 Main results

This paper revisits the component models from a probabilistic and statistical perspective. We attempt to explore stationarity of the underlying time series and investigate the sampling behavior of the maximum likelihood estimator of the GARCH-MIDAS model which features heavy tails.

A typical approach to handle stationarity of a process is to relate it with a stochastic difference equation via Markovian representation. Suppose that Y_t is R^d-valued random vector ($d \geq 1$) and it satisfies the following stochastic difference equation

$$Y_t = A_t Y_{t-1} + B_t, \quad t \in \mathbb{Z} \tag{3.1}$$

where A_t is $R^{d\times d}$-valued random matrix and B_t is R^d-valued random vector, and $\{(A_t, B_t)\}$ are strictly stationary and ergodic. The stationarity of Y_t in the form (3.1) has been studied extensively in the literature, see for instance Pham (1985), Pham (1986), Brandt (1986), Bougerol (1987), Meyn and Caines (1991), Bougerol and Picard (1992b), Bougerol and Picard (1992a), Glasserman and Yao (1995), and among others.

The top Lyapunov exponent associated with $\{A_t\}$ is $\gamma(A) = \inf_{t\in\mathbb{N}} E(\frac{1}{t}\log\|A_t A_{t-1}\dots A_1\|)$ provided that $E\log^+\|A_0\| < \infty$. A more tractable expression of $\gamma(A)$, due to Furstenberg and Kesten (1960) and the subadditive ergodic theory of Kingman (1973), is

$$\gamma(A) = \lim_{t\to\infty}\frac{1}{t}E\log\|A_t A_{t-1}\dots A_1\| \overset{a.s.}{=} \lim_{t\to\infty}\frac{1}{t}\log\|A_t A_{t-1}\dots A_1\| \tag{3.2}$$

The stability of model (3.1) closely relates to $\gamma(A)$. Suppose further that $E\log^+\|B_0\| < \infty$. Bougerol and Picard (1992b) shows that $\gamma(A) < 0$ is a *necessary and sufficient (N&S)* condition under which (3.1) has a unique stationary ergodic solution *if the model is irreducible and* $\{(A_t, B_t)\}$ *are iid*. The irreducibility of model (3.1) is equivalent to the rank condition of the controllability matrix of the linear system (see Meyn and Caines (1991), Kristensen (2009)). However for a general situation where $\{(A_t, B_t)\}$ are just strictly stationary and ergodic, Glasserman and Yao (1995) shows that $\gamma(A) < 0$ is only a *sufficient* condition.

The models we will discuss next are similar in spirit to (3.1). What is new about our work is we will show in Section 4 that $\gamma(A) < 0$ is also a necessary condition for strictly stationary ergodic $\{(A_t, B_t)\}$.

Moreover, we give explicit conditions - both necessary and sufficient - for $\gamma(A) < 0$. Our conditions are expressed in terms of model parameters. This is important, because it makes easier to track stationarity of the underlying process for given parameters and to manage the model inference. The evaluation of Lyapunov exponents, in turn, requires to study the limiting behavior of product of random matrices. The study on matrix product goes back to Furstenberg and Kesten (1960), Furstenberg (1963) (See Goldsheid (1991) for a comprehensive review). Kesten and Spitzer (1984) studied the convergence of product of iid nonnegative matrices. Cohen and Newman (1984) focused on a product of iid matrices whose entries are symmetric stable random variables, while Newman (1986) extended to the case when the entries are normal distributed. Peres (1992) examined a product of positive matrices with Markovian dependence, see also Yao (2001). An explicit formulae for Lyapunov exponents of infinite product of random matrices are not available in general. Therefore, most of the discussion regarding chaotic behavior via Lyapunov exponents focuses on simulation approach, see for instance Lu and Smith (1997), Whang and Linton (1999), Vanneste (2010), among others.

Last but not least, we pointed out in Section 4 that the GARCH-MIDAS model with rolling window RV has heavily tailed marginal distribution, which is a desirable property for financial time series modeling. And Section 5 studies the maximum likelihood estimation of the GARCH-MIDAS model. We establish the consistency and asymptotic normality of the estimators.

We close this section with some notation. For matrix $A \in \mathbb{R}^{n\times n}$, we say $A \geq 0$ if $ent_{i,j}A \geq 0$. For matrices A and B, we say $A \leq B$ if $B - A \geq 0$. I_N represents a $N \times N$ identify matrix. The spectral radius of matrix A is defined as $\rho(A) = \max_i(|\lambda_i|)$ where $\lambda_i \in \mathbb{C}$ is the eigenvalue of A. Since Lyapunov exponents are invariant to multiplicative matrix norms, we consider $\|V\| = \max_{i=1,2,\ldots,n} |v_j|$ 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 $M \in \mathbb{R}^{n\times n}$.

4 Stationarity

This section explores the stationarity of three models: the component model of Engle and Lee, and the GARCH-MIDAS models with rolling-window RV and with fixed time span RV.

4.1 Component model of Engle and Lee

The component model of Engle and Lee (1999) consists of two additive GARCH(1,1) components and the conditional volatility therefore has a structure of GARCH(2,2) (see Engle and Lee (1999)). Namely,

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} \tag{4.1}$$

where $\alpha_0 = \omega(1-\alpha-\beta) > 0, \alpha_1 = \phi+\alpha > 0$, $\alpha_2 = -(\phi(\alpha+\beta)+\alpha\rho) < 0$, $\beta_1 = \rho+\beta-\phi > 0$, and $\beta_2 = \phi(\alpha+\beta) - \rho\beta < 0$. Therefore model (2.1) is also referred to as restricted GARCH (2,2). Engle and Lee (1999) has pointed out that the return process r_t defined in (2.1) is weakly stationary if $\rho < 1$ and $\alpha + \beta < 1$. Albeit the component model has a GARCH structure, it does not have all nonnegative coefficients – a condition which is always assumed in studying the stationarity of GARCH models, see, for instance, Nelson (1990), Bougerol and Picard (1992a), Chen and An (1998), Carrasco and Chen (2002), Francq and Zakoian (2006), Meitz and Saikkonen (2008), Lindner (2009), Kristensen (2009), among others.

Define $Y_t = (h_{t+1}, h_t, r_t^2)'$, $B = (\alpha_0, 0, 0)'$, and

$$A_t \equiv A(\varepsilon_t^2) = \begin{pmatrix} \beta_1 + \alpha_1\varepsilon_t^2 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ \varepsilon_t^2 & 0 & 0 \end{pmatrix}. \tag{4.2}$$

(r_t, h_t) is strictly stationary ergodic if and only if $Y_t = A_t Y_{t-1} + B$ has 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 and only if $\Phi(Z)$ and $\Theta(Z)$ have no common roots and all the roots to $\Phi(Z)$ lie outside the unit circle (see Meyn and Caines (1991), Kristensen (2009)). We have the following

Proposition 4.1. *Y_t is irreducible if $\alpha + \beta < \rho < 1$, $\phi < \beta$.*

Note that A_t's are iid. Y_t is strictly stationary ergodic if and only if the top Lyapunov exponent associated with $\{A(\varepsilon_t^2)\}$, denoted by $\gamma(A)$, is negative according to Bougerol and Picard (1992b). Next proposition provides a sufficient condition, which is explicitly stated in terms of the parameters, to guarantee $\gamma(A) < 0$. Note that *if A_t's were nonnegative*, then $\gamma(A) \leq \rho(EA_1)$ (see Kesten and Spitzer (1984)). It is *not* the case however. Therefore in proposition 4.2, we try to find a 'mirror image' of A_t and then bound $\gamma(A)$ by the spectral radius of the 'image'.

Proposition 4.2. *r_t^2 defined in (2.1) is strictly stationary if $\alpha + \beta < \rho < 1$, $\phi < \beta$, $(1 + \rho)(1 + \alpha + \beta) \leq 2$, and they are positive.*

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 $a = \frac{1-(\alpha_1+\beta_1)}{4} > 0$ (since $\alpha_1 + \beta_1 = \alpha + \beta + \rho < 1$ by assumption). Let $\pi = \frac{1+\alpha_1+\beta_1}{2} < 1$ and $B > 0$ be such that $\frac{\alpha_0+1}{B} < 1 - \pi$. It is easy to verify that $E[V(Y_t)|Y_{t-1}] \leq \alpha_0 + \pi V(Y_{t-1})$. Define $K = \{k \in \mathbb{R}^3 : V(k) \leq B\}$, then $E[V(Y_t)|Y_{t-1} = y]$ is bounded for $y \in K$. On K^c,

$$E[V(Y_t)|Y_{t-1} = y] \leq \alpha_0 + \pi V(y) \leq (\frac{\alpha_0 + 1}{B} + \pi)V(y) - 1$$

It follows from Mokkadem (1990) or Theorem 1 of Carrasco and Chen (2002) that r_t is also geometrically ergodic under the assumptions stated in Proposition 4.2, and hence β-mixing.

