Supplementary Appendix for Fitting Vast Dimensional Time-Varying Covariance Models

Cavit Pakel
Department of Economics, Bilkent University
and
Neil Shephard
Department of Economics, Harvard University
& Department of Statistics, Harvard University
and
Kevin Sheppard
Department of Economics, University of Oxford
and
Robert F. Engle
Stern Business School, New York University

October 4, 2019

A Notation and some results

In what follows, K is some generic finite term (which can be scalar or matrix-valued) and ϕ is some generic scalar such that $0 < \phi < 1$; the values of both terms may change from line to line. Convergence to zero in probability and convergence to zero almost surely are denoted by $o_p(1)$ and $o_{a.s.}(1)$, respectively. Given the parameter vectors γ_j and λ ($j = 1, ..., N$) we define $\theta_j = (\gamma_j', \lambda')'$, $\gamma = (\gamma_1', ..., \gamma_N')'$, $\gamma_0 = (\gamma_{10}', ..., \gamma_{N0}')'$ and $\hat{\gamma} = (\hat{\gamma}_1', ..., \hat{\gamma}_N')'$. We similarly define $\theta = (\gamma', \lambda')'$, $\theta_0 = (\gamma_0', \lambda_0')'$ and $\hat{\theta} = (\hat{\gamma}', \hat{\lambda}')'$. Whenever a variable is generated based on an arbitrarily chosen initial value, we denote it in the subscript; e.g. $H_{jt,h}(\gamma_j, \lambda)$ is based on the initial value h_j. We also use the following shorthand notation:

$$\sup_{\lambda} \equiv \sup_{\lambda \in \Theta_\lambda}, \qquad \sup_{\theta_j} \equiv \sup_{\gamma_j \in \Theta_\gamma, \lambda \in \Theta_\lambda} \qquad \text{and} \qquad \sup_{\theta} \equiv \sup_{\gamma_1 \in \Theta_\gamma, ..., \gamma_N \in \Theta_\gamma, \lambda \in \Theta_\lambda}.$$

Using this notation, the composite likelihood function based on arbitrarily chosen initial values h_j is given by

$$\begin{aligned}
l_{NT,h}(\gamma_1, ..., \gamma_N, \lambda) &= \frac{1}{N} \sum_{j=1}^{N} l_{jT,h}(\gamma_j, \lambda), \\
l_{jT,h}(\gamma_j, \lambda) &= \frac{1}{T} \sum_{t=1}^{T} l_{jt,h}(\gamma_j, \lambda), \\
l_{jt,h}(\gamma_j, \lambda) &= \log(\det(H_{jt,h}(\gamma_j, \lambda))) + tr(X_{jt} X_{jt}' H_{jt,h}^{-1}(\gamma_j, \lambda)).
\end{aligned}$$

The composite likelihood based on the stationary solution $H_{jt}(\gamma_j, \lambda)$ is given by

$$\begin{aligned}
l_{NT}(\gamma_1, ..., \gamma_N, \lambda) &= \frac{1}{N} \sum_{j=1}^{N} l_{jT}(\gamma_j, \lambda) \\
l_{jT}(\gamma_j, \lambda) &= \frac{1}{T} \sum_{t=1}^{T} l_{jt}(\gamma_j, \lambda), \\
l_{jt}(\gamma_j, \lambda) &= \log(\det(H_{jt}(\gamma_j, \lambda))) + tr(X_{jt} X_{jt}' H_{jt}^{-1}(\gamma_j, \lambda)). \qquad \text{(A.1)}
\end{aligned}$$

In the following, we will make use of several matrix algebra results. Firstly, if A is

$(m \times n)$ and B is $(n \times m)$ then, by result 4.1.1(8b) of Lütkepohl (1996),

$$tr\,(AB) = vec\,(A')'\,vec\,(B)\,.$$

By this result and the definition of Euclidian norm, we can show that for any symmetric A

$$||vec\,(A)|| = ||A||\,. \tag{A.2}$$

Again, by the definition of $||\cdot||$, it can be shown that if A is $(m \times 1)$, then

$$||AA'|| = ||A'A|| = ||A||^2\,. \tag{A.3}$$

Moreover, by equations (B.3) and (B.4) of PR, if one or both of A and B are square, then

$$||AB|| \leq ||A||\,||B||\,. \tag{A.4}$$

Also, by equation (B.8) of PR, for an $(n \times n)$ matrix $A \geq 0$ and an $(n \times n)$ matrix $B > 0$,

$$0 < tr\left((A+B)^{-1}\right) \leq tr\left(B^{-1}\right). \tag{A.5}$$

Next, by result 4.1.2(2) of Lütkepohl (1996), if A and B are both $(m \times n)$ then,

$$|tr\,(A'B)| \leq ||A||\,||B||\,. \tag{A.6}$$

By result 2.4.(11b) of Lütkepohl (1996), if A is $(m \times m)$ and B is $(n \times n)$, then

$$tr\,(A \otimes B) = tr\,(A)\,tr\,(B)\,. \tag{A.7}$$

Moreover, by result 2.4(5) of Lütkepohl (1996) if A is $(m \times n)$, B is $(p \times q)$, C is $(n \times r)$ and D is $(q \times s)$, then

$$(A \otimes B)\,(C \otimes D) = AC \otimes BD. \tag{A.8}$$

Finally, by result 2.2.(16) of Lütkepohl (1996), if A is $(m \times n)$, B is $(n \times p)$ and C is $(p \times q)$, then

$$vec\,(ABC) = (C' \otimes A)\, vec\,(B)\,.$$

B Scalar BEKK

In this part, we prove Theorem 4.1. Section B.1 provides the definitions that are used throughout Section B, Section B.2 proves consistency and Section B.3 deals with asymptotic normality. Section B.4 provides the formula for the asymptotic variance of $\hat{\lambda}$. Finally, the required lemmas are stated and proved in Section B.5.

B.1 Definitions

Throughout Section B we have

$$\begin{aligned} H_{jt}(\gamma_j, \lambda) &= \Gamma_j\,(1-\alpha-\beta) + \alpha X_{j,t-1}X'_{j,t-1} + \beta H_{j,t-1}(\gamma_j, \lambda), \\ H_{jt,h}(\gamma_j, \lambda) &= \Gamma_j\,(1-\alpha-\beta) + \alpha X_{j,t-1}X'_{j,t-1} + \beta H_{j,t-1,h}(\gamma_j, \lambda), \end{aligned}$$

where $\gamma_j = vec(\Gamma_j)$, $\lambda = (\alpha, \beta)'$, and $H_{j0,h}(\gamma_j, \lambda) = h_j > 0$. Here h_j is some fixed initial value.

