Asymptotic Theory for Rotated Multivariate GARCH Models*

Manabu Asai
Faculty of Economics, Soka University, Japan

Chia-Lin Chang
Department of Applied Economics & Department of Finance
National Chung Hsing University, Taiwan

Michael McAleer
Department of Finance, Asia University, Taiwan
Discipline of Business Analytics, University of Sydney Business School, Australia
Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
The Netherlands
Department of Economic Analysis and ICAE, Complutense University of Madrid, Spain
Institute of Advanced Sciences, Yokohama National University, Japan

Laurent Pauwels
Discipline of Business Analytics, University of Sydney Business School, Australia

EI2018-38

October 2018

*The authors are most grateful to Yoshi Baba for very helpful comments and suggestions. The first author acknowledges the financial support of the Japan Ministry of Education, Culture, Sports, Science and Technology, Japan Society for the Promotion of Science, and the Australian Academy of Science. The second author thanks the Ministry of Science and Technology (MOST) for financial support. The third author is most grateful for the financial support of the Australian Research Council, Ministry of Science and Technology (MOST), Taiwan, and the Japan Society for the Promotion of Science.

Abstract

In this paper, we derive the statistical properties of a two step approach to estimating multivariate GARCH rotated BEKK (RBEKK) models. By the definition of rotated BEKK, we estimate the unconditional covariance matrix in the first step in order to rotate observed variables to have the identity matrix for its sample covariance matrix. In the second step, we estimate the remaining parameters via maximizing the quasi-likelihood function. For this two step quasi-maximum likelihood (2sQML) estimator, we show consistency and asymptotic normality under weak conditions. While second-order moments are needed for consistency of the estimated unconditional covariance matrix, the existence of finite sixth-order moments are required for convergence of the second-order derivatives of the quasi-log-likelihood function. We also show the relationship of the asymptotic distributions of the 2sQML estimator for the RBEKK model and the variance targeting (VT) QML estimator for the VT-BEKK model. Monte Carlo experiments show that the bias of the 2sQML estimator is negligible, and that the appropriateness of the diagonal specification depends on the closeness to either of the Diagonal BEKK and the Diagonal RBEKK models.

Keywords: BEKK, Rotated BEKK, Diagonal BEKK, Variance targeting, Multivariate GARCH, Consistency, Asymptotic normality.

JEL Classification: C13, C32.

1 Introduction

The BEKK model of Baba, Engle, Kraft and Kroner (1985) and Engle and Kroner (1995) is widely used for estimating and forecasting time-varying conditional covariance dynamics, especially in the empirical analysis of multiple asset returns of financial time series (see the surveys of Bauwens et al. (2006), Laurent et al. (2012), McAleer (2005), and Silvennoinen and Teräsvirta (2009)), among others). The BEKK model is a natural extension of the ARCH/GARCH models of Engle (1982) and Bollerslev (1986). One of the features of the BEKK model is that it guarantees the positive definiteness of the covariance matrix. However, BEKK does not satisfy appropriate regularity conditions, so that the corresponding estimators do not possess asymptotic properties, except under restrictive conditions (see Chang and McAleer (2018), Comte and Lieberman (2003), and McAleer et al. (2008)). To cope with this problem, Hafner and Preminger (2009) showed asymptotic properties for the quasi-maximum likelihood (QML) estimator under moderate regularity conditions.

As for other multivariate GARCH models, a drawback of the BEKK model is that it contains a large number of parameters, even for moderate dimensions. To reduce the number of parameters, the so-called scalar BEKK and diagonal BEKK specifications are occasionally used in empirical analyses (see also Chang and McAleer (2018)). Recently Noureldin et al. (2014) suggested the rotated BEKK (RBEKK) model to handle the high-dimensional BEKK model. They suggest estimating the unconditional covariance matrix of the observed variables in the first step, in order to rotate the variables to have unit sample variance and zero sample correlation coefficients. In the second step, Noureldin et al. (2014) consider simplified BEKK models for QML estimation. We call this procedure two step QML (2sQML) estimation. One of the major advantages of the RBEKK model is that it can save on the number of parameters in the optimization step, while another is that it is more natural to consider simplified specifications after the rotation than to simplify the structure directly without the rotation.

The 2sQML is closely related to the concept of the variance targeting (VT) specification analyzed by Francq et al. (2011) and Pedersen and Rahbek (2014), among others. The VT-QML estimation also use the estimated unconditional covariance matrix in the first step, in order to reduce the number of parameters in the QML maximization step. Pedersen and Rahbek (2014) show the consistency and the asymptotic normality of the VT-QML estimator under the finite sixth order moments. As Noureldin et al. (2014) discuss the general framework for the asymptotic distribution of the 2sQML estimator for the RBEKK model, it is worth examining the detailed moment condition, as in Pedersen and Rahbek (2014).

In this paper, we show the consistency and asymptotic normality of the 2sQML estimator for the RBEKK models by extending the approach of Pedersen and Rahbek (2014). For asymptotic normality, we need to impose sixth-order moment restrictions, as in Hafner and Preminger (2009) and Pedersen and Rahbek (2014). We also derive the asymptotic relationship between the VT-QML estimator for the BEKK and the 2sQML estimator for RBEKK. We conduct Monte Carlo experiments to check the finite sample properties of the 2sQML estimator, and to compare the performance of the estimated diagonal BEKK and diagonal RBEKK models. All proofs of propositions and corollaries are given in the Appendix.

We use the following notation throughout the paper. 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 $\phi < 1$.

2 Rotated BEKK-GARCH Model

As in Hafner and Preminger (2009) and Pedersen and Rahbek (2014), we focus on a simple specification of the BEKK model that is defined by:

$$X_t = H_t^{1/2} Z_t, \tag{1}$$

$$H_t = C^* + A^* X_{t-1} X_{t-1}' A^{*\prime} + B^* H_{t-1} B^{*\prime}, \tag{2}$$

where $t = 1, \ldots, T$, A^* and B^* are d-dimensional square matrices, C^* is a d-dimensional positive definite matrix, and Z_t $(d \times 1)$ is an $i.i.d.(0, I_d)$ sequence of random variables.

We start from the following assumption.

Assumption 1.

(a) The distribution of Z_t is absolutely continuous with respect to Lebesgue measure on $\Re^d$, and zero is an interior point of the support of the distribution.

(b) The matrices A^* and B^* satisfy $\rho((A^* \otimes A^*) + (B^* \otimes B^*)) < 1$.

By Theorem 2.4 of Boussama et al. (2011), Assumption 1 implies the existence of a unique stationary and ergodic solution to the model in (1) and (2). Furthermore, the stationary solution has finite second-order moments, $E||X_t||^2 < \infty$, and variance $V(X_t) = E(Ht) = \Omega$, with positive definite Ω, which is the solution to:

$$\Omega = C^* + A^* \Omega A' + B^* \Omega B'. \tag{3}$$

Lemma 2.4 and Proposition 4.3 of Boussama et al. (2011) indicate that the necessary and sufficient conditions for (3) to have a solution of a positive definite matrix is Assumption 1(b). As in Pedersen and Rahbek (2014), we obtain the variance targeting specification by substituting C^* in

(3) to the model (2), giving:

$$H_t = \Omega - A^*\Omega A^{*\prime} - B^*\Omega B^{*\prime} + A^* X_{t-1} X_{t-1}' A^{*\prime} + B^* H_{t-1} B^{*\prime}. \tag{4}$$

Based on the specification, Noureldin et al. (2014) suggested the Rotated RBEKK (RBEKK) model, which is obtained by setting $A^* = \Omega^{1/2} A \Omega^{-1/2}$ and $B^* = \Omega^{1/2} B \Omega^{-1/2}$ in (2), A and B are d-dimensional square matrices. The transformation yields:

$$H_t = \Omega^{1/2} \underline{H}_t \Omega^{1/2}, \quad \underline{H}_t = (I_d - AA' - BB') + A\tilde{X}_{t-1}\tilde{X}_{t-1}' A' + B\underline{H}_{t-1} B', \tag{5}$$

with the rotated vector $\tilde{X}_t = \Omega^{-1/2} X_t$, which gives $E(\tilde{X}_t \tilde{X}_t') = I_d$. As discussed in Noureldin et al. (2014), the specification gives an natural interpretation for considering diagonal matrices A and B for reducing the number of parameters. Rather than the special case with the diagonal matrices, we consider general A and B for the asymptotic theory. With respect to the initial values, we consider estimation conditional on the initial values X_0 and $\underline{H}_0 = h$, where h is a positive definite matrix. By the structure, it is natural to replace Assumption 1(b) with the following:

Assumption 2. *The matrices A and B satisfy $\rho((A \otimes A) + (B \otimes B)) < 1$.*

Lemma 2 in Appendix A.2 shows that Assumption 2 is equivalent to Assumption 1(b).