4.2 The GARCH-MIDAS models

We next consider the GARCH-MIDAS with rolling window RV, i.e. model (2.3). Note that $\tau_t = m + \theta\sum_{l=1}^{N+K-1} c_l r_{t-l}^2$, where c_l's are combinations of the weights $\varphi_k(\omega)$ and satisfy $\sum_{l=1}^{N+K-1} c_l = N\sum_{k=1}^{K}\varphi_k(\omega) = N$. Introducing $\epsilon_t = \sqrt{g_t}\varepsilon_t$, $r_t = \sqrt{\tau_t}\epsilon_t$ and hence r_t can be viewed as a semi-strong ARCH$(N + K - 1)$ with multiplicative GARCH error. And the GARCH error is strictly stationary ergodic under the conditions $\alpha > 0, \beta > 0$ and $\alpha + \beta < 1$, but it may not have finite second moment.

Consider the following stochastic difference equation

$$Y_t = A_t(\vec{c})Y_{t-1} + B_t. \tag{4.3}$$

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

$$A_t(\vec{c}) = \begin{pmatrix} \theta g_t \varepsilon_t^2 c_1 & \ldots & \theta g_t \varepsilon_t^2 c_{N+K-2} & \theta g_t \varepsilon_t^2 c_{N+K-1} \\ 1 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & 0 \\ \vdots & & & \\ 0 & \ldots & 1 & 0. \end{pmatrix}. \tag{4.4}$$

Therefore, the stationarity of process (2.3) is equivalent to the stability of dynamic system (4.3). Denoted by $\gamma(A)$ the top Lyapunov exponent defined on $\{A_t(\vec{c})\}$. Note that $\{A_t(\vec{c})\}$ is strictly stationary ergodic. It follows from Glasserman and Yao (1995) that Equation (4.3) has a unique stationary ergodic solution if $\gamma(A) < 0$. We will show next that $\gamma(A) < 0$ is also a *necessary* condition. Define $S_{t,n} = A_t(\vec{c})A_{t-1}(\vec{c}) \ldots A_{t-n+1}(\vec{c})$ for $n > 0$ and $S_{t,0} = 1$. Then

$$Y_t = S_{t,k}Y_{t-k} + \sum_{n=0}^{k-1} S_{t,n}B_{t-n}. \tag{4.5}$$

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

$$H_{N+K-1} = \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}, \quad G_{N+K-1} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}. \tag{4.6}$$

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

Proposition 4.3. *Suppose that $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$. Model (4.3) has a unique strictly stationary ergodic solution if and only if $\gamma(A) < 0$.*

Note that Model (2.3) is also equivalent to a stochastic difference equation as follows:

$$X_t = \breve{A}_t X_{t-1} + B. \tag{4.7}$$

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

$$
\breve{A}_t = \begin{pmatrix} \theta c_1 g_t \varepsilon_t^2 & \theta c_2 & \ldots & \theta c_{N+K-2} & \theta c_{N+K-1} \\ g_t \varepsilon_t^2 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & 1 & 0. \end{pmatrix}. \tag{4.8}
$$

The results stated in Proposition 4.3 can also apply to Model (4.7) for top Lyapunov exponent defined on $\{\breve{A}_t\}$ – using a similar argument. However since it is easier to work with (4.4) for the purpose of top Lyapunov exponent evaluation which will be discussed next, we do not pursue expression (4.7) here.

The top Lyapunov exponent controls chaotic behavior of dynamic system. Next we look for tractable conditions - both sufficient and necessary - to render a negative γ. We analyze three cases: (1) $K = 1, N = 1$, (2) $K = 1, N > 1$, and (3) $K > 1$, $N \geq 1$. The simplest situation is when $KN = 1$. In this case, $\gamma(A) = E \log(\theta g_0 \varepsilon_0^2)$. Therefore,

Proposition 4.4. *Suppose that $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$. For $K = N = 1$, model (2.3) has a unique strictly stationary ergodic solution if $\theta \leq 1$.*

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

$$
A_t(\iota) = \begin{pmatrix} \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 & \ldots & \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}. \tag{4.9}
$$

where ι represents a vector of 1's. Define $M(a) = aH_N + G_N$ for $a > 0$. And hence $A_t(\iota) = M(\theta g_t \varepsilon_t^2)$. It is easy to verify that the matrix $M(a)$ is primitive and hence irreducible and aperiodic. It follows from the Perron-Frobenius theory that the spectral radius of $M(a)$, denoted by $\rho(M(a))$, is the maximal positive root of $f_a(\lambda) = det(\lambda I_N - M(a))$ – the characteristic equation of $M(a)$, and it is simple. Matrix of this type is encountered a lot when one expresses an autoregressive model using a Markovian representation. We first look at a few properties regarding the spectral radius of $M(a)$.

Lemma 4.1. *Consider matrix $M(a)$ with $a > 0$.*

1. *For any $k > 0$, $\rho(M(a)) \leq k$ if and only if $f_a(k) \geq 0$;*

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 $\zeta = (1-\alpha-\beta)/(1-\beta)$. Note that $g_t \geq \zeta$ with probability one.

Proposition 4.5. *Suppose that $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$. For $K = 1$ and $N > 1$, the top Lyapunov exponent associated with $A_t(\iota)$ is negative if $\theta \leq \zeta^{N-1}/(1+\zeta+\ldots+\zeta^{N-1})$.*

Proposition 4.5 can be extended to a more general situation.

Proposition 4.6. *Suppose that $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$. For $K > 1$ and $N \geq 1$, the top Lyapunov exponent associated with $A_t(\vec{c})$ is negative if $\theta \leq \zeta^{K+N-2}/(c_1\zeta^{K+N-2} + c_2\zeta^{K+N-3} + \ldots + c_{K+N-2}\zeta + c_{K+N-1})$.*

Therefore, combining propositions 4.4, 4.5, and 4.6, we have the following

Proposition 4.7. *Suppose that $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$, $m > 0$. Model (4.3) has a unique strictly stationary ergodic solution if*

$$\theta \leq \zeta^{K+N-2}/(c_1\zeta^{K+N-2} + c_2\zeta^{K+N-3} + \ldots + c_{K+N-2}\zeta + c_{K+N-1}), \tag{4.10}$$

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

We next give the necessary condition for $\gamma < 0$ which might be more useful under certain circumstances.

Proposition 4.8. *Suppose that $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$, $m > 0$. Model (4.3) has a unique strictly stationary ergodic solution only if $\theta < 1/(N\zeta)$.*

Note that under the assumptions of Proposition 4.7, the unique strictly stationary ergodic solution to equation (4.3) is

$$Y_t = \sum_{n=0}^{\infty} S_{t,n} B_{t-n}, \quad w.p.1 \tag{4.11}$$

This solution is not integrable. Consider $K = N = 1$. Then

$$r_t^2 = m \sum_{n=0}^{\infty} \theta^n \left(\prod_{j=0}^{n} g_{t-j} \varepsilon_{t-j}^2 \right).$$

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 $u_{t-1}^{k_1}, u_{t-2}^{k_2}, u_{t-3}^{k_3}, \ldots, u_{t-n}^{k_n}$ and $u_t = \alpha \varepsilon_t^2 + \beta$, $0 \leq k_i \leq n$. Therefore,

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

$Er_t^2 < \infty$ if and only if $\limsup_{n\to\infty} \left[m\theta^n \sum_{k=1}^{n+1} E \left(G_k^{(n)} \prod_{j=1}^{n} \varepsilon_{t-j}^2 \right) E(g_0^k) \right]^{1/(n+1)} < 1$ due to Cauchy root test. Therefore, the necessary condition for $Er_t^2 < \infty$ is $E(u_0^{n+1}) < 1$ for any n (see Bollerslev (1986), and He and Terasvirta (1999)). This is impossible because ε_t has unbounded support – ε_t is $N(0,1)$. For the general situation, 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 $e_1 = (1, 0, \ldots, 0)^T$. By a similar argument one can show that Y_t defined in (4.11) is not in L_1.

