# ON BUFFERED THRESHOLD GARCH MODELS

Pak Hang Lo, Wai Keung Li, Philip L.H. Yu and Guodong Li

*University of Hong Kong*

**Supplementary Material**

This online supplementary material gives the proofs of Theorems 1 and 2

## S1 Proof of Theorem 1

Let $\mathcal{B}$ be the class of Borel sets of $\mathbb{R}^+$ and $\mathcal{U} = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$. For the temporarily homogeneous Markov chain $\{\boldsymbol{\sigma}_t^2\}$ defined as $\boldsymbol{\sigma}_t^2 = \mathbf{g}(\boldsymbol{\sigma}_{t-1}^2, \varepsilon_{t-1})$, we denote its state space by $(\mathbb{R}^+ \times \{0,1\}, \mathcal{B} \times \mathcal{U})$, and set its transition probability function as

$$P(\mathbf{x}, \mathbf{A}) = \int_{\mathbf{A}_\varepsilon} f(y)dy \quad \text{for } \mathbf{x} \in \mathbb{R}^+ \times \{0,1\} \text{ and } \mathbf{A} \in \mathcal{B} \times \mathcal{U},$$

where $\mathbf{A}_\varepsilon = \{y : \mathbf{g}(\mathbf{x}, y) \in \mathbf{A}\}$ and $f(\cdot)$ is the density of $\varepsilon_t$. From Theorem 1 of Feigin and Tweedie (1985) and Theorem 4 of Tweedie (1983), it is sufficient to show the following claims:

(i) $\{\boldsymbol{\sigma}_t^2\}$ is a Feller Markov chain;

(ii) $\{\boldsymbol{\sigma}_t^2\}$ is $\phi$-irreducible for some measure $\phi$ on the state space $(\mathbb{R}^+ \times \{0,1\}, \mathcal{B} \times \mathcal{U})$;

(iii) There exists a compact set $C \subset \mathbb{R}^+ \times \{0,1\}$ such that $\phi(C) > 0$ and a nonnegative continuous function (or test function) $V : \mathbb{R}^+ \times \{0,1\} \to \mathbb{R}$ such that

$$V(\mathbf{x}) \geq 1, \text{ for any } \mathbf{x} \in C,$$

and, for some $0 < c < 1$,

$$E\{V(\boldsymbol{\sigma}_t^2) | \boldsymbol{\sigma}_{t-1}^2 = \mathbf{x}\} \leq cV(\mathbf{x}), \text{ for any } \mathbf{x} \in C^c,$$

where $C^c$ is the complement of $C$.

We first prove Claim (i). Note that

$$\begin{aligned}\sigma_t^2 =&(\omega^{(1)} + \alpha^{(1)}\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta^{(1)}\sigma_{t-1}^2)I(\varepsilon_{t-1} \le r_L/\sigma_{t-1}) \\ &+ (\omega^{(1)} + \alpha^{(1)}\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta^{(1)}\sigma_{t-1}^2)R_{t-1}I(r_L/\sigma_{t-1} < \varepsilon_{t-1} \le r_L/\sigma_{t-1}) \\ &+ (\omega^{(2)} + \alpha^{(2)}\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta^{(2)}\sigma_{t-1}^2)(1 - R_{t-1})I(r_L/\sigma_{t-1} < \varepsilon_{t-1} \le r_L/\sigma_{t-1}) \\ &+ (\omega^{(2)} + \alpha^{(2)}\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta^{(2)}\sigma_{t-1}^2)I(\varepsilon_{t-1} > r_U/\sigma_{t-1})\end{aligned} \tag{S1.1}$$

and, for a bounded and continuous function $h(\cdot,\cdot)$,

$$\begin{aligned}&E\{h(\sigma_t^2, R_t)|(\sigma_{t-1}^2, R_{t-1}) = (x_1, x_2)\} \\ &\quad = E\{h(\omega^{(1)} + \alpha^{(1)}x_1\varepsilon_{t-1}^2 + \beta^{(1)}x_1, 1)I(\varepsilon_{t-1} \le r_L/x_1)\} \\ &\quad\quad + x_2E\{h(\omega^{(1)} + \alpha^{(1)}x_1\varepsilon_{t-1}^2 + \beta^{(1)}x_1, 1)I(r_L/x_1 < \varepsilon_{t-1} \le r_U/x_1)\} \\ &\quad\quad + (1 - x_2)E\{h(\omega^{(2)} + \alpha^{(2)}x_1\varepsilon_{t-1}^2 + \beta^{(2)}x_1, 0)I(r_L/x_1 < \varepsilon_{t-1} \le r_U/x_1)\} \\ &\quad\quad + E\{h(\omega^{(2)} + \alpha^{(2)}x_1\varepsilon_{t-1}^2 + \beta^{(2)}x_1, 0)I(\varepsilon_{t-1} \ge r_U/x_1)\}.\end{aligned} \tag{S1.2}$$