In the next section, we consider the two step QML (2sQML) estimation for the RBEKK model (1) and (5), as in Noureldin et al. (2014) and Pedersen and Rahbek (2014).

3 Two Step QML Estimation

Let $\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Re^{3d^2}$, denote the parameter vector of the RBEKK model, which is defined by $\boldsymbol{\theta} = (\boldsymbol{\omega}', \boldsymbol{\lambda}')'$, where $\boldsymbol{\omega} = \text{vec}(\Omega)$ and $\boldsymbol{\lambda} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ with $\boldsymbol{\alpha} = \text{vec}(A)$ and $\boldsymbol{\beta} = \text{vec}(B)$. We also define the parameter space $\Theta = \Theta_\omega \times \Theta_\lambda \subset \Re^{d^2} \times \Re^{2d^2}$. As in Hafner and Preminger (2009) and Pedersen and Rahbek (2014), we emphasize the dependence of H_t and $\underline{H}_t$ on the parameters $\boldsymbol{\omega}$ and $\boldsymbol{\lambda}$, by writing $H_t(\boldsymbol{\omega}, \boldsymbol{\lambda})$ and $\underline{H}_t(\boldsymbol{\omega}, \boldsymbol{\lambda})$, respectively. We also place emphasis on the initial value of the covariance matrix, h, by denoting $H_{t,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})$ and $\underline{H}_{t,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})$. Now we restate the RBEKK model as:

$$X_t = H_t^{1/2}(\boldsymbol{\omega}, \boldsymbol{\lambda}) Z_t, \quad H_t(\boldsymbol{\omega}, \boldsymbol{\lambda}) = \Omega^{1/2} \underline{H}_t(\boldsymbol{\omega}, \boldsymbol{\lambda}) \Omega^{1/2}, \tag{6}$$

$$\underline{H}_t(\boldsymbol{\omega}, \boldsymbol{\lambda}) = (I_d - AA' - BB') + A\Omega^{-1/2} X_{t-1} X_{t-1}' \Omega^{-1/2} A' + B\underline{H}_{t-1}(\boldsymbol{\omega}, \boldsymbol{\lambda}) B', \tag{7}$$

with given initial values X_0 and $\underline{H}_{0,h}(\boldsymbol{\omega}, \boldsymbol{\lambda}) = h$.

As mentioned above, we consider 2sQML estimation which constitutes two steps. In the first step, we estimate $\boldsymbol{\omega}$ by the sample covariance matrix, while the second step conducts QML estimation by optimizing the log-likelihood function for $\boldsymbol{\lambda}$ conditional on the estimates of $\boldsymbol{\omega}$. For the RBEKK model, the Gaussian log-likelihood function is given by:

$$L_{T,h}(\boldsymbol{\omega}, \boldsymbol{\lambda}) = T^{-1} \sum_{t=1}^{T} l_{t,h,}(\boldsymbol{\omega}, \boldsymbol{\lambda}), \tag{8}$$

with the tth contribution to the log-likelihood given as:

$$l_{t,h}(\boldsymbol{\omega}, \boldsymbol{\lambda}) = -\frac{1}{2} \log\left(\det\left(H_{t,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})\right)\right) - \frac{1}{2} \operatorname{tr}\left(X_t X_t' H_{t,h}^{-1}(\boldsymbol{\omega}, \boldsymbol{\lambda})\right), \tag{9}$$

excluding the constant. In the first step, we estimate the unconditional covariance matrix by:

$$\hat{\boldsymbol{\omega}} = \operatorname{vec}\left(\hat{\Omega}\right) = \operatorname{vec}\left(T^{-1} \sum_{t=1}^{T} X_t X_t'\right), \tag{10}$$

in order to rotate X_t and $H_{t,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})$ as:

$$\tilde{X}_t = \hat{\Omega}^{1/2} X_t, \quad \underline{H}_{t,h}(\hat{\boldsymbol{\omega}}, \boldsymbol{\lambda}).$$

By the definition, we have $T^{-1} \sum_{t=1}^{T} \tilde{X}_t \tilde{X}_t' = I_d$. The conditional log-likelihood function is given by:

$$-\frac{1}{2T} \sum_{t=1}^{T} \left[\log\left(\det\left(\underline{H}_{t,h}(\boldsymbol{\omega}, \boldsymbol{\lambda})\right)\right) + \operatorname{tr}\left(\tilde{X}_t \tilde{X}_t' \underline{H}_{t,h}^{-1}(\boldsymbol{\omega}, \boldsymbol{\lambda})\right)\right],$$

which is equivalent to $L_{T,h}(\hat{\boldsymbol{\omega}}, \boldsymbol{\lambda}) + 0.5T \log(\det(\hat{\Omega}))$. Hence, the second step estimator is given by:

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda} \in \Theta_\lambda} L_{T,h}(\hat{\boldsymbol{\omega}}, \boldsymbol{\lambda}). \tag{11}$$

We derive the asymptotic theory for the 2sQML estimator, which consists of (10) and (11).

Following Comte and Lieberman (2003), Hafner ad Preminger (2009), and Pedersen and Rahbek (2014), we make the following conventional assumptions.

Assumption 3.

(a) The process $\{X_t\}$ is strictly stationary and ergodic.

(b) The true parameter $\boldsymbol{\theta}_0 \in \Theta$ and Θ is compact.

(c) For $\boldsymbol{\lambda} \in \Theta_\lambda$, if $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$, then $\underline{H}_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}) \neq \underline{H}_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)$ almost surely, for all $t \geq 1$.

For Assumption 3(a), Assumptions 1(a) and 2 imply the existence of a strictly stationary ergodic solution $\{X_t\}$ in the RBEKK model. Regarding Assumption 3(a), one of the conditions is that the first element in the matrices A and B should be strictly positive, which is a sufficient condition for parameter identification, as shown in Engle and Kroner (1995).

We now state the following result regarding consistency of the 2sQML estimator.

Proposition 1. *Under Assumptions 1(a), 2, and 3, as $T \to \infty$, $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$.*

Assumptions 2(a) and 2(b) imply the finite second-order moments of X_t, which are necessary for estimating Ω with the sample covariance matrix. As shown by Hafner ad Preminger (2009), the consistency of the QML estimator for the BEKK model (1) and (2) do not require the finite second-order moment of X_t.

We make the following assumption for the asymptotic normality of the 2sQML estimator.

Assumption 4.

(a) $E[||X_t||^6] < \infty$.

(b) $\boldsymbol{\theta}_0$ *is in the interior of* Θ.

As in Pedersen and Rahbek (2014), we need to assume finite six-order moments in order to show that the second-order derivatives of the log-likelihood function converge uniformly on the parameter space. This is different from the univariate case, which only requires finite fourth-order moments (see Francq at al. (2011)).

