Supplementary Appendix for "Estimating Rotated Multivariate GARCH Models"

March 2021

Appendix B

Appendix B gives the proofs of Propositions 1 and 2, which show the asymptotic property for the 2sQML estimator suggested for estimating the parameters for the rotated multivariate BEKK model of Noureldin et al. (2014).

We use the following notation in Appendix B. For a matrix A, we define $A^{\otimes 2} = (A \otimes A)$. With $\xi_1, \ldots, \xi_n$, the n eigenvalues of a matrix A, $\rho(A) = \max_{i \in \{1,\ldots,n\}} |\xi_i|$, is the spectral radius of A. The Frobenius norm of the matrix, or vector A, is defined as $||A|| = \sqrt{\mathrm{tr}(A'A)}$. For a positive matrix A, we define the square root, $A^{1/2}$, by the spectral decomposition of A. By K and ϕ, we denote strictly positive generic constants with $\phi < 1$.

B.1 Proof of Proposition 1

To prove the consistency of the 2sQML estimator, we need to accommodate the estimate of Ω in $A^* = \Omega^{1/2} A \Omega^{-1/2}$ and $B^* = \Omega^{1/2} B \Omega^{-1/2}$ by modifying the proof of Theorem 4.1 of Pedersen and Rahbek (2014).

By the ergodic theorem under Assumption 3(a) and $E[||X_t||^2] < \infty$, as $T \to \infty$, we obtain:

$$\hat{\boldsymbol{\omega}} \xrightarrow{a.s.} \boldsymbol{\omega}_0. \tag{B.1}$$

For the consistency of $\hat{\boldsymbol{\lambda}}$, we apply the technique used in the proof of Theorem 4.1 of Pedersen and Rahbek (2014). For this purpose we first give the following lemma.

Lemma B.1. *Under Assumptions 1(a), 2, and 3, as $T \to \infty$,*

$$\sup_{\boldsymbol{\lambda} \in \Theta_\lambda} |L_T(\boldsymbol{\omega}_0, \boldsymbol{\lambda}) - L_{T,h}(\hat{\boldsymbol{\omega}}, \boldsymbol{\lambda})| \xrightarrow{a.s.} 0. \tag{B.2}$$

Proof. We can apply the technique used in the proof of Lemma B.1 of Pedersen and Rahbek (2014), by considering bounds regarding $\underline{H}_t$. By recursion, we obtain:

$$\begin{aligned} &\mathrm{vec}\left(\underline{H}_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda})\right) - \mathrm{vec}\left(\underline{H}_{t,h}(\hat{\boldsymbol{\omega}}, \boldsymbol{\lambda})\right) \\ &\quad = \sum_{i=0}^{t-1} (B^{\otimes 2})^i A^{\otimes 2} \left\{ (\Omega^{-1})^{\otimes 2} - (\hat{\Omega}^{-1})^{\otimes 2} \right\} \mathrm{vec}\left(X_{t-i-1} X'_{t-i-1}\right) + (B^{\otimes 2})^t \mathrm{vec}\left(\underline{H}_0 - h\right). \end{aligned} \tag{B.3}$$

By Proposition 4.5 of Boussama et al. (2011), the assumption, $\rho\left(A^{\otimes 2}+B^{\otimes 2}\right)<1$ on Θ, indicates $\rho\left(B^{\otimes 2}\right)<1$ on Θ. Hence, for any i and for some $0<\phi<1$:

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|(B^{\otimes 2})^i\right\|\leq K\phi^i. \tag{B.4}$$

For equation (B.3), by the compactness of Θ, (B.1), and (B.4), we obtain:

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|\mathrm{vec}\left(\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right)-\mathrm{vec}\left(\underline{H}_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right)\right\|\leq K\phi^t+o(1)\ \text{ a.s.,} \tag{B.5}$$

as $T\to\infty$, as in (B.16) of Pedersen and Rahbek (2014). We can also show:

$$\begin{aligned}\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|\underline{H}_{t,h}^{-1}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right\|&\leq\sup_{\boldsymbol{\theta}\in\Theta}\left\|\underline{H}_{t,h}^{-1}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right\|\leq K,\\ \sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|\underline{H}_{t,h}^{-1}(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right\|&\leq\sup_{\boldsymbol{\theta}\in\Theta}\left\|\underline{H}_{t,h}^{-1}(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right\|\leq K,\end{aligned} \tag{B.6}$$

by the approach used in (B.13) of Pedersen and Rahbek (2014).

