Statistical Inference for Volatility Component Models

Fangfang Wang[*] Eric Ghysels[†]

September 23, 2008

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. The paper studies the distributional properties of various volatility component models. Sufficient conditions for the existence or/and uniqueness of (strictly) stationary (ergodic) solutions with mixing property to the volatility component models are derived. Hence, the paper revisits the component models from a statistical perspective and attempts to explore the stationarity and mixing properties 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 local consistency and asymptotic normality are established as well.

[*]Department of Statistics, University of North Carolina at Chapel Hill. Email: fangfanw@email.unc.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 class models originated by Engle [15] is the most popular. The basic structure of ARCH is very much similar to ARMA, the appearance is deceiving. Indeed, there is a considerable literature on the stationarity, mixing and moment properties of various ARCH-type models, see e.g. Carrasco and Chen [10], He and Teräsvirta [22].

The prime focus has been on the GARCH(p,q) model - in particular GARCH(1,1) - originated by Bollerslev [7]. Yet, empirical evidence suggests that volatility dynamics is better described by component models. Engle and Lee [17] introduced a GARCH model with a long and short run component, and several others have proposed related two-factor volatility models, see e.g. Ding and Granger [14], Alizadeh et al. [3], Chernov et al. [12] and Adrian and Rosenberg [1] among many others. The volatility component model of Engle and Lee [17] decomposed the equity conditional variance as the sum of the short-run (transitory) and long-run (trend) components.

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 [4], Berkes et al. [5]. Chen and Gupta [11], Horvath et al.[23], Horvath et al. [24], Inclan and Tiao [25], Kokoszka and Leipus [28], Kulperger and Yu [29], 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 [13]. This paper focuses exclusively on component models. For some component models, like the restricted GARCH(2,2) model of Engle and Lee [17] which consist of two GARCH(1,1) components, the literature has not well covered the conditions that characterize non-stationarity issues of the components. Moreover, the component models that have been suggested recently are not of the additive ARCH-type, but instead consist of a multiplicative structure. The first to suggest a component structure that accommodates non-stationarity of volatility is Engle and Rangel [18], later extended by Engle, Ghysels and Sohn [16]. 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 framework. Exponential quadratic Spline is used to estimate the long memory structure of low-frequency volatility 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 economic implications of component models and their empirical application have been studied intensively in Engle and Lee [17], Engle and Rangel [18], Engle, Ghysels and Sohn [16]. This paper revisits the component models from a statistical perspective and attempts to explore the stationarity and mixing properties 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.

Although most of our focus is on the aforementioned multiplicative models, we start with filling a gap in the literature pertaining to additive component models, that is the original Engle and Lee model. The dynamic structure of the conditional variance in their model can be reduced to a restricted GARCH(2,2) model with certain coefficients negative, which, to some extent, distinguishes itself from the classic GARCH model. Hence, the existing regularity conditions for GARCH models need to be extended to handle the constrained additive component models. Under certain regularity conditions on the parameters, the transitory component mean-reverts to zero and the trend converges to the unconditional variance but at a much slower rate. While, the resulting volatility process is covariance stationary, as pointed out by Engle and Lee [17], the mapping from component models to GARCH involves nonlinear transformations of the parameter space.

The GARCH-MIDAS model of Engle, Ghysels and Sohn [16] 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. [20]) regression and MIDAS 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 GARCH-MIDAS model has two basic specifications. In terms of the structure of low-frequency volatility, they are classified as: (1) GARCH-MIDAS model with fixed time span realized volatility (RV) where the low-frequency component is constant within a fixed time span, say a month or a quarter but the high-frequency component is varying from day to day; (2) GARCH-MIDAS model with rolling window realized volatility (RV) where both low-frequency and high-frequency components change at a daily basis.

In particular, we are looking for regularity conditions under which the models could admit covariance stationary or strictly stationary ergodic solutions with/without β-mixing property. By linking the models with multivariate stochastic difference equations, we study the covariance stationary property through a reversed martingale argument and the strict stationarity property in terms of the top Lyapounov exponent. The dilemma is how to evaluate theoretically the top Lyapounov exponents which are defined on (i) a sequence of i.i.d. matrices with certain negative entries and (ii) a sequence of strictly stationary ergodic matrices with positive entries. In addition, we derive the locally consistent estimates of the GARCH-MIDAS model with rolling window realized volatility specification and study their asymptotic behaviors by means of Cramér-Wold device.

The rest of paper is organized as follows: we revisit the volatility component model of Engle and Lee in section 2, and give the condition under which it is strictly stationary ergodic and β-mixing. Section 3 focuses on the stationarity properties of the two GARCH-MIDAS specifications. The consistent estimates with asymptotic behaviors of GARCH-MIDAS model with rolling window RV are studied in section 4. Section 5 gives the concluding remarks. In the appendix, we list the theorems and lemmas cited from others' work for quick reference.

2 Volatility component model of Engle and Lee

The volatility component model of Engle and Lee [17] structures the daily return r_t as

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

where $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and the parameter space is

$$\mathcal{P} = \{(\alpha, \beta, \omega, \rho, \phi) \in (\mathbb{R}^{5+})^{\circ} : \alpha + \beta < \rho < 1, \phi < \beta\}$$

which ensures the conditional variance h is nonnegative (see [17] for the proof of nonnegativity of h).

According to the model, the conditional variance is the sum of long-run (trend) variance τ and the short-run (transitory) variance g. The condition $0 < \alpha + \beta < \rho < 1$ guarantees that the short-run volatility mean-reverts to zero at a geometric rate of $\alpha + \beta$ and long-run volatility converges to $\omega/(1-\rho)$ with a much slower rate.

Engle and Lee [17] provided sufficient conditions for the covariance stationarity of $\{r_t\}$ with parameter space $\mathcal{P}$ by linking it to an ARMA(2,2) process, i.e.

$$\begin{aligned} r_t^2 &= \omega(1-\alpha-\beta) + (\alpha+\beta+\rho)r_{t-1}^2 - (\rho\alpha + \rho\beta)r_{t-2}^2 \\ &\quad + \eta_t - (\rho + \beta - \phi)\eta_{t-1} - [(\phi(\alpha+\beta) - \beta\rho)]\eta_{t-2} \end{aligned}$$

where $\eta_t = r_t^2 - h_t$ (see [17]). Here we shall present conditions for strict stationarity and β-mixing. For the time being, we assume the process to extend infinitely into the past. Later, we will consider the scenario of closing the system by assigning an initial distribution at time point 0.

The volatility component model of Engle and lee is also referred to as the restricted GARCH(2,2) model because the dynamics of conditional variance h can be cast into the framework of a GARCH(2,2) process as

$$\begin{aligned} r_t &= \sqrt{h_t}\varepsilon_t \\ h_t &= \alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} \end{aligned} \tag{2.2}$$

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 $\beta_2 = \phi(\alpha+\beta) - \rho\beta < 0$. The distinct feature of this 'new' model is its similarity to a GARCH(2,2) setting but of having negative coefficients (α_2 and β_2 are negative). So the existing results about classic GARCH(2,2) model [8] can not be applied to the volatility component model of Engle and Lee.

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

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

the restricted GARCH(2,2) process (2.2) of Engle and Lee is equivalent to the solution to a stochastic difference equation defined through

$$Y_t = A(\varepsilon_t)Y_{t-1} + B. \tag{2.3}$$

with iid coefficients.

There is a vast literature on the existence/uniqueness of the strictly stationary solution to the stochastic difference equation of the form

$$Y_t = A_t Y_{t-1} + B_t, \;\; t \in \mathbb{Z} \tag{2.4}$$

where Y_t and B_t are $\mathbb{R}^n$-valued random vectors, A_t is a $\mathbb{R}^{n\times n}$-valued random matrix, and $\{(A_t, B_t), t \in \mathbb{Z}\}$ is a strictly stationary ergodic sequence. Vervaat [34] and Brandt [9] analyzed the stochastic difference equation for the scaler case, i.e. $n = 1$ with assumption that the coefficients are iid and strictly stationary ergodic respectively. Bougerol and Picard [8] studied the problem with A_t and B_t being iid. Glasserman and Yao [21] extended the results for the general strictly stationary ergodic sequence. For the vector case, the problem of strictly stationary ergodic solution to (2.4) is closely related to the associated top Lyapounov exponent which is defined as

Definition 2.1 *Let $\{A_t, t \in \mathbb{Z}\}$ be a strictly stationary and ergodic sequence of $\mathbb{R}^{n\times n}$-valued random matrices, such that $E\log^+\|A_0\| < \infty$. Then the top Lyapounov exponent associated with $\{A_t, t \in \mathbb{Z}\}$ is defined as*

$$\gamma = \inf_{t\in\mathbb{N}} E(\frac{1}{t+1}\log\|A_tA_{t-1}\dots A_0\|).$$

Combining subadditive ergodic theory of Kingman [27] due to the submultiplicativity of matrix norm and the work of Furstenberg and Kesten [19], we could derive a well-known property of the top Lyapounov exponent which is stated as

Theorem 2.1 ([19], [27]) *If $\{A_t, t \in \mathbb{Z}\}$ is a strictly stationary ergodic sequence of $\mathbb{R}^{n\times n}$-valued random matrices, such that $E\log^+\|A_0\| < \infty$, then*

$$-\infty \le \gamma < \infty$$

$$\lim_{t\to\infty}\frac{1}{t+1}\log\|A_tA_{t-1}\dots A_0\| = \gamma \text{ almost surely}$$

and

$$\lim_{t\to\infty}\frac{1}{t+1}E\log\|A_tA_{t-1}\dots A_0\| = \gamma$$

The top Lyapounov exponent is independent of the choice of underlying matrix norm $\|.\|$ since all the norms on the finite norm space are equivalent. For ease of analysis, we consider the Frobenius norm in particular throughout this paper. Next proposition gives a sufficient condition for the strict stationarity of the restricted GARCH(2,2) model of Engle and Lee when we assume the whole system starts from the negative infinity.