We also use the following notation: first, we let $V_{jt} = vec\left(Z_{jt}Z'_{jt} - I_2\right)$ and define the $(4N \times 1)$ vector $V_t^N = (V'_{1t}, ..., V'_{Nt})'$. Next, we define Q_t^N as a $(4N \times 4N)$ block diagonal matrix where the j^{th} (4×4) diagonal block is given by $Q_{jt} = D(H_{jt}^{1/2})^{\otimes 2}$ with $D = (1-\alpha_0-\beta_0)^{-1}(1-\beta_0)$. We furthermore define

$$\begin{aligned} W_{jt} &= \left[-\sum_{i=0}^{\infty} \beta_0^i vec(X_{j,t-1-i}X'_{j,t-1-i} - \Gamma_{j0})\right]' \left(H_{jt}^{-1/2}(\gamma_{j0}, \lambda_0)\right)^{\otimes 2}, \\ \tilde{W}_{jt} &= \left[-\sum_{i=0}^{\infty} \beta_0^i vec(H_{j,t-1-i}\,(\gamma_{j0}, \lambda_0) - \Gamma_{j0})\right]' \left(H_{jt}^{-1/2}(\gamma_{j0}, \lambda_0)\right)^{\otimes 2}, \end{aligned}$$

and collect these objects for all $j = 1, ..., N$ in the matrices $W_t^N = N^{-1}\,(W_{1t}, W_{2t}, ..., W_{Nt})$ and $\tilde{W}_t^N = N^{-1}(\tilde{W}_{1t}, \tilde{W}_{2t}, ..., \tilde{W}_{Nt})$. Finally, for V_t^N, W_t^N, $\tilde{W}_t^N$ and Q_t^N defined as such, we let

$$\Omega_0 = E\left[((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')' V_t^N (V_t^N)' ((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')\right].$$

B.2 Consistency

By Lemma B.1 and the triangle inequality

$$\sup_{\lambda} |l_{NT}(\gamma_0, \lambda) - l_{NT,h}(\hat{\gamma}, \lambda)| \leq \frac{1}{N} \sum_{j=1}^{N} \sup_{\lambda} |l_{jT}(\gamma_{j0}, \lambda) - l_{jT,h}(\hat{\gamma}_j, \lambda)| = o_{a.s.}(1), \quad \text{(B.1)}$$

as $T \to \infty$. Similarly, by Lemma B.2

$$\sup_{\theta} |l_{NT}(\gamma, \lambda) - E\left[l_{Nt}(\gamma, \lambda)\right]| \leq \frac{1}{N} \sum_{j=1}^{N} \sup_{\theta_j} |l_{jT}(\gamma_j, \lambda) - E\left[l_{jt}(\gamma_j, \lambda)\right]| = o_{a.s.}(1), \quad \text{(B.2)}$$

as $T \to \infty$. Note that in both (B.1) and (B.2) we are able to retain the rate $o_{a.s.}(1)$ from Lemmas B.1 and B.2 after averaging across j, since N is fixed. Finally, by Lemma B.3

$$E[l_{Nt}(\gamma_0, \lambda)] > E[l_{Nt}(\gamma_0, \lambda_0)] \quad \text{for any } \lambda \neq \lambda_0. \quad \text{(B.3)}$$

Then, by the same line of arguments as in the Proof of Theorem 4.1 of PR, for any $\varepsilon > 0$ we have almost surely for sufficiently large T that

$$\begin{aligned} E[l_{Nt}(\gamma_0, \hat{\lambda})] &< l_{NT}(\gamma_0, \hat{\lambda}) + \varepsilon/5 \\ &< l_{NT,h}(\hat{\gamma}, \hat{\lambda}) + 2\varepsilon/5 \\ &< l_{NT,h}(\hat{\gamma}, \lambda_0) + 3\varepsilon/5 \\ &< l_{NT}(\gamma_0, \lambda_0) + 4\varepsilon/5 \\ &< E\left[l_{Nt}(\gamma_0, \lambda_0)\right] + \varepsilon, \end{aligned} \quad \text{(B.4)}$$

where the first and last inequalities follow from (B.2), the second and fourth inequalities follow from (B.1), and the third inequality follows from the definition of $\hat{\lambda}$. By standard arguments, it follows from (B.4) and the identification condition (B.3) that $\hat{\lambda} - \lambda_0 \overset{a.s.}{\to} 0$ as

$T \to \infty$. Notice moreover that, by the same arguments as those leading to equation (A.6) of PR, Assumptions 4.3 and 4.6 are sufficient to obtain $\hat{\gamma}_j - \gamma_0 \overset{a.s.}{\to} 0$ for all j as $T \to \infty$. The desired result follows.

B.3 Asymptotic normality

We proceed along the same lines as in the Proof of Theorem 4.2 of PR. First, by a mean value expansion we obtain

$$\begin{aligned} 0 &= \frac{\partial l_{NT,h}(\gamma_0, \lambda_0)}{\partial \lambda} + K_{NT,h}(\theta^*)\left((\hat{\gamma}_1 - \gamma_{10})', ..., (\hat{\gamma}_N - \gamma_{N0})'\right)' \\ &\quad + J_{NT,h}(\theta^*)(\hat{\lambda} - \lambda_0), \end{aligned} \tag{B.5}$$

where θ^* is some mean value between $\hat{\theta}$ and θ_0,

$$K_{NT,h}(\theta^*) = \frac{\partial^2 l_{NT,h}(\theta^*)}{\partial \lambda \partial \gamma'} \quad \text{and} \quad J_{NT,h}(\theta^*) = \frac{\partial^2 l_{NT,h}(\theta^*)}{\partial \lambda \partial \lambda'}.$$

Let

$$\begin{aligned} K_{NT}(\theta^*) &= \frac{\partial^2 l_{NT}(\theta^*)}{\partial \lambda \partial \gamma'}, \quad J_{NT}(\theta^*) = \frac{\partial^2 l_{NT}(\theta^*)}{\partial \lambda \partial \lambda'}, \\ K_N(\theta_0) &= E\left[\frac{\partial^2 l_{Nt}(\theta_0)}{\partial \lambda \partial \gamma'}\right] \quad \text{and} \quad J_N(\theta_0) = E\left[\frac{\partial^2 l_{Nt}(\theta_0)}{\partial \lambda \partial \lambda'}\right]. \end{aligned} \tag{B.6}$$

Note that while the above terms depend on N, N plays no role in the asymptotic analysis since it remains fixed as $T \to \infty$.