Now, we turn to the difference of the likelihood function as in (B.2). By the technique of the proof of Lemma B.1 of Pedersen and Rahbek (2014), we obtain:

$$\begin{aligned}&\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda})-L_{T,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right|\\ &\qquad\leq\left|\log\left(\frac{\det(\Omega_0)}{\det(\hat{\Omega})}\right)\right|+\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|\log\left(\frac{\det(\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}))}{\det(\underline{H}_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda}))}\right)\right|\\ &\qquad\quad+\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|\mathrm{tr}\left(X_tX_t'\left(H_t^{-1}(\boldsymbol{\omega}_0,\boldsymbol{\lambda})-H_{t,h}^{-1}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right)\right)\right|\\ &\qquad\leq\left|\log\left(\frac{\det(\Omega_0)}{\det(\hat{\Omega})}\right)\right|+dK\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})-\underline{H}_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right\|\\ &\qquad\quad+K\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|H_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})-H_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right\|\,||X_t||^2.\end{aligned}$$

Noting that:

$$\begin{aligned}&\mathrm{vec}\left(H_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right)-\mathrm{vec}\left(H_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right)\\ &\qquad=\left(\Omega_0^{\otimes 2}-\hat{\Omega}_0^{\otimes 2}\right)\mathrm{vec}\left(\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right)+\hat{\Omega}^{\otimes 2}\left(\mathrm{vec}\left(\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right)-\mathrm{vec}\left(\underline{H}_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right)\right),\end{aligned}$$

and (B.5), we obtain:

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda})-L_{T,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right|\leq K\frac{1}{T}\sum_{t=1}^{T}\phi^t+K\frac{1}{T}\sum_{t=1}^{T}\phi^t||X_t||^2+o(1)\ \text{ a.s.}$$

As in the proof of Lemma B.1 of Pedersen and Rahbek (2014), it is shown that (B.2) holds. □

By the structure of the RBEKK model as a special case of the BEKK model, Lemmas B.2-B.4 of Pedersen and Rahbek (2014) also hold under Assumptions 1(a), 2, and 3. Using Lemma B.2 with the above Lemma B.1 and the definition of $\hat{\boldsymbol{\lambda}}$, we obtain:

$$\begin{aligned}
&E[l_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)] < L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)+\frac{\varepsilon}{5}, \quad L_T(\boldsymbol{\omega}_0,\hat{\boldsymbol{\lambda}}) < E[l_t(\boldsymbol{\omega}_0,\hat{\boldsymbol{\lambda}})]+\frac{\varepsilon}{5},\\
&L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0) < L_{T,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda}_0)+\frac{\varepsilon}{5}, \quad L_{T,h}(\hat{\boldsymbol{\omega}},\hat{\boldsymbol{\lambda}}) < L_T(\boldsymbol{\omega}_0,\hat{\boldsymbol{\lambda}})+\frac{\varepsilon}{5},\\
&L_{T,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda}_0) < L_{T,h}(\hat{\boldsymbol{\omega}},\hat{\boldsymbol{\lambda}})+\frac{\varepsilon}{5},
\end{aligned}$$

for any $\varepsilon > 0$ almost surely for large enough T. Hence, for any $\varepsilon > 0$,

$$E[l_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)] < E[l_t(\boldsymbol{\omega}_0,\hat{\boldsymbol{\lambda}})]+\varepsilon.$$

By applying the arguments of the proof of Theorem 2.1 in Newey and McFadden (1994), it follows that as $T\to\infty$, $\hat{\boldsymbol{\lambda}} \xrightarrow{a.s.} \boldsymbol{\lambda}_0$. Combined with (B.1), we obtain as $T\to\infty$, $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$.

B.2 Proof of Proposition 2

For notational convenience, let $H_{0t} = H_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)$. We use the following lemma to show the asymptotic normality of the 2sQML estimator.