Proposition 2.1 *For the volatility component model of Engle and Lee with the parameter space $\mathcal{P}$, $\{r_t, h_t\}$ is strictly stationary ergodic if $\alpha < \phi$, $2\alpha + \beta + \phi < \rho < 5\alpha + \beta$ and $\alpha + \beta + \rho < 1$.*

The proof of Proposition 2.1 needs the following lemmas.

Lemma 2.1 *Let $\{F_t, t \in \mathbb{Z}\}$ be a sequence of iid random matrices such that $P(F_t F_{t-1} \ldots F_0 \geq 0, \text{infinitely often}) = 1$ and suppose that $E(\log^+ \|F_0\|) < \infty$ and $\rho(E(F_0)) < 1$. Then the top Lyapounov exponent associated with this sequence is strictly negative.*

Proof of Lemma 2.1 *Define $M_k = F_{-1} \ldots F_{-k}$, then $E(M_k) = F^k$ where $F = E(F_0)$. Under the assumption that $\rho(F) < 1$, $\sum_k F^k < \infty$. Since F_j is iid and $P(F_t F_{t-1} \ldots F_0 \geq 0, i.o.) = 1$, by Fubini's theorem, we have $\sum_k M_k < \infty$ almost surely. Therefore, almost surely $\lim_{k\to\infty} M_k = 0$, or $\lim_{k\to\infty} \|M_k\| = 0$. Let $\tilde{F}_k$ be the transpose of F_{-k}. Since $\|\tilde{F}_k \ldots \tilde{F}_1\| = \|F_{-1} \ldots F_{-k}\|$, the top Lyapounov exponent associated with $\{F_t, t \in \mathbb{Z}\}$ is strictly negative, following from Lemma 3.4 of Bougerol and Picard [8] (See Lemma 6.2 in Appendix).*

Lemma 2.2 *Suppose that $\alpha < \phi$, and $2\alpha + \beta + \phi < \rho < 5\alpha + \beta$. If further express h_t in model (2.2) with parameter space $\mathcal{P}$ as an infinite distributed lag of r_t^2, then all the coefficients are positive, i.e.*

$$h_t = \omega^* + \sum_{k=0}^{\infty} \phi_k r_{t-k-1}^2$$

with $\omega^* \geq 0, \phi_k \geq 0 \; \forall k$.

Proof of Lemma 2.2 *Let Z_1 and Z_2 be the roots of $Z^2 - \beta_1 Z - \beta_2$. WLOG, assume $|Z_1| \geq |Z_2|$. By theorem 2 of Nelson and Cao [32] (see Appendix), to show $\omega^* \geq 0, \phi_k \geq 0$ it is equivalent to prove that*

(1) Z_1, Z_2 are real, and $|Z_1| < 1$ $|Z_2| < 1$;

(2) $\alpha_0/(1 - Z_1 - Z_2 + Z_1 Z_2) \geq 0$;

(3) $\alpha_1 Z_1 + \alpha_2 > 0$ and $\alpha_1 Z_2 + \alpha_2 \neq 0$;

(4) $\phi_k \geq 0$ for $k = 0, 1, 2$.

Conditions (1) & (2) have been checked by Engle and Lee (see Appendix of [17]). We only need to justify conditions (3) & (4) under the restrictions specified. Since

$$\begin{aligned}
\alpha_1 Z_1 + \alpha_2 &= \tfrac{\phi+\alpha}{2}[\rho + \beta - \phi + \sqrt{(\rho - \beta - \phi)^2 + 4\alpha\phi} - 2\tfrac{\alpha\phi+\alpha\rho+\phi\beta}{\phi+\alpha}] \\
&= \tfrac{\phi+\alpha}{2}[\rho - (\beta + 2\alpha + \phi) + \sqrt{(\rho - \beta - \phi)^2 + 4\alpha\phi} \\
&\quad -\tfrac{2\alpha}{\phi+\alpha}(\rho - \alpha - \beta)]
\end{aligned}$$

Note that under the restrictions, $\rho - (\beta + 2\alpha + \phi) > 0$, $\frac{2\alpha}{\phi+\alpha} < 1$, and the polynomial $g(\phi) = (\rho - \beta - \phi)^2 + 4\alpha\phi - (\rho - \alpha - \beta)^2 = \phi^2 - 2\phi(\rho - \beta - 2\alpha) + 2\alpha(\rho - \beta) - \alpha^2 > 0$ due to the fact that $\Delta = (\rho - \beta - 2\alpha)^2 - 2\alpha(\rho - \beta) - \alpha^2 = (\rho - \beta - 5\alpha)(\rho - \beta - \alpha) < 0$. Therefore, $\alpha_1 Z_1 + \alpha_2 > 0$. Meanwhile,

$$\begin{aligned}
\alpha_1\beta_1/2 + \alpha_2 &= (\phi + \alpha)(\rho + \beta - \phi)/2 - (\phi\alpha + \phi\beta + \alpha\rho) \\
&= -\frac{1}{2}[\phi(\phi - \alpha) + \phi(5\alpha + \beta - \rho) + \alpha(\rho - \beta - \phi)] < 0
\end{aligned}$$

thus $\alpha_1 Z_2 + \alpha_2 \neq 0$.

Next to check condition (4). Since

$$\begin{aligned}
\phi_0 &= \alpha_1 > 0 \\
\phi_1 &= \beta_1\alpha_1 + \alpha_2 \\
&= (\phi + \alpha)(\rho + \beta - \phi) - (\phi\alpha + \phi\beta + \alpha\rho) \\
&= \phi(\rho - \phi - \alpha) + \alpha(\beta - \phi) > 0 \\
\phi_2 &= \beta_1\phi_1 + \beta_2\phi_0 \\
&= (\rho + \beta - \phi)\phi_1 + [\phi(\alpha + \beta) - \rho\beta](\phi + \alpha) \\
&= (\beta - \phi)\phi_1 + \phi(\phi + \alpha)(\alpha + \beta) + \rho\phi(\rho - 2\alpha - \beta - \phi) > 0
\end{aligned}$$

Condition (4) is also satisfied. Therefore $\omega^* \geq 0$ and $\phi_k \geq 0\ \forall k$.

Proof of Proposition 2.1 *According to Theorem 3.1 of Glasserman and Yao (see Appendix), the statement is true if*

$$E(\log \|A(\varepsilon_0)\|)^+ < \infty \text{ and } \gamma < 0.$$

Under Frobenius norm,

$$\|A(\varepsilon_0)\|^2 = (\beta_1 + \alpha_1 \varepsilon_t^2)^2 + (\beta_2)^2 + (\alpha_2)^2 + 1 + (\varepsilon_t^2)^2 > 1$$

so $E(\log \|A(\varepsilon_0)\|)^+ < \infty$.

Define $M_{t,k} = A(\varepsilon_t)A(\varepsilon_{t-1}) \dots A(\varepsilon_{t-k})$, *then* $E(M_{t,k}) = M^k$ *where*

$$M = \begin{pmatrix} \beta_1 + \alpha_1 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of M *is* $0, \alpha + \beta$ *and* ρ, *from where we know* $\rho(M) < 1$ *by assumption. Using Lemma 2.2, it could be derived that each component of* $M_{t,k}$ *is nonnegative. Further applying Lemma 2.1, the top Lyapounov exponent* γ *associated with* $\{A(\varepsilon_t), t \in \mathbb{Z}\}$ *is strictly negative.*

In Proposition 2.1, the model is assumed to extend infinitely into the past. Next we consider the system (2.2) starting from time 0 with initial values g_0 and τ_0 defined on the probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$ such that $P(0 < \tau_0 < \infty) = P(0 < \tau_0 + g_0 < \infty) = 1$. Now the process (2.3) can be viewed as a time-homogeneous Markov process, which puts us in the setting of the polynomial random coefficient autoregressive model mentioned in Carrasco and Chen [10]. Starting from there, we could derive the mixing property of volatility component model.

Based on the work of Mokkadem [31], Carrasco and Chen [10] studied the conditions for the stationarity, mixing and moment properties of various ARCH-type models. Again, we consider Theorem 4.3 of Mokkadem [31] or Theorem 1 of Carrasco and Chen [10] (see Appendix), and we have the following,

Proposition 2.2 *Consider the volatility component model of Engle and Lee with the parameter space* $\mathcal{P}$, *with* $\alpha < \phi$, $2\alpha + \beta + \phi < \rho < 5\alpha + \beta$, $\alpha + \beta + \rho < 1$ *and the distribution induced by* $\tau_0 + g_0$ *invariant, then* $E[h_t] < \infty$, $E[r_t^2] < \infty$, $\{r_t, h_t\}$ *is strictly stationary and* β*-mixing with exponential decay.*

Proof of Proposition 2.2 *By Theorem 4.3 of Mokkadem [31] or Theorem 1 of Carrasco and Chen [10] (see Appendix), as long as assumptions (A.1-A.5) are verified, the statement is true. (A.1) and (A.2) are satisfied straightforwardly. Hence, we need to check (A.3), (A.4) and (A.5).*

- *Assumption (A.3): Note*

$$A(0) = \begin{pmatrix} \beta_1 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and its characteristic function is $det(\lambda I_3 - A(0)) = \lambda(\lambda^2 - \beta_1\lambda - \beta_2)$. *Let* $f(\lambda) = \lambda^2 - \beta_1\lambda - \beta_2$. *Since*

$$\begin{aligned} \beta_1^2 + 4\beta_2 &= (\rho - \phi - \beta)^2 + 4\alpha\phi > 0, \\ f(\beta_1/2) &= -(\beta_1^2/4 + \beta_2) < 0, \\ f(0) &= -\beta_2 > 0, \\ f(1) &= (1-\beta)(1+\phi-\rho) - \phi\alpha > 0, \end{aligned}$$

hence $\rho[A(0)] < 1$.