Now, following the same arguments as in the Proof of Theorem 4.2 of PR, by Lemmas B.7 and B.8, and $(\hat{\theta} - \theta_0) = o_{a.s.}(1)$ as proved in Section B.2 we have that $J_{NT}(\theta^*)$ is invertible with probability approaching one. Moreover, by Lemmas B.11, B.12 and B.13, the expansion in (B.5) yields

$$\begin{aligned} 0 &= \sqrt{T}\frac{\partial l_{NT}(\gamma_0, \lambda_0)}{\partial \lambda} + K_{NT}(\theta^*)\sqrt{T}\left((\hat{\gamma}_1 - \gamma_{10})', ..., (\hat{\gamma}_N - \gamma_{N0})'\right)' \\ &\quad + J_{NT}(\theta^*)\sqrt{T}(\hat{\lambda} - \lambda_0) + o_p(1), \end{aligned}$$

and rearranging this yields

$$\sqrt{T}\begin{bmatrix}\hat{\gamma}-\gamma_0\\ \hat{\lambda}-\lambda_0\end{bmatrix}=\begin{bmatrix}I_{4N} & 0_{4N\times 2}\\ -J_{NT}^{-1}(\theta^*)K_{NT}(\theta^*) & -J_{NT}^{-1}(\theta^*)\end{bmatrix}\sqrt{T}\begin{bmatrix}\hat{\gamma}-\gamma_0\\ \frac{\partial l_{NT}(\gamma_0,\lambda_0)}{\partial\lambda}\end{bmatrix}+o_p\left(1\right). \quad \text{(B.7)}$$

By Lemma B.7 and the consistency of $\hat{\theta}$,

$$\begin{bmatrix}I_{4N} & 0_{4N\times 2}\\ -J_{NT}^{-1}(\theta^*)K_{NT}(\theta^*) & -J_{NT}^{-1}(\theta^*)\end{bmatrix}\xrightarrow{p}\begin{bmatrix}I_{4N} & 0_{4N\times 2}\\ -J_{N}^{-1}(\theta_0)K_{N}(\theta_0) & -J_{N}^{-1}(\theta_0)\end{bmatrix} \quad \text{(B.8)}$$

as $T\to\infty$. Combining (B.7) and (B.8), and using Lemma B.11, we finally obtain

$$\sqrt{T}\begin{bmatrix}\hat{\gamma}-\gamma_0\\ \hat{\lambda}-\lambda_0\end{bmatrix}\xrightarrow{d}\begin{bmatrix}I_{4N} & 0_{4N\times 2}\\ -J_{N}^{-1}(\theta_0)K_{N}(\theta_0) & -J_{N}^{-1}(\theta_0)\end{bmatrix}N\left(0,\Omega_0\right),$$

where Ω_0 is as defined in Section B.1.

B.4 Asymptotic variance of $\hat{\lambda}$

In this part, we calculate the asymptotic variance of $\hat{\lambda}-\lambda_0$ for the pairwise scalar-BEKK composite likelihood estimator of Theorem 4.1. First, we partition the asymptotic variance matrix Ω_0 of Section B.1. Using the same definitions as in Sections B.1 and B.3, let

$$\underset{4N\times 4N}{A}=E\left[Q_t^N V_t^N (V_t^N)' Q_t^{N\prime}\right],\quad \underset{4N\times 2}{B}=E\left[Q_t^N V_t^N (V_t^N)' W_t^{N\prime}\quad Q_t^N V_t^N (V_t^N)' \tilde{W}_t^{N\prime}\right],$$
$$\underset{2\times 4N}{C}=E\begin{bmatrix}W_t^N V_t^N (V_t^N)' Q_t^{N\prime}\\ \tilde{W}_t^N V_t^N (V_t^N)' Q_t^{N\prime}\end{bmatrix},\quad \text{and}\quad \underset{2\times 2}{D}=E\begin{bmatrix}W_t^N V_t^N (V_t^N)' W_t^{N\prime} & W_t^N V_t^N (V_t^N)' \tilde{W}_t^{N\prime}\\ \tilde{W}_t^N V_t^N (V_t^N)' W_t^{N\prime} & \tilde{W}_t^N V_t^N (V_t^N)' \tilde{W}_t^{N\prime}\end{bmatrix},$$

which yields

$$\Omega_0=\begin{bmatrix}\underset{4N\times 4N}{A} & \underset{4N\times 2}{B}\\ \underset{2\times 4N}{C} & \underset{2\times 2}{D}\end{bmatrix}.$$

Now, by Theorem 4.1, the asymptotic variance for $\hat{\theta}$ is given by

$$\begin{aligned}
& \begin{bmatrix} I_{4N} & 0_{4N\times 2} \\ -J_N^{-1}(\theta_0) K_N(\theta_0) & -J_N^{-1}(\theta_0) \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I_{4N} & -K_N(\theta_0)' J_N^{-1}(\theta_0) \\ 0_{2\times 4N} & -J_N^{-1}(\theta_0) \end{bmatrix} \\
= & \begin{bmatrix} I_{4N} & 0_{4N\times 2} \\ -J_N^{-1}(\theta_0) K_N(\theta_0) & -J_N^{-1}(\theta_0) \end{bmatrix} \begin{bmatrix} A & -AK_N(\theta_0)' J_N^{-1}(\theta_0) - BJ_N^{-1}(\theta_0) \\ C & -CK_N(\theta_0)' J_N^{-1}(\theta_0) - DJ_N^{-1}(\theta_0) \end{bmatrix} \\
= & \begin{bmatrix} A & -AK_N(\theta_0)' J_N^{-1}(\theta_0) - BJ_N^{-1}(\theta_0) \\ -J_N^{-1}(\theta_0) K_N(\theta_0) A - J_N^{-1}(\theta_0) C & J_N^{-1}(\theta_0) Z J_N^{-1}(\theta_0) \end{bmatrix}
\end{aligned}$$

where

$$Z = K_N(\theta_0) AK_N(\theta_0)' + K_N(\theta_0) B + CK_N(\theta_0)' + D.$$

Remember that $\hat{\lambda}$ corresponds to the bottom (2×1) block of $\hat{\theta}$. Hence, the asymptotic variance of $\hat{\lambda}$ is given by the bottom-right (2×2) block of the asymptotic variance matrix for $\hat{\theta}$. Therefore,

$$Var(\hat{\lambda} - \lambda_0) = J_N^{-1}(\theta_0) Z J_N^{-1}(\theta_0).$$

B.5 Lemmas

Remark 1 *Throughout Section B.5 we assume that Assumptions 4.1-4.7 hold.*

Lemma B.1 $\sup_\lambda |l_{jT}(\gamma_{j0}, \lambda) - l_{jT,h}(\hat{\gamma}_j, \lambda)| = o_{a.s.}(1)$ *as* $T \to \infty$, *for any* $j = 1, ..., N$.