Lemma B.2. *Under Assumptions 1(a), 2-4, as $T\to\infty$,*

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0 \\ \partial L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)/\partial\boldsymbol{\lambda} \end{pmatrix} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Upsilon_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\mathrm{vec}\left(Z_tZ_t'-I_d\right)+o_p(1), \tag{B.7}$$

where

$$\begin{aligned}
&\Upsilon_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\\
&=\begin{pmatrix}\Upsilon_{\omega t}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\\ \Upsilon_{\alpha t}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\\ \Upsilon_{\beta t}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\end{pmatrix} = \begin{pmatrix}\left(\Omega_0^{1/2}\right)^{\otimes 2}\left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)^{-1}\left(I_{d^2}-B_0^{\otimes 2}\right)\left(\Omega_0^{-1/2}H_{0t}^{1/2}\right)^{\otimes 2}\\ \frac{1}{2}\left[\sum_{i=0}^{\infty}(B_0^{\otimes 2})^i N_{t-1-i}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\right]'\left(\Omega_0^{1/2}H_{0t}^{-1/2}\right)^{\otimes 2}\\ \frac{1}{2}\left[\sum_{i=0}^{\infty}(B_0^{\otimes 2})^i \tilde{N}_{t-1-i}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\right]'\left(\Omega_0^{1/2}H_{0t}^{-1/2}\right)^{\otimes 2}\end{pmatrix}
\end{aligned} \tag{B.8}$$

with

$$\begin{aligned}
N_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0) &= \left[A_0(\Omega_0^{-1/2}X_tX_t'\Omega_0^{-1/2}-I_d)\otimes I_d\right]+\left[I_d\otimes A_0(\Omega_0^{-1/2}X_tX_t'\Omega_0^{-1/2}-I_d)\right]C_{dd},\\
\tilde{N}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0) &= \left[B_0(\underline{H}_{0t}-I_d)\otimes I_d\right]+\left[I_d\otimes B_0(\underline{H}_{0t}-I_d)\right]C_{dd}.
\end{aligned} \tag{B.9}$$

Proof. By (A.4), we obtain:

$$\frac{\partial\mathrm{vec}(\underline{H}_{0t})}{\partial\boldsymbol{\alpha}'}=\sum_{i=0}^{\infty}(B_0^{\otimes 2})^i N_{t-1-i}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0), \quad \frac{\partial\mathrm{vec}(\underline{H}_{0t})}{\partial\boldsymbol{\beta}'}=\sum_{i=0}^{\infty}(B_0^{\otimes 2})^i \tilde{N}_{t-1-i}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0).$$

Hence, by (A.1)-(A.3), we obtain the result for $\sqrt{T}\,\partial L_T(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)/\partial\boldsymbol{\lambda}$ stated in (B.7).

Now, we consider $\hat{\boldsymbol{\omega}}$ in the vector form as:

$$\hat{\boldsymbol{\omega}} = \frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\mathrm{vec}\left(Z_t Z_t' - I_d\right) + \mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T} H_{0t}\right), \tag{B.10}$$

with

$$\mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T} H_{0t}\right) = \left(\Omega_0^{1/2}\right)^{\otimes 2}\mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T}\underline{H}_{0t}\right).$$

Furthermore,

$$\begin{aligned}\mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T}\underline{H}_{0t}\right) &= \mathrm{vec}(I - A_0A_0' - B_0B_0')\\ &\quad + \left(A_0\Omega_0^{-1/2}\right)^{\otimes 2}\mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T}X_tX_t' + \frac{1}{T}(X_0X_0' - X_TX_t')\right)\\ &\quad + B_0^{\otimes 2}\mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T}\underline{H}_{0t} + \frac{1}{T}(\underline{H}_{00} - \underline{H}_{0T})\right),\end{aligned}$$

yielding:

$$\begin{aligned}\mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T}\underline{H}_{0t}\right) &= \left(I_{d^2} - B_0^{\otimes 2}\right)^{-1}\mathrm{vec}(I - A_0A_0' - B_0B_0')\\ &\quad + \left(I_{d^2} - B_0^{\otimes 2}\right)^{-1}\left(A_0\Omega_0^{-1/2}\right)^{\otimes 2}\left(\hat{\boldsymbol{\omega}} + \frac{1}{T}\mathrm{vec}(X_0X_0' - X_TX_t')\right)\\ &\quad + \left(I_{d^2} - B_0^{\otimes 2}\right)^{-1}B_0^{\otimes 2}\frac{1}{T}\mathrm{vec}(\underline{H}_{00} - \underline{H}_{0T}).\end{aligned} \tag{B.11}$$

As $\rho(B_0^{\otimes 2}) < 1$, it follows that $\left(I_{d^2} - B_0^{\otimes 2}\right)$ is invertible.