- *Assumption (A.4): From the proofs of Lemma 2.1 and Proposition 2.1,*

$$\sum_{k=1}^{\infty} A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})B < \infty$$

almost surely and

$$A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})$$

converges almost surely to the 0 matrix.

- *Assumption (A.5): 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* $\frac{\alpha_0+1}{B} < 1 - \pi$. *Note*

$$\begin{aligned} E[V(Y_t)|Y_{t-1}] &= E[h_{t+1} + ah_t + ar_t^2|Y_{t-1}] \\ &= E[\alpha_0 + (\beta_1 + a)h_t + (\alpha_1 + a)r_t^2 + \beta_2 h_{t-1} + \alpha_2 r_{t-1}^2|Y_{t-1}] \\ &= \alpha_0 + (\beta_1 + \alpha_1 + 2a)h_t^2 + \beta_2 h_{t-1} + \alpha_2 r_{t-1}^2 \\ &\le \alpha_0 + \pi V(Y_{t-1}) \end{aligned}$$

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,

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

Assumption (A.5) is also satisfied.

3 Stationarity of GARCH-MIDAS process

The spline-GARCH model of Engle and Rangel [18] and the GARCH-MIDAS model of Engle, Ghysels and Sohn [16] assume the conditional volatility to be the product of long-run and short-run volatility. To be specific, the spline-GARCH model is defined through the following three equations

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

where

- $\varepsilon_t \overset{iid}{\sim} N(0,1)$
- $\{0 = t_0 < t_1 < t_2 < \ldots < t_k = T\}$ is a partition of the time horizon T in k equally spaced intervals.

The high-frequency component g follows a mean-reverting unit GARCH(1,1) process. The low-frequency component τ is deterministic, and it equals the unconditional variance, ie $E(r_t - \mu)^2 = \tau_t$ from where we could see the conditional volatility process is not mean-reverting and is not stationary as well.

GARCH-MIDAS model, as an extension of spline-GARCH model, keeps the structure of short-run component g but modifies the long-run component τ as stochastic. According to the way the low-frequency component is structured, GARCH-MIDAS model has two basic specifications: GARCH-MIDAS model with fixed time span realized volatilities (RV) and GARCH-MIDAS model with rolling window realized volatility (RV).

For the fixed time span RV setting, the dynamics of long-run and short-run components are specified as

$$\begin{array}{rcl} r_{i,t} & = & \mu + \sqrt{\tau_t g_{i,t}} \varepsilon_{i,t}, \quad 2 \leq i \leq N_t, t \in \mathbb{Z} \\ g_{i,t} & = & (1-\alpha-\beta) + \alpha \frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta g_{i-1,t} \\ \tau_t & = & m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, \;\; RV_t = \sum_{i=1}^{N_t} r_{i,t}^2 \end{array} \tag{3.1}$$

where

- r_{it} is the log return on day i of period (say month, quarter, etc.) t.
- N_t is the number of days in period t, but in this paper we assume $N_t = N$(a predetermined number) for any t.
- $\varepsilon_{i,t} \overset{iid}{\sim} N(0,1) \;\; \forall i, t$.
- $E(g_{1,t}|\mathcal{F}_{t-1}) = 1$, which is equivalent to $E(g_{i,t}|\mathcal{F}_{t-1}) = 1 \;\; (1 \leq i \leq N_t)$, an assumption used in Engle, Ghysels and Sohn [16].
- $\varphi_k(\omega)$ are nonnegative functions of ω such that $\sum_{k=1}^{N} \varphi_k(\omega) = 1$.
- $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ $m > 0$.

For the rolling window RV setting, the long-run component dynamics is simplified,

$$\begin{array}{rcl} r_t & = & \mu + \sqrt{\tau_t g_t} \varepsilon_t, t \in \mathbb{Z} \\ g_t & = & (1-\alpha-\beta) + \alpha \frac{(r_{t-1}-\mu)^2}{\tau_{t-1}} + \beta g_{t-1} \\ \tau_t & = & m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, RV_t = \sum_{j=0}^{N-1} r_{t-j}^2 \end{array} \tag{3.2}$$

where

- r_t is the log return on day t,
- $\varepsilon_t \overset{iid}{\sim} N(0,1)$,
- N is the length of a certain period of interest with value predetermined,
- $\varphi_k(\omega)$ are nonnegative functions of ω such that $\sum_{k=1}^{N} \varphi_k(\omega) = 1$,

- $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$, $m > 0$.

The appeal of GARCH-MIDAS model is that the structure of long-run component is stochastic which makes it possible to study the statistical property of the conditional volatility process.

3.1 GARCH-MIDAS model with fixed time span RV

We start with the fixed time span RV setting and assume $\mu = 0$. Model (3.1) is simplified as

$$\begin{aligned} r_{i,t} &= \sqrt{\tau_t g_{i,t}}\varepsilon_{i,t}, \quad 1 \le i \le N, t \in \mathbb{Z} \\ g_{i,t} &= (1-\alpha-\beta) + \alpha\frac{(r_{i-1,t})^2}{\tau_t} + \beta g_{i-1,t} \\ \tau_t &= m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, \ \ RV_t = \sum_{i=1}^{N} r_{i,t}^2 \end{aligned} \tag{3.3}$$

Proposition 3.1 *Suppose that $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$, $\{r_{i,t}\}$ defined in (3.3) is a White Noise if $0 < \theta < 1/N$.*

As an immediate consequence, we have

Corollary 3.1 *GARCH-MIDAS process with fixed time span RV defined in (3.1) is covariance stationary with mean μ and variance $\frac{m}{1-\theta N}$ if $0 < \theta < 1/N$.*

Proof of Proposition 3.1 *To show that $\{r_{i,t}\}$ is a White Noise, we need to verify the following three conditions:*

(i) $E(r_{i,t}) = 0$

(ii) $Cov(r_{i,t}, r_{j,s}) = 0$ *for* $j \neq i$ *or* $t \neq s$

(iii) $Var(r_{i,t})$ *is a finite constant.*

(i) is true since $E(r_{i,t}) = E(\sqrt{\tau_t g_{i,t}}\varepsilon_{i,t}) = 0$ and (ii) also holds due to the property of $\varepsilon_{i,t}$. Now we need to check the third condition.

For ease of reference, let $\eta \equiv \alpha + \beta$, $\Psi_{i-1,t} \equiv \alpha\varepsilon_{i-1,t}^2 + \beta$. Then $g_{i,t} = 1 - \eta + \Psi_{i-1,t} g_{i-1,t}$ and

$$\begin{aligned} E_{t-1}[\tau_t g_{i,t}] &= \tau_t[(1-\eta) + \eta E_{t-1} g_{i-1,t}] \\ &\vdots \\ &= \tau_t[(1-\eta^{i-1}) + \eta^{i-1} E_{t-1} g_{1,t}] \\ &= \tau_t \end{aligned}$$

where $E_{t-1}[.]$ *is equivalent to* $E[.|F_{N,t-1}]$.

It follows that $E_{t-s}[\tau_t g_{i,t}] = E_{t-s}[\tau_t]$ *for* $s \geq 1$, *and*

$$Var_{t-s}[r_{i,t}] = E_{t-s}[\tau_t g_{i,t}\varepsilon^2_{i,t}] = E_{t-s}[\tau_t g_{i,t}] = E_{t-s}[\tau_t]$$

therefore,

$$Var[r_{i,t}] = Var[E_{t-s}(r_{i,t})] + E[Var_{t-s}(r_{i,t})] = E[\tau_t]$$

Next we need to show that $E[\tau_t]$ *exists and is finite. Notice that*

$$\begin{aligned} E_{t-K-1}[\tau_t] &= m + \theta \textstyle\sum_{k=1}^{K} \varphi_k(\omega) E_{t-K-1}[RV_{t-k}] \\ &= m + \theta \textstyle\sum_{k=1}^{K} \varphi_k(\omega) E_{t-K-1}[\sum_{i=1}^{N} r^2_{i,t-k}] \\ &= m + \theta N \textstyle\sum_{k=1}^{K} \varphi_k(\omega) E_{t-K-1}(\tau_{t-k}) \end{aligned} \tag{3.4}$$

Introduce $Y_t = (\tau_t, \tau_{t-1}, \ldots, \tau_{t-K+1})^T$. *(3.4) is equivalent to*

$$E_{t-K-1}(Y_t) = AE_{t-K-1}(Y_{t-1}) + B \tag{3.5}$$

where

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

Moreover, we have

$$E_{t-s}(Y_t) = AE_{t-s}(Y_{t-1}) + B, \forall s \geq K+1 \tag{3.6}$$

by iteration,

$$E_{t-s}(Y_t) = A^{s-K} E_{t-s}(Y_{t-s+K}) + (I + A + \ldots + A^{s-K-1})B \tag{3.7}$$

Set $t = s$, *equ(3.7) becomes*

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

Since if $0 < \theta < 1/N$, $\lim_{s\to\infty} A^s = 0$*(∗), then* $\lim_{s\to\infty} E_0(Y_s) = (I - A)^{-1}B$. *It follows that* $E(Y_s)$ *is finite (elementary-wise) when* s *is sufficiently large.*

Together with equ(3.6), we know $E(Y_s)$ *is finite for every* s. *Fix* t, *and let* s *go to infinity in (3.7). By the property of reversed martingale, we have*

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

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* $0 < \theta < 1/N$. *Note*

$$f(\lambda) = det(\lambda I_K - A) = \lambda^K - N\theta\varphi_1\lambda^{K-1} - N\theta\varphi_2\lambda^{K-2} - \ldots - N\theta\varphi_K$$

Since

$$|f(\lambda)| \geq 1 - N\theta\varphi_1 - \ldots - N\theta\varphi_K = 1 - N\theta > 0 \ \ if \ |\lambda| \geq 1,$$

$\rho(A) = \max_j |\lambda_j|$ *should be strictly less than* 1 *which implies that* $\lim_{s\to\infty} A^s = 0$.