Denote $g_h(x_1, \varepsilon_{t-1}) = h(\omega^{(1)}+\alpha^{(1)}x_1\varepsilon_{t-1}^2+\beta^{(1)}x_1, 1)$ and $C_h = \sup_{x_1,x_2}|h(x_1, x_2)| < \infty$. Due to the dominated convergence theorem and the fact that $x_1 > \min\{\omega^{(1)}, \omega^{(2)}\} > 0$, it holds that

$$\begin{aligned}&|E\{g_h(x_1, \varepsilon_{t-1})I(\varepsilon_{t-1} \le r_L/x_1)\} - E\{g_h(x_1^*, \varepsilon_{t-1})I(\varepsilon_{t-1} \le r_L/x_1^*)\}| \\ &\quad \le E|g_h(x_1, \varepsilon_{t-1}) - g_h(x_1^*, \varepsilon_{t-1})| + C_h \cdot E|I(\varepsilon_{t-1} \le r_L/x_1) - I(\varepsilon_{t-1} \le r_L/x_1^*)| \\ &\quad = \int |g_h(x_1, y) - g_h(x_1^*, y)|f(y)dy + C_h \int_{r_L/x_1^*}^{r_L/x_1} f(y)dy \to 0\end{aligned}$$

as $|x_1^* - x_1| \to 0$, i.e. $E\{g_h(x_1, \varepsilon_{t-1})I(\varepsilon_{t-1} \le r_L/x_1)\}$ is continuous with respect to $x_1$. Similarly we can show that the other three terms at the right hand side of (S1.2) are continuous with respect to $x_1$. As a result, $E\{h(\sigma_t^2, R_t)|(\sigma_{t-1}^2, R_{t-1}) = (x_1, x_2)\}$ is continuous with respect to $x_1 \in \mathbb{R}^+$, and hence with respect to $(x_1, x_2) \in \mathbb{R}^+ \times \{0, 1\}$. Thus, the Markov chain $\{\boldsymbol{\sigma}_t^2\}$ is a Feller chain.

We next prove the irreducibility at Claim (ii), and first consider the case with $r_L \le r_U \le 0$. Note that, if $\varepsilon_j > 0$ for all $0 \le j \le t-1$, then the process will stay at the upper regime up to time $t$ and, by (S1.1),

$$\sigma_t^2 = \omega^{(2)}+(\alpha^{(2)}\varepsilon_{t-1}^2+\beta^{(2)}) \left[\omega^{(2)} \sum_{i=1}^{t-1}\prod_{j=2}^{i}(\alpha^{(2)}\varepsilon_{t-j}^2+\beta^{(2)}) + \sigma_0^2\prod_{j=2}^{t-2}(\alpha^{(2)}\varepsilon_{t-j}^2+\beta^{(2)})\right]. \tag{S1.3}$$

From the assumptions of this theorem, there exist a $\tau > 0$ and a $0 < \rho < 1$ such that $\alpha^{(1)}\tau^2+\beta^{(1)} \le \rho$ and $\alpha^{(2)}\tau^2+\beta^{(2)} \le \rho$. Let $M = \omega^{(2)}[1+(1-\rho)^{-1}\beta^{(2)}]+1$, and denote by $\mu_M$ the restriction of the Lebesgue measure on $(M, M^*)$, where $M^* > M$ is a fixed value, and we will introduce its selection in the proof for Claim (iii). From (S1.3), it can be verified that, if $0 < \varepsilon_j < \tau$ with $0 \le j \le t-2$ and $\varepsilon_{t-1} > 0$, then

$$\sigma_t^2 \le L_{\sigma,t} + \left(\frac{\omega^{(2)}}{1-\rho} + \sigma_0^2\rho^{t-1}\right)\alpha^{(2)}\varepsilon_{t-1}^2$$

where

$$L_{\sigma,t} = \omega^{(2)} + \left(\frac{\omega^{(2)}}{1-\rho} + \sigma_0^2\rho^{t-1}\right)\beta^{(2)}.$$

Thus, conditional on $\sigma_0^2 = x_1$, $0 < \varepsilon_j < \tau$ with $0 \le j \le t-2$ and $\varepsilon_{t-1} > 0$, the random variable $\sigma_t^2$ admits a density, $f_{\sigma,t}(\cdot)$, positive on $[L_{\sigma,t}, +\infty)$. For any $B \subset \mathcal{B}$ and any $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^+ \times \{0,1\}$, there exists a $t^* > 0$ such that $L_{\sigma,t^*} < M$, and then

$$\begin{aligned}
&P\{\sigma_{t^*}^2 \in B|(\sigma_0^2, R_0) = \mathbf{x}\}\\
&\quad\ge P\{\sigma_{t^*}^2 \in B|0 < \varepsilon_j < \tau \text{ with } 0 \le j \le t^*-2, \varepsilon_{t^*-1} > 0, (\sigma_0^2, R_0) = \mathbf{x}\}\\
&\qquad\cdot P\{0 < \varepsilon_j < \tau \text{ with } 0 \le j \le t^*-2, \varepsilon_{t^*-1} > 0\}\\
&\quad= \int_{B\bigcap(M,M^*)} f_{\sigma,t^*}(y)dy\left[\int_0^\tau f(y)dy\right]^{t^*-1}\int_0^{+\infty} f(y)dy > 0
\end{aligned}$$