After inserting (B.10) in (B.11), we can transform the equation to obtain:

$$\begin{aligned}&[I - A_0^{\otimes 2} - B_0^{\otimes 2}]\left(\Omega_0^{-1/2}\right)^{\otimes 2}\hat{\boldsymbol{\omega}}\\ &\quad = \mathrm{vec}(I - A_0A_0' - B_0B_0')\\ &\qquad + \left(I_{d^2} - B_0^{\otimes 2}\right)\left(\Omega_0^{-1/2}\right)^{\otimes 2}\frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\mathrm{vec}\left(Z_tZ_t' - I_d\right)\\ &\qquad + \left[\left(A_0\Omega_0^{-1/2}\right)^{\otimes 2}\frac{1}{T}\mathrm{vec}(X_0X_0' - X_TX_t') + B_0^{\otimes 2}\frac{1}{T}\mathrm{vec}(\underline{H}_{00} - \underline{H}_{0T})\right],\end{aligned}$$

which gives

$$\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0 = \left(\Omega_0^{1/2}\right)^{\otimes 2} [I - A_0^{\otimes 2} - B_0^{\otimes 2}]^{-1} \left(I_{d^2} - B_0^{\otimes 2}\right) \left(\Omega_0^{-1/2}\right)^{\otimes 2}$$
$$\times \frac{1}{T} \sum_{t=1}^{T} \left(H_{0t}^{1/2}\right)^{\otimes 2} \mathrm{vec}\left(Z_t Z_t' - I_d\right)$$
$$+ \left(\Omega_0^{1/2}\right)^{\otimes 2} [I - A_0^{\otimes 2} - B_0^{\otimes 2}]^{-1}$$
$$\times \left[\left(A_0 \Omega_0^{-1/2}\right)^{\otimes 2} \frac{1}{T} \mathrm{vec}(X_0 X_0' - X_T X_t') + B_0^{\otimes 2} \frac{1}{T} \mathrm{vec}(\underline{H}_{00} - \underline{H}_{0T})\right].$$

For any $\varepsilon > 0$, by the Markov's inequality:

$$P\left(\left\|\left(A_0 \Omega_0^{-1/2}\right)^{\otimes 2} \frac{1}{\sqrt{T}} \mathrm{vec}(X_0 X_0' - X_T X_t') + B_0^{\otimes 2} \frac{1}{\sqrt{T}} \mathrm{vec}(\underline{H}_{00} - \underline{H}_{0T})\right\| > \varepsilon\right) \leq \frac{KE||X_t||^2}{\sqrt{T}\varepsilon} \to 0,$$

as $T \to \infty$, which yields:

$$\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0 = \left(\Omega_0^{1/2}\right)^{\otimes 2} [I - A_0^{\otimes 2} - B_0^{\otimes 2}]^{-1} \left(I_{d^2} - B_0^{\otimes 2}\right) \left(\Omega_0^{-1/2}\right)^{\otimes 2}$$
$$\times \frac{1}{T} \sum_{t=1}^{T} \left(H_{0t}^{1/2}\right)^{\otimes 2} \mathrm{vec}\left(Z_t Z_t' - I_d\right) + o_p(T^{-1/2}).$$

Therefore, (B.7) holds. □

We use the approach in the proof of Proposition 4.2 of Pedersen and Rahbek (2014). By Assumption 4(b) and the definition of $\hat{\boldsymbol{\lambda}}$ in (11), we apply the mean value theorem in order to obtain:

$$0 = \frac{\partial L_{T,h}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)}{\partial \boldsymbol{\lambda}} + K_{T,h}(\boldsymbol{\theta}^\dagger)(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0) + J_{T,h}(\boldsymbol{\theta}^\dagger)(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0), \tag{B.12}$$

where

$$\frac{\partial L_{T,h}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)}{\partial \boldsymbol{\lambda}} = \left.\frac{\partial L_{T,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$
$$K_{T,h}(\boldsymbol{\theta}^\dagger) = \left.\frac{\partial^2 L_{T,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\dagger}, \quad J_{T,h}(\boldsymbol{\theta}^\dagger) = \left.\frac{\partial^2 L_{T,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\dagger},$$