Proof of Lemma B.1. The proof can be obtained by a straightforward modification of the Proof of Lemma B.1 of PR, replacing A and B by $\sqrt{\alpha}$ and $\sqrt{\beta}$, respectively. In particular, given Assumption 4.4 and $\hat{\gamma}_j \overset{a.s.}{\to} \gamma_{j0}$, by the same arguments as in PR we obtain in our case

$$\sup_\lambda ||H_{jt}^{-1}(\gamma_{j0}, \lambda)|| \leq \sup_{\theta_j} ||H_{jt}^{-1}(\gamma_j, \lambda)|| \leq K, \text{ and } \sup_\lambda ||H_{jt,h}^{-1}(\hat{\gamma}_j, \lambda)|| \leq \sup_{\theta_j} ||H_{jt,h}^{-1}(\gamma_j, \lambda)|| \leq K; \tag{B.9}$$

see the arguments leading to equations (B.12) and (B.13) of PR. Notice that all the above bounds hold independent of t. Now, in the scalar case equation (B.14) of PR becomes

$$vec(H_{jt}(\gamma_{j0},\lambda))-vec(H_{jt,h}(\hat{\gamma}_j,\lambda))=\sum_{i=0}^{t-1}\beta^i\,(1-\alpha-\beta)\,(\gamma_{j0}-\hat{\gamma}_j)+\beta^t vec(H_{j0}(\gamma_{j0},\lambda)-h_j).$$

We note that by Assumption 4.2, $\sup_\lambda|\beta^i|\leq\phi^i$ and $\sup_\lambda|\,(1-\alpha-\beta)\,(1-\beta)^{-1}\,|\leq K$. Moreover, by Assumptions 4.3 and 4.6, the ergodic theorem yields that $||\hat{\gamma}_j-\gamma_{j0}||=o_{a.s.}(1)$ for all j as $T\rightarrow\infty$. In addition, by Assumption 4.4 compactness of Θ_λ is guaranteed. Therefore, by the same arguments as those leading to equation (B.16) of PR we obtain that as $T\rightarrow\infty$

$$\sup_\lambda||vec(H_{jt}(\gamma_{j0},\lambda))-vec(H_{jt,h}(\hat{\gamma}_j,\lambda))||\leq K\phi^t+o_{a.s.}(1). \tag{B.10}$$

Next, we obtain a series of results in (B.11), (B.12) and (B.13). In particular,

$$\begin{aligned}
&\sup_\lambda|l_{jT}(\gamma_{j0},\lambda)-l_{jT,h}(\hat{\gamma}_j,\lambda)|\\
=\;&\frac{1}{T}\sum_{t=1}^{T}\sup_\lambda\left|\log\left(\frac{\det(H_{jt}(\gamma_{j0},\lambda))}{\det(H_{jt,h}(\hat{\gamma}_j,\lambda))}\right)\right|\\
&+\frac{1}{T}\sum_{t=1}^{T}\sup_\lambda\left|tr\left[X_{jt}X_{jt}'\left(H_{jt}^{-1}(\gamma_{j0},\lambda)-H_{jt,h}^{-1}(\hat{\gamma}_j,\lambda)\right)\right]\right|,
\end{aligned}\tag{B.11}$$

which corresponds to equation (B.11) of PR. Similarly, given that $\sup_\lambda\left|\left|H_{jt,h}^{-1}(\hat{\gamma}_j,\lambda)\right|\right|\leq K$ and $\sup_\lambda\left|\left|H_{jt}^{-1}(\gamma_{j0},\lambda)\right|\right|\leq K$ by (B.9), we obtain

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\sup_\lambda\left|\log\left(\frac{\det(H_{jt}(\gamma_{j0},\lambda))}{\det(H_{jt,h}(\hat{\gamma}_j,\lambda))}\right)\right|\\
\leq\;&\frac{2}{T}\sum_{t=1}^{T}\sup_\lambda||(H_{jt}(\gamma_{j0},\lambda)-H_{jt,h}(\hat{\gamma}_j,\lambda))||\times\sup_\lambda\left|\left|H_{jt,h}^{-1}(\hat{\gamma}_j,\lambda)\right|\right|\\
\leq\;&\frac{K}{T}\sum_{t=1}^{T}\sup_\lambda||(H_{jt}(\gamma_{j0},\lambda)-H_{jt,h}(\hat{\gamma}_j,\lambda))||,
\end{aligned}\tag{B.12}$$

and

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\sup_{\lambda}\left|tr\left[X_{jt}X_{jt}'\left(H_{jt}^{-1}(\gamma_{j0},\lambda)-H_{jt,h}^{-1}(\hat{\gamma}_j,\lambda)\right)\right]\right| \\
\leq \; & \frac{K}{T}\sum_{t=1}^{T}\sup_{\lambda}||(H_{jt}(\gamma_{j0},\lambda)-H_{jt,h}(\hat{\gamma}_j,\lambda))||\times||X_{jt}||^2\,; \qquad \text{(B.13)}
\end{aligned}
$$

see the derivations for the corresponding terms on pages 43-44 of PR. Substituting (B.12) and (B.13) in (B.11), and then using (B.10) yields

$$
\sup_{\lambda}|l_{jT}(\gamma_{j0},\lambda)-l_{jT,h}(\hat{\gamma}_j,\lambda)|\leq K\frac{1}{T}\sum_{t=1}^{T}\phi^t+\frac{K}{T}\sum_{t=1}^{T}\phi^t\,||X_{jt}||^2+o_{a.s.}(1),
$$

as on page 44 of PR. Now, given that $0<\phi<1$, $T^{-1}\sum_{t=1}^{T}\phi^t\to 0$ as $T\to\infty$. Moreover, by the same arguments as in PR (see the last equation in their Proof of Lemma B.1), $T^{-1}\sum_{t=1}^{T}\phi^t\,||X_{jt}||^2\overset{a.s.}{\to}0$ for all j as $T\to\infty$. Hence, $\sup_{\lambda}|l_{jT}(\gamma_{j0},\lambda)-l_{jT,h}(\hat{\gamma}_j,\lambda)|=o_{a.s.}(1)$ for all j as $T\to\infty$, as stated. ■

Lemma B.2 *For any* $j=1,...,N$,

$$
\sup_{\theta_j}|l_{jT}(\gamma_j,\lambda)-E\left[l_{jt}(\gamma_j,\lambda)\right]|\overset{a.s.}{\to}0,\;\; \textit{as } T\to\infty.
$$

Proof of Lemma B.2. The result follows by the same arguments as in the Proof of Lemma B.2 of PR, applied to each pair $j=1,...,N$. ■

Lemma B.3 $E[l_{jt}(\gamma_{j0},\lambda)]>E[l_{jt}(\gamma_{j0},\lambda_0)]$ *for any* $\lambda\neq\lambda_0$. *In addition,* $E\left[|l_{jt}\left(\gamma_{j0},\lambda_0\right)|\right]<\infty$.