if $\mu_M(B) > 0$. Define the measure $\mu = \mu_M \times \mu_1$ on the space $(\mathbb{R}^+ \times \{0,1\}, \mathcal{B}\times\mathcal{U})$, where $\mu_1$ is a measure on $(\{0,1\},\mathcal{U})$ with $\mu_1(\{0\}) = \mu_1(\{1\}) > 0$. Hence, the process $\{\boldsymbol{\sigma}_t^2\}$ is $\mu$-irreducible. Similarly, we can show the irreducibility for the case $0 \le r_L \le r_U$ by using the structure at the lower regime.

For the case of $r_L < 0 < r_U$, the process will stay at the upper regime up to time $t$ if $R_0 = 0$ and $\varepsilon_j > 0$ for all $0 \le j \le t-1$, while it will keep staying at the lower regime if $R_0 = 1$ and $\varepsilon_j < 0$ for all $0 \le j \le t-1$. As a result, we can show the irreducibility similarly, and hence finish the proof for Claim (ii).

Finally we prove Claim (iii). Consider the test function $V(\mathbf{x}) = 1 + |x_1|$, where $\mathbf{x} = (x_1, x_2)'$. From (S1.1), we have that

$$\sigma_t^2 \le \max\{\omega^{(1)}, \omega^{(2)}\} + \max\{\alpha^{(1)}, \alpha^{(2)}\}\sigma_{t-1}^2\varepsilon_{t-1}^2 + \max\{\beta^{(1)}, \beta^{(2)}\}\sigma_{t-1}^2,$$

and

$$E\{V(\boldsymbol{\sigma}_t^2)|\boldsymbol{\sigma}_{t-1}^2 = \mathbf{x}\} \le \max\{\omega^{(1)}, \omega^{(2)}\} + c|x_1|,$$

where $c = \max\{\alpha^{(1)}, \alpha^{(2)}\} + \max\{\beta^{(1)}, \beta^{(2)}\} < 1$. Let

$$C = \left\{\mathbf{x} : |x_1| \le \max\left(\frac{\omega^{(1)} - 1}{c}, \frac{\omega^{(2)} - 1}{c}, M + 0.5\right)\right\}$$

and $C^c$ be its complement, where $M$ is defined as in the proof for Claim (ii). It can be easily verified that

(a) $V(\mathbf{x}) \ge 1$ when $\mathbf{x} \in C$, and

(b) $E\{V(\boldsymbol{\sigma}_t^2)|\boldsymbol{\sigma}_{t-1}^2 = \mathbf{x}\} \le cV(\mathbf{x})$ when $\mathbf{x} \in C^c$.

Let

$$M^* = \max\left(\frac{\omega^{(1)} - 1}{c}, \frac{\omega^{(2)} - 1}{c}, M\right) + 1,$$

and it holds that $M < \max\{c^{-1}(\omega^{(1)} - 1), c^{-1}(\omega^{(2)} - 1), M + 0.5\} < M^*$. Thus,

$$\mu(C) = \mu_1(\{0, 1\}) \left[\max\left(\frac{\omega^{(1)} - 1}{c}, \frac{\omega^{(2)} - 1}{c}, M + 0.5\right) - M\right] > 0,$$

where $\mu$ is the irreducibility measure constructed previously. As a result, we finish the proof for Claim (iii), and hence the proof of Theorem 1.

# S2 Proof of Theorem 2

We first denote $R_t = R_t(r_L, r_U, d)$, $R_{0t} = R_t(r_{0L}, r_{0U}, d_0)$ and $\widetilde{R}_t = \widetilde{R}_t(r_L, r_U, d)$ for simplicity. Moreover, let $\|\cdot\|$ be the Euclidean norm, $\mathcal{F}_t$ be the $\sigma$-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, ...\}$, and $C$ be a generic constant which may vary from line to line but independent of time $t$ and the parameter space.

We follow the standard arguments in Huber (1967) to show the strong consistency of $\widetilde{\boldsymbol{\lambda}}_n$, and it is sufficient to verify the following three claims:

(i) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},a\le r_L\le r_U\le b,d\in D}|n^{-1}\sum_{t=1}^{n}[\widetilde{l}_t(\boldsymbol{\lambda})-l_t(\boldsymbol{\lambda})]|\to 0$ with probability one as $n\to\infty$, where the parameter vector $\boldsymbol{\lambda}=(\boldsymbol{\theta}',r_L,r_U,d)'$.