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

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, Q_0\Gamma_0 Q_0'\right),$$

where

$$Q_0 = \begin{pmatrix} I_{d^2} & O_{d^2\times 2d^2} \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix},$$

with the non-singular matrix J_0 *and the matrix* K_0 *stated in (A.17), and the non-singular matrix* Γ_0 *stated in (A.21), and* Q_0*.*

Given the asymptotic distribution of $\hat{\boldsymbol{\theta}}$, we can show the asymptotic distribution of the 2sQML estimator of (Ω, A^*, B^*) in the VT representation of the BEKK. Define $\boldsymbol{\theta} = (\boldsymbol{\omega}', \boldsymbol{\lambda}^{*\prime})'$, where $\boldsymbol{\lambda}^* = (\boldsymbol{\alpha}^{*\prime}, \boldsymbol{\beta}^{*\prime})'$ with $\boldsymbol{\alpha}^* = \text{vec}(A^*)$ and $\boldsymbol{\beta}^* = \text{vec}(B^*)$.

Corollary 1. *Under the assumptions of Proposition 2, as* $T \to \infty$*:*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \xrightarrow{d} N\left(\mathbf{0}, Q_0^*\Gamma_0^* Q_0^{*\prime}\right),$$

where

$$Q_0^* = \begin{pmatrix} I_{d^2} & O_{d^2\times 2d^2} \\ -J_0^{*-1}K_0^* & -J_0^{*-1} \end{pmatrix},$$

with the non-singular matrix J_0^* and the matrix K_0^* stated in (A.23), and the non-singular matrix Γ_0^* stated in (A.28).

As implied in the proof of Corollay 1, the asymptotic covariance matrix is equivalent to the one derived by Theorem 4.2 of Pedersen and Rahbek (2014). Combining Corollary 4.1 of Pedersen and Rahbek (2014) and Corollary 1, we provide the asymptotic distribution of the 2sQML estimator for (C^*, A^*, B^*) in the original BEKK model. Define $\boldsymbol{c}^* = \mathrm{vec}(C^*)$.

Corollary 2. *Under the assumptions of Proposition 2, as $T \to \infty$,*

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{c}}^* - \boldsymbol{c}^* \\ \hat{\boldsymbol{\alpha}}^* - \boldsymbol{\alpha}^* \\ \hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^* \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, S_0' R_0 Q_0 \Gamma_0 Q_0' R_0 S_0\right),$$

where

$$S_0 = \begin{pmatrix} I_{d^2} - (\Omega_0^{1/2} A_0 \Omega_0^{-1/2})^{\otimes 2} - (\Omega_0^{1/2} B_0 \Omega_0^{-1/2})^{\otimes 2} & O_{d^2\times d^2} & O_{d^2\times d^2} \\ -(I_{d^2} + C_{dd})((\Omega_0^{1/2} A_0 \Omega_0^{1/2}) \otimes I_d) & I_{d^2} & O_{d^2\times d^2} \\ -(I_{d^2} + C_{dd})((\Omega_0^{1/2} B_0 \Omega_0^{1/2}) \otimes I_d) & O_{d^2\times d^2} & I_{d^2} \end{pmatrix},$$

with R_0 defined by (A.22).

We can estimate Γ_0, K_0, and J_0 by the sample outer-product of the gradient and Hessian matrices, as:

$$\hat{\Gamma} = \frac{1}{T}\sum_{t=1}^{T} \hat{\gamma}_t \hat{\gamma}_t', \quad \hat{K} = \frac{1}{T}\sum_{t=1}^{T} \hat{K}_t, \quad \hat{J} = \frac{1}{T}\sum_{t=1}^{T} \hat{J}_t,$$

where

$$\hat{\gamma}_t = \begin{pmatrix} \mathrm{vec}(X_t X_t') - \hat{\boldsymbol{\omega}} \\ \frac{\partial l_{t,h}(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \end{pmatrix}, \quad \hat{K}_t = \left.\frac{\partial^2 l_{t,h}(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad \hat{J}_t = \left.\frac{\partial^2 l_{t,h}(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

By Proposition 1, we can estimate S_0 and R_0 via the 2sQML estimate, $\hat{\boldsymbol{\theta}}$.

4 Monte Carlo Experiments

In this section, we illustrate the theoretical results in the previous section via Monte Carlo experiments. We consider bivariate RBEKK models ($d = 2$) for the data generating processes

(DGPs). As we assume finite sixth-order moments for asymptotic normality, we use the sufficient condition for the BEKK-ARCH models, given in Theorem C.1 of Pedersen and Rahbek (2014) (see Avarucci et al. (2013) for an extensive discussions on higher-order moment restrictions on BEKK-ARCH models). For the sufficient condition, we restrict the parameter to satisfy $\rho(A_0^* \otimes A_0^*) < (1/15)^{1/3} \approx 0.4055$. Note that $\rho(A_0^* \otimes A_0^*) = \rho(A_0 \otimes A_0)$ by Lemma 2. We use $\underline{H}_1 = I_2$ for the initial value, in order to generate $T = 500$ observations. We set the number of replications as 2000.

In the first experiment, we consider the following structure in (5):

$$\Omega_0 = \begin{pmatrix} s_{01} & 0 \\ 0 & s_{02} \end{pmatrix} \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \begin{pmatrix} s_{01} & 0 \\ 0 & s_{02} \end{pmatrix}, \quad A_0 = \begin{pmatrix} A_{0,11} & 0 \\ 0 & A_{0,22} \end{pmatrix},$$

with $B_0 = O_{2\times 2}$. We consider two kinds of parameter sets:

$$\text{DGP1:} \quad (s_{01}, s_{02}) = (1, 0.9), \quad \rho_0 = 0.5, \quad (A_{0,11}, A_{0,22}) = (0.6, 0.4),$$

$$\text{DGP2:} \quad (s_{01}, s_{02}) = (0.8, 1.1), \quad \rho_0 = -0.3, \quad (A_{0,11}, A_{0,22}) = (0.6, -0.3),$$

which are used to obtain (C_0^*, A_0^*) for the DGPs by (1) and (2). The values of (Ω_0, A_0) and the corresponding values of (C_0^*, A_0^*) are given in Table 1 and Table 2, respectively. While DGP1 describes the positive unconditional correlation, DGP2 uses the negative correlation. By the specification, we can verify that $\rho(A_0^* \otimes A_0^*) = 0.3969$. From this setting, we examine the finite sample property of the 2sQML estimator for (Ω, A). Table 1 shows the sample mean, standard error, and root mean squared error of the 2sQML estimator. Table 1 indicates that the bias of the estimators is negligible, even for $T = 500$.

We also check the effects of the transformation from $(\hat{\Omega}, \hat{A}, \hat{B})$ to $(\hat{C}^*, \hat{A}^*, \hat{B}^*)$, as shown by Corollary 2. Table 2 shows the sample mean, standard error, and root mean squared error of the transformed estimator. As in Table 1, the bias of the estimators is negligible.