In fact r_t^2 defined by (2.3) is not covariance stationary. Consider (4.5) with $k = t$. Note that for $K = N = 1$, we have

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

A similar argument yields

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

And $Er_t^2 < \infty$ only if $E(g_1^l r_0^2) < \infty$ and $E(g_1^l) < \infty$ for $1 \leq l \leq n+1$ and $0 \leq n \leq t-1$. Therefore, in general, we have

Proposition 4.9. *Consider r_t defined by model (2.3). 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 $t > t_0$.*

Proposition 4.9 implies the marginal distribution of the stationary version of r_t is heavily tailed. This is a desirable property in modeling financial time series. Though the strictly stationary ergodic solution is not integrable, we will show next that it is 'fractionally' integrable. We first present a general result.

Proposition 4.10. *Consider the stochastic difference equation (3.1), 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 4.1. *There exists* $0 < \delta^* < 1$ *such that* $E(r_t^2)^{\delta^*} < \infty$ *for* r_t *defined by model (2.3), and hence* $E(\log r_t^2) < \infty$*.*

Since Er_t^2 does not exist, Corollary 4.1 has important implication in statistical inference, especially in studying the asymptote of maximum likelihood estimators. Most of the discussion in Section 5 will rely on Corollary 4.1.

We close this section by discussing the GARCH-MIDAS model with fixed time span RV. Note that $r_{t,i}$ defined by model (2.4) is not strictly stationary, but we will show that it forms a white noise.

Proposition 4.11. *Suppose that* $\alpha > 0$*,* $\beta > 0$*,* $\alpha + \beta < 1$*,* $\theta > 0$ *and* $m > 0$*.* $\{r_{t,i}\}$ *defined in (2.4) is a white noise with variance* $m/(1 - \theta N)$ *if* $0 < \theta < 1/N$*.*

5 Model estimation

In this section, we will investigate maximum likelihood estimation of multiplicative component model: the GARCH-MIDAS model with rolling-window RV. The (quasi-) maximum likelihood estimation of conditionally heteroscedastic time series has been discussed in Weiss (1986), Bollerslev and Wooldridge (1992), Lee and Hansen (1994), Lumsdaine (1996), Berkes, Horváth, and Kokoszka (2003), Francq and Zakoian (2004), Straumann and Mikosch (2006), among many others. Due to the complexity of the GARCH-MIDAS model – the volatility is driven by two components and Er_t^2 does not exist, the standard results can not carry over to this model directly.

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

$$g_t(\Phi) = (1 - \alpha - \beta) + \alpha \frac{r_{t-1}^2}{\tau_{t-1}(\Phi)} + \beta g_{t-1}(\Phi) \tag{5.1}$$

$$\tau_t(\Phi) = m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k} \tag{5.2}$$

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

$$r_t = \sqrt{g_t(\Phi_0)\tau_t(\Phi_0)}\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0,1) \tag{5.3}$$

for $t \in \mathbb{Z}$. Given a finite record of r_t: $\{r_t, 1 \le t \le T\}$ where $T \gg N + K$, define

$$\begin{array}{rcl} \tilde{g}_t & = & (1 - \alpha - \beta) + \alpha \frac{r_{t-1}^2}{\tilde{\tau}_{t-1}} + \beta \tilde{g}_{t-1} \\ \tilde{\tau}_t & = & m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k} \end{array} \tag{5.4}$$

for $t = N + K + 1, \ldots, T$, and let $\tilde{g}_{N+K}$ be an arbitrary number. The maximum likelihood estimation of Φ_0 is done by minimizing

$$\tilde{L}_T(\Phi) = \frac{1}{T - N - K} \sum_{t=N+K+1}^{T} \tilde{l}_t = \frac{1}{T - N - K} \sum_{t=N+K+1}^{T} \log \tilde{V}_t + \frac{r_t^2}{\tilde{V}_t} \tag{5.5}$$

over $\Phi \in \mathcal{U}$, where $\tilde{V}_t = \tilde{g}_t \tilde{\tau}_t$. The estimator is then denoted by $\tilde{\Phi}_T$.

Property of $\tilde{\Phi}_T$ closely relates to the choice of parameter space $\mathcal{U}$. Define

$$\begin{aligned} \mathcal{U}_1 = \{\Phi = (\alpha, \beta, m, \theta, \omega)' \in \mathcal{R}^5 : \alpha > 0, \beta > 0, m > 0, \alpha + \beta < 1, \\ 0 < \theta \le \frac{\zeta^{K+N-2}}{c_1\zeta^{K+N-2} + c_2\zeta^{K+N-3} + \ldots + c_{K+N-2}\zeta + c_{K+N-1}}\} \end{aligned} \tag{5.6}$$

where $\zeta = (1 - \alpha - \beta)/(1 - \beta)$. r_t^2 implied by (5.3) is strictly stationary ergodic if $\mathcal{U} = \mathcal{U}_1$. For the rest of the paper, we assume that

Assumption 5.1. *$\mathcal{U}$ is a compact subset of $\mathcal{U}_1$.*

Therefore we exclude the nonstationary case. Jensen and Rahbek (2004b), Jensen and Rahbek (2004a) studied the asymptotic inference for non-stationary ARCH(1) and GARCH(1,1), and Linton, Pan, and Wang (2010) explored the non-stationary semi-strong GARCH(1,1).

The model under study can be viewed as a semi-strong ARCH with stationary GARCH error (see section 4.2) which shares some similarity with the semi-strong GARCH(1,1) discussed in Linton, Pan, and Wang (2010). However, the extension of the results in Linton, Pan, and Wang (2010) to the GARCH-MIDAS is not straightforward. The GARCH error is parameterized and the condition imposed in Linton, Pan, and Wang (2010) is too strong. Due to the complex structure, we will leave the inference on non-stationary GARCH-MIDAS as a future research.

Consider a companion estimator, denoted as $\hat{\Phi}_T$, which minimizes

$$L_T(\Phi) = \frac{1}{T-(N+K)} \sum_{t=N+K+1}^{T} l_t(\Phi) = \frac{1}{T-(N+K)} \sum_{t=N+K+1}^{T} \log V_t(\Phi) + \frac{r_t^2}{V_t(\Phi)} \tag{5.7}$$

for $\Phi \in \mathcal{U}$ where $V_t(\Phi) = g_t(\Phi)\tau_t(\Phi)$ (or simply $V_t = g_t\tau_t$), and $g_t(\Phi)$ and $\tau_t(\Phi)$ are defined in (5.1) and (5.2). Clearly, $\tau_t = \tilde{\tau}_t$. Although $\hat{\Phi}_T$ is not feasible, it is theoretically tractable. We will show $E\sup_{\Phi\in\mathcal{U}} |\log V_t(\Phi) + r_t^2/V_t(\Phi)|$ is finite, though Er_t^2 does not exist. First of all, we present some useful results. Since they are pretty elementary, proofs are skipped.

Lemma 5.1. *Suppose that $\{a_i\}_{i=1}^n$ are nonnegative. Let $a = \sum_{i=1}^n a_i$ and $a > 0$. Then*

$$\log(\sum_{i=1}^n a_i) \le \log n + \sum_{i=1}^n \frac{a_i}{a}\log a_i \le \log n + \sum_{i=1}^n \log^+ a_i \tag{5.8}$$

$$\log^+(\prod_{i=1}^n a_i) \le \sum_{i=1}^n \log^+ a_i \tag{5.9}$$

Lemma 5.2. *Define $l_t(\Phi) = \log V_t(\Phi) + r_t^2/V_t(\Phi)$. Under Assumption 5.1, $E\sup_{\Phi\in\mathcal{U}} |\log(g_t)|$, $E\sup_{\Phi\in\mathcal{U}} |\log(\tau_t)|$, $E\sup_{\Phi\in\mathcal{U}} r_t^2/V_t$ are finite, and hence $E\sup_{\Phi\in\mathcal{U}} |l_t(\Phi)| < \infty$.*

Next we need to show that Φ_0 is identifiably unique. In other words, $E[\log g_t(\Phi) + \log\tau_t(\Phi) + r_t^2/(g_t(\Phi)\tau_t(\Phi))]$ is minimized at Φ_0. And we need the following assumption:

Assumption 5.2. *For each $k = 1, 2, \ldots, K$, $\omega \mapsto \varphi_k(\omega)$ is a C^2-diffeomorphism.*

Lemma 5.3. *Under Assumptions 5.1 and 5.2, $El_t(\Phi) > El_t(\Phi_0)$ for $\Phi \in \mathcal{U}$ and $\Phi \neq \Phi_0$.*

Lemmas 5.2 and 5.3 imply that $\hat{\Phi}_T$ is identifiably unique and it converges to the target Φ_0 almost surely, using a standard compactness argument (see, for instance, Gallant and White (1988), Straumann and Mikosch (2006)). To show the consistency of $\tilde{\Phi}_T$, we need the following lemma:

Lemma 5.4. *Under assumption 5.1,*

$$\lim_{T\to\infty}\sup_{\Phi\in\mathcal{U}}|L_T(\Phi)-\tilde{L}_T(\Phi)|=0 \quad \textit{almost surely.} \tag{5.10}$$

Therefore we have the following statement regarding $\hat{\Phi}_T$ and $\tilde{\Phi}_T$:

Proposition 5.1. *Under Assumptions 5.1 and 5.2, both $\hat{\Phi}_T$ and $\tilde{\Phi}_T$ are identifiably unique and they converge to Φ_0 with probability one.*

Further, we will show that $\tilde{\Phi}_T$ is also asymptotic normal, with an additional assumption:

Assumption 5.3. $\Phi_0 \in \mathcal{U}^0$, *the interior of set* $\mathcal{U}$.