(ii) $E[l_t(\boldsymbol{\lambda})]\ge E[l_t(\boldsymbol{\lambda}_0)]$ for all $\boldsymbol{\lambda}$, and the equality holds if and only if $\boldsymbol{\lambda}=\boldsymbol{\lambda}_0$.

(iii) For any $\boldsymbol{\lambda}$,

$$E\sup_{\boldsymbol{\lambda}^*\in U_\lambda(\eta)}|l_t(\boldsymbol{\lambda}^*)-l_t(\boldsymbol{\lambda})|\to 0\quad \text{as } \eta\to 0,$$

where $U_\lambda(\eta)=\{\boldsymbol{\lambda}^*:\|\boldsymbol{\lambda}^*-\boldsymbol{\lambda}\|<\eta\}$. Thus, $E[l_t(\boldsymbol{\lambda})]$ is a continuous function of $\boldsymbol{\lambda}$.

We first show Claim (i). Let

$$B_t(\boldsymbol{\lambda})=\begin{pmatrix}\beta_1^{(1)} & \cdots & \beta_p^{(1)}\\ & I_{p-1} & \mathbf{0}_{(p-1)\times 1}\end{pmatrix}R_t+\begin{pmatrix}\beta_1^{(2)} & \cdots & \beta_p^{(2)}\\ & I_{p-1} & \mathbf{0}_{(p-1)\times 1}\end{pmatrix}(1-R_t),$$

and denote it by $\widetilde{B}_t(\boldsymbol{\lambda})$ when $R_t$ in $B_t(\boldsymbol{\lambda})$ is replaced by $\widetilde{R}_t$, where $I_k$ is the $k\times k$ identity matrix, and $\mathbf{0}_{k\times 1}$ is a $k$-dimensional zero vector. From Lemma A.1 in Li and Li (2008), it is implied by Assumption 1 that

$$\sup_{\boldsymbol{\lambda}}\|\prod_{j=0}^{i-1}B_{t-j}(\boldsymbol{\lambda})\|_S=O(\rho^i)\quad\text{and}\quad\sup_{\boldsymbol{\lambda}}\|\prod_{j=0}^{i-1}\widetilde{B}_{t-j}(\boldsymbol{\lambda})\|_S=O(\rho^i),\tag{S2.4}$$

where $0<\rho<1$, and $\|\cdot\|_S$ is the spectral norm. Note that

$$\begin{aligned}\widetilde{\sigma}_t^2(\boldsymbol{\lambda})=&\left[\omega^{(1)}+\sum_{i=1}^{q}\alpha_i^{(1)}y_{t-i}^2+\sum_{j=1}^{p}\beta_j^{(1)}\widetilde{\sigma}_{t-j}^2(\boldsymbol{\lambda})\right]\widetilde{R}_t\\&+\left[\omega^{(2)}+\sum_{i=1}^{q}\alpha_i^{(2)}y_{t-i}^2+\sum_{j=1}^{p}\beta_j^{(2)}\widetilde{\sigma}_{t-j}^2(\boldsymbol{\lambda})\right][1-\widetilde{R}_t],\quad 1\le t\le n,\end{aligned}\tag{S2.5}$$

where the initial values $(\widetilde{\sigma}_0^2(\boldsymbol{\lambda}),...,\widetilde{\sigma}_{1-p}^2(\boldsymbol{\lambda}))'=\widetilde{\boldsymbol{\sigma}}_0^2$ are nonnegative random variables or even non-random. We then can show that

$$\sup_{\boldsymbol{\lambda}}|\sigma_t^2(\boldsymbol{\lambda})|\le C\sum_{j=0}^{\infty}\rho^j z_{t-j}\quad\text{and}\quad\sup_{\boldsymbol{\lambda}}|\widetilde{\sigma}_t^2(\boldsymbol{\lambda})|\le C\left(\sum_{j=0}^{\infty}\rho^j z_{t-j}+\rho^t\|\widetilde{\boldsymbol{\sigma}}_0^2\|\right),\tag{S2.6}$$

where $z_t = 1 + \sum_{i=1}^{q} y_{t-i}^2$. Define

$$\widehat{\sigma}_t^2(\boldsymbol{\lambda}) = \left[\omega^{(1)} + \sum_{i=1}^{q} \alpha_i^{(1)} y_{t-i}^2 + \sum_{j=1}^{p} \beta_j^{(1)} \widehat{\sigma}_{t-j}^2(\boldsymbol{\lambda})\right] R_t + \left[\omega^{(2)} + \sum_{i=1}^{q} \alpha_i^{(2)} y_{t-i}^2 + \sum_{j=1}^{p} \beta_j^{(2)} \widehat{\sigma}_{t-j}^2(\boldsymbol{\lambda})\right] [1 - R_t], \quad 1 \leq t \leq n, \tag{S2.7}$$