3.2 GARCH-MIDAS model with rolling window RV

For the rolling window RV setting, we still consider the simple case of $\mu = 0$ first. Model (3.2) becomes

$$\begin{aligned} r_t &= \sqrt{\tau_t g_t}\varepsilon_t, t \in \mathbb{Z} \\ g_t &= (1-\alpha-\beta) + \alpha\tfrac{(r_{t-1})^2}{\tau_{t-1}} + \beta g_{t-1} \\ \tau_t &= m + \theta\textstyle\sum_{k=1}^K \varphi_k(\omega) RV_{t-k}, RV_t = \sum_{j=0}^{N-1} r_{t-j}^2 \end{aligned} \tag{3.8}$$

Further, the dynamics of r_t^2 could be reduced to

$$r_t^2 = m g_t \varepsilon_t^2 + \theta g_t \varepsilon_t^2 \textstyle\sum_{l=1}^{N+K-1} c_l r_{t-l}^2 \tag{3.9}$$

where c_l's are certain combinations of φ_k's and they satisfy

$$\sum_{l=1}^{N+K-1} c_l = N \sum_{k=1}^{K} \varphi_k(\omega) = N.$$

Under the assumptions $\alpha > 0, \beta > 0$ and $\alpha + \beta < 1$, model (3.9) can be linked to a multivariate stochastic difference equation with strictly stationary ergodic coefficients through Markovian representation ([2], [34]). In other words, the stationarity property of the process $\{r_t^2, t \in \mathbb{Z}\}$ is equivalent to the existence of stationary solution to the following stochastic difference equation

$$Y_t = A_t(\tilde{c})Y_{t-1} + B_t. \tag{3.10}$$

where

$$Y_t = (r_t^2, r_{t-1}^2, \ldots, r_{t-N-K+2}^2)',$$

$$A_t(\tilde{c}) = \begin{pmatrix} \theta g_t \varepsilon_t^2 c_1 & \ldots & \theta g_t \varepsilon_t^2 c_{N+K-2} & \theta g_t \varepsilon_t^2 c_{N+K-1} \\ 1 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & 0 \\ \vdots & & & \\ 0 & \ldots & 1 & 0, \end{pmatrix} \tag{3.11}$$

$$B_t = (m g_t \varepsilon_t^2, 0, \ldots, 0)',$$

$$\tilde{c} = (c_1, c_2, \ldots, c_{N+K-1})'.$$

Again, we are put in the setting of model (2.4) with strictly stationary ergodic coefficients. If we could find conditions to meet the assumptions in Theorem 3.1 of Glasserman and Yao [21], then model (3.11) will have a unique strictly stationary solution. But the problem is how to evaluate the top Lyapounov exponent associated with the stationary ergodic matrices. We approach this problem in three steps: (1) $K = 1, N = 1$ (2) $K = 1, N > 1$ (3) $K > 1$ and $N \geq 1$ due to the complicated structure of $A_t(\tilde{c})$.

When $KN = 1$, $A_t(\tilde{c})$ is just a scaler and the top Lyapounov exponent is easy to compute. The sufficient condition of stationary solution comes directly from Theorem 1 of Brandt [9] or Theorem 3.1 of Glasserman and Yao [21] (see Appendix).

Proposition 3.2 *When $KN = 1$, under the assumptions that $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$, model (3.8) has a unique strictly stationary ergodic solution if $\theta < 1$.*

Proof of Proposition 3.2 *When $KN = 1$, r_t^2 defined in model (3.8) is reduced to $r_t^2 = m g_t \varepsilon_t^2 + \theta g_t \varepsilon_t^2 r_{t-1}^2$. Notice that when $\alpha > 0, \beta > 0, \alpha + \beta < 1$, $\{g_t \varepsilon_t^2, t \in \mathbb{Z}\}$ is strictly stationary ergodic. If $0 < \theta < 1$,*

$$E \log(\theta g_0 \varepsilon_0^2) \leq \log E(\theta g_0 \varepsilon_0^2) = \log \theta < 0,$$

$$E \log(m g_0 \varepsilon_0^2) \leq \log E(m g_0 \varepsilon_0^2) = \log m < \infty$$

the conclusion follows from Theorem 1 of Brandt [9] or Theorem 3.1 of Glasserman and Yao [21] (see Appendix) directly.

When $K = 1$ and $N > 1$, the weight function vanishes and $A_t(\tilde{c})$ is simplified as

$$A_t(\tilde{1}) = \begin{pmatrix} \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 & \dots & \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \doteq A_t. \qquad (3.12)$$

Introduce

$$H = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, G = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

and define $M(a) = aH + G$, then

$$A_t = M(\theta g_t \varepsilon_t^2).$$

Matrix of this type is encountered a lot when one expresses an autoregressive model using a Markovian representation. The next lemma gives the basic properties of matrix $M(a)$ with a positive.

Lemma 3.1 *For matrix $M(a)$ with $a > 0$, we have the following properties:*

1. *Let $f(\lambda)$ be the characteristic function of $M(a)$. For any positive number k, $\rho(M(a)) < k$ if $f(k) > 0$;*
2. *$\rho(M(a))$ is increasing in a and it is a concave function of a.*

Proof of Lemma 3.1 *1. If $|\lambda| \geq k$,*

$$\begin{aligned} |f(\lambda)| &= |\lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \dots - a\lambda^2 - a\lambda - a| \\ &\geq |\lambda^N|(1 - \frac{a}{|\lambda|} - \frac{a}{|\lambda^2|} - \dots - \frac{a}{|\lambda^N|}) \\ &\geq k^N(1 - \frac{a}{k} - \frac{a}{k^2} - \dots - \frac{a}{k^N}) \\ &= f(k) > 0 \end{aligned}$$

therefore, $\rho(M(a)) < k$.

2. *Note A is nonnegative and irreducible. By Perron-Frobenius theory, $\rho(A)$ is the maximal positive root of $f(\lambda) = det(\lambda I - A)$. It is simple and $\rho(A) \geq |\lambda|$ for each root λ of $f(\lambda) = 0$*

   For ease of reference, we use $\lambda(a)$ or λ for $\rho(A)$. Since $f(\lambda) = \lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \ldots - a\lambda^2 - a\lambda - a = 0$,

$$a = \frac{\lambda^N}{\lambda^{N-1} + \lambda^{N-2} + \ldots + \lambda^2 + \lambda + 1} = \lambda - 1 + g(\lambda) \qquad (3.13)$$

   where $g(\lambda) = \frac{1}{h(\lambda)}$ and $h(\lambda) = \lambda^{N-1} + \lambda^{N-2} + \ldots + \lambda^2 + \lambda + 1$

   Since λ is a smooth function of a, to prove λ is a concave function of a is equivalent to show that $\frac{d^2\lambda(a)}{da^2} < 0$. Taking derivative on both sides of (3.13) with respect to a, we could have

$$1 = (1 + g^{'})\lambda^{'} \qquad (3.14)$$

   where $g^{'} = \frac{dg(\lambda)}{d\lambda}$ and $\lambda^{'} = \frac{d\lambda(a)}{da}$

   Furthermore

$$0 = (1 + g^{'})\lambda^{''} + g^{''}(\lambda^{'})^2 \qquad (3.15)$$

   On the other hand, put $f(\lambda) = 0$ 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

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

therefore,

$$\Delta = \frac{N\lambda^{N-2}[(N-1)\lambda^{N+1} - (N+1)\lambda^{N} + (N+1)\lambda - (N-1)]}{(\lambda-1)^3} \tag{3.16}$$

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 $D(1) = D^{'}(1) = D^{''}(1) = 0$ *and* $D^{''} < 0$ *for* $0 < \lambda < 1$*, while on* $\lambda > 1$, $D^{''} > 0$*, which 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.*

Proposition 3.3 *For* $K = 1$ *and* $N > 1$*, if* $\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$ *and* $\theta < \frac{\eta^{N-1}}{1+\eta+...+\eta^{N-1}}$*, the top Lyapounov exponent* γ *associated with* A_t *(defined in (3.12)) is negative.*

Proof of Proposition 3.3 *Note when* $\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$*, under Frobenius norm*

$$\begin{aligned}\|A_0\| &= \sqrt{tr(A_0^* A_0)} \\ &= \sqrt{N\theta^2 g_0^2 \varepsilon_0^4 + N - 1} \geq 1\end{aligned}$$

$$\begin{aligned}E(\log\|A_0\|)^+ &= E\log\sqrt{N\theta^2 g_0^2 \varepsilon_0^4 + N - 1} \\ &\leq \frac{1}{2}\log[N\theta^2 E(g_0^2\varepsilon_0^4) + N - 1] < \infty\end{aligned}$$

Furthermore,

$$A_t = g_t(\theta\varepsilon_t^2 H + \frac{1}{g_t}G) \leq g_t(\theta\varepsilon_t^2 H + \frac{1}{\eta}G).$$

Let $\tilde{A}_t \doteq \theta\varepsilon_t^2 H + \frac{1}{\eta}G$, *we have*

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

It follows that

$$\gamma \leq E \log g_0 + \lim_t \frac{1}{1+t} E \log \|\tilde{A}_t \tilde{A}_{t-1} \dots \tilde{A}_0\|.$$

Let $\tilde{\gamma}$ *be the top Lyapounov exponent associated with sequence* $\{\tilde{A}_t, t \in \mathbb{Z}\}$, *then* $\gamma \leq \tilde{\gamma}$.