Lemma 5.5. *Define* $\Sigma(\Phi) \doteq E\left(V_t^{-2}(\Phi)\nabla V_t(\Phi)\nabla V_t(\Phi)'\right)$. *Under Assumptions 5.1 and 5.2,*

1. $\Sigma(\Phi)$ *exists and it is positive definite at* Φ_0.
2. $\sqrt{T}\nabla L_T(\Phi_0) \Longrightarrow N(0, E(\varepsilon_t^2-1)^2\Sigma(\Phi_0))$.
3. *Let* $B(\Phi_0, 1/N) = \{\Phi \in \mathbb{R}^5 : \|\Phi-\Phi_0\| < 1/N\}$. *Under additional Assumption 5.3,*

$$\lim_{T\to\infty}\limsup_{N\to\infty}\sup_{\Phi\in B(\Phi_0,1/N)\cap\mathcal{U}}\|H(L_T)(\Phi)-\Sigma(\Phi_0)\|=0 \quad \textit{a.s.} \tag{5.11}$$

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

Lemma 5.5 allows us to establish the asymptotic normality of $\tilde{\Phi}_T$.

Proposition 5.2. *Under assumptions 5.1, 5.2, and 5.3,*

$$\sqrt{T}(\tilde{\Phi}_T-\Phi_0) \Longrightarrow N(0, 2\Sigma(\Phi_0)^{-1}).$$

6 Conclusion

This paper investigated the distributional properties of three component models: the restricted GARCH(2,2) model of Engle and Lee (1999), the GARCH-MIDAS models of Engle, Ghysels, and Sohn (2008) with rolling window RV and with fixed time span RV. The restricted GARCH(2,2) structured the conditional variance as a sum of low-frequency and high-frequency components. It was shown that, under certain regularity conditions, it was strictly stationary ergodic and β-mixing. The GARCH-MIDAS models consist of two multiplicative volatility components. For GARCH-MIDAS model with fixed time span RV, we showed that it could admit a covariance stationary solution in a specific parameter space. We also derived sufficient and necessary conditions for the existence and uniqueness of strictly stationary ergodic solution to the GARCH-MIDAS model with rolling window RV. Further, this paper studied the asymptote of maximum likelihood estimation of GARCH-MIDAS model with rolling window RV which features heavy tailed marginal distribution.

Appendix

Proof of Proposition 4.1: 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, ϕ are not 0, $\Phi(Z)$ and $\Theta(Z)$ have no common roots, and hence Y_t is irreducible. ■

Proof of Proposition 4.2: We only need to show $\gamma(A) < 0$. Let

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

where

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

Note that by induction $\|M_t\| \leq \|\tilde{M}_t\|$ for any t and $\lim_{t\to\infty}\frac{1}{t}\log\|\tilde{M}_t\| < 0$ if $\beta_1+\alpha_1-\beta_2-\alpha_2 \leq 1$ (see Bougerol and Picard (1992a)). Therefore $\gamma(A) < 0$ if $(1+\rho)(1+\alpha+\beta) \leq 2$ ■

Proof of Proposition 4.3: Note that $1 \leq \|A_0(\vec{c})\| \leq \theta g_0\varepsilon_0^2\|H_{N+K-1}D(\vec{c})\| + \|G_{N+K-1}\| \leq N\theta g_0\varepsilon_0^2 + 1$. And hence $E(\log\|A_0(\vec{c})\|) \leq E\log(N\theta g_0\varepsilon_0^2 + 1) \leq E(N\theta g_0\varepsilon_0^2) < \infty$.

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

Suppose that (4.3) has a unique strictly stationary ergodic solution. Let $t = 0$ in (4.5). 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,

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

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

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

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 4.1: 1. Note that $f_a(\lambda) = \lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \ldots - a\lambda^2 - a\lambda - a$. For $|\lambda| > k$,

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

And $\rho(M(a))$ 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 $f(\lambda) = \lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \ldots - a\lambda^2 - a\lambda - a = 0$,

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

where $g(\lambda) = \frac{1}{h(\lambda)}$ and $h(\lambda) = \lambda^{N-1} + \lambda^{N-2} + \ldots + \lambda^2 + \lambda + 1$. 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 (.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, $0 = (1 + g')\lambda'' + g''(\lambda')^2$. 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 $\Delta = 2(h'(\lambda))^2 - h(\lambda)h''(\lambda) > 0$. Note that

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

$$h^{''}(\lambda) = \frac{N(N-1)\lambda^{N-2}}{\lambda - 1} - \frac{2N\lambda^{N-1}}{(\lambda-1)^2} + \frac{2(\lambda^N - 1)}{(\lambda-1)^3}.$$

Therefore,

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

Define

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

Then $D^{'}(\lambda) = (N-1)(N+1)\lambda^N - (N+1)N\lambda^{N-1} + (N+1)$ and $D^{''}(\lambda) = (N-1)N(N+1)\lambda^{N-2}(\lambda-1)$. 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 4.5: 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 iid and irreducible and ε has unbounded support, then $\tilde{\gamma} < 0$ if $\rho[E(\tilde{A}_0)] \le 1$ (see Kesten and Spitzer (1984), theorem 2). Note that $E(\tilde{A}_0) = \frac{1}{\zeta}M(\theta\zeta)$ and $\rho[E(\tilde{A}_0)] \le 1$ is equivalent to $\rho[M(\theta\zeta)] \le \zeta$, or equivalently $f_{\theta\zeta}(\zeta) = det(\zeta I - M(\theta\zeta)) \ge 0$ (by Lemma 4.1) which is satisfied if $\theta \le \zeta^{N-1}/(1 + \zeta + \ldots + \zeta^{N-1})$. ∎

Proof of Proposition 4.6: Note that $A_t(\vec{c}) = \theta g_t \varepsilon_t^2 H_{N+K-1} D(\vec{c}) + G_{N+K-1}$, where $D(\vec{c}) = Diag(c_1, \ldots, c_{N+K-1})$. $1 \le \|A_0(\vec{c})\| \le \theta g_0 \varepsilon_0^2 \|H_{N+K-1} D(\vec{c})\| + \|G_{N+K-1}\| \le N\theta g_0 \varepsilon_0^2 + 1$. And 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 4.5, 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

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

The matrix $\theta\zeta H_{N+K-1}D(\vec{c}) + G_{N+K-1}$ is primitive. Similarly to the discussion in the proof of Lemma 4.1(1), $\rho(\theta\zeta H_{N+K-1}D(\vec{c}) + G_{N+K-1}) \leq \zeta$ if and only if $f(\zeta) \geq 0$, or equivalently, $\theta \leq \zeta^{K+N-2}/(c_1\zeta^{K+N-2} + c_2\zeta^{K+N-3} + \ldots + c_{K+N-2}\zeta + c_{K+N-1})$. ■

Proof of Proposition 4.8: Note that the top Lyapunov exponent associated with $A_t(\vec{c})$, denoted by γ, should be strictly negative. And 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,

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

where $S_{t,t}(i,j)$ is the $(i,j)^{th}$ entry of $S_{t,t}$. 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 iid. Then we have

$$\sum_{i,j=1}^{N+K-1} (EL_1)^t(i,j) = E\sum_{i,j=1}^{N+K-1} (L_tL_{t-1}\ldots L_1)(i,j) < (N+K-1)e^{t\gamma/2} \qquad (.4)$$

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

Let $f(\lambda) = \det(\lambda I_{N+K-1} - E(L_t))$. And $f(\lambda) = \lambda^{N+K-1} - \theta\zeta c_1\lambda^{N+K-2} - \theta\zeta c_2\lambda^{N+K-3} - \ldots - \theta\zeta c_{N+K-2}\lambda - \theta\zeta c_{N+K-1}$. A similar argument to the proof of Lemma 4.1 yields that $f(1) > 0$, i.e., $\theta < 1/(N\zeta)$. ■

Proof of Proposition 4.10: Note that $E\|A_0\|^\delta < \infty$ and $E\|B_0\|^\delta < \infty$ implies that $E\log^+\|A_0\| < \infty$, and $E\log\|B_0\| < \infty$, and hence the equation $Y_t = A_tY_{t-1} + B_t$ has a unique strictly stationary ergodic solution. The solution is $Y_t = \sum_{n=0}^{\infty} S_{t,n}B_{t-n}$, where $S_{t,n}$ $= A_tA_{t-1}\ldots A_{t-n+1}$ for $n > 0$ and $S_{t,0} = 1$.