We examine the effects of the diagonal specification for the BEKK and RBEKK models when the true model is full BEKK. For this purpose, we consider several measures for checking the distance from the diagonal BEKK and RBEKK models to the full BEKK model. Define the non-diagonal indices as:

$$\gamma = \|A^* - \mathrm{diag}(A^*)\| + \|B^* - \mathrm{diag}(B^*)\| \quad \text{(Diagonal BEKK)},$$

$$\gamma^r = \left\|A^* - \Omega^{1/2}\mathrm{diag}(\Omega^{-1/2}A^*\Omega^{1/2})\Omega^{-1/2}\right\| + \left\|B^* - \Omega^{1/2}\mathrm{diag}(\Omega^{-1/2}B^*\Omega^{1/2})\Omega^{-1/2}\right\| \qquad (12)$$

$$\text{(Diagonal RBEKK)},$$

where $\mathrm{diag}(Y)$ creates a diagonal matrix from a square matrix Y. By the non-diagonal indices, we can calculate the theoretical distance of the diagonal BEKK and RBEKK models. For the remaining measures, we use the estimated values of the parameters of the diagonal BEKK and RBEKK models. The maximized log-likelihood $L_{T,h}(\hat{\boldsymbol{\theta}})$ is used, as is the average of the Frobenius norm of the difference of conditional covariance matrices:

$$\frac{1}{T}\sum_{t=1}^{T}\left\|H_{t,h}(\hat{\boldsymbol{\theta}}) - H_{t,h}(\boldsymbol{\theta}_0)\right\|.$$

Note that the last measure uses the true values used in the DGPs.

By using these measures, the following Monte Carlo simulations investigate the effects of the diagonal specification for the BEKK and RBEKK models when the true model is full BEKK. For this purpose, consider the specification for (4) with $B_0^* = O_{2\times 2}$:

$$A_0^* = wD_1 + (1-w)\Omega_0^{1/2}D_2\Omega_0^{-1/2}, \qquad (13)$$

for $0 \le w \le 1$, where D_0 and D_1 are diagonal matrices. When $w = 1$, the specification reduces to the diagonal BEKK model, while it becomes the diagonal RBEKK model for $w = 0$. Except for these endpoints, the full BEKK specification gives a non-diagonal structure for A_0^* in (4) and A_0

in (5). For the specification in (4), the non-diagonal indices give linear functions of w:

$$\gamma_w = \xi(1-w), \quad \xi = \left\| \Omega_0^{1/2} D_2 \Omega_0^{-1/2} - \mathrm{diag}(\Omega_0^{1/2} D_2 \Omega_0^{-1/2}) \right\|,$$

$$\gamma_w^r = \xi_r w, \quad \xi_r = \left\| D_1 - \Omega_0^{1/2} \mathrm{diag}(\Omega_0^{-1/2} D_1 \Omega_0^{1/2}) \Omega_0^{-1/2} \right\|,$$

so as to calculate the theoretical distances. Consider the parameter settings for the DGPs as:

$$\mathrm{DGP3}_w: \ (\Omega_0, A_0) \text{ in DGP1, with } D_1 = D_2 = A_0 \text{ in (13),}$$

$$\mathrm{DGP4}_w: \ (\Omega_0, A_0) \text{ in DGP2, with } D_1 = D_2 = A_0 \text{ in (13).}$$

Set $w = 0, 0.1, \ldots, 1$ to examine 11 cases, with $T = 500$, and the number of replications set to 2000. We estimate the diagonal RBEKK model by the 2sQML method, while VT-QML is used for the diagonal BEKK model.

Figures 1 and 2 show the sample means of the average bias for the conditional covariance matrices and the sample means of the maximized log-likelihood function for DGP3 and DGP4, respectively. As expected from the structure, the superiority of the diagonal models depends on the structure of the true BEKK model. If w is closer to zero, the diagonal RBEKK model is preferred. The non-diagonality indices are

$$\mathrm{DGP3}_w: \ \gamma_w = 0.0106(1-w), \quad \gamma_w^r = 0.0155w, \quad \text{crossing at } w^\dagger = 0.406,$$

$$\mathrm{DGP4}_w: \ \gamma_w = 0.0203(1-w), \quad \gamma_w^r = 0.0221w, \quad \text{crossing at } w^\dagger = 0.479,$$

and these theoretical values of $w^\dagger$ correspond to the intersections shown in Figures 1 and 2, respectively. Note that the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) lead to the same conclusion, as the numbers of parameters in these two models are the same.

5 Conclusion

For the RBEKK-GARCH model, we have shown consistency and asymptotic normality of the 2sQML estimator under weak conditions. The 2sQML estimation uses the unconditional covariance matrix for the first step, and rotates the observed vector to have the identity matrix for its sample covariance matrix. The second step conducts QML estimation for the remaining parameters. While we require second-order moments for consistency due to the estimation of the covariance matrix, we need finite sixth-order moments for asymptotic normality, as in Pedersen and Rahbek (2014). We also showed the asymptotic relation of the 2sQML estimator for the RBEKK model and the VT-QML estimator for the VT-BEKK model. Monte Carlo results showed that the finite sample properties of the 2sQML estimator are satisfactory, and that the adequacy of the diagonal RBEKK depends on the structure of the true parameters.

As an extension of the dynamic conditional correlation (DCC) model of Engle (2002), Noureldin et al. (2014) suggested the rotated DCC models (for a caveat about the regularity conditions underlying DCC, see McAleer (2018)). We may apply the rotation for different kinds of correlation models suggested by McAleer et al. (2008) and Tse and Tsui (2002). Together with such extensions, the derivation of the asymptotic theory for the rotated models is an important direction for future research.

References

Avarucci, M., E. Beutner, and P. Zaffaroni (2013), "On Moment Conditions for Quasi-Maximum Likelihood Estimation of Multivariate ARCH models", *Econometric Theory*, **29**, 545–566.

Baba, Y., R. Engle, D. Kraft and K. Kroner (1985), "Multivariate Simultaneous Generalized ARCH", Unpublished Paper, University of California, San Diego. [Published as Engle and Kroner (1995)]

Bauwens L., S. Laurent, and J. K. V. Rombouts (2006), "Multivariate GARCH Models: A Survey", *Journal of Applied Econometrics*, **21**, 79–109.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, **31**, 307–327.

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.

Chang, C.-L. and M. McAleer (2018), "The Fiction of Full BEKK: Pricing Fossil Fuels and Carbon Emissions", to appear in *Finance Research Letters*.

Comte, F. and O. Lieberman (2003), "Asymptotic Theory for Multivariate GARCH Processes", *Journal of Multivariate Analysis*, **84**, 61–84.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", *Econornetrica*, **50**, 987–1007.

Engle, R. F. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business & Economic Statistics*, **20**, 339–350.

Engle, R.F. and K.F. Kroner (1995), "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, **11**, 122–150.

Francq, C., L. Horváth, and J.-M. Zakoïan (2011), "Merits and Drawbacks of Variance Targeting in GARCH Models", *Journal of Financial Econometrics*, **9**, 619–656.

Hafner, C. M. and A. Preminger (2009), "On Asymptotic Theory for Multivariate GARCH Models", *Journal of Multivariate Analysis*, **100**, 2044–2054.

Laurent, S., J. V. K. Rombouts, and F. Violante (2012), "On the Forecasting Accuracy of Multivariate GARCH Models", *Journal of Applied Econometrics*, **27**, 934–955.

Lütkephol, H. (1996), *Handbook of Matrices*, New York: John Wiley & Sons.

McAleer, M. (2005), "Automated Inference and Learning in Modeling Financial Volatility", *Econometric Theory*, **21**, 232–261.

McAleer, M. (2018), "Stationarity and Invertibility of a Dynamic Correlation Matrix", *Kybernetika*, **54**(2), 363–374.

McAleer, M., F. Chan, S. Hoti and O. Lieberman (2008), "Generalized Autoregressive Conditional Correlation", *Econometric Theory*, **24**, 1554–1583.

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.

Silvennoinen, A., and T. Teräsvirta (2009), “Multivariate GARCH Models”, In T. G. Andersen, R.A. Davis, J.-P. Kreiss, and T. Mikosch (eds.), *Handbook of Financial Time Series*, 201–229, New York: Springer.

Tse, Y. K. and A. K. C. Tsui (2002), “A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations”, *Journal of Business & Economic Statistics*, **20**, 351–361.