Proof of Lemma B.3. That $E\left[|l_{jt}\left(\gamma_{j0},\lambda_0\right)|\right]<\infty$ holds is implied by Lemma B.4. As for the first result, the arguments made in the Proof of Lemma B.3 in PR are directly applicable to our case. This is because the arguments made there are independent of the underlying particular volatility model. Indeed, the 2-dimensional case we consider here is a special case of Lemma B.3 of PR which is concerned with a d-dimensional problem. Apart from that the only change is the addition of the pair index j.

In particular, let $\xi_{1,jt}$ and $\xi_{2,jt}$ be the eigenvalues of $H_{jt}(\gamma_{j0},\lambda_0)H_{jt}^{-1}(\gamma_{j0},\lambda)$ for some particular j and t. Suppose that $\lambda \neq \lambda_0$. The same arguments as in the Proof of Lemma B.3 in PR yield

$$E\left[tr\left(X_{jt}X_{jt}'\left(H_{jt}^{-1}(\gamma_{j0},\lambda)-H_{jt}^{-1}(\gamma_{j0},\lambda_0)\right)\right)\right] = E\left[\sum_{i=1}^{2}(\xi_{i,jt}-1)\right], \tag{B.14}$$

$$\log\det\left(H_{jt}(\gamma_{j0},\lambda)H_{jt}^{-1}(\gamma_{j0},\lambda_0)\right) = -\sum_{i=1}^{2}\log\xi_{i,jt}, \tag{B.15}$$

where (B.14) follows since Z_{jt} is independent of the information set $\mathcal{F}_{j,t-1} = \sigma\left(X_{j,t-1}, X_{j,t-2}, ...\right)$. Using the definition of $l_{jt}(\gamma_j,\lambda)$ in (A.1) with (B.14) and (B.15) yields

$$E\left[l_{jt}(\gamma_{j0},\lambda)-l_{jt}(\gamma_{j0},\lambda_0)\right] = E\left[\sum_{i=1}^{2}\left(\xi_{i,jt}-1-\log\xi_{i,jt}\right)\right] \geq 0,$$

since the eigenvalues $\xi_{1,jt}$ and $\xi_{2,jt}$ are both positive. The inequality will be strict unless $\xi_{1,jt} = \xi_{2,jt} = 1$ almost surely, which is equivalent to $H_{jt}(\gamma_{j0},\lambda) = H_{jt}(\gamma_{j0},\lambda_0)$ being true almost surely. As Assumption 4.5 rules this out, we conclude that $E[l_{jt}(\gamma_{j0},\lambda)] > E[l_{jt}(\gamma_{j0},\lambda_0)]$ for any $\lambda \neq \lambda_0$. ■

Lemma B.4 *$E[\sup_{\theta_j}|l_{jt}(\gamma_j,\lambda)|] < \infty$ for all j.*

Proof of Lemma B.4. By the definition of $l_{jt}(\gamma_j,\lambda)$ in (A.1),

$$E\left[\sup_{\theta_j}|l_{jt}(\gamma_j,\lambda)|\right] \leq \sqrt{2}E\left[\sup_{\theta_j}||H_{jt}(\gamma_j,\lambda)||\right] + \sqrt{2}E\left[||X_{jt}||^2\sup_{\theta_j}\left|\left|H_{jt}^{-1}(\gamma_j,\lambda)\right|\right|\right];$$

see also the last display in the Proof of Lemma B.4 of PR. Now, we know that $\sup_{\theta_j}||H_{jt}^{-1}(\gamma_j,\lambda)|| \leq K$ independent of j and t by (B.9). Moreover, $E[\sup_{\theta_j}||H_{jt}(\gamma_j,\lambda)||] < \infty$ by Lemma B.5. Lastly, $E[||X_{jt}||^2] < \infty$ by Assumption 4.6. The stated result follows. ■

Lemma B.5 *Let θ_i be the i^{th} entry of θ, where $i = 1, ..., 4N+2$. Then*

$$E\left[\sup_{\theta_j}\left|\left|\frac{\partial H_{jt,h}(\gamma_j,\lambda)}{\partial\theta_i}\right|\right|^3\right] < \infty, \quad E\left[\sup_{\theta_j}\left|\left|\frac{\partial^2 H_{jt,h}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}}\right|\right|^3\right] < \infty,$$

$$E\left[\sup_{\theta_j}||vec(H_{jt}(\gamma_j,\lambda))||^3\right] < \infty, \quad E\left[\sup_{\theta_j}\left|\left|\frac{\partial vec(H_{jt}(\gamma_j,\lambda))}{\partial\theta_i}\right|\right|^3\right] < \infty,$$

$$E\left[\sup_{\theta_j}\left|\left|\frac{\partial^2 vec(H_{jt}(\gamma_j,\lambda))}{\partial\theta_i\partial\theta_{i'}}\right|\right|^3\right]<\infty,$$

for all j,i and i'.

Proof of Lemma B.5. We consider the proof of $E[\sup_{\theta_j}||\partial H_{jt,h}(\gamma_j,\lambda)/\partial\theta_i||^3]<\infty$. First note that by a standard recursion argument,

$$\begin{aligned} vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right) &= (1-\alpha-\beta)\,\gamma_j+\alpha vec\left(X_{j,t-1}X'_{j,t-1}\right)+\beta vec\left(H_{j,t-1}(\gamma_j,\lambda)\right) \\ &= \frac{\gamma_j\left(1-\alpha-\beta\right)\left(1-\beta^t\right)}{1-\beta} \\ &\quad +\sum_{i=0}^{t-1}\beta^i\alpha vec\left(X_{j,t-1}X'_{j,t-1}\right)+\beta^t vec\left(h_j\right). \end{aligned} \tag{B.16}$$

Using (B.16), we will consider each component of $E[\sup_{\theta_j}||\partial H_{jt,h}\left(\gamma_j,\lambda\right)/\partial\theta_i||^3]$ individually. First, since h_j is a constant, we have

$$\sup_{\theta_j}\left|\left|\frac{\partial vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right)}{\partial\alpha}\right|\right|^3=\sup_{\theta_j}\left|\left|-\frac{\gamma_j\left(1-\beta^t\right)}{1-\beta}+\sum_{i=0}^{t-1}\beta^i vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right)\right|\right|^3.$$