where the initial values $(\widehat{\sigma}_0^2(\boldsymbol{\lambda}), ..., \widehat{\sigma}_{1-p}^2(\boldsymbol{\lambda}))' = \widetilde{\boldsymbol{\sigma}}_0^2$ are the same as those for $\widetilde{\sigma}_t^2(\boldsymbol{\lambda})$. Accordingly, let $\widehat{l}_t(\boldsymbol{\lambda}) = y_t^2/\widehat{\sigma}_t^2(\boldsymbol{\lambda}) + \log[\widehat{\sigma}_t^2(\boldsymbol{\lambda})]$. Similarly, it can be shown that

$$\sup_{\boldsymbol{\lambda}} |\widehat{\sigma}_t^2(\boldsymbol{\lambda})| \leq C \left( \sum_{j=0}^{\infty} \rho^j z_{t-j} + \rho^t \|\widetilde{\boldsymbol{\sigma}}_0^2\| \right) \tag{S2.8}$$

and

$$\sup_{\boldsymbol{\lambda}} |\widehat{\sigma}_t^2(\boldsymbol{\lambda}) - \sigma_t^2(\boldsymbol{\lambda})| = \sup_{\boldsymbol{\lambda}} \left| \mathbf{l}_p' \prod_{j=0}^{t-1} B_{t-j}(\boldsymbol{\lambda}) [\widetilde{\boldsymbol{\sigma}}_0^2 - \boldsymbol{\sigma}_0^2(\boldsymbol{\lambda})] \right| \leq C(\|\widetilde{\boldsymbol{\sigma}}_0^2\| + \sup_{\boldsymbol{\lambda}} \|\boldsymbol{\sigma}_0^2(\boldsymbol{\lambda})\|)\rho^t, \tag{S2.9}$$

where $\mathbf{l}_p = (1, 0, ..., 0)'$ is a $p$-dimensional vector, and $\boldsymbol{\sigma}_0^2(\boldsymbol{\lambda}) = (\sigma_0^2(\boldsymbol{\lambda}), ..., \sigma_{1-p}^2(\boldsymbol{\lambda}))'$. Note that $E(\sum_{t=1}^{n} \rho^t y_t^2) \leq \rho E(y_t^2)/(1-\rho)$. Hence, by (S2.6), (S2.9), the compactness of $\boldsymbol{\Theta}$ and the fact that $\log(x) \leq x - 1$, we can show that

$$\begin{aligned} &\sup_{\boldsymbol{\lambda}} \left| \frac{1}{n} \sum_{t=1}^{n} \widehat{l}_t(\boldsymbol{\lambda}) - l_t(\boldsymbol{\lambda}) \right| \\ &\leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\lambda}} \left\{ \frac{|\widehat{\sigma}_t^2(\boldsymbol{\lambda}) - \sigma_t^2(\boldsymbol{\lambda})|}{\widehat{\sigma}_t^2(\boldsymbol{\lambda}) \sigma_t^2(\boldsymbol{\lambda})} y_t^2 + \left| \log\left(1 + \frac{\widehat{\sigma}_t^2(\boldsymbol{\lambda}) - \sigma_t^2(\boldsymbol{\lambda})}{\sigma_t^2(\boldsymbol{\lambda})}\right) \right| \right\} \\ &\leq C(\|\widetilde{\boldsymbol{\sigma}}_0^2\| + \sup_{\boldsymbol{\lambda}} \|\boldsymbol{\sigma}_0^2(\boldsymbol{\lambda})\|) \left( \frac{1}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^{n} \rho^t y_t^2 + \frac{1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^{n} \rho^t \right) \\ &\to 0 \end{aligned} \tag{S2.10}$$

with probability one as $n \to \infty$, where $\underline{\omega} = \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \{\omega^{(1)}, \omega^{(2)}\} > 0$.

Note that $0 \leq t_0 \leq n$ and, from the proof of Theorem 2 in Li et al. (2015), it holds that

$$P(\lim_{n \to \infty} t_0 = \infty) = 0. \tag{S2.11}$$

Without loss of generality, we assume that $t_0 > p$. When $t > t_0$, it holds that $\widetilde{R}_t = R_t$ and, by (S2.4), (S2.5) and (S2.7),

$$\sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})| = \sup_{\boldsymbol{\lambda}} \left| \mathbf{l}_p' \prod_{j=0}^{t-t_0-1} B_{t-j}(\boldsymbol{\lambda}) \begin{pmatrix} \widetilde{\sigma}_{t_0}^2(\boldsymbol{\lambda}) - \widehat{\sigma}_{t_0}^2(\boldsymbol{\lambda}) \\ \vdots \\ \widetilde{\sigma}_{t_0-p+1}^2(\boldsymbol{\lambda}) - \widehat{\sigma}_{t_0-p+1}^2(\boldsymbol{\lambda}) \end{pmatrix} \right|$$
$$\leq C\rho^{t-t_0} \sum_{t=1}^{t_0} \sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})|,$$

which implies that

$$\frac{1}{n} \sum_{t=t_0+1}^{n} \sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})| \leq \frac{C}{1-\rho} \cdot \frac{1}{n} \sum_{t=1}^{t_0} \sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})|, \tag{S2.12}$$

and

$$\frac{1}{n} \sum_{t=t_0+1}^{n} y_t^2 \sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})| \leq C\rho^{-t_0} \sum_{t=1}^{\infty} \rho^t y_t^2 \cdot \frac{1}{n} \sum_{t=1}^{t_0} \sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})|. \tag{S2.13}$$