Appendix

A.1 Derivatives of Log-Likelihood Function

Although Pedersen and Rahbek (2014) demonstrate the derivatives with respect to Ω, A^*, and B^*, they are not applicable as A^* and B^* in (2) depend on $\Omega^{1/2}$ and $\Omega^{-1/2}$ in the RBEKK model (6) and (7), respectively. Related to this issue, we need the following lemma to show the derivatives of the log-likelihood function.

Lemma 1.

$$\frac{\partial \text{vec}\left(\Omega^{1/2}\right)}{\partial \boldsymbol{\omega}'} = \left[\left(\Omega^{1/2} \otimes I_d\right) + \left(I_d \otimes \Omega^{1/2}\right)\right]^{-1},$$
$$\frac{\partial \text{vec}\left(\Omega^{-1/2}\right)}{\partial \boldsymbol{\omega}'} = -\left[\left(\Omega^{-1/2} \otimes I_d\right) + \left(I_d \otimes \Omega^{-1/2}\right)\right]^{-1} \left(\Omega^{-1}\right)^{\otimes 2}.$$

Proof. By the product rule, it is straightforward to obtain:

$$\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\omega}'} = \frac{\partial \text{vec}\left(\Omega^{1/2}\Omega^{1/2}\right)}{\partial \boldsymbol{\omega}'} = \left[\left(\Omega^{1/2} \otimes I_d\right) + \left(I_d \otimes \Omega^{1/2}\right)\right] \frac{\partial \text{vec}\left(\Omega^{1/2}\right)}{\partial \boldsymbol{\omega}'}.$$

Since $\Omega^{1/2}$ is positive definite, we obtain the result. A similar application produces:

$$\frac{\partial \text{vec}\left(\Omega^{-1}\right)}{\partial \boldsymbol{\omega}'} = \frac{\partial \text{vec}\left(\Omega^{-1/2}\Omega^{-1/2}\right)}{\partial \boldsymbol{\omega}'} = \left[\left(\Omega^{-1/2} \otimes I_d\right) + \left(I_d \otimes \Omega^{-1/2}\right)\right] \frac{\partial \text{vec}\left(\Omega^{-1/2}\right)}{\partial \boldsymbol{\omega}'}.$$

By the derivative of the inverse of the symmetric matrix shown by 10.6.1(1) of Lütkephol (1996), we obtain the second result. □

The gradient and Hessian of the log likelihood function are given by:

$$\frac{\partial L_T}{\partial \boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t}{\partial \boldsymbol{\theta}}, \quad \frac{\partial^2 L_T}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

Applying the chain rule and product rule, we obtain:

$$\begin{aligned} \frac{\partial l_t}{\partial \boldsymbol{\theta}} &= \frac{\partial \text{vec}(H_t)'}{\partial \boldsymbol{\theta}} \frac{\partial l_t}{\partial \text{vec}(H_t)}, \\ \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} &= \frac{\partial^2 \text{vec}(H_t)'}{\partial \theta_i \partial \theta_j} \frac{\partial l_t}{\partial \text{vec}(H_t)} + \frac{\partial \text{vec}(H_t)'}{\partial \theta_i} \frac{\partial^2 l_t}{\partial \text{vec}(H_t) \partial \text{vec}(H_t)} \frac{\partial \text{vec}(H_t)}{\partial \theta_j} \end{aligned} \tag{A.1}$$

where θ_i $(i = 1, \ldots, 3d^2)$ is the ith element of $\boldsymbol{\theta}$,

$$
\begin{aligned}
\frac{\partial l_t}{\partial H_t} &= -\frac{1}{2} H_t^{-1} + \frac{1}{2} H_t^{-1} X_t X_t' H_t^{-1}, \\
\frac{\partial^2 l_t}{\partial \mathrm{vec}(H_t) \partial \mathrm{vec}(H_t)} &= \frac{1}{2} \left[I_{d^2} - \{(H_t^{-1} X_t X_t) \otimes I_d\} - \{I_d \otimes (H_t^{-1} X_t X_t)\} \right] (H_t^{-1})^{\otimes 2}.
\end{aligned} \tag{A.2}
$$

The first equation of (A.2) uses 10.3.2(23) and 10.3.3(10) of Lütkephol (1996), while we applied 10.6.1(1) for the second equation.

By Lemma 1, the product rule, and the chain rule, we obtain the first derivatives:

$$
\begin{aligned}
\frac{\partial \mathrm{vec}(H_t)}{\partial \boldsymbol{\omega}'} &= \left[\left(\Omega^{1/2} \underline{H}_t \otimes I_d \right) + \left(I_d \otimes \Omega^{1/2} \underline{H}_t \right) \right] \left[\left(\Omega^{1/2} \otimes I_d \right) + \left(I_d \otimes \Omega^{1/2} \right) \right]^{-1} \\
&\quad + \left(\Omega^{1/2} \right)^{\otimes 2} \frac{\partial \mathrm{vec}(\underline{H}_t)}{\partial \boldsymbol{\omega}'}, \\
\frac{\partial \mathrm{vec}(H_t)}{\partial \boldsymbol{\lambda}'} &= \left(\Omega^{1/2} \right)^{\otimes 2} \frac{\partial \mathrm{vec}(\underline{H}_t)}{\partial \boldsymbol{\lambda}'},
\end{aligned} \tag{A.3}
$$

and

$$
\begin{aligned}
\frac{\partial \mathrm{vec}(\underline{H}_t)}{\partial \boldsymbol{\omega}'} &= B^{\otimes 2} \frac{\partial \mathrm{vec}(\underline{H}_{t-1})}{\partial \boldsymbol{\omega}'} - A^{\otimes 2} \left[(I_d \otimes \Omega^{-1/2} X_{t-1} X_{t-1}') + (\Omega^{-1/2} X_{t-1} X_{t-1}' \otimes I_d) \right] \\
&\quad \times \left[\left(\Omega^{-1/2} \otimes I_d \right) + \left(I_d \otimes \Omega^{-1/2} \right) \right]^{-1} \left(\Omega^{-1} \right)^{\otimes 2}, \\
\frac{\partial \mathrm{vec}(\underline{H}_t)}{\partial \boldsymbol{\alpha}'} &= B^{\otimes 2} \frac{\partial \mathrm{vec}(\underline{H}_{t-1})}{\partial \boldsymbol{\alpha}'} + \left(A \left\{ \Omega^{-1/2} X_{t-1} X_{t-1}' \Omega^{-1/2} - I_d \right\} \otimes I_d \right) \\
&\quad + \left(I_d \otimes A \left\{ \Omega^{-1/2} X_{t-1} X_{t-1}' \Omega^{-1/2} - I_d \right\} \right) C_{dd}, \\
\frac{\partial \mathrm{vec}(\underline{H}_t)}{\partial \boldsymbol{\beta}'} &= B^{\otimes 2} \frac{\partial \mathrm{vec}(\underline{H}_{t-1})}{\partial \boldsymbol{\beta}'} + \left(B \left\{ \underline{H}_{t-1} - I_d \right\} \otimes I_d \right) + \left(I_d \otimes B \left\{ \underline{H}_{t-1} - I_d \right\} \right) C_{dd},
\end{aligned} \tag{A.4}
$$

where C_{dd} is the commutation matrix, which consists of one and zero satisfying $\mathrm{vec}(A') = C_{dd}\mathrm{vec}(A)$.