Since $\tilde{A}_t$*'s are iid and nonnegative, according to Lemma 2.1, if* $\rho[E(\tilde{A}_0)] < 1$, *then* $\gamma \leq \tilde{\gamma} < 0$.

Note $E(\tilde{A}_0) = \frac{1}{\eta} M(\theta\eta)$ *and* $\rho[E(\tilde{A}_0)] < 1$ *is equivalent to* $\rho(M(\theta\eta)) < \eta$. *Its sufficient condition is*

$$f(\eta) = det(\eta I_N - M(\theta\eta)) > 0,$$

by Lemma 3.1 which is satisfied if $\theta < \frac{\eta^{N-1}}{1+\eta+\ldots+\eta^{N-1}}$.

Proposition 3.4 *When* $K > 1$ *and* $N \geq 1$, *the top Lyapounov exponent associated with* $A_t(\tilde{c})$ *defined in (3.11) is negative if* $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$ *and* $\theta < \frac{\eta^{K+N-2}}{1+\eta+\ldots+\eta^{K+N-2}}$.

Proof of Proposition 3.4 *Under the Frobenius norm,*

$$\begin{aligned} \|A_0\| &= \sqrt{tr(A_0^* A_0)} \\ &= \sqrt{\theta^2 g_0^2 \varepsilon_0^4 (c_1^2 + \ldots + c_{N+K-1}^2) + N + K - 2} \geq 1. \end{aligned}$$

And when $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$,

$$\begin{aligned} E(\log \|A_0(\tilde{c})\|)^+ &= E \log \sqrt{\theta^2 g_0^2 \varepsilon_0^4 (c_1^2 + \ldots + c_{N+K-1}^2) + N + K - 2} \\ &\leq \frac{1}{2} \log[\theta^2 (c_1^2 + \ldots + c_{N+K-1}^2) E(g_0^2 \varepsilon_0^4) + N + K - 2] < \infty \end{aligned}$$

The top Lyapounov exponent associated with $A_t(\tilde{c})$ *is*

$$\gamma(\tilde{c}) = \lim_{t\to\infty} \frac{1}{t+1} E \log \|A_t(\tilde{c}) A_{t-1}((\tilde{c})) \dots A_0(\tilde{c})\|$$

Define

$$g_n(\tilde{c}) = \|A_t(\tilde{c})A_{t-1}(\tilde{c})\dots A_0(\tilde{c})\|^2$$

Since g_n is a polynomial in $\tilde{c}$ and all the entries in the matrices are nonnegative, the coefficients of c_j $(1 \le j \le K+N-1)$ are positive which implies that, for every n, $g_n(\tilde{c})$ is nondecreasing in each c_j. In other words, $g(\tilde{c}) \le g(\tilde{1})$. It follows from Proposition 3.3 that

$$\gamma(\tilde{c}) \le \gamma(\tilde{1}) < 0 \text{ if } \theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}.$$

Combining the above results, we have

Proposition 3.5 *Suppose that $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$, $m > 0$, and $KN > 1$. The sufficient condition for the existence and uniqueness of a strictly stationary ergodic solution to model (3.8) is $\theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}$ and $\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$.*

Proof of Proposition 3.5 *Under the assumption $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$,*

$$\begin{aligned} E(\log\|A_0\|)^+ &= E\log\sqrt{\theta^2 g_0^2\varepsilon_0^4(c_1^2+\dots+c_{N+K-1}^2)+N+K-2} \\ &\le \frac{1}{2}\log[\theta^2(c_1^2+\dots+c_{N+K-1}^2)E(g_0^2\varepsilon_0^4)+N+K-2] < \infty \\ E(\log\|B_0\|)^+ &= E(\log m + \log g_0 + \log\varepsilon_0^2)^+ < \infty \end{aligned}$$

Further $\gamma < 0$ is derived from Proposition 3.3 and 3.4 if $\theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}$. Applying Theorem 3.1 of Glasserman and Yao [21] (see Appendix), there exists a unique strictly stationary ergodic solution to model (3.8).

Corollary 3.2 *GARCH-MIDAS model with rolling window RV in (3.2) has a unique strictly stationary ergodic solution if*

1. *$\theta < 1$ when $KN = 1$*
2. *or $\theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}$ and $\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$ when $KN > 1$.*

The resulting process is nonanticipative (or causal). In addition, the low-frequency volatility component τ is strictly stationary ergodic as well.

4 Asymptotic properties of GARCH-MIDAS model

The last property in section 3 tells us that GARCH-MIDAS model with rolling-window RV (3.8) has a unique strictly stationary ergodic solution under certain regularity conditions. In this section, we will follow this line and study the consistency and asymptotic behavior of maximum likelihood estimates (MLE) of this model.

The parameter space we will consider in this section is

$$
\begin{aligned}
\mathcal{U} = &\{\Phi = (\alpha, \beta, m, \theta, \omega)' \in \mathcal{R}^5 : \alpha > 0, \beta > 0, m > 0 \\
&(\alpha+\beta)^2 + 2\alpha^2 < 1, 0 < \theta < \frac{\eta^{K+N-2}}{1+\eta+\ldots+\eta^{K+N-2}}\}
\end{aligned}
$$

Suppose that $\Phi_0 = (\alpha_0, \beta_0, m_0, \theta_0, \omega_0)$ is the true parameter such that $\Phi_0 \in \mathcal{U}$. Given a sequence of $\{r_t, 1 \le t \le T\}$ where $T \gg N + K$ which are generated by the following dynamics

$$
\begin{aligned}
r_t &= \sqrt{g_t(\Phi_0)\tau_t(\Phi_0)}\varepsilon_t, \quad t \in \mathbb{Z} \\
g_t(\Phi_0) &= (1-\alpha_0-\beta_0) + \alpha_0 \tfrac{r_{t-1}^2}{\tau_{t-1}(\Phi_0)} + \beta_0 g_{t-1}(\Phi_0) \\
\tau_t(\Phi_0) &= m_0 + \theta_0 \textstyle\sum_{k=1}^{K} \varphi_k(\omega_0) RV_{t-k}, RV_t = \sum_{j=0}^{N-1} r_{t-j}^2
\end{aligned}
\tag{4.1}
$$

the MLE of Φ_0 (denoted as $\hat{\Phi}_T$) is the minimizer of

$$
L_T(\Phi) = \frac{1}{T} \sum_{t=N+K}^{T} [\log g_t(\Phi) + \log \tau_t(\Phi) + \frac{r_t^2}{g_t(\Phi)\tau_t(\Phi)}]
$$

For ease of reference, we use $\{\phi_i, 1 \le i \le 5\}$ to refer to the parameter set $\{\alpha, \beta, m, \theta, \omega\}$ when there is no confusion. Introduce

$$
l_t(\Phi) \equiv \log g_t(\Phi) + \log \tau_t(\Phi) + \frac{r_t^2}{g_t(\Phi)\tau_t(\Phi)}.
$$

The gradient of $L_T(\Phi)$ is

$$
\nabla L_T(\Phi) = \frac{1}{T} \sum_{t=N+K}^{T} \nabla l_t(\Phi) = \frac{1}{T} \sum_{t=N+K}^{T} (s_t^\alpha, s_t^\beta, s_t^m, s_t^\theta, s_t^\omega)'(\Phi)
$$

with $s_t^{\phi_i}(\Phi) = \frac{\partial l_t(\Phi)}{\partial \phi_i}, i = 1, \ldots, 5$. The hessian matrix of $L_T(\Phi)$ is

$$H(L_T)(\Phi) = (\frac{\partial^2 L_T(\Phi)}{\partial \phi_i \partial \phi_j})_{1 \leq i,j \leq 5} = \frac{1}{T} \sum_{t=N+K}^{T} H(l_t)(\Phi) \tag{4.2}$$

As a convention, if a function is expressed without specifying Φ, we assume that it is evaluated at the true parameter Φ_0.

The following main result establishes the existence and uniqueness of the consistent and asymptotically normal estimator $\hat{\Phi}_T$.

Proposition 4.1 *Assume $\{r_t, 1 \leq t \leq T\}$ is generated from model (4.1) with $\Phi_0 \in \mathcal{U}$. Then there exists a fixed open neighborhood $N(\Phi_0) \subset \overline{N(\Phi_0)} \subset \mathcal{U}$ of Φ_0 such that with probability tending to 1 as T goes to ∞, $L_T(\Phi)$ has a unique minimum $\hat{\Phi}_T$ in $N(\Phi_0)$ such that*

$$\hat{\Phi}_T \xrightarrow{P} \Phi_0$$

and

$$\sqrt{T}(\hat{\Phi}_T - \Phi_0) \Rightarrow N(0, \Sigma_I^{-1} \Sigma_S \Sigma_I^{-1})$$

where $\Sigma_I = E(H(l_1)), \Sigma_S = E(\nabla l_1 \nabla l_1')$.

The next proposition gives the consistent estimate of the asymptotic covariance matrix $\Sigma_I^{-1} \Sigma_S \Sigma_I^{-1}$.