By the ergodic theorem, we have that

$$\frac{1}{t_0} \sum_{t=1}^{t_0} \sum_{j=0}^{\infty} \rho^j z_{t-j} \to \sum_{j=0}^{\infty} \rho^j E(z_{t-j}) \quad \text{and} \quad \frac{1}{t_0} \sum_{t=1}^{t_0} y_t^2 \sum_{j=0}^{\infty} \rho^j z_{t-j} \to \sum_{j=0}^{\infty} \rho^j E(y_t^2 z_{t-j})$$

with probability one as $t_0 \to \infty$. This, together with (S2.6), (S2.8) and (S2.11), implies that

$$\frac{1}{n} \sum_{t=1}^{t_0} \sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})| \leq \frac{2Ct_0}{n} \cdot \frac{1}{t_0} \sum_{t=1}^{t_0} \sum_{j=0}^{\infty} \rho^j z_{t-j} + \frac{2C\rho}{1-\rho} \|\widetilde{\boldsymbol{\sigma}}_0^2\| n^{-1} \to 0 \tag{S2.14}$$

and

$$\frac{1}{n} \sum_{t=1}^{t_0} y_t^2 \sup_{\boldsymbol{\lambda}} |\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \widehat{\sigma}_t^2(\boldsymbol{\lambda})| \leq \frac{2Ct_0}{n} \cdot \frac{1}{t_0} \sum_{t=1}^{t_0} y_t^2 \sum_{j=0}^{\infty} \rho^j z_{t-j} + 2C\|\widetilde{\boldsymbol{\sigma}}_0^2\| \cdot \frac{1}{n} \sum_{t=1}^{\infty} \rho^t y_t^2 \to 0 \tag{S2.15}$$

with probability one as $n \to \infty$. By a method similar to (S2.10), together with (S2.12)-(S2.15), we can show that $\sup_{\boldsymbol{\lambda}} |n^{-1} \sum_{t=1}^{n} [\widetilde{l}_t(\boldsymbol{\lambda}) - \widehat{l}_t(\boldsymbol{\lambda})]| \to 0$, and then $\sup_{\boldsymbol{\lambda}} |n^{-1} \sum_{t=1}^{n} [\widetilde{l}_t(\boldsymbol{\lambda}) - l_t(\boldsymbol{\lambda})]| \to 0$ with probability one as $n \to \infty$. This completes the proof of Claim (i).

We now prove Claim (ii). Note that $x-1-\log x \geq 0$ for $x>0$, and the equality holds only when $x=1$. We then have that

$$E[l_t(\boldsymbol{\lambda})-l_t(\boldsymbol{\lambda}_0)] = E\left(\frac{\sigma_t^2(\boldsymbol{\lambda}_0)}{\sigma_t^2(\boldsymbol{\lambda})} - 1 - \log\frac{\sigma_t^2(\boldsymbol{\lambda}_0)}{\sigma_t^2(\boldsymbol{\lambda})}\right) \geq 0,$$

and the equality holds if and only if $\sigma_t^2(\boldsymbol{\lambda}) = \sigma_t^2(\boldsymbol{\lambda}_0)$ with probability one. It is then sufficient for Claim (ii) to show that $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ under the assumption of $\sigma_t^2(\boldsymbol{\lambda}) = \sigma_t^2(\boldsymbol{\lambda}_0)$ with probability one for all $t$.

Note that, with probability one,

$$\begin{aligned}0 &= \sigma_t^2(\boldsymbol{\lambda}) - \sigma_t^2(\boldsymbol{\lambda}_0)\\ &= [\omega^{(1)}R_t + \omega^{(2)}(1-R_t)] - [\omega_0^{(1)}R_{0t} + \omega_0^{(2)}(1-R_{0t})]\\ &\quad + \sum_{i=1}^{q}\{[\alpha_i^{(1)}R_t + \alpha_i^{(2)}(1-R_t)] - [\alpha_{0i}^{(1)}R_{0t} + \alpha_{0i}^{(2)}(1-R_{0t})]\}y_{t-i}^2\\ &\quad + \sum_{j=1}^{p}\{[\beta_j^{(1)}R_t + \beta_j^{(2)}(1-R_t)] - [\beta_{0j}^{(1)}R_{0t} + \beta_{0j}^{(2)}(1-R_{0t})]\}\sigma_{t-j}^2(\boldsymbol{\lambda}_0).\end{aligned} \tag{S2.16}$$