Then, by the triangle and Minkowski's inequalities, and Assumptions 4.2, 4.3 and 4.6

$$\begin{aligned} & E\left[\sup_{\theta_j}\left|\left|\frac{\partial vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right)}{\partial\alpha}\right|\right|^3\right] \\ = & E\left[\sup_{\theta_j}\left|\left|-\frac{\gamma_j\left(1-\beta^t\right)}{1-\beta}+\sum_{i=0}^{t-1}\beta^i vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right)\right|\right|^3\right] \\ \le & E\left[\left(K+\sum_{i=0}^{t-1}\sup_{\theta_j}\left|\left|\beta^i vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right)\right|\right|\right)^3\right] \\ \le & \left(K+\sum_{i=0}^{t-1}\left\{E\left[\left(\sup_{\theta_j}\left|\left|\beta^i vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right)\right|\right|\right)^3\right]\right\}^{1/3}\right)^3 \\ \le & \left(K+\sum_{i=0}^{t-1}\sup_{\theta_j}\beta^i\left\{E\left[||X_{jt}||^6\right]\right\}^{1/3}\right)^3 \\ \le & \left(K+K\sum_{i=0}^{t-1}\phi^i\right)^3 \end{aligned}$$

$$< \quad \infty, \tag{B.17}$$

for any j ane t. Next, letting $A_t = \frac{t\beta^{t-1}(1-\alpha-\beta)(1-\beta)+\alpha\left(1-\beta^t\right)}{(1-\beta)^2}$,

$$\sup_{\theta_j}\left|\left|\frac{\partial vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right)}{\partial\beta}\right|\right|^3 = \sup_{\theta_j}\left|\left|-\gamma_j A_t + \sum_{i=1}^{t-1} i\beta^{i-1}\alpha vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right) + t\beta^{t-1}vec\left(h_j\right)\right|\right|^3.$$

Then, by using the same approach as in the derivation of (B.17),

$$\begin{aligned}
& \sup_{\theta_j}\left|\left|\frac{\partial vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right)}{\partial\beta}\right|\right|^3 \\
\leq\ & E\left[\left(K + \sup_{\theta_j}\left|\left|t\beta^{t-1}vec\left(h_j\right)\right|\right| + \sup_{\theta_j}\left|\left|\sum_{i=1}^{t-1} i\beta^{i-1}\alpha vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right)\right|\right|\right)^3\right] \\
\leq\ & \left(K + Kt\phi^{t-1} + \left\{E\left[\left(\sum_{i=1}^{t-1}\sup_{\theta_j}\left|\left|i\beta^{i-1}\alpha vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right)\right|\right|\right)^3\right]\right\}^{1/3}\right)^3 \\
\leq\ & \left(K + Kt\phi^{t-1} + \sum_{i=1}^{t-1}\left\{E\left[\left(\sup_{\theta_j}\left|\left|i\beta^{i-1}\alpha vec\left(X_{j,t-1-i}X'_{j,t-1-i}\right)\right|\right|\right)^3\right]\right\}^{1/3}\right)^3 \\
\leq\ & \left(K + Kt\phi^{t-1} + \sum_{i=1}^{t-1}\left\{E\left[\left(\sup_{\theta_j}\left(i\beta^{i-1}\alpha\right) \times \left|\left|X_{jt}\right|\right|^2\right)^3\right]\right\}^{1/3}\right)^3 \\
\leq\ & \left(K + Kt\phi^{t-1} + \sum_{i=1}^{t-1} i\phi^i\left\{E\left[\left|\left|X_{jt}\right|\right|^6\right]\right\}^{1/3}\right)^3 \\
\leq\ & \left(K + Kt\phi^{t-1} + K\sum_{i=1}^{t-1} i\phi^i\right)^3 \\
<\ & \infty,
\end{aligned} \tag{B.18}$$

for any j and t. Finally

$$E\left[\sup_{\theta_j}\left|\left|\frac{\partial vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right)}{\partial vec\left(\gamma_j\right)'}\right|\right|^3\right] = \left[\sup_{\theta_j}\left|\left|I_4\frac{\left(1-\alpha-\beta\right)\left(1-\beta^t\right)}{1-\beta}\right|\right|^3\right] < \infty, \tag{B.19}$$

by Assumption 4.2. Hence, by (B.17)-(B.19) we have $E\left[\sup_{\theta_j}\left|\left|\partial vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right)/\partial\theta_i\right|\right|^3\right] <$

∞ for any j and t. We note that this conclusion does not change in case θ_i correponds to a parameter which is not included in $(\gamma_j', \lambda')'$; this is because then $\partial vec\left(H_{jt,h}\left(\gamma_j,\lambda\right)\right)/\partial\theta_i$ is automatically equal to zero. The remaining results can be obtained similarly. ■

Lemma B.6 *Let θ_i be the i^{th} entry of θ, where $i = 1, ..., 4N+2$. Then for all j, i and i' we have $E[\sup_{\theta_j} ||\partial^2 l_{jt}\left(\gamma_j,\lambda\right)/\partial\theta_i\partial\theta_{i'}||] < \infty$.*

Proof of Lemma B.6. This is a straightforward modification of Lemma B.5 of PR. Let

$$
\begin{aligned}
A_{jt}(\gamma_j,\lambda) &= tr\left(H_{jt}^{-1}(\gamma_j,\lambda)\frac{\partial^2 H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i\partial\theta_{i'}}\right), \\
B_{jt}(\gamma_j,\lambda) &= tr\left(H_{jt}^{-1}(\gamma_j,\lambda)\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_{i'}}H_{jt}^{-1}(\gamma_j,\lambda)\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i}\right), \\
C_{jt}(\gamma_j,\lambda) &= tr\left(H_{jt}^{-1}(\gamma_j,\lambda)X_{jt}X_{jt}'H_{jt}^{-1}(\gamma_j,\lambda)\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_{i'}}H_{jt}^{-1}(\gamma_j,\lambda)\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i}\right), \\
D_{jt}(\gamma_j,\lambda) &= tr\left(H_{jt}^{-1}(\gamma_j,\lambda)X_{jt}X_{jt}'H_{jt}^{-1}(\gamma_j,\lambda)\frac{\partial^2 H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i\partial\theta_{i'}}\right).
\end{aligned}
$$