with $\boldsymbol{\theta}^\dagger$ between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Instead of $L_{T,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})$, we also use $L_T(\boldsymbol{\omega}, \boldsymbol{\lambda})$ to denote $\partial L_T(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda}$, $K_T(\boldsymbol{\theta}^\dagger)$, and $J_T(\boldsymbol{\theta}^\dagger)$. Moreover, define:

$$K_0 = E\left(\frac{\partial^2 l_t(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}'}\right), \quad J_0 = E\left(\frac{\partial^2 l_t(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'}\right). \tag{B.13}$$

By the techniques used in the proofs of Lemmas B.5-B.7 of Pedersen and Rahbek (2014), under Assumptions 1(a), 2-4, we show that:

$$E\left[\sup_{\boldsymbol{\theta} \in \Theta} \left|\frac{\partial^2 l_t(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \theta_i \partial \theta_j}\right|\right] < \infty, \tag{B.14}$$

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|\frac{\partial^2 L_T(\boldsymbol{\omega},\boldsymbol{\lambda})}{\partial\theta_i\partial\theta_j} - E\left[\frac{\partial^2 l_t(\boldsymbol{\omega},\boldsymbol{\lambda})}{\partial\theta_i\partial\theta_j}\right]\right| \xrightarrow{a.s.} 0, \tag{B.15}$$

for all $i,j=1,\ldots,3d^2$, and that J_0 is non-singular. With the consistency of $\hat{\boldsymbol{\theta}}$, the above results imply that $J_T(\boldsymbol{\theta}^\dagger)$ is invertible with probability approaching one.

As a straightforward extension of Lemma B.11 of Pedersen and Rahbek (2014), we can show that:

$$\left|\sqrt{T}\left(\frac{\partial L_{T,h}(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)}{\partial\lambda_i} - \frac{\partial L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)}{\partial\lambda_i}\right)\right| \xrightarrow{p} 0,$$

for $i=1,\ldots,2d^2$, and

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|\frac{\partial^2 L_T(\boldsymbol{\omega},\boldsymbol{\lambda})}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 l_{t,h}(\boldsymbol{\omega},\boldsymbol{\lambda})}{\partial\theta_i\partial\theta_j}\right| \xrightarrow{a.s.} 0,$$

for $i,j=1,\ldots,3d^2$. Applying the above result to (B.12) that $J_T(\boldsymbol{\theta}^\dagger)$ is invertible with probability approaching to one, we obtain:

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) = \begin{pmatrix} I_{d^2} & O_{d^2\times 2d^2} \\ -J_T^{-1}(\boldsymbol{\theta}^\dagger)K_T(\boldsymbol{\theta}^\dagger) & -J_T^{-1}(\boldsymbol{\theta}^\dagger)\end{pmatrix}\sqrt{T}\begin{pmatrix}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0) \\ \partial L(\boldsymbol{\omega},\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\end{pmatrix} + o_p(1).$$

By (B.15) and Proposition 1:

$$\begin{pmatrix} I_{d^2} & O_{d^2\times 2d^2} \\ -J_T^{-1}(\boldsymbol{\theta}^\dagger)K_T(\boldsymbol{\theta}^\dagger) & -J_T^{-1}(\boldsymbol{\theta}^\dagger)\end{pmatrix} \xrightarrow{p} \begin{pmatrix} I_{d^2} & O_{d^2\times 2d^2} \\ -J_0^{-1}K_0 & -J_0^{-1}\end{pmatrix}.$$

By the same argument used in the proof of Lemma B.10 of Pedersen and Rahbek (2014), as $T\to\infty$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Upsilon_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\text{vec}\left(Z_tZ_t'-I_d\right) \xrightarrow{d} N(0,\Gamma_0), \tag{B.16}$$

where

$$\Gamma_0 = E\left[\Upsilon_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\text{vec}\left(Z_tZ_t'-I_d\right)\left(\text{vec}\left(Z_tZ_t'-I_d\right)\right)'\Upsilon_t'(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)\right], \tag{B.17}$$

with $\Upsilon_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}_0)$ defined by (B.8). By Lemma B.2, (B.16), and the Slutzky theorem, we can obtain the asymptotic normality of the 2sQML estimator.