By conditioning the above equation on the $\sigma$-field $\mathcal{F}_{t-2}$ and from Assumption 2, we have that

$$[\alpha_1^{(1)}R_t + \alpha_1^{(2)}(1-R_t)] - [\alpha_{01}^{(1)}R_{0t} + \alpha_{01}^{(2)}(1-R_{0t})] = 0.$$

Note that $E[R_tR_{0t}] \geq P(y_{t-d} < r_L, y_{t-d_0} < r_{0L}) > 0$ and $E[(1-R_t)(1-R_{0t})] \geq P(y_{t-d} > r_U, y_{t-d_0} > r_{0U}) > 0$. It then can be shown that $\alpha_1^{(1)} = \alpha_{01}^{(1)}$, $\alpha_1^{(2)} = \alpha_{01}^{(2)}$ and $R_t = R_{0t}$ if $\alpha_{01}^{(1)} + \alpha_{01}^{(2)} > 0$, or $\alpha_1^{(1)} = \alpha_1^{(2)} = 0$ if $\alpha_{01}^{(1)} = \alpha_{01}^{(2)} = 0$. From the definition of $\sigma_t^2(\boldsymbol{\lambda}_0)$ at (**??**) and by conditioning equation (S2.16) on the $\sigma$-field $\mathcal{F}_{t-3}$, we can further obtain that

$$\begin{aligned}0 =& \{[\beta_1^{(1)}R_t + \beta_1^{(2)}(1-R_t)] - [\beta_{01}^{(1)}R_{0t} + \beta_{01}^{(2)}(1-R_{0t})]\}\cdot[\alpha_{01}^{(1)}R_{0t} + \alpha_{01}^{(2)}(1-R_{0t})]\\ &+ [\alpha_2^{(1)}R_t + \alpha_2^{(2)}(1-R_t)] - [\alpha_{02}^{(1)}R_{0t} + \alpha_{02}^{(2)}(1-R_{0t})],\end{aligned}$$

which implies that $\alpha_2^{(1)} = \alpha_{02}^{(1)}$, $\alpha_2^{(2)} = \alpha_{02}^{(2)}$, $\beta_1^{(1)} = \beta_{01}^{(1)}$ and $\beta_1^{(2)} = \beta_{01}^{(2)}$ if $\alpha_{01}^{(1)} + \alpha_{01}^{(2)} > 0$, or $\alpha_2^{(1)} = \alpha_{02}^{(1)}$, $\alpha_2^{(2)} = \alpha_{02}^{(2)}$ and $R_t = R_{0t}$ if $\alpha_{01}^{(1)} = \alpha_{01}^{(2)} = 0$ and $\alpha_{02}^{(1)} + \alpha_{02}^{(2)} > 0$, or $\alpha_2^{(1)} = \alpha_2^{(2)} = 0$ if $\alpha_{01}^{(1)} = \alpha_{01}^{(2)} = \alpha_{02}^{(1)} = \alpha_{02}^{(2)} = 0$. Similarly, we can show that $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}_0^{(1)}$, $\boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)}$ and $R_t = R_{0t}$.

The fact of $R_t = R_{0t}$ leads to

$$
\begin{aligned}
0 &= P(R_t = 0, R_{0t} = 1) \\
&\geq P(y_{t-d_0} \leq r_{0L}, y_{t-d} > r_U) + P(y_{t-d_0} \leq r_{0L}, r_U \geq y_{t-d} > r_L, y_{t-d-1} > r_U) \\
&\quad + P(r_{0L} < y_{t-d_0} \leq r_{0U}, y_{t-d_0-1} \leq r_{0L}, y_{t-d} > r_U),
\end{aligned}
$$

which implies that $d = d_0$, $r_L \geq r_{0L}$ and $r_U \geq r_{0U}$. Similarly, we have that $r_L \leq r_{0L}$ and $r_U \leq r_{0U}$ from $P(R_t = 1, R_{0t} = 0) = 0$. Thus, $d = d_0$, $r_L = r_{0L}$ and $r_U = r_{0U}$, and we then complete the proof of Claim (ii).

We now consider proving Claim (iii). Let $\boldsymbol{\lambda}^* = (\boldsymbol{\theta}^{*\prime}, r_L^*, r_U^*, d)' \in U_\lambda(\eta)$, and denote $\boldsymbol{\lambda}_1^* = (\boldsymbol{\theta}', r_L^*, r_U^*, d)'$ and $R_t^* = R_t(r_L^*, r_U^*, d)$, where $\boldsymbol{\lambda} = (\boldsymbol{\theta}', r_L, r_U, d)'$. Note that

$$
\sigma_t^2(\boldsymbol{\lambda}_1^*) - \sigma_t^2(\boldsymbol{\lambda}) = \xi_t(\boldsymbol{\lambda})(R_t^* - R_t) + \sum_{j=1}^{p}[\beta_j^{(1)} R_t^* + \beta_j^{(2)}(1 - R_t^*)][\sigma_{t-j}^2(\boldsymbol{\lambda}_1^*) - \sigma_{t-j}^2(\boldsymbol{\lambda})],
$$