Consider $Y_0 = \sum_{n=0}^{\infty} S_{0,n}B_{-n}$. Note that $\lim_{n\to\infty}\frac{1}{n}\log\|S_{0,n}\| = \gamma < 0$ almost surely. There exists n_0 such that

$$e^{3n\gamma/2} \leq \|S_{0,n}\| \leq e^{n\gamma/2}$$

almost surely for $n > n_0$. Let $r = \min(\delta/(n_0+1), 2)$. For $1 \le n \le n_0$, $nr < \delta$

$$E\|S_{0,n}\|^r \le E\left[\|A_0\|^r \ldots \|A_{-n+1}\|^r\right] \le E\|A_0\|^{nr} \le (E\|A_0\|^\delta)^{nr/\delta} < \infty.$$

Because $r/2 \le 1$, we have

$$\begin{aligned} E\|Y_0\|^{r/2} &\le \sum_{n=0}^{\infty} E\|S_{0,n}\|^{r/2}\|B_{-n}\|^{r/2} \\ &\le \sum_{n=0}^{n_0-1}(E\|S_{0,n}\|^r)^{1/2}(E\|B_0\|^r)^{1/2} + \sum_{n=n_0}^{\infty} e^{n\gamma r/4}E\|B_0\|^{r/2} \\ &\le \left(\sum_{n=0}^{n_0-1}(E\|A_0\|^\delta)^{nr/\delta} + \sum_{n=n_0}^{\infty} e^{n\gamma r/4}\right)(E\|B_0\|^\delta)^{r/(2\delta)} \end{aligned}$$

and $E\|Y_0\|^{\delta/2} < \infty$. It follows that $E(\log Y_t) = \frac{1}{\delta/2}E(\log(Y_t)^{\delta/2}) \le \frac{1}{\delta/2}\log E(Y_t)^{\delta/2} < \infty$. ■

Proof of Corollary 4.1: 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 4.10 immediately with $\delta = 1$. ■

Proof of Proposition 4.11: Note that $E(r_{t,i}) = 0$, $Cov(r_{t,i}, r_{s,j}) = 0$ for $j \ne i$, $t \ne s$. We only need to show that $Var(r_{t,i})$ is a finite constant.

Let $\xi = \alpha + \beta$, $\Psi_{t,i} = \alpha\varepsilon_{t,i}^2 + \beta$. Then $g_{t,i} = 1 - \xi + \Psi_{t,i-1}g_{t,i-1}$ and

$$E_{t-1}[\tau_t g_{t,i}] = \tau_t[1 - \xi + \xi E_{t-1}g_{t,i-1}] = \ldots = \tau_t[1 - \xi^{i-1} + \xi^{i-1}E_{t-1}g_{t,1}] = \tau_t$$

by assumption, where $E_{t-1}[\cdot] = E[\cdot|\sigma(\varepsilon_{s,j}, j = 1, 2, \ldots, N, s \le t-1)]$. It follows that $E_{t-s}[\tau_t g_{t,i}] = E_{t-s}[\tau_t]$ for $s \ge 1$, and $Var_{t-s}[r_{t,i}]= E_{t-s}[\tau_t g_{t,i}\varepsilon_{t,i}^2] = E_{t-s}[\tau_t g_{t,i}] = E_{t-s}[\tau_t]$. Therefore, $Var[r_{t,i}] = Var[E_{t-s}(r_{t,i})] + E[Var_{t-s}(r_{t,i})] = E[\tau_t]$.

Next we need to show that $E[\tau_t]$ exists and is a finite constant. Note that

$$E_{t-K-1}[\tau_t] \quad = \quad m + \theta N \textstyle\sum_{k=1}^{K}\varphi_k(\omega)E_{t-K-1}(\tau_{t-k}) \tag{.5}$$

Introduce $Y_t = (\tau_t, \tau_{t-1}, \ldots, \tau_{t-K+1})^T$. (.5) is equivalent to

$$E_{t-K-1}(Y_t) = AE_{t-K-1}(Y_{t-1}) + B \tag{.6}$$

where

$$A = \begin{pmatrix} N\theta\varphi_1 & N\theta\varphi_2 & \dots & N\theta\varphi_{K-1} & N\theta\varphi_K \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} m \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Moreover, we have

$$E_{t-s}(Y_t) = AE_{t-s}(Y_{t-1}) + B, \forall s \geq K+1. \tag{.7}$$

By backwards iteration,

$$E_{t-s}(Y_t) = A^{s-K}E_{t-s}(Y_{t-s+K}) + (I + A + \ldots + A^{s-K-1})B \tag{.8}$$

Set $t = s$,(.8) becomes

$$E_0(Y_s) = A^{s-K}E_0(Y_K) + (I + A + \ldots + A^{s-K-1})B$$

Since $\lim_{s\to\infty} A^s = 0$ if $0 < \theta < 1/N$ $(*)$, we have $\lim_{s\to\infty} E_0(Y_s) = (I - A)^{-1}B$. It follows that $E(Y_s)$ is finite (entry-wise) when s is sufficiently large. Together with (.7), $E(Y_s)$ is finite for every s. Fix t, and let s go to infinity in (.8). By the property of reversed martingale,

$$E(Y_t) = \lim_{s\to\infty} E_{t-s}(Y_t) = (I - A)^{-1}B = \frac{m}{1 - N\theta}\iota$$

where ι is a vector of 1's, and $Var[r_{it}] = E[\tau_t] = \frac{m}{1-N\theta}$.

Now we need to verify $(*)$: $\lim_{s\to\infty} A^s = 0$ if $0 < \theta < 1/N$. Consider the characteristic equation of A: $f(\lambda) = \lambda^K - N\theta\varphi_1\lambda^{K-1} - N\theta\varphi_2\lambda^{K-2} - \ldots - N\theta\varphi_K$. Since $|f(\lambda)| \geq 1 - N\theta\varphi_1 - \ldots - N\theta\varphi_K = 1 - N\theta > 0$ if $|\lambda| \geq 1$, $\rho(A) < 1$ which implies $\lim_{s\to\infty} A^s = 0$. ■

Proof of Lemma 5.2: Note that $g_t \geq 1 - \alpha - \beta$. (5.1) and Corollary 4.1 imply that $g_t = (1 - \alpha - \beta)/(1 - \beta) + \alpha \sum_{k=0}^{\infty} \beta^k \frac{r_{t-1-k}^2}{\tau_{t-1-k}}$. Thus, for δ^* defined in Corollary 4.1,

$$(g_t)^{\delta^*} \leq \frac{(1 - \alpha - \beta)^{\delta^*}}{(1 - \beta)^{\delta^*}} + \sum_{k=0}^{\infty} (\alpha\beta^k/m)^{\delta^*}(r_{t-1-k}^{2\delta^*}) \leq K + K\sum_{k=0}^{\infty} (\bar{\beta})^{k\delta^*}(r_{t-1-k}^{2\delta^*}) \tag{.9}$$

for some constants $K > 0$ and $0 < \bar{\beta} < 1$ due to the fact that $\mathcal{U}$ is compact. Therefore

$$E \sup_{\Phi \in \mathcal{U}} |\log(g_t)| \leq \frac{1}{\delta^*} E|\log(\sup_{\Phi \in \mathcal{U}} g_t^{\delta^*})| < \infty.$$