Then,

$$
\frac{\partial^2 l_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i\partial\theta_{i'}} = A_{jt}(\gamma_j,\lambda) - B_{jt}(\gamma_j,\lambda) + 2C_{jt}(\gamma_j,\lambda) - D_{jt}(\gamma_j,\lambda), \tag{B.20}
$$

which is the same as equation (B.19) of PR, except for the addition of the pair index j. Now, using (A.3), (A.4), (A.6), (B.9) and Hölder's inequality,

$$
E\left[\sup_{\theta_j}||A_{jt}(\gamma_j,\lambda)||\right] \leq KE\left[\sup_{\theta_j}\left|\left|\frac{\partial^2 H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i\partial\theta_{i'}}\right|\right|\right], \tag{B.21}
$$

$$
\begin{aligned}
E\left[\sup_{\theta_j}||B_{jt}(\gamma_j,\lambda)||\right] &\leq K\left\{E\left[\sup_{\theta_j}\left|\left|\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_{i'}}\right|\right|^2\right]\right\}^{1/2} \\
&\quad \times\left\{E\left[\sup_{\theta_j}\left|\left|\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i}\right|\right|^2\right]\right\}^{1/2},
\end{aligned} \tag{B.22}
$$

$$
\begin{aligned}
E\left[\sup_{\theta_j}||C_{jt}(\gamma_j,\lambda)||\right] &\leq K\left\{E\left[||X_{jt}||^6\right]\right\}^{1/3}\left\{E\left[\sup_{\theta_j}\left|\left|\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_{i'}}\right|\right|^3\right]\right\}^{1/3} \\
&\quad \times\left\{E\left[\sup_{\theta_j}\left|\left|\frac{\partial H_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i}\right|\right|^3\right]\right\}^{1/3},
\end{aligned} \tag{B.23}
$$

$$E\left[\sup_{\theta_j}||D_{jt}(\gamma_j,\lambda)||\right] \leq K\left\{E\left[||X_{jt}||^4\right]\right\}^{1/2}\left\{E\left[\sup_{\theta_j}\left|\left|\frac{\partial^2 H_{jt}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}}\right|\right|^2\right]\right\}^{1/2} \quad \text{(B.24)}$$

By Assumption 4.6 we already have $E[||X_{jt}||^6] < \infty$. Moreover, by Lemma B.5 we have $\{E[\sup_{\theta_j}||\partial H_{jt}(\gamma_j,\lambda)/\partial\theta_i||^3]\}^{1/3} < \infty$ and $\{E[\sup_{\theta_j}||\partial^2 H_{jt}(\gamma_j,\lambda)/\partial\theta_i\partial\theta_{i'}||^2]\}^{1/2} < \infty$. Then, from (B.20) and (B.21)-(B.24) it follows that $E[\sup_{\theta_j}||\partial^2 l_{jt}(\gamma_j,\lambda)/\partial\theta_i\partial\theta_{i'}||] < \infty$, independent of j, as stated. Notice that this result continues to hold if θ_i and/or $\theta_{i'}$ are not included in $(\gamma_j',\lambda')'$; this is because, in that case we automatically have $\partial^2 l_{jt}(\gamma_j,\lambda)/\partial\theta_i\partial\theta_{i'} = 0$. ■

Lemma B.7 *Let θ_i be the i^{th} entry of θ, where $i = 1, ..., 4N+2$. Then,*

$$\sup_{\theta}\left|\frac{\partial^2 l_{NT}(\gamma,\lambda)}{\partial\theta_i\partial\theta_{i'}} - E\left[\frac{\partial^2 l_{Nt}(\gamma,\lambda)}{\partial\theta_i\partial\theta_{i'}}\right]\right| \overset{a.s.}{\to} 0 \quad \text{as } T\to\infty.$$

Proof of Lemma B.7. We invoke Lemma B.6 and use the same arguments as in the Proof of Lemma B.6 of PR to obtain

$$\sup_{\theta_j}\left|\frac{\partial^2 l_{jT}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}} - E\left[\frac{\partial^2 l_{jt}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}}\right]\right| \overset{a.s.}{\to} 0, \quad \text{(B.25)}$$

for each pair j as $T\to\infty$ (notice that, if θ_i and/or $\theta_{i'}$ is any parameter that is not included in $(\gamma_j',\lambda')'$, then $\partial^2 l_{jT}(\gamma_j,\lambda)/\partial\theta_i\partial\theta_{i'} = 0$ and the above statement holds automatically). Then, by the triangle inequality and (B.25), and since N is fixed, we obtain

$$\sup_{\theta}\left|\frac{\partial^2 l_{NT}(\gamma,\lambda)}{\partial\theta_i\partial\theta_{i'}} - E\left[\frac{\partial^2 l_{Nt}(\gamma,\lambda)}{\partial\theta_i\partial\theta_{i'}}\right]\right| \leq \frac{1}{N}\sum_{j=1}^{N}\sup_{\theta_j}\left|\frac{\partial^2 l_{jT}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}} - E\left[\frac{\partial^2 l_{jt}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}}\right]\right| = o_{a.s.}(1),$$

as stated. ■

Lemma B.8 $E\left[\partial^2 l_{Nt}(\gamma_0,\lambda_0)/\partial\lambda\partial\lambda'\right]$ *is non-singular.*

Proof of Lemma B.8. The Proof of Lemma C.2 is directly applicable here, with a few straightforward modifications due to the change in parameter dimensions. In particular, given that the parameter vector λ for the scalar BEKK model in (12) is (2×1), we have to re-define h_{jt}, k_{jt} and c as follows:

$$h_{jt} = (h_{jt,1}, h_{jt,2}), \qquad k_{jt} = (k_{jt,1}, k_{jt,2}), \qquad \text{and} \qquad c = (c_1, c_2).$$

As in the Proof of Lemma C.2, λ_i is the i^{th} entry of λ; but now $i = 1, 2$. In line with this, the summations $\sum_{i=1}^{8}$ have to be replaced by $\sum_{i=1}^{2}$, as well. Moreover, $A^{\otimes 2}$ and $B^{\otimes 2}$ should be replaced by α and β. Consequently, the following modified definitions follow:

$$\begin{aligned}
\omega_j &= (1-\alpha-\beta)\,\gamma_j, \qquad \omega_{j0} = (1-\alpha_0-\beta_0)\,\gamma_{j0}, \\
\tilde{\omega}_{j0} &= \sum_{i=1}^{2} c_i \frac{\partial}{\partial \lambda_i}\omega_j \bigg|_{\theta_j=\theta_{j0}}, \qquad \tilde{A}_0 = \sum_{i=1}^{2} c_i \frac{\partial}{\partial \lambda_i}\alpha \bigg|_{\theta_j=\theta_{j0}}, \qquad \text{and} \qquad \tilde{B}_0 = \sum_{i=1}^{2} c_i \frac{\partial}{\partial \lambda_i}\beta \bigg|_{\theta_j=\theta_{j0}}.
\end{aligned}$$

With these modifications, the Proof of Lemma C.2 becomes valid for the case at hand, and the desired result follows. ■

Lemma B.9 *For $\hat{\gamma}_j$ as defined in (13), we have*

$$\begin{bmatrix} \hat{\gamma}-\gamma_0 \\ \partial l_{NT}(\gamma_0,\lambda_0)/\partial\lambda \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} Q_t^N \\ W_t^N \\ \tilde{W}_t^N \end{bmatrix} V_t^N + o_p\left(\frac{1}{\sqrt{T}}\right) \quad \text{as } T \to \infty,$$

where Q_t^N, W_t^N, $\tilde{W}_t^N$ and V_t^N are as defined in Section B.1.