Similarly, the second derivatives of H_t are given by:

$$
\begin{aligned}
\frac{\partial^2 \mathrm{vec}(H_t)}{\partial \omega_i \partial \omega_j} &= \left[\left(\Omega^{1/2} \frac{\underline{H}_t}{\partial \omega_i} \otimes I_d \right) + \left(I_d \otimes \Omega^{1/2} \frac{\underline{H}_t}{\partial \omega_i} \right) \right] \left[\left(\Omega^{1/2} \otimes I_d \right) + \left(I_d \otimes \Omega^{1/2} \right) \right]^{-1} \boldsymbol{e}^{(j)} \\
&\quad + \left(\Omega^{1/2} \right)^{\otimes 2} \frac{\partial^2 \mathrm{vec}(\underline{H}_t)}{\partial \omega_i \partial \omega_j} \qquad (i, j = 1, \ldots, d^2), \\
\frac{\partial \mathrm{vec}(H_t)}{\partial \lambda_i \partial \lambda_j} &= \left(\Omega^{1/2} \right)^{\otimes 2} \frac{\partial \mathrm{vec}(\underline{H}_t)}{\partial \lambda_i \partial \lambda_j} \qquad (i, j = 1, \ldots, 2d^2),
\end{aligned}
$$

$$\frac{\partial^2 \text{vec}(H_t)}{\partial\lambda_i \partial\omega_j} = \left[\left(\Omega^{1/2}\frac{\underline{H}_t}{\partial\lambda_i} \otimes I_d\right) + \left(I_d \otimes \Omega^{1/2}\frac{\underline{H}_t}{\partial\lambda_i}\right)\right]\left[\left(\Omega^{1/2} \otimes I_d\right) + \left(I_d \otimes \Omega^{1/2}\right)\right]^{-1} \boldsymbol{e}^{(j)},$$
$$+ \left(\Omega^{1/2}\right)^{\otimes 2} \frac{\partial^2 \text{vec}(\underline{H}_t)}{\partial\lambda_i \partial\omega_j} \qquad (i = 1, \ldots, 2d^2,\ j = 1, \ldots, d^2),$$

where $\boldsymbol{e}^{(j)}$ is a $d^2 \times 1$ vector of zeros except for the jth element, which takes one. We have omitted the derivatives of $\underline{H}_t$.

A.2 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).

Before we proceed, we show the equivalence of Assumptions 1(b) and 2.

Lemma 2. *For the RBEKK model defined by (4) and (5), it can be shown that:*

$$\rho((A^* \otimes A^*) + (B^* \otimes B^*)) = \rho((A \otimes A) + (B \otimes B)).$$

Proof. Noting that

$$(A^* \otimes A^*) + (B^* \otimes B^*) = (\Omega^{1/2} \otimes \Omega^{1/2})\left\{(A \otimes A) + (B \otimes B)\right\}(\Omega^{-1/2} \otimes \Omega^{-1/2}),$$

5.2.1(8) of Lütkephol (1996) indicates that the eigenvalues of $(A^* \otimes A^*) + (B^* \otimes B^*)$ are the same as those of $(A \otimes A) + (B \otimes B)$, which proves the lemma. □

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

$$\hat{\boldsymbol{\omega}} \xrightarrow{a.s.} \boldsymbol{\omega}_0. \tag{A.5}$$

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

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda} |L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda}) - L_{T,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})| \xrightarrow{a.s.} 0. \tag{A.6}$$

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:

$$\begin{aligned}&\text{vec}\left(\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right) - \text{vec}\left(\underline{H}_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right)\\ &\quad= \sum_{i=0}^{t-1}(B^{\otimes 2})^i A^{\otimes 2}\left\{(\Omega^{-1})^{\otimes 2} - (\hat{\Omega}^{-1})^{\otimes 2}\right\}\text{vec}\left(X_{t-i-1}X'_{t-i-1}\right) + (B^{\otimes 2})^t\text{vec}\left(\underline{H}_0 - h\right).\end{aligned} \tag{A.7}$$

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 $0 < \phi < 1$:

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|(B^{\otimes 2})^i\right\| \le K\phi^i. \tag{A.8}$$

For equation (A.7), by the compactness of Θ, (A.5), and (A.8), we obtain:

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|\text{vec}\left(\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right) - \text{vec}\left(\underline{H}_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right)\right\| \le K\phi^t + o(1) \;\; \text{a.s.}, \tag{A.9}$$

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

$$\begin{aligned}\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|\underline{H}_{t,h}^{-1}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right\| &\le \sup_{\boldsymbol{\theta}\in\Theta}\left\|\underline{H}_{t,h}^{-1}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})\right\| \le K,\\ \sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left\|\underline{H}_{t,h}^{-1}(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right\| &\le \sup_{\boldsymbol{\theta}\in\Theta}\left\|\underline{H}_{t,h}^{-1}(\boldsymbol{\omega}_0,\boldsymbol{\lambda})\right\| \le K,\end{aligned} \tag{A.10}$$

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

Now, we turn to the difference of the likelihood function as in (A.6). By the technique of the proof of Lemma B.1 of Pedersen and Rahbek (2014), we obtain:

$$\begin{aligned}&\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}|L_T(\boldsymbol{\omega}_0,\boldsymbol{\lambda}) - L_{T,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda})|\\ &\qquad\le \left|\log\left(\frac{\det(\Omega_0)}{\det(\hat{\Omega})}\right)\right| + \frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|\log\left(\frac{\det(\underline{H}_t(\boldsymbol{\omega}_0,\boldsymbol{\lambda}))}{\det(\underline{H}_{t,h}(\hat{\boldsymbol{\omega}},\boldsymbol{\lambda}))}\right)\right|\end{aligned}$$

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

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 (A.9), we obtain:

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

As in the proof of Lemma B.1 of Pedersen and Rahbek (2014), it is shown that (A.6) 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 3 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 (A.5), we obtain as $T \to \infty$, $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$.

A.3 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 4. *Under Assumptions 1(a), 2-4, as $T \to \infty$,*

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0 \\ \partial L_T(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda} \end{pmatrix} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Upsilon_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \mathrm{vec}\left(Z_t Z_t' - I_d\right) + o_p(1), \tag{A.11}$$

where

$$\begin{aligned} &\Upsilon_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \\ &= \begin{pmatrix} \Upsilon_{\omega t}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \\ \Upsilon_{\alpha t}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \\ \Upsilon_{\beta t}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \end{pmatrix} = \begin{pmatrix} \left(\Omega_0^{1/2}\right)^{\otimes 2} \left(I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right)^{-1} \left(I_{d^2} - B_0^{\otimes 2}\right) \left(\Omega_0^{-1/2} H_{0t}^{1/2}\right)^{\otimes 2} \\ \frac{1}{2}\left[\sum_{i=0}^{\infty} (B_0^{\otimes 2})^i N_{t-1-i}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)\right]' \left(\Omega_0^{1/2} H_{0t}^{-1/2}\right)^{\otimes 2} \\ \frac{1}{2}\left[\sum_{i=0}^{\infty} (B_0^{\otimes 2})^i \tilde{N}_{t-1-i}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)\right]' \left(\Omega_0^{1/2} H_{0t}^{-1/2}\right)^{\otimes 2} \end{pmatrix} \end{aligned} \tag{A.12}$$

with

$$\begin{aligned} N_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) &= \left[A_0(\Omega_0^{-1/2} X_t X_t' \Omega_0^{-1/2} - I_d) \otimes I_d\right] + \left[I_d \otimes A_0(\Omega_0^{-1/2} X_t X_t' \Omega_0^{-1/2} - I_d)\right] C_{dd}, \\ \tilde{N}_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) &= \left[B_0(\underline{H}_{0t} - I_d) \otimes I_d\right] + \left[I_d \otimes B_0(\underline{H}_{0t} - I_d)\right] C_{dd}. \end{aligned} \tag{A.13}$$

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 (A.11).