Proposition 4.2 *With the same regularity conditions as Proposition 4.1, we have*

$$\hat{\Sigma}_{I(T)}^{-1} \hat{\Sigma}_{S(T)} \hat{\Sigma}_{I(T)}^{-1} \xrightarrow{P} \Sigma_I^{-1} \Sigma_S \Sigma_I^{-1}$$

where $\hat{\Sigma}_{I(T)} = \frac{1}{T} \sum_{t=N+K}^{T} H(l_t)(\hat{\Phi}_T)$ and $\hat{\Sigma}_{S(T)} = \frac{1}{T} \sum_{t=N+K}^{T} \nabla l_t \nabla l_t'(\hat{\Phi}_T)$.

4.1 Proofs of Proposition 4.1 and Proposition 4.2

To establish the consistency and asymptotic normality of $\hat{\Phi}_T$, we need the following helpful lemmas.

Lemma 4.1 *Let $\{X_n, \mathcal{F}_n : n \geq 1\}$ be a strictly stationary ergodic martingale difference sequence such that $\sigma^2 = E(X_1^2) < \infty$. Then*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j \Rightarrow N(0, \sigma^2).$$

Proof of Lemma 4.1 *Define* $X_{nj} = \frac{X_j}{\sigma\sqrt{n}}$, $1 \le j \le n$. *Note for any* ε,

$$P(\max_{j\le n} |X_{nj}| > \varepsilon) \le P(\sum_{j\le n} X_{nj}^2 I(|X_{nj}| > \varepsilon) > \varepsilon^2) \le \frac{1}{\varepsilon^2} E(\sum_{j\le n} X_{nj}^2 I(|X_{nj}| > \varepsilon))$$

$$\max_{j\le n} X_{nj}^2 \le \varepsilon^2 + \sum_{j\le n} X_{nj}^2 I(|X_{nj}| > \varepsilon)$$

Since

$$\begin{aligned} E(\sum_{j\le n} X_{nj}^2 I(|X_{nj}| > \varepsilon)) &= \frac{1}{\sigma^2 n} \sum_{j\le n} E(X_j^2 I(|X_j| > \varepsilon\sigma\sqrt{n})) \\ &= \frac{1}{\sigma^2} E(X_1^2 I(|X_1| > \varepsilon\sigma\sqrt{n})) \to 0 \text{ as } n \to \infty \end{aligned}$$

due to the fact that $P\{|X_1| > \varepsilon\sigma\sqrt{n}\} \to 0$ *as* $n \to \infty$. *Therefore* $\{\max_{j\le n} |X_{nj}|\}$ *is uniformly bounded in* L_2 *norm and* $\max_{j\le n} |X_{nj}| \xrightarrow{P} 0$.

Note also that

$$\sum_j X_{nj}^2 = \frac{1}{\sigma^2 n} \sum_j X_j^2 \to 0 \text{ almost surely, as } n \to \infty$$

by Birkhoff's Ergodic Theorem. It follows from martingale central limit theorem of Mcleish [30] that $\frac{1}{\sqrt{n}} \sum_{j=1}^n X_j \Rightarrow N(0, \sigma^2)$.

Lemma 4.1 presents a fact for a one-dimensional situation. To extend it to higher dimensions, we need to use Cramér-Wold Device of Bilingsley [6]. Moreover, we could derive the following result.

Lemma 4.2 *Under the assumptions in Proposition 4.1,*

$$\sqrt{T}\nabla L_T(\Phi_0) \Rightarrow N(0, \Sigma_S)$$

where

$$\Sigma_S = E(\nabla l_1 \nabla l_1') = \begin{pmatrix} E(s_1^{\alpha 2}) & E(s_1^\alpha s_1^\beta) & E(s_1^\alpha s_1^m) & E(s_1^\alpha s_1^\theta) & E(s_1^\alpha s_1^\omega) \\ * & E(s_1^{\beta 2}) & E(s_1^\beta s_1^m) & E(s_1^\beta s_1^\theta) & E(s_1^\beta s_1^\omega) \\ * & * & E(s_1^{m2}) & E(s_1^m s_1^\theta) & E(s_1^m s_1^\omega) \\ * & * & * & E(s_1^{\theta 2}) & E(s_1^\theta s_1^\omega) \\ * & * & * & * & E(s_1^{\omega 2}) \end{pmatrix} \tag{4.3}$$

Remark 4.1 *Σ_S is symmetric. We only display its upper triangular part here for brevity. In the rest of paper, we will express a symmetric matrix this way.*

Proof of Lemma 4.2 *According to Cramér-Wold Device, it is sufficient to show that for any $t = (t_1, t_2, t_3, t_4, t_5)' \in \mathbb{R}^5$,*

$$\sqrt{T}t'\nabla L_T(\Phi_0) \Rightarrow t'Z$$

where $Z \sim N(0, \Sigma_S)$.

Notice that

$$\sqrt{T}t'\nabla L_T(\Phi_0) = \frac{1}{\sqrt{T}} \sum_{t=N+K}^{T} t_1 s_t^\alpha + t_2 s_t^\beta + t_3 s_t^m + t_4 s_t^\theta + t_5 s_t^\omega.$$

Let

$$s_t = t_1 s_t^\alpha + t_2 s_t^\beta + t_3 s_t^m + t_4 s_t^\theta + t_5 s_t^\omega.$$

The strictly stationary ergodic solutions g_t and r_t are measurable functions of $\{\varepsilon_j : 1 \le j \le t\}$, so is τ_t. It follows that s_t^α,s_t^β,s_t^m,s_t^θ,s_t^ω are also measurable functions of $\{\varepsilon_j : 1 \le j \le t\}$. Therefore $\{s_t\}$ is a strictly stationary and ergodic process (due to Stout [33]).

$$\begin{aligned} E(s_t|\mathcal{F}_{t-1}) &= t_1 E(s_t^\alpha|\mathcal{F}_{t-1}) + t_2 E(s_t^\beta|\mathcal{F}_{t-1}) + t_3 E(s_t^m|\mathcal{F}_{t-1}) \\ &\quad + E(s_t^\theta|\mathcal{F}_{t-1}) + E(s_t^\omega|\mathcal{F}_{t-1}) \\ &= 0 \\ E(s_t^2) &\le t_1^2 E(s_t^{\alpha 2}) + t_2^2 E(s_t^{\beta 2}) + t_3^2 E(s_t^{m2}) + t_4^2 E(s_t^{\theta 2}) + t_5^2 E(s_t^{\omega 2}) \end{aligned} \tag{4.4}$$

Now we need to show that $E(s_t^2) < \infty$. Since

$$\frac{\partial l_t(\Phi)}{\partial \phi_i} = (1 - \frac{r_t^2}{g_t(\Phi)\tau_t(\Phi)})(\frac{\partial \tau_t/\partial \phi_i}{\tau_t}(\Phi) + \frac{\partial g_t/\partial \phi_i}{g_t}(\Phi)),$$

evaluated at the true parameters

$$s_t^{\phi_i} = \frac{\partial l_t}{\partial \phi_i} = (1 - \varepsilon_t^2)(\frac{\partial \tau_t/\partial \phi_i}{\tau_t} + \frac{\partial g_t/\partial \phi_i}{g_t}), i = 1, \ldots, 5.$$

Note also

$$s_t^{\phi_i 2} \le 2(1 - \varepsilon_t^2)^2[(\frac{\partial \tau_t/\partial \phi_i}{\tau_t})^2 + (\frac{\partial g_t/\partial \phi_i}{g_t})^2].$$

For $i=1$, ie $\phi_1=\alpha$, $\frac{\partial\tau_t}{\partial\alpha}=0$, and $\frac{\partial g_t}{\partial\alpha}=\sum_{j=1}^{\infty}\beta^{j-1}g_{t-j}\varepsilon_{t-j}^2-\frac{1}{1-\beta}$, we have

$$E(s_t^{\alpha 2})\le\frac{2E(1-\varepsilon_t^2)^2}{(1-\alpha_0-\beta_0)^2}[(\frac{1}{(1-\beta_0)^2})+\sum_{j=1}^{\infty}\beta_0^{2j-2}Eg_1^2E\varepsilon_1^4]<\infty.$$

For $i=2$, ie $\phi_2=\beta$, $\frac{\partial\tau_t}{\partial\beta}=0$, and $\frac{\partial g_t}{\partial\beta}=\sum_{j=1}^{\infty}\beta^{j-1}g_{t-j}-\frac{1}{1-\beta}$, we have

$$E(s_t^{\beta 2})\le\frac{2E(1-\varepsilon_t^2)^2}{(1-\alpha_0-\beta_0)^2}[(\frac{1}{(1-\beta_0)^2})+\sum_{j=1}^{\infty}\beta_0^{2j-2}Eg_1^2E\varepsilon_1^4]<\infty.$$

For $i=3$, ie $\phi_3=m$, $\frac{\partial\tau_t/\partial m}{\tau_t}=\frac{1}{\tau_t}$, $\frac{\partial g_t}{\partial m}=-\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{\partial\tau_{t-1}/\partial m}{\tau_{t-1}})=-\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{1}{\tau_{t-1}})$, we have

$$E(s_t^{m2})\le 2E(1-\varepsilon_t^2)^2[\frac{1}{m_0^2}+\frac{E(g_1^2)E(\varepsilon_1^4)}{(1-\alpha_0-\beta_0)^2m^2}]<\infty.$$

For $i=4$, ie $\phi_4=\theta$, $\frac{\partial\tau_t/\partial\theta}{\tau_t}=\frac{1}{\theta_0}(1-\frac{m_0}{\tau_t})$, $\frac{\partial g_t}{\partial\theta}=-\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{\partial\tau_{t-1}/\partial\theta}{\tau_{t-1}})$, we have

$$E(s_t^{\theta 2})\le 2E(1-\varepsilon_t^2)^2[\frac{1}{\theta_0^2}+\frac{E(g_1^2)E(\varepsilon_1^4)}{(1-\alpha_0-\beta_0)^2\theta_0^2}]<\infty.$$