Note also that

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

Together with Corollary 4.1 and the fact that $\mathcal{U}$ is compact, we have $E\sup_{\Phi \in \mathcal{U}} |\log(\tau_t)| < \infty$. Moreover, note that

$$\frac{\tau_t(\Phi_0)}{\tau_t} = \frac{m_0}{\tau_t} + \sum_{l=1}^{N+K-1} \frac{\theta_0 c_l(\omega_0) r_{t-l}^2}{\tau_t} \leq \frac{m_0}{m} + \sum_{l=1}^{N+K-1} \frac{\theta_0 c_l(\omega_0)}{\theta c_l(\omega)} \tag{.10}$$

Therefore,

$$\frac{r_t^2}{V_t(\Phi)} = \frac{\tau_t(\Phi_0) g_t(\Phi_0) \varepsilon_t^2}{\tau_t g_t} \leq \frac{1}{1-\alpha-\beta} \left(\frac{m_0}{m} + \sum_{l=1}^{N+K-1} \frac{\theta_0 c_l(\omega_0)}{\theta c_l(\omega)} \right) g_t(\Phi_0) \varepsilon_t^2.$$

It follows that $E\sup_{\Phi \in \mathcal{U}} \frac{r_t^2}{V_t} < \infty$, and hence $E\sup_{\Phi \in \mathcal{U}} |l_t(\Phi)| < \infty$. ■

Proof of Lemma 5.3: Note that $El_t(\Phi) - El_t(\Phi_0) = E(-\log(\frac{V_t(\Phi_0)}{V_t(\Phi)}) + \frac{V_t(\Phi_0)}{V_t(\Phi)} - 1) \geq 0$, and equality holds if and only if $V_t(\Phi_0) = V_t(\Phi)$.

Next we will show that $V_t(\Phi_0) = V_t(\Phi)$ implies $\Phi_0 = \Phi$. Note that

$$\begin{aligned} \tau_t g_t &= [m + \theta \sum_{l=2}^{K+N-1} c_l r_{t-l}^2 + \theta c_1 r_{t-1}^2][\eta + \eta(1 + \sum_{n \geq 2} \prod_{j=2}^{n} u_{t-j}) u_{t-1}] \\ &= [m + \theta \sum_{l=2}^{K+N-1} c_l r_{t-l}^2 + \theta c_1 r_{t-1}^2][\eta + \eta(1 + \sum_{n \geq 2} \prod_{j=2}^{n} u_{t-j})\beta + \eta(1 + \sum_{n \geq 2} \prod_{j=2}^{n} u_{t-j})\alpha \varepsilon_{t-1}^2] \end{aligned} \tag{.11}$$

where $\eta = 1 - \alpha - \beta$ 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

$$\theta c_1 \eta (1 + \sum_{n \geq 2} \prod_{j=2}^{n} u_{t-j}) \alpha V_{t-1} = \theta_0 c_1^0 \eta_0 (1 + \sum_{n \geq 2} \prod_{j=2}^{n} u_{t-j}^0) \alpha_0 V_{t-1}^0 \tag{.12}$$

where the superscript ^{0} to indicate that the function is evaluated at Φ_0. Let

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

(.12) indicates that $F_{t-2} = F_{t-2}^0$, and hence

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

It follows that

$$(\beta_0 \frac{\theta_0 c_1^0 \eta_0 \alpha_0}{\theta c_1 \eta \alpha} - \beta \frac{\theta_0 c_1^0 \eta_0 \alpha_0^2}{\theta c_1 \eta \alpha^2}) [1 + \sum_{n\geq 3} \prod_{j=3}^{n} u_{t-j}^0] = 1 - \frac{\theta_0 c_1^0 \eta_0 \alpha_0}{\theta c_1 \eta \alpha} \tag{.13}$$

This is true only if

$$(\beta_0 \frac{\theta_0 c_1^0 \eta_0 \alpha_0}{\theta c_1 \eta \alpha} - \beta \frac{\theta_0 c_1^0 \eta_0 \alpha_0^2}{\theta c_1 \eta \alpha^2}) = 0 \tag{.14}$$
$$1 - \frac{\theta_0 c_1^0 \eta_0 \alpha_0}{\theta c_1 \eta \alpha} = 0 \tag{.15}$$

Therefore $\theta_0 c_1^0 \eta_0 \alpha_0 = \theta c_1 \eta \alpha$, and $\alpha / \beta = \alpha_0 / \beta_0$, and hence

$$\alpha [1 + \sum_{n\geq 3} \prod_{j=3}^{n} u_{t-j}] = \alpha_0 [1 + \sum_{n\geq 3} \prod_{j=3}^{n} u_{t-j}^0] \tag{.16}$$

Note that the left hand side of (.16) is

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

Similarly we have

$$\alpha^2 [1 + \sum_{n\geq 4} \prod_{j=4}^{n} u_{t-j}] = \alpha_0^2 [1 + \sum_{n\geq 4} \prod_{j=4}^{n} u_{t-j}^0] \tag{.17}$$

$$\alpha + \alpha\beta[1 + \sum_{n\geq 4}\prod_{j=4}^{n} u_{t-j}] = \alpha_0 + \alpha_0\beta_0[1 + \sum_{n\geq 4}\prod_{j=4}^{n} u^0_{t-j}] \tag{.18}$$

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 5.2, one can show that $m = m_0$, $\theta = \theta_0$, and $\omega = \omega_0$. ■

Proof of Lemma 5.4: 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.

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

$$\log(\frac{g_t}{\tilde{g}_t}) \leq \frac{g_t}{\tilde{g}_t} - 1 \leq \frac{1}{1-\alpha-\beta}\beta^{t-K-N} g_{K+N} \leq C_* \bar{\beta}^t g_{K+N} \tag{.19}$$

$$|\frac{r_t^2}{\tau_t g_t}\frac{g_t - \tilde{g}_t}{\tilde{g}_t}| \leq \frac{1}{(1-\alpha-\beta)^2 m}\beta^{t-K-N} r_t^2 g_{K+N} \leq C_* \bar{\beta}^t r_t^2 g_{K+N} \tag{.20}$$

for some constant $C_* > 0$, and $\bar{\beta} \in (0,1)$ which satisfies the restriction of the set $\mathcal{U}$. 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 $\epsilon > 0$,

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

where δ^* defined in Corollary 4.1 and hence $E(r_t^2)^{\delta^*}$ is finite. It can be deduced from (.9) that $E \sup_{\phi\in\mathcal{U}} g_{K+N}^{\delta^*}$ is also finite. Therefore

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

It follows from 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 5.1: Note that $\mathcal{U}$ is a compact set, and Φ_0 is the unique minimizer of $El_0(\Phi)$. We only need to show that $L_T(\Phi)$ and $\tilde{L}_T(\Phi)$ converge to $El_0(\Phi)$ a.s. uniformly on $\mathcal{U}$ (see, for instance, Theorem 3.3 of Gallant and White (1988)). This is true because l_t

is strictly stationary ergodic, $E \sup_{\theta \in \mathcal{U}} |l_t(\theta)| < \infty$, and an application of the uniform SLLN and Lemma 5.4 yields the result. ■

Proof of Lemma 5.5: (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

$$|\tau_t^{-1} \partial_m \tau_t| \le 1/m, \quad |\tau_t^{-1} \partial_\theta \tau_t| \le 1/\theta \tag{.21}$$

$$|\tau_t^{-1} \partial_\omega \tau_t| \le \tau_t^{-1} \theta \sum_{l=1}^{N+K-1} |c_l'(\omega)| r_{t-l}^2 \le \sum_{l=1}^{N+K-1} |c_l'(\omega)| / c_l(\omega) \tag{.22}$$

$$|g_t^{-1} \partial_m g_t| \le g_t^{-1} \alpha \sum_{k=0}^{\infty} \beta^k (r_{t-1-k}^2 / \tau_{t-1-k}) |\tau_{t-1-k}^{-1} \partial_m \tau_{t-1-k}| \le 1/m \tag{.23}$$