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

$$\hat{\boldsymbol{\omega}} = \frac{1}{T} \sum_{t=1}^{T} \left(H_{0t}^{1/2}\right)^{\otimes 2} \mathrm{vec}\left(Z_t Z_t' - I_d\right) + \mathrm{vec}\left(\frac{1}{T} \sum_{t=1}^{T} H_{0t}\right), \tag{A.14}$$

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,

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

yielding:

$$\operatorname{vec}\left(\frac{1}{T}\sum_{t=1}^{T}\underline{H}_{0t}\right) = \left(I_{d^2} - B_0^{\otimes 2}\right)^{-1}\operatorname{vec}(I - A_0A_0' - B_0B_0') \\ + \left(I_{d^2} - B_0^{\otimes 2}\right)^{-1}\left(A_0\Omega_0^{-1/2}\right)^{\otimes 2}\left(\hat{\boldsymbol{\omega}} + \frac{1}{T}\operatorname{vec}(X_0X_0' - X_TX_t')\right) \\ + \left(I_{d^2} - B_0^{\otimes 2}\right)^{-1}B_0^{\otimes 2}\frac{1}{T}\operatorname{vec}(\underline{H}_{00} - \underline{H}_{0T}). \tag{A.15}$$

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

After inserting (A.14) in (A.15), we can transform the equation to obtain:

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

which gives

$$\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0 = \left(\Omega_0^{1/2}\right)^{\otimes 2}[I - A_0^{\otimes 2} - B_0^{\otimes 2}]^{-1}\left(I_{d^2} - B_0^{\otimes 2}\right)\left(\Omega_0^{-1/2}\right)^{\otimes 2} \\ \times\frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\operatorname{vec}\left(Z_tZ_t' - I_d\right) \\ + \left(\Omega_0^{1/2}\right)^{\otimes 2}[I - A_0^{\otimes 2} - B_0^{\otimes 2}]^{-1} \\ \times\left[\left(A_0\Omega_0^{-1/2}\right)^{\otimes 2}\frac{1}{T}\operatorname{vec}(X_0X_0' - X_TX_t') + B_0^{\otimes 2}\frac{1}{T}\operatorname{vec}(\underline{H}_{00} - \underline{H}_{0T})\right].$$

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

$$P\left(\left\|\left(A_0\Omega_0^{-1/2}\right)^{\otimes 2}\frac{1}{\sqrt{T}}\text{vec}(X_0X_0' - X_TX_t') + B_0^{\otimes 2}\frac{1}{\sqrt{T}}\text{vec}(\underline{H}_{00} - \underline{H}_{0T})\right\| > \varepsilon\right) \leq \frac{KE||X_t||^2}{\sqrt{T}\varepsilon} \to 0,$$

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

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

Therefore, (A.11) 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:

$$0 = \frac{\partial L_{T,h}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)}{\partial \boldsymbol{\lambda}} + K_{T,h}(\boldsymbol{\theta}^\dagger)(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0) + J_{T,h}(\boldsymbol{\theta}^\dagger)(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0), \tag{A.16}$$

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:

$$K_0 = E\left(\frac{\partial^2 l_t(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\omega}'}\right), \quad J_0 = E\left(\frac{\partial^2 l_t(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right). \tag{A.17}$$

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:

$$E\left[\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial^2 l_t(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \theta_i \partial \theta_j}\right|\right] < \infty, \tag{A.18}$$

$$\sup_{\boldsymbol{\lambda}\in\Theta_\lambda}\left|\frac{\partial^2 L_T(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \theta_i \partial \theta_j} - E\left[\frac{\partial^2 l_t(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \theta_i \partial \theta_j}\right]\right| \xrightarrow{a.s.} 0, \tag{A.19}$$

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 (A.16) 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 (A.19) 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 $T \to \infty$:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Upsilon_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \mathrm{vec}\left(Z_t Z_t' - I_d \right) \xrightarrow{d} N(0, \Gamma_0), \tag{A.20}$$

where

$$\Gamma_0 = E\left[\Upsilon_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \mathrm{vec}\left(Z_t Z_t' - I_d \right) \left(\mathrm{vec}\left(Z_t Z_t' - I_d \right) \right)' \Upsilon_t'(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \right], \tag{A.21}$$

with $\Upsilon_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)$ defined by (A.12). By Lemma 4, (A.20), and the Slutzky theorem, we can obtain the asymptotic normality of the 2sQML estimator.

A.4 Proof of Corollary 1

By the definition of A^* and B^* and the rule of vectorization, $\boldsymbol{\alpha}^* = (\Omega^{-1/2} \otimes \Omega^{1/2})\boldsymbol{\alpha}$ and $\boldsymbol{\beta}^* = (\Omega^{-1/2} \otimes \Omega^{1/2})\boldsymbol{\beta}$. Hence, $\boldsymbol{\theta}^* = R\boldsymbol{\theta}$, where:

$$R = \begin{pmatrix} I_{d^2} & O_{d^2 \times 2d^2} \\ O_{2d^2 \times d^2} & P \end{pmatrix}, \quad P = \begin{pmatrix} (\Omega^{-1/2} \otimes \Omega^{1/2}) & O_{d^2 \times d^2} \\ O_{d^2 \times d^2} & (\Omega^{-1/2} \otimes \Omega^{1/2}) \end{pmatrix}. \tag{A.22}$$

Note that $P' = P$ and $R' = R$. We also define P_0 and R_0 which correspond to the true value Ω_0. By Proposition 2 and the delta method, $\sqrt{T}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \xrightarrow{d} N(\mathbf{0}, R_0 Q_0 \Gamma_0 Q_0' R)$.

In the following, we will show the equivalence of the asymptotic covariance matrix. First, consider the second derivatives of the tth contribution to the likelihood function in order to obtain:

$$\frac{\partial^2 l_t}{\partial \boldsymbol{\lambda}^* \partial \boldsymbol{\omega}'} = P^{-1} \frac{\partial^2 l_t}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}'}, \quad \frac{\partial^2 l_t}{\partial \boldsymbol{\lambda}^* \partial \boldsymbol{\lambda}^{*\prime}} = P^{-1} \frac{\partial^2 l_t}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} P^{-1}.$$

Define

$$K_0^* = E\left(\frac{\partial^2 l_t}{\partial \boldsymbol{\lambda}^* \partial \boldsymbol{\omega}'} \right), \quad J_0^* = E\left(\frac{\partial^2 l_t}{\partial \boldsymbol{\lambda}^* \partial \boldsymbol{\lambda}^{*\prime}} \right). \tag{A.23}$$

Then, we obtain $K_0^* = P_0^{-1} K_0$ and $J_0^* = P_0^{-1} J_0 P_0^{-1}$. For Q_0^* defined by Corollary 1:

$$Q_0^* = \begin{pmatrix} I_{d^2} & O_{d^2 \times 2d^2} \\ O_{2d^2 \times d^2} & P \end{pmatrix} \begin{pmatrix} I_{d^2} & O_{d^2 \times 2d^2} \\ -J_0^{-1} K_0 & -J_0^{-1} \end{pmatrix} \begin{pmatrix} I_{d^2} & O_{d^2 \times 2d^2} \\ O_{2d^2 \times d^2} & P \end{pmatrix} = R_0 Q_0 R_0. \tag{A.24}$$