For $i=5$, ie $\phi_5=\omega$, $\frac{\partial\tau_t/\partial\omega}{\tau_t}=\frac{\sum\varphi_k^{'}RV_{t-k}}{\tau_t}\le\frac{\max_k\varphi_k^{'}(\omega_0)}{\min_k\varphi_k(\omega_0)}$ (without loss of generality, we could assume $\{\varphi_k, 1\le k\le K\}$ are all positive) and $\frac{\partial g_t}{\partial\omega}=-\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{\partial\tau_{t-1}/\partial\omega}{\tau_{t-1}})$. We have

$$E(s_t^{\omega 2})\le 2E(1-\varepsilon_t^2)^2[1+\frac{E(g_1^2)E(\varepsilon_1^4)}{(1-\alpha_0-\beta_0)^2}](\frac{\max_k\varphi_k^{'}(\omega_0)}{\min_k\varphi_k(\omega_0)})^2<\infty.$$

Therefore $E(s_1^2)<\infty$. Applying Lemma 4.1, we get

$$\sqrt{T}t^{'}\frac{\partial}{\partial\phi}L_T(\Phi_0)\Rightarrow N(0,t^{'}\Omega t)\quad\forall t\in\mathbb{R}^5.$$

The following lemma evaluates the probabilistic property of the Hessian matrix of L_T with value taken at $\Phi=\Phi_0$.

Lemma 4.3 *Under the assumptions in Proposition 4.1*

$$H(L_T)(\Phi_0) \xrightarrow{P} \Sigma_I$$

where

$$\Sigma_I = E(H(l_1)) = \begin{pmatrix} E(\frac{\partial^2 l_1(\Phi_0)}{\partial \alpha^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \alpha \partial \beta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \alpha \partial m}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \alpha \partial \theta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \alpha \partial \omega}) \\ * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \beta^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \beta \partial m}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \beta \partial \theta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \beta \partial \omega}) \\ * & * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial m^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial m \partial \theta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial m \partial \omega}) \\ * & * & * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \theta^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \theta \partial \omega}) \\ * & * & * & * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial \omega^2}) \end{pmatrix} \quad (4.5)$$

Proof of Lemma 4.3 *Introduce* $D_T = (d^T_{i,j})_{1 \le i,j \le 5} = H(L_T)(\Phi_0)$ *and each element in* Σ_I *is denoted by* σ^2_{ij}. *We need to show that*

$$\lim_{T \to \infty} P(\|D_T - \Sigma_I\| > \epsilon) = 0 \quad \forall \epsilon > 0$$

where $\|.\|$ *is an arbitrary matrix norm.*

All norms on the finite dimensional norm space are equivalent, which implies that all the matrix norms on $\mathbb{C}^{n \times n}$ *should be equivalent. Thus, we only need to show the result is true for Frobenious norm. Under Frobenious norm,*

$$\|D_T - \Sigma_I\|^2 = trace[(D_T - \Sigma_I)^*(D_T - \Sigma_I)] = \sum_{i,j=1}^{5} (d^T_{i,j} - \sigma^2_{i,j})^2$$

Note

$$d^T_{i,j} = \frac{1}{T} \sum_{t=N+K}^{T} \frac{\partial^2 l_t(\Phi_0)}{\partial \phi_i \partial \phi_j},$$

and $\frac{\partial^2 l_t(\Phi_0)}{\partial \phi_i \partial \phi_j}$ *is a measurable function of* $\{\varepsilon_s, s \le t\}$, *hence is strictly stationary ergodic. By Birkhoff's ergodic theorem,*

$$d^T_{i,j} \xrightarrow{P} \sigma^2_{i,j}$$

ie.

$$P(\|D_T - \Sigma_I\| > \epsilon) \le \sum_{i,j=1}^{5} P(|d^T_{i,j} - \sigma^2_{i,j}| > \frac{\epsilon}{\sqrt{5}}) \to 0.$$

Therefore,

$$D_T \xrightarrow{P} \Sigma_I.$$

Next, we want to show the third derivatives of L_T is locally bounded in a 'weak' sense, i.e.,

Lemma 4.4 *Let $N(\Phi_0)$ be an arbitrary open set of Φ_0 such that $N(\Phi_0) \subset \overline{N(\Phi_0)} \subset \mathcal{U}$. Then there exists a random variable c_T which satisfies*

$$\max_{i,j,k=1,\ldots,5} \sup_{\Phi \in N(\Phi_0)} |\frac{\partial^3 L_T(\Phi)}{\partial\phi_i \partial\phi_j \partial\phi_k}| \leq c_T$$

and

$$c_T \xrightarrow{P} c \text{ for some constant } c.$$

Proof of Lemma 4.4

$$\frac{\partial^3 L_T(\Phi)}{\partial\phi_i \partial\phi_j \partial\phi_k} = \frac{1}{T} \sum_{t=N+K}^{T} \frac{\partial^3 l_t(\Phi)}{\partial\phi_i \partial\phi_j \partial\phi_k}$$

Note $|\frac{\partial^3 l_t(\Phi)}{\partial\phi_i \partial\phi_j \partial\phi_k}|$ is continuous in Φ, there exists an open neighborhood $N(\Phi_0)$ of Φ_0 such that $\overline{N(\Phi_0)} \subset \mathcal{U}$ and further, there exists a point $\tilde{\Phi}_t^{i,j,k} \in \overline{N(\Phi_0)}$ such that

$$|\frac{\partial^3 L_T(\Phi)}{\partial\phi_i \partial\phi_j \partial\phi_k}| \leq \frac{1}{T} \sum_{t=N+K}^{T} w_t^{i,j,k}$$

where

$$w_t^{i,j,k} = |\frac{\partial^3 l_t(\tilde{\Phi}_t^{i,j,k})}{\partial\phi_i \partial\phi_j \partial\phi_k}|.$$

Therefore

$$\max_{i,j,k=1,\ldots,5} \sup_{\phi \in N(\phi_0)} |\frac{\partial^3 L_T(\Phi_0)}{\partial\phi_i \partial\phi_j \partial\phi_k}| \leq \frac{1}{T} \sum_{t=N+K}^{T} (\sum_{i,j,k=1}^{5} w_t^{i,j,k}).$$

Further, let $w_t = \sum_{i,j,k=1}^{5} w_t^{i,j,k}$. Since $\{w_t\}$ is a strictly stationary ergodic sequence, by Birkhoff's ergodic theorem,

$$\frac{1}{T} \sum_{t=N+K}^{T} w_t \xrightarrow{P} E(w_1).$$

With the above established results, we can complete the proof of Proposition 4.1.

Proof of Proposition 4.1 *Combine the lemmas 4.2, 4.3, and 4.4, and apply Lemma 1 of Jensen and Rahbek [26] (see Appendix). The existence and uniqueness of the consistent and asymptotic normal estimator $\hat{\Phi}_T$ are ensured.*

The proof of Proposition 4.2 needs one more lemma.

Lemma 4.5 *Let $\{x_n(\theta), n = 1, 2, \ldots\}$ be a sequence of random variables defined on probability space $\{\Omega, \mathcal{F}, P\}$ such that x_n is uniformly continuous in θ and for each fixed θ, $x_n(\theta) \xrightarrow{P} x(\theta)$. Suppose that $\hat{\theta}_n \xrightarrow{P} \theta_0$, then*

$$x_n(\hat{\theta}_n) \xrightarrow{P} x(\theta_0).$$

Proof of Lemma 4.5 *For any $\varepsilon > 0$,*

$$\begin{aligned} P(|x_n(\hat{\theta}_n) - x(\theta_0)| > \varepsilon) \leq & P(|x_n(\theta_0) - x(\theta_0)| > \varepsilon) + P(|\hat{\theta}_n - \theta_0| > \varepsilon) \\ & + P(|x_n(\hat{\theta}_n) - x_n(\theta_0)| > \varepsilon, |\hat{\theta}_n - \theta_0| < \varepsilon) \end{aligned}$$

The conclusion follows from the inequality immediately.

Proof of Proposition 4.2 *For each $\Phi \in \overline{N(\Phi_0)}$, $H(l_t)(\Phi), \nabla l_t \nabla l_t'(\Phi)$ are strictly stationary ergodic,*

$$\frac{1}{T} \sum H(l_t)(\Phi) \xrightarrow{P} E(H(l_1)(\Phi)),$$

and

$$\frac{1}{T} \sum \nabla l_t \nabla l_t'(\Phi) \xrightarrow{P} E(\nabla l_1 \nabla l_1'(\Phi))$$

due to Birkhoff's ergodic theorem. Also consider the fact that $\hat{\Phi}_T \xrightarrow{P} \Phi$, and $H(l_t)(\Phi), \nabla l_t \nabla l_t'(\Phi)$ are uniformly continuous in $\Phi \in \overline{N(\Phi_0)}$. Therefore,

$$\frac{1}{T} \sum H(l_t)(\hat{\Phi}_T) \xrightarrow{P} E(H(l_1))$$

and

$$\frac{1}{T} \sum \nabla l_t \nabla l_t'(\hat{\Phi}_T) \xrightarrow{P} E(\nabla l_1 \nabla l_1').$$

Applying continuous mapping theorem,

$$\hat{\Sigma}_{I(T)}^{-1} \hat{\Sigma}_{S(T)} \hat{\Sigma}_{I(T)}^{-1} \xrightarrow{P} \Sigma_I^{-1} \Sigma_S \Sigma_I^{-1}.$$

5 Conclusion

This paper focused on the distributional properties of two volatility component models: the restricted GARCH(2,2) model of Engle and Lee, the GARCH-MIDAS model of Engle, Ghysels and Sohn. The restricted GARCH(2,2) model structured the conditional variance as the sum of low-frequency and high-frequency stochastic components. It was shown that, under certain regularity conditions on the parameter space, it was strictly stationary ergodic and β-mixing. In the GARCH-MIDAS model, the conditional volatility was characterized as the multiplicative effects of low-frequency and high-frequency stochastic components. It was an extension of Spline-GARCH model of Engle and Rangel where the low-frequency volatility was fitted by an exponential quadratic spline, a deterministic structure. For GARCH-MIDAS model with fixed time span realized volatility, we showed that it could admit a covariance stationary solution in a specific parameter space. We also derived sufficient conditions for the existence and uniqueness of strictly stationary ergodic solution to the GARCH-MIDAS model with rolling window realized volatility. Further, this paper showed that its maximum likelihood estimates were locally consistent and asymptotically normal. Its asymptotic variance-covariance matrix and associated consistent estimate were also specified.