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

$$|g_t^{-1} \partial_\alpha g_t| \le 1/(1-\beta) g_t^{-1} + g_t^{-1} \sum_{k=0}^{\infty} \beta^k (r_{t-1-k}^2 / \tau_{t-1-k}) \le 1/(1-\alpha-\beta) + 1/\alpha$$

$$\begin{aligned}
|g_t^{-1} \partial_\beta g_t| &\le \alpha/(1-\beta)^2 g_t^{-1} + g_t^{-1} \alpha \sum_{k=1}^{\infty} k \beta^{k-1} (r_{t-1-k}^2 / \tau_{t-1-k}) \\
&\le \frac{\alpha}{(1-\beta)^2 (1-\alpha-\beta)} + \sum_{k=1}^{\infty} k \frac{\alpha \beta^{k-1} (r_{t-1-k}^2 / \tau_{t-1-k})}{(1-\alpha-\beta)/(1-\beta) + \alpha \beta^k (r_{t-1-k}^2 / \tau_{t-1-k})} \\
&\le \frac{\alpha}{(1-\beta)^2 (1-\alpha-\beta)} + \beta^{-1} \sum_{k=1}^{\infty} k \left(\frac{\alpha \beta^k (r_{t-1-k}^2 / \tau_{t-1-k})}{(1-\alpha-\beta)/(1-\beta)} \right)^\delta
\end{aligned} \tag{.24}$$

where $\delta = \delta^*/2$ and δ^* defined in Corollary 4.1. And (.24) indicates that

$$(E|g_t^{-1} \partial_\beta g_t|^2)^{1/2} \le \frac{\alpha}{(1-\beta)^2 (1-\alpha-\beta)} + \beta^{-1} \sum_{k=1}^{\infty} k \sqrt{E \left(\frac{\alpha \beta^k (r_{t-1-k}^2 / \tau_{t-1-k})}{(1-\alpha-\beta)/(1-\beta)} \right)^{\delta^*}} \tag{.25}$$

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

$$p_1(g_{t-1}\varepsilon_{t-1}^2 - 1) + p_2(g_{t-1} - 1) + g_t\sum_{k=3}^{5} p_k\frac{\partial_k\tau_t}{\tau_t} - (\alpha g_{t-1}\varepsilon_{t-1}^2 + \beta g_{t-1})\sum_{k=3}^{5} p_k\frac{\partial_k\tau_{t-1}}{\tau_{t-1}} = 0 \quad (.26)$$

almost surely at Φ_0. Rewrite (.26) as

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

and let $H_{t,2} = m + \theta\sum_{l\geq 2} c_l r_{t-l}^2$, and we have

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

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

$$\theta c_1 F_{t-2} + \alpha g_{t-1}\sum_k p_k\partial_k(\theta c_1) = 0 \quad (.27)$$

Note that (.27) indicates that $(\theta c_1 p_1 + \alpha p_4 c_1 + \alpha p_5\theta c_1')\tau_{t-1} = \theta c_1\alpha\sum_{k=3}^{5} p_k\partial_k\tau_{t-1}$. Therefore

$$\begin{aligned}(\theta c_1 p_1 + \alpha p_4 c_1 + \alpha p_5\theta c_1')m &= \theta c_1\alpha p_3\\ (\theta c_1 p_1 + \alpha p_4 c_1 + \alpha p_5\theta c_1')\theta c_l &= \theta c_1\alpha(p_4 c_l + p_5\theta c_l'),\end{aligned}$$

for $l = 1, 2, \ldots, N + K - 1$. Note that $\sum_l c_l = N$ and $\sum_l c_l' = 0$. We have

$$c_1 p_1 = 0, \quad \theta p_3 = m p_4, \quad p_5 c_l' = 0 \quad \text{for any } l. \quad (.28)$$

Taking expectation on both sides of (.26), we have $\sum_{k=3}^{5} p_k E(g_1\partial_k\tau_1/\tau_1) = 0$. Together with (.28), we have $p_4 E(g_t) = 0$ and hence $p_4 = 0$. Therefore, $p \equiv 0$ and this contradicts the assumption that $|p| = 1$. ■

(2) Note that

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

$\partial_i l_t$ is strictly stationary ergodic. $\partial_i l_t(\theta_0) = (1-\varepsilon_t^2)(\partial_i\tau_t/\tau_t + \partial_i g_t/g_t)$, and $E(\partial_i l_t(\theta_0))^2$ is finite due to Lemma 5.5. Note also that $E(\partial_i l_t(\theta_0)|\mathcal{F}_{t-1}) = 0$. Hence $\{\partial_i l_t(\theta_0), t \in \mathbb{Z}\}$ is a martingale difference sequence with finite second moment. And the asymptotic normality follows from the martingale central limit theorem, and the Cramer-Wold device. ■

(3) Since $\Phi_0 \in \mathcal{U}^0$, $B_N \equiv B(\Phi_0, 1/N) \cap \mathcal{U}$ is not empty for sufficiently large N. 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 (5.11), first we need to show $E\sup_{\Phi\in B_N}|\partial_i\partial_j l_t(\Phi)| < \infty$. Note that $|\tau_t^{-1}\tau_t(\Phi_0)|$, $|\tau_t^{-1}\partial_i\tau_t|$, $|g_t^{-1}\partial_\alpha g_t|$, $|g_t^{-1}\partial_\theta g_t|$, $|g_t^{-1}\partial_\omega g_t|$, and $|g_t^{-1}\partial_m g_t|$ are bounded on $\mathcal{U}$ (see (.10) and Lemma 5.5). And in an argument similar to the proof of Lemma 5.5, one can easily show that $|\tau_t^{-1}\partial_{ij}\tau_t|$ is bounded on $\mathcal{U}$ 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 4.1. For $\Phi \in B_N$,

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

where $\eta = 1-\alpha-\beta$, and the superscript ^{0} indicates that the quantity is evaluated at Φ_0. Note that $\frac{\eta_0}{\eta}$, $\frac{\alpha_0}{\alpha^{1-\delta}\eta^\delta m^\delta}$ are bounded on $\mathcal{U}$, and (.10) indicates that $\frac{\tau_{t-1-k}}{\tau_{t-1-k}^0}$ is bounded on $\mathcal{U}$ as well. Meanwhile, for $\Phi \in B_N$,

$$\frac{\beta_0}{\beta^{(1-\delta)}} \le \frac{\beta_0}{\beta_0^{(1-\delta)} - N^{-(1-\delta)}} \le \frac{\beta_0}{\beta_0^{(1-\delta)} - N_0^{-(1-\delta)}} < 1$$

Let $\rho_0 = \frac{\beta_0}{\beta_0^{(1-\delta)} - N_0^{-(1-\delta)}}$. It is deduced from (.29) that $\sup_{\Phi \in B_N} \frac{g_t^0}{g_t} \leq K + K \sum_{k=0}^{\infty} \rho_0^k r_{t-1-k}^{2\delta}$ for some constant $K > 0$. Considering Corollary 4.1, we have

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

Take δ appearing in (.24) as $\delta^*/4$, then in a similar argument we have

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

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, in an argument similar to the proof of Lemma 5.5. Note also that

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

Therefore $E \sup_{\Phi \in B_N} |g_t^{-1} \partial_{\beta\beta} g_t| < \infty$ and $E \sup_{\Phi \in B_N} |g_t^{-1} \partial_{\beta m} g_t| < \infty$ due to (.21), Corollary 4.1 and $\mathcal{U}$ 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 $\partial_i \partial_j l_t$ is strictly stationary ergodic,

$$\sup_{\Phi \in B_N} \|H(L_T)(\Phi) - \Sigma(\Phi_0)\| \leq \sup_{\Phi \in B_N} \|H(L_T)(\Phi) - EH(l_1)(\Phi)\| + E \sup_{\Phi \in B_N} \|H(l_1)(\Phi) - H(l_1)(\Phi_0)\|,$$

and $E\sup_{\Phi\in B_N}\|H(l_t)(\Phi)\|$ is O(1) uniformly in t, (5.11) holds due to the dominated convergence theorem and uniform SLLN. ■

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

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

Together with (.19) and (.20), we have, for $\Phi\in\mathcal{U}$,

$$|\partial_i l_t-\partial_i\tilde{l}_t| \le C_*\bar{\beta}^t r_t^2 g_{K+N}|\frac{\partial_i V_t}{V_t}| \tag{.35}$$

$$+\,C_*\bar{\beta}^t(|\partial_i g_t|g_{K+N}+t|g_{K+N}-\tilde{g}_{K+N}|+|\partial_i g_{K+N}|)|1-\frac{r_t^2}{\tilde{V}_t}| \tag{.36}$$