Next we define some quantities, as in Lemma B.8 of Pedersen and Rahbek (2014), as:

$$\Upsilon_t^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) = \begin{pmatrix} \Upsilon_{\omega t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) \\ \Upsilon_{\alpha t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) \\ \Upsilon_{\beta t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) \end{pmatrix} = \begin{pmatrix} \left(I_{d^2} - (A_0^*)^{\otimes 2} - (B_0^*)^{\otimes 2} \right)^{-1} \left(I_{d^2} - (B_0^*)^{\otimes 2} \right) \left(H_{0t}^{1/2} \right)^{\otimes 2} \\ \frac{1}{2} \left[\sum_{i=0}^{\infty} ((B_0^*)^{\otimes 2})^i M_{t-1-i}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) \right]' \left(H_{0t}^{-1/2} \right)^{\otimes 2} \\ \frac{1}{2} \left[\sum_{i=0}^{\infty} ((B_0^*)^{\otimes 2})^i \tilde{M}_{t-1-i}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) \right]' \left(H_{0t}^{-1/2} \right)^{\otimes 2} \end{pmatrix}, \tag{A.25}$$

with

$$\begin{aligned} M_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) &= \left[A_0^*(X_t X_t' - \Omega_0) \otimes I_d \right] + \left[I_d \otimes A_0^*(X_t X_t' - \Omega_0) \right] C_{dd}, \\ \tilde{M}_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) &= \left[B_0^*(H_{0t} - \Omega_0) \otimes I_d \right] + \left[I_d \otimes B_0^*(H_{0t} - \Omega_0) \right] C_{dd}. \end{aligned} \tag{A.26}$$

We show that:

$$\Upsilon_t^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) = R_0^{-1}\Upsilon_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0), \tag{A.27}$$

Noting that:

$$I_{d^2} - (A_0^*)^{\otimes 2} - (B_0^*)^{\otimes 2} = (\Omega_0^{1/2})^{\otimes 2}\left[I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right](\Omega_0^{-1/2})^{\otimes 2},$$

$$I_{d^2} - (B_0^*)^{\otimes 2} = (\Omega_0^{1/2})^{\otimes 2}[I_{d^2} - B_0^{\otimes 2}](\Omega_0^{-1/2})^{\otimes 2},$$

we can verify that $\Upsilon_{\omega t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) = \Upsilon_{\omega t}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)$.

For $\Upsilon_{\alpha t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*)$ and $\Upsilon_{\beta t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*)$, we obtain:

$$\left[(B_0^*)^{\otimes 2}\right]^i = \left[(\Omega_0^{1/2}B_0\Omega_0^{-1/2})^{\otimes 2}\right]^i = \left[(\Omega_0^{1/2})^{\otimes 2}\left(B_0^{\otimes 2}\right)(\Omega_0^{-1/2})^{\otimes 2}\right]^i = (\Omega_0^{1/2})^{\otimes 2}\left(B_0^{\otimes 2}\right)^i(\Omega_0^{-1/2})^{\otimes 2}.$$

By 9.3.2(5)(a) of Lütkephol (1996), $(\Omega_0^{-1/2} \otimes \Omega_0^{1/2})C_{dd} = C_{dd}(\Omega_0^{1/2} \otimes \Omega_0^{-1/2})$. Hence

$$M_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) = (\Omega_0^{1/2})^{\otimes 2}N_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)(\Omega_0^{1/2} \otimes \Omega_0^{-1/2}),$$

$$\tilde{M}_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) = (\Omega_0^{1/2})^{\otimes 2}\tilde{N}_t(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)(\Omega_0^{1/2} \otimes \Omega_0^{-1/2}).$$

Combining these two results, we show that $\Upsilon_{\alpha t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) = (\Omega_0^{1/2}\otimes\Omega_0^{-1/2})\Upsilon_{\alpha t}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)$ and $\Upsilon_{\alpha t}^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*) = (\Omega_0^{1/2} \otimes \Omega_0^{-1/2})\Upsilon_{\beta t}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)$. Hence, (A.27) holds.

Define:

$$\Gamma_0^* = E\left[\Upsilon_t^*(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*)\text{vec}\left(Z_tZ_t' - I_d\right)\left(\text{vec}\left(Z_tZ_t' - I_d\right)\right)'\Upsilon_t^{*\prime}(\boldsymbol{\omega}_0, \boldsymbol{\lambda}_0^*)\right], \tag{A.28}$$

from which we obtain $\Gamma_0^* = R_0^{-1}\Gamma_0R_0^{-1}$. Combined with (A.24), it follows that $R_0Q_0\Gamma_0Q_0'R_0 = Q_0^*\Gamma_0^*Q_0^{*\prime}$. $\square$

Table 1: Finite Sample Properties of 2sQML Estimator for the RBEKK-ARCH Model

	DGP1				DGP2			
Parameters	True	Mean	Std. Dev.	RMSE	True	Mean	Std. Dev.	RMSE
Ω_{11}	1.00	0.9998	0.1085	0.1085	0.640	0.6413	0.0725	0.0725
Ω_{21}	0.54	0.5391	0.0671	0.0671	−0.264	−0.2650	0.0383	0.0383
Ω_{22}	0.81	0.8090	0.0662	0.0662	1.210	1.2093	0.0843	0.0843
A_{11}	0.60	0.5882	0.0642	0.0652	0.600	0.5892	0.0675	0.0683
A_{21}	0.00	0.0018	0.0614	0.0614	0.000	−0.0004	0.0623	0.0623
A_{12}	0.00	0.0007	0.0622	0.0622	0.000	−0.0003	0.0617	0.0617
A_{22}	0.40	0.3925	0.0702	0.0706	−0.300	−0.2988	0.0741	0.0741

Table 2: Finite Sample Properties of 2sQML Estimator for the BEKK-ARCH Model

	DGP1				DGP2			
Parameters	True	Mean	Std. Dev.	RMSE	True	Mean	Std. Dev.	RMSE
C^*_{11}	0.6579	0.6561	0.0577	0.0577	0.4149	0.4143	0.0383	0.0383
C^*_{21}	0.3964	0.3934	0.0471	0.0472	−0.2104	−0.2091	0.0438	0.0438
C^*_{22}	0.6625	0.6568	0.0527	0.0530	1.0958	1.0836	0.0812	0.0821
A^*_{11}	0.6249	0.6129	0.0784	0.0793	0.6212	0.6108	0.0709	0.0716
A^*_{21}	0.0706	0.0703	0.0724	0.0724	−0.1644	−0.1634	0.0970	0.0970
A^*_{12}	−0.0794	−0.0777	0.0845	0.0845	0.1187	0.1175	0.0484	0.0484
A^*_{22}	0.3751	0.3678	0.0859	0.0862	−0.3212	−0.3204	0.0771	0.0771

Figure 1: Comparison of Diagonal Specifications for the BEKK and RBEKK Models: DGP3

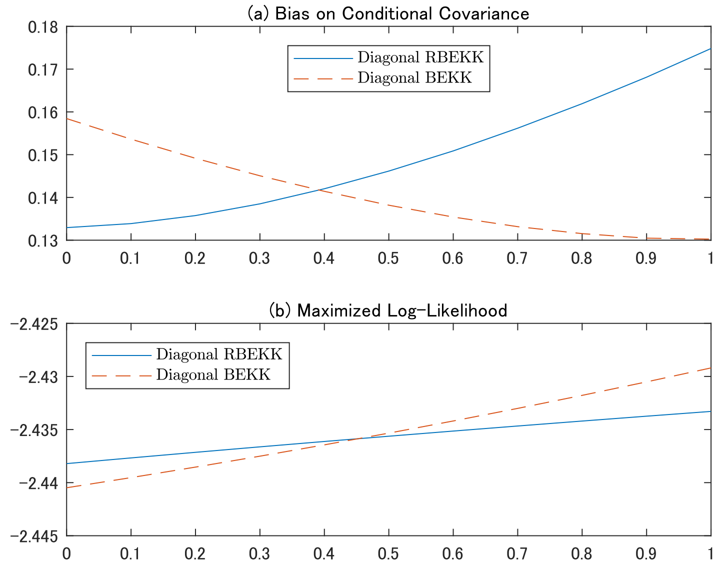

Figure 2: Comparison of Diagonal Specifications for the BEKK and RBEKK Models: DGP4