and then

$$
\sup_{\boldsymbol{\lambda}^* \in U_\lambda(\eta)} |\sigma_t^2(\boldsymbol{\lambda}_1^*) - \sigma_t^2(\boldsymbol{\lambda})| \leq C \sum_{j=0}^{\infty} \rho^j |\xi_{t-j}(\boldsymbol{\lambda})| \sup_{\boldsymbol{\lambda}^* \in U_\lambda(\eta)} |R_t^* - R_t|,
$$

where $\xi_t(\boldsymbol{\lambda}) = (\omega^{(1)} - \omega^{(2)}) + \sum_{i=1}^{q}(\alpha_i^{(1)} - \alpha_i^{(2)}) y_{t-i}^2 + \sum_{j=1}^{p}(\beta_j^{(1)} - \beta_j^{(2)})\sigma_{t-j}^2(\boldsymbol{\lambda})$. Moreover, from the proof of Theorem 2 in Li et al. (2015),

$$
E \sup_{\boldsymbol{\lambda}^* \in U_\lambda(\eta)} |R_t(r_L^*, r_U^*, d) - R_t(r_L, r_U, d)| \to 0
$$

as $\eta \to 0$. By a method similar to (S2.10), together with Hölder inequality and $E|y_t|^{4+\delta} < \infty$, we have that

$$
\begin{aligned}
&E \sup_{\boldsymbol{\lambda}^* \in U_\lambda(\eta)} |l_t(\boldsymbol{\lambda}_1^*) - l_t(\boldsymbol{\lambda})| \\
&\leq E\left[\left(\frac{1}{\underline{\omega}} + \frac{y_t^2}{\underline{\omega}^2}\right) \sup_{\boldsymbol{\lambda}^* \in U_\lambda(\eta)} |\sigma_t^2(\boldsymbol{\lambda}_1^*) - \sigma_t^2(\boldsymbol{\lambda})|\right] \\
&\leq C\left\{E\left[\left(\frac{1}{\underline{\omega}} + \frac{y_t^2}{\underline{\omega}^2}\right)\sum_{j=0}^{\infty} \rho^j |\xi_{t-j}(\boldsymbol{\lambda})|\right]^{1+\delta/4}\right\}^{4/(4+\delta)} \left\{E \sup_{\boldsymbol{\lambda}^* \in U_\lambda(\eta)} |R_t^* - R_t|\right\}^{\delta/(4+\delta)} \\
&\to 0
\end{aligned}
\tag{S2.17}
$$

as $\eta \to 0$. Consider

$$\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} = \mathbf{x}_t(\boldsymbol{\lambda}) + \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} B_{t-j}(\boldsymbol{\lambda}) \mathbf{x}_{t-i}(\boldsymbol{\lambda}),$$

where $\mathbf{x}_{1t}(\boldsymbol{\lambda}) = (1, y_{t-1}^2, ..., y_{t-q}^2, \sigma_{t-1}^2(\boldsymbol{\lambda}), ..., \sigma_{t-p}^2(\boldsymbol{\lambda}))'$ and $\mathbf{x}_t(\boldsymbol{\lambda}) = (\mathbf{x}_{1t}'(\boldsymbol{\lambda})R_t, \mathbf{x}_{1t}'(\boldsymbol{\lambda})(1-R_t))'$. By (S2.6) and the compactness of $\boldsymbol{\Theta}$, we can show that

$$E \sup_{\boldsymbol{\lambda}^* \in U_\lambda(\eta)} |l_t(\boldsymbol{\lambda}^*) - l_t(\boldsymbol{\lambda}_1^*)| \leq \eta \cdot E \sup_{\boldsymbol{\lambda}} \left| \left( \frac{1}{\sigma_t^2(\boldsymbol{\lambda})} - \frac{y_t^2}{\sigma_t^4(\boldsymbol{\lambda})} \right) \frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} \right| = O(\eta),$$

which, together with (S2.17), implies Claim (iii).

Following the standard argument for the strong consistency in Huber (1967), together with Claims (i), (ii) and (iii), we can show that $\widetilde{\boldsymbol{\lambda}}_n \to \boldsymbol{\lambda}_0$ with probability one; see also Francq and Zakoïan (2004) and Straumann and Mikosch (2006). Hence, we finish the proof.

## References


Feigin, P. D. and Tweedie, R. L. (1985). Random coefficient autoregressive processes: a markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis* **6** 1–14.

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

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, Berkeley.

Li, G., Guan, B., Li, W. K. and Yu, P. L. H. (2015). Hysteretic autoregressive time series models. *Biometrika* **102** 717–723.

Li, G. and Li, W. K. (2008). Testing for threshold moving average with conditional heteroscedasticity. *Statistica Sinica* **18** 647–665.

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

Tweedie, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In *Probability, Statistics and Analysis* (J. F. C. Kingman and G. E. H. Reuter, eds.). Cambridge University Press, Cambridge.