Appendix C

Appendix C provides the results of Monte Carlo experiments for DGP1 and DGP2 with the multivariate standardized t distribution with the degree of freedom, nu, denoted by $St(\nu)$. The sample size is $T=500$ and the number of replications is 2000.

Table C.1: Finite Sample Properties of the 2sQML Estimator for the Diagonal RBEKK Model with Heavy-Tailed Distributions

	DGP1 with $St(7)$				DGP2 with $St(7)$			
Parameters	True	Mean	Std. Dev.	RMSE	True	Mean	Std. Dev.	RMSE
Ω_{11}	1.00	0.9811	0.4550	0.4553	0.640	0.6336	0.3517	0.3518
Ω_{21}	0.54	0.5238	0.3317	0.3320	-0.264	-0.2633	0.0915	0.0915
Ω_{22}	0.81	0.7781	0.6309	0.6316	1.210	1.2046	0.2264	0.2264
A_{11}	0.60	0.5748	0.0693	0.0737	0.600	0.5744	0.0746	0.0789
A_{22}	0.40	0.3861	0.0507	0.0525	-0.300	-0.3024	0.0655	0.0655
B_{11}	0.70	0.6866	0.1104	0.1112	0.700	0.6858	0.1091	0.1100
B_{22}	0.90	0.8860	0.0968	0.0978	-0.900	-0.8552	0.1419	0.1488

Table C.2: Finite Sample Properties of the 2sQML Estimator for the Diagonal RBEKK Model without Sixth Moments

	DGP1 with $St(5)$				DGP2 with $St(5)$			
Parameters	True	Mean	Std. Dev.	RMSE	True	Mean	Std. Dev.	RMSE
Ω_{11}	1.00	0.9712	0.6597	0.6602	0.640	0.6272	0.3968	0.3969
Ω_{21}	0.54	0.5180	0.5967	0.5969	-0.264	-0.2635	0.1184	0.1184
Ω_{22}	0.81	0.7975	1.7237	1.7233	1.210	1.2110	0.4820	0.4821
A_{11}	0.60	0.5662	0.0786	0.0855	0.600	0.5673	0.0848	0.0908
A_{22}	0.40	0.3824	0.0616	0.0614	-0.300	-0.3003	0.0839	0.0839
B_{11}	0.70	0.6833	0.1235	0.1246	0.700	0.6705	0.1628	0.1654
B_{22}	0.90	0.8742	0.1525	0.1546	-0.900	-0.8223	0.2583	0.2697

Table C.3: Finite Sample Properties of the 2sQML Estimator for the Diagonal RBEKK Model without Fourth Moments

	DGP1 with $St(3)$				DGP2 with $St(3)$			
Parameters	True	Mean	Std. Dev.	RMSE	True	Mean	Std. Dev.	RMSE
Ω_{11}	1.00	1.0159	1.2195	1.2197	0.640	0.7357	7.1907	7.1919
Ω_{21}	0.54	0.5497	0.5517	0.5518	-0.264	-0.2713	1.0842	1.0844
Ω_{22}	0.81	0.6084	0.9935	1.0135	1.210	1.1652	0.8867	0.8876
A_{11}	0.60	0.5157	0.1183	0.1453	0.60	0.5179	0.1221	0.1471
A_{22}	0.40	0.3522	0.1304	0.1388	-0.30	-0.2663	0.1617	0.1651
B_{11}	0.70	0.5193	0.4442	0.4794	0.70	0.5166	0.4366	0.4735
B_{22}	0.90	0.6222	0.5935	0.6551	-0.90	-0.5525	0.6241	0.7142

References

Boussama, F., F. Fuchs, and R. Stelzer (2011), "Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models", *Stochastic Processes and their Applications*, **121**, 2331–2360.

Newey,W. K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing", In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, 2111-2245. Amsterdam: Elsevier.

Noureldin, D., N. Shephard, and K. Sheppard (2014), "Multivariate Rotated ARCH Models", *Journal of Econometrics*, **179**, 16–30.

Pedersen, R. S. and A. Rahbek (2014), "Multivariate Variance Targeting in the BEKK-GARCH Model", *Econometrics Journal*, **17**, 24–55.