6 Appendix

In the appendix, we list the theorems and lemmas used throughout this paper for readers' quick reference.

Theorem 6.1 (Theorem 3.1 of Glasserman and Yao [21]) *Suppose* $\{(A_t, B_t), t \in \mathbb{Z}\}$ *is a strictly stationary ergodic process and one of the conditions*

$$E(\log \|A_0\|)^+ < \infty, \gamma < 0, E(\log \|B_0\|)^+ < \infty$$

or

$$P(A_t \dots A_0 = 0) > 0 \text{ for some } n \geq 0$$

is satisfied, then

$$y_t = \sum_{j=1}^{\infty} [A_t \dots A_{t-j+1}] B_{t-j}, \;\; t \in \mathbb{Z} \tag{6.1}$$

converges almost surely. It is the only strictly stationary ergodic solution of $Y_t = A_t Y_{t-1} + B_t$.

Theorem 6.2 (Lemma 3.4 of Bougerol and Picard [8]) *Let* $\{F_t, t \in \mathbb{Z}\}$ *be a strictly stationary ergodic sequence of* $\mathbb{R}^{n\times n}$*-valued random matrices and suppose that* $E(\log^+ \|F_0\|) < \infty$ *and that*

$$\lim_{t\to\infty} \|F_t F_{t-1} \dots F_1\| = 0.$$

Then the top Lyapounov exponent associated with this sequence is strictly negative.

Theorem 6.3 *(Theorem 4.3 of Mokkadem [31]; Theorem 1 of Carrasco and Chen [10]) Given a polynomial random coefficient vector autoregressive model defined as*

$$Y_t = A(\varepsilon_t)Y_{t-1} + B(\varepsilon_t) \tag{6.2}$$

where $\{Y_t, t \in \mathbb{Z}^+\}$ *is a sequence of* $\mathbb{R}^m$*-valued random process,* $\{\varepsilon_t\}$ *is a* $\mathbb{R}^p-$*valued iid sequence,* $A(.)$ *is a* $m \times m$ *matrix-valued polynomial function and* $B(.)$ *is a* $m \times 1$ *vector-valued polynomial function. If further it satisfies the following assumptions:*

- *(A.1) The marginal probability distribution of* ε_t *is absolutely continuous with respect to Lebesgue measure on* $\mathbb{R}^p$ *and zero is in the interior of its support.*
- *(A.2)* $A(.)$ *and* $B(.)$ *are measurable with respect to the sigma-field generated by* ε_t.
- *(A.3) The spectral radius of* $A(0)$*, denoted by* $\rho[A(0)]$*, is less than 1.*
- *(A.4) The series* $\sum_{k=1}^{\infty}[A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})]B(\varepsilon_{t-k-1})$ *converges almost surely. The sequence* $A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})$ *converges (as* $k \to \infty$*) to the 0 matrix almost surely.*
- *(A.5) There exists a positive function* V *on* $\mathbb{R}^m$*, a compact set* K *of* $\mathbb{R}^m$ *with nonempty interior, and some positive numbers* $\delta > 0, \nu > 0$*, and* $0 < \lambda < 1$ *such that*

(i) $E[V(Y_t)|Y_{t-1} = y] \leq \lambda V(y) - \nu$ *if* $x \notin K$

(ii) $E[V(Y_t)|Y_{t-1} = y] \leq \delta$ *if* $x \in K$

Then $\{Y_t\}$ *is Markov geometrically ergodic and* $E[V(Y_t)] < \infty$*. Moreover, if* Y_0 *is initialized from an invariant distribution,* $\{Y_t\}$ *is strictly stationary and* β*-mixing with exponential decay.*

Theorem 6.4 *(Theorem 2 of Nelson and Cao [32]) For a GARCH(2,q) as below*

$$r_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{2} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j r_{t-j}^2$$

where ε_t*'s are iid and* $E(\varepsilon_t) = 0, var(\varepsilon_t) = 1$*. Let* z_1 *and* z_2 *be the roots of* $1 - \sum_{i=1}^{2} \beta_i z^{-i}$ *such that* $|z_2| \leq |z_1| \leq 1$ *and if* $z_1 = -z_2$*, we take* $z_1 > 0$*. Suppose further that* $1 - \sum_{i=1}^{2} \beta_i z^i$ *and* $\sum_{j=1}^{q} \alpha_j z^{j-1}$ *have no common roots. If we write* σ_t^2 *in ARCH(*∞*) form:*

$$\sigma_t^2 = \omega^\star + \sum_{k=0}^{\infty} \phi_k r_{t-k-1}^2,$$

then $\omega^\star \geq 0$ *and* $\phi_k \geq 0 (\forall k)$ *if and only if*

(B.1) $\omega^\star = \omega/(1 - z_1 - z_2 + z_1 z_2) \geq 0$

(B.2) z_1 *and* z_2 *are real*

(B.3) $z_1 > 0$

(B.4) $\sum_{j=0}^{q-1} z_1^{-j} \alpha_{j+1} > 0$

(B.5) $\phi_k \geq 0$ *for* $k = 0$ *to* q*.*

Theorem 6.5 *(Lemma 1 of Jensen and Rahbek [26]) Consider* $L_T(\Phi)$*, which is a function of the observations* $\{X_t\}_{1 \leq t \leq T}$ *and the parameter* $\Phi \in O \subseteq \mathbb{R}^k$*. Let* Φ_0 *be is an interior point of* O*. Assume that* $L_T(.) : \mathbb{R}^k \to \mathbb{R}$ *is three times continuously differentiable in* Φ *and that*

(A.1) As $T \to \infty$, $\sqrt{T}\nabla L_T(\Phi_0) \Rightarrow N(0, \Sigma_S)$, $\Sigma_S > 0$.

(A.2) As $T \to \infty$, $H(L_T)(\Phi_0) \xrightarrow{P} \Sigma_I > 0$.

(A.3) $\max_{i,j,h=1,\dots,k} \sup_{\Phi \in N(\Phi_0)} |\frac{\partial^3 L_T(\Phi)}{\partial\phi_i \partial\phi_j \partial\phi_k}| \le c_T$

where $N(\Phi_0)$ is a neighborhood of Φ_0 and $0 \le c_T \xrightarrow{P} c$, $0 < c < \infty$. Then there exists a fixed open neighborhood $U(\Phi_0) \subseteq N(\Phi_0)$ of Φ_0 such that

(B.1) *With probability tending to one as $T \to \infty$, there exists a minimum point $\hat{\Phi}_T$ of $L_T(\Phi)$ in $U(\Phi_0)$. In particular, $\hat{\Phi}_T$ is unique and solves $\nabla L_T(\hat{\Phi}_T) = 0$*

(B.2) As $T \to \infty$, $\hat{\Phi}_T \xrightarrow{P} \Phi_0$.

(B.3) As $T \to \infty$, $\sqrt{T}(\hat{\Phi}_T - \Phi_0) \Rightarrow N(0, \Sigma_I^{-1}\Sigma_S\Sigma_I^{-1})$

References

[1] T. Adrian and J. Rosenberg. Stock returns and volatility: Pricing the short-run and long-run components of market risk. *Journal of Finance* (forthcoming), 2008.

[2] H. Akaike. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics*, 26(1):363–387, 1974.

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

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

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

[6] P. Bilingsley. *Probability and Measure*. John Wiley and Sons, Series in Probability and Mathematical Statistics, New York, 1995.

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

[8] P. Bougerol and N. Picard. Strict Stationarity of Generalized Autoregressive Processes. *The Annals of Probability*, 20(4):1714–1730, 1992.

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

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

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

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

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

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

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

[16] R.F. Engle, E. Ghysels, and B. Sohn. On the Economic Sources of Stock Market Volatility. University of North Carolina at Chapel Hill, Manuscript, 2008.

[17] R.F. Engle and G.G.J. Lee. 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*, pages 475–497, 1999.

[18] R.F. Engle and J.G. Rangel. The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes. *Review of Financial Studies*, 21:1187–1222, 2008.

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

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

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

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

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

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

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

[26] S.T. Jensen and A. Rahbek. Asymptotic Inference For Nonstationary Garch. *Econometric Theory*, 20(06):1203–1226, 2004.

[27] JFC Kingman. Subadditive Ergodic Theory. *The Annals of Probability*, 1(6):883–899, 1973.

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

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

[30] DL McLeish. Dependent central limit theorems and invariance principles. *Ann. Probab*, 2(4):620–628, 1974.

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

[32] D.B. Nelson and C.Q. Cao. Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, 10(2):229–235, 1992.

[33] W.F. Stout. *Almost sure convergence*. New York: Academic Press, 1974.

[34] W. Vervaat. On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Advances in Applied Probability*, 11:750–783, 1979.