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

$$E(\sup_{\Phi\in\mathcal{U}}|\partial_i l_t(\Phi)-\partial_i\tilde{l}_t(\Phi)|)^\delta \le C_*^\delta t^\delta\bar{\beta}^{\delta t}E(t,\delta) \tag{.37}$$

where

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

Note that $Er_t^{2\delta_*}$, $E(\sup_{\Phi\in\mathcal{U}}g_t)^{\delta_*}$ are finite (see Corollary 4.1, (.9)). And in a similarly argument and using the fact that $\tau_t^{-1}\partial_i\tau_t$ are bounded on $\mathcal{U}$, $E(\sup_{\Phi\in\mathcal{U}}\partial_i g_t)^{\delta_*}$ and $E(\sup_{\Phi\in\mathcal{U}}V_t^{-1}\partial_i V_t)^{\delta_*}$ are finite. Take $\delta=\delta_*/3$, then $E(t,\delta_*/3)$ is bounded, say by C_{**}. It follows that, for any

$\epsilon > 0$,

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

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

Proof of Proposition 5.2: Note that $\nabla L_T(\tilde{\Phi}_T) - \nabla L_T(\hat{\Phi}_T) = H(L_T)(\bar{\Phi}_T)(\tilde{\Phi}_T - \hat{\Phi}_T)$ where $\bar{\Phi}_T \in \mathcal{U}$ 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 5.5 (4). On the other hand, since $\lim_{T\to 0} \bar{\Phi}_T = \Phi_0$ a.s., $H(L_T)(\bar{\Phi}_T)$ converges to $\Sigma(\Phi_0)$ a.s. due to Lemma 5.5 (3), and hence $H(L_T)(\bar{\Phi}_T)$ is positive definite for sufficiently large T. Therefore we have

$$
\sqrt{T}(\tilde{\Phi}_T - \hat{\Phi}_T) = (H(L_T)(\bar{\Phi}_T))^{-1}(\nabla L_T(\tilde{\Phi}_T) - \nabla \tilde{L}_T(\tilde{\Phi}_T))
$$

converges to 0 in probability. An application of Slutsky's theorem yields that $\sqrt{T}(\tilde{\Phi}_T - \Phi_0)$ converges to $N(0, 2\Sigma(\Phi_0)^{-1})$ in distribution. ■

References

Adrian, T., and J. Rosenberg (2008): "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *Journal of Finance* (forthcoming).

Alizadeh, S., M. Brandt, and F. Diebold (2002): "Range-Based Estimation of Stochastic Volatility Models," *The Journal of Finance*, 57(3), 1047–1091.

Andreou, E., and E. Ghysels (2002): "Detecting multiple breaks in financial market volatility dynamics," *Journal of Applied Econometrics*, 17(5), 579–600.

Berkes, I., E. Gombay, L. Horváth, and P. Kokoszka (2004): "Sequential changepoint detection in GARCH(p,q) models," *Econometric Theory*, 20(06), 1140–1167.

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

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), 307–327.

Bollerslev, T., and J. Wooldridge (1992): "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric reviews*, 11(2), 143–172.

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

Bougerol, P., and N. Picard (1992a): "Stationarity of GARCH processes and of some nonnegative time series," *Journal of Econometrics*, 52(1-2), 115–127.

——— (1992b): "Strict Stationarity of Generalized Autoregressive Processes," *The Annals of Probability*, 20(4), 1714–1730.

Brandt, A. (1986): "The Stochastic Equation $Y_{n+1} = A_nY_n + B_n$ with Stationary Coefficients," *Advances in Applied Probability*, 18(1), 211–220.

Carrasco, M., and X. Chen (2002): "Mixing and moment properties of various GARCH and stochastic volatility models," *Econometric Theory*, 18(01), 17–39.

Chen, J., and A. Gupta (1997): "Testing and Locating Variance Changepoints With Application to Stock Prices," *Journal of the American Statistical Association*, 92(438), 739–747.

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

Chernov, M., A. Ronald Gallant, E. Ghysels, and G. Tauchen (2003): "Alternative models for stock price dynamics," *Journal of Econometrics*, 116(1-2), 225–257.

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

Dahlhaus, R., and S. Rao (2006): "Statistical inference for time-varying ARCH processes," *Annals of Statistics*, 34(3), 1075–1114.

Ding, Z., and C. Granger (1996): "Modeling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73(1), 185–215.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

Engle, R., E. Ghysels, and B. Sohn (2008): "On the Economic Sources of Stock Market Volatility," University of North Carolina at Chapel Hill, Manuscript.

Engle, R., and G. Lee (1999): "A Permanent and Transitory Component Model of Stock Return Volatility," *R. Engle and H. White (ed.) Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger, Oxford University Press*, pp. 475–497.

Engle, R., and J. Rangel (2008): "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187–1222.

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

Francq, C., and J. Zakoian (2006): "Mixing properties of a general class of GARCH (1, 1) models without moment assumptions on the observed process," *Econometric Theory*, 22(5), 815.

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

Furstenberg, H., and H. Kesten (1960): "Products of Random Matrices," *The Annals of Mathematical Statistics*, 31(2), 457–469.

Gallant, A., and H. White (1988): *A unified theory of estimation and inference for nonlinear dynamic models*. Basil Blackwell Oxford.

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

Glasserman, P., and D. Yao (1995): "Stochastic Vector Difference Equations with Stationary Coefficients," *Journal of Applied Probability*, 32(4), 851–866.

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

He, C., and T. Terasvirta (1999): "Properties of moments of a family of GARCH processes," *Journal of Econometrics*, 92(1), 173–192.

Horvath, L., P. Kokoszka, and G. Teyssière (2001): "Empirical process of the squared residuals of an ARCH sequence," *Annals of Statistics*, 29(2), 445–469.

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

Inclan, C., and G. Tiao (1994): "Use of cumulative sums of squares for retrospective detection of changes of variance," *Journal of the American Statistical Association*, 89(427), 913–923.

Jensen, S., and A. Rahbek (2004a): "Asymptotic inference for nonstationary GARCH," *Econometric Theory*, 20, 1203–1226.

——— (2004b): "Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case," *Econometrica*, 72(2), 641–646.

Kesten, H., and F. Spitzer (1984): "Convergence in distribution of products of random matrices," *Probability Theory and Related Fields*, 67(4), 363–386.

Kingman, J. (1973): "Subadditive Ergodic Theory," *The Annals of Probability*, 1(6), 883–899.

Kokoszka, P., and R. Leipus (2000): "Change-point estimation in ARCH models," *Bernoulli*, 6(3), 513–540.

Kristensen, D. (2009): "On stationarity and ergodicity of the bilinear model with applications to GARCH models," *Journal of Time Series Analysis*, 30(1), 125–144.

Kulperger, R., and H. Yu (2005): "High moment partial sum processes of residuals in GARCH models and their applications," *Annals of Statistics*, 33(5), 2395–2422.

Lee, S., and 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.

Linton, O., J. Pan, and H. Wang (2010): "Estimation for a nonstationary semi-strong GARCH (1, 1) model with heavy-tailed errors," *Econometric Theory*, 26(1), 28.

Lu, Z., and 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*, pp. 575–596.

Meitz, M., and 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(05), 1291–1320.

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

Mokkadem, A. (1990): "Propriétés de mélange des processus autorégressifs polynomiaux," *Annales de l'I. H. P. Probabilités et statistiques*, 26(2), 219–260.

Nelson, D. (1990): "Stationarity and persistence in the GARCH (l, l) model," *Econometric theory*, 6(3), 318–334.

Newman, C. (1986): "The distribution of Lyapunov exponents: Exact results for random matrices," *Communications in mathematical physics*, 103(1), 121–126.

Peres, Y. (1992): "Domains of analytic continuation for the top Lyapunov exponent," *Ann. Inst. H. Poincaré Probab. Statist*, 28(1), 131–148.

Pham, D. (1985): "Bilinear markovian representation and bilinear models," *Stochastic Processes and their applications*, 20(2), 295–306.

——— (1986): "The mixing property of bilinear and generalised random coefficient autoregressive models," *Stochastic Processes and their Applications*, 23(2), 291–300.

Seneta, E. (1981): *Non-negative matrices and Markov chains*. Springer Verlag.

Straumann, D., and T. Mikosch (2006): "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach," *The Annals of Statistics*, 34(5), 2449–2495.

Vanneste, J. (2010): "Estimating generalized Lyapunov exponents for products of random matrices," *Physical Review E*, 81(3), 036701.

Weiss, A. (1986): "Asymptotic theory for ARCH models: estimation and testing," *Econometric theory*, pp. 107–131.

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

Yao, J. (2001): "On square-integrability of an AR process with Markov switching," *Statistics & probability letters*, 52(3), 265–270.