Proof of Lemma B.9. To keep the notation concise, in what follows we let $H_{jt} = H_{jt}(\gamma_{j0}, \lambda_0)$. We start with the proof for the score, $\partial l_{NT}(\gamma_0, \lambda_0)/\partial\lambda$. Given the definition of $l_{jt}(\gamma_j, \lambda)$ in (A.1) we have

$$\begin{aligned}
\frac{\partial l_{jt}(\gamma_j,\lambda)}{\partial \alpha} &= \left[vec\left(H_{jt}^{-1}\right)\right]' \frac{\partial vec\left(H_{jt}(\gamma_j,\lambda)\right)}{\partial \alpha} - vec\left(X_{jt}X_{jt}'\right)'\left(H_{jt}^{-1}\right)^{\otimes 2} \frac{\partial vec\left(H_{jt}(\gamma_j,\lambda)\right)}{\partial \alpha} \\
&= \left\{\left[vec\left(H_{jt}^{-1}\right)\right]' - vec\left(X_{jt}X_{jt}'\right)'\left(H_{jt}^{-1}\right)^{\otimes 2}\right\} \frac{\partial vec\left(H_{jt}(\gamma_j,\lambda)\right)}{\partial \alpha}.
\end{aligned}$$

We also have,

$$\begin{aligned}
\frac{\partial vec\left(H_{jt}(\gamma_j,\lambda)\right)}{\partial \alpha} &= \frac{\partial vec}{\partial \alpha}\left[\Gamma_j + \alpha(X_{j,t-1}X_{j,t-1}' - \Gamma_j) + \beta(H_{j,t-1}(\gamma_j,\lambda) - \Gamma_j)\right] \\
&= \frac{\partial vec}{\partial \alpha}\alpha(X_{j,t-1}X_{j,t-1}' - \Gamma_j) + \frac{\partial vec}{\partial \alpha}\beta(H_{j,t-1}(\gamma_j,\lambda) - \Gamma_j)
\end{aligned}$$

$$= vec(X_{j,t-1}X'_{j,t-1} - \Gamma_j) + \beta \frac{\partial vec\left(H_{j,t-1}\left(\gamma_j, \lambda\right)\right)}{\partial \alpha},$$

and by recursion this yields,

$$\begin{aligned}
\frac{\partial vec\left(H_{jt}(\gamma_j, \lambda)\right)}{\partial \alpha} &= vec(X_{j,t-1}X'_{j,t-1} - \Gamma_j) + \beta \frac{\partial vec\left(H_{j,t-1}\left(\gamma_j, \lambda\right)\right)}{\partial \alpha} \\
&= \sum_{i=0}^{\infty} \beta^i vec(X_{j,t-1-i}X'_{j,t-1-i} - \Gamma_j) + \lim_{i \to \infty} \beta^i \frac{\partial vec\left(H_{j,t-i}\left(\gamma_j, \lambda\right)\right)}{\alpha} \\
&= \sum_{i=0}^{\infty} \beta^i vec(X_{j,t-1-i}X'_{j,t-1-i} - \Gamma_j),
\end{aligned}$$

since $0 < \beta < 1$ by Assumption 4.2. Therefore,

$$\begin{aligned}
\frac{\partial l_{jt}(\gamma_j, \lambda)}{\partial \alpha} &= \left\{\left[vec\left(H_{jt}^{-1}\right)\right]' - vec\left(X_{jt}X'_{jt}\right)'\left(H_{jt}^{-1}\right)^{\otimes 2}\right\} \sum_{i=0}^{\infty} \beta^i vec(X_{j,t-1-i}X'_{j,t-1-i} - \Gamma_j) \\
&= \left\{\left[vec\left(I_2\right) - vec(Z_{jt}Z'_{jt})\right]'\left(H_{jt}^{-1/2}\right)^{\otimes 2}\right\} \sum_{i=0}^{\infty} \beta^i vec(X_{j,t-1-i}X'_{j,t-1-i} - \Gamma_j).
\end{aligned}$$

Notice that this is a scalar term and so its transpose is equal to itself. Hence,

$$\frac{\partial l_{jT}(\gamma_j, \lambda)}{\partial \alpha} = -\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=0}^{\infty} \beta^i vec(X_{j,t-1-i}X'_{j,t-1-i} - \Gamma_j)\right)' \left(H_{jt}^{-1/2}\right)^{\otimes 2} vec(Z_{jt}Z'_{jt} - I_2),$$

and consequently,

$$\frac{\partial l_{NT}(\gamma, \lambda)}{\partial \alpha} = \frac{1}{NT} \sum_{j=1}^{N} \sum_{t=1}^{T} \underbrace{\left[-\sum_{i=0}^{\infty} \beta^i \left(vec(X_{j,t-1-i}X'_{j,t-1-i} - \Gamma_j)\right)' (H_{jt}^{-1/2})^{\otimes 2} vec(Z_{jt}Z'_{jt} - I_2)\right]}_{\partial l_{jt}(\gamma_j, \lambda)/\partial \alpha}. \tag{B.26}$$

Similarly,

$$\frac{\partial l_{jt}(\gamma_j, \lambda)}{\partial \beta} = \left\{\left[vec\left(H_{jt}^{-1}\right)\right]' - vec\left(X_{jt}X'_{jt}\right)'\left(H_{jt}^{-1}\right)^{\otimes 2}\right\} \frac{\partial vec\left(H_{jt}(\gamma_j, \lambda)\right)}{\partial \beta},$$

where

$$\begin{aligned}
\frac{\partial vec\left(H_{jt}(\gamma_j,\lambda)\right)}{\partial\beta} &= \frac{\partial vec}{\partial\beta}\left[\Gamma_j+\alpha(X_{j,t-1}X'_{j,t-1}-\Gamma_j)+\beta(H_{j,t-1}\left(\gamma_j,\lambda\right)-\Gamma_j)\right] \\
&= vec(H_{j,t-1}\left(\gamma_j,\lambda\right)-\Gamma_j)+\beta\frac{\partial vec(H_{j,t-1}\left(\gamma_j,\lambda\right))}{\partial\beta} \\
&= \sum_{i=0}^{\infty}\beta^i vec(H_{j,t-1-i}\left(\gamma_j,\lambda\right)-\Gamma_j)+\lim_{i\to\infty}\beta^i\frac{\partial vec(H_{j,t-i}\left(\gamma_j,\lambda\right))}{\partial\beta} \\
&= \sum_{i=0}^{\infty}\beta^i vec(H_{j,t-1-i}\left(\gamma_j,\lambda\right)-\Gamma_j).
\end{aligned}$$

Therefore, by the same arguments as for the score with respect to α,

$$\frac{\partial l_{NT}(\gamma,\lambda)}{\partial\beta}=\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\underbrace{\left[-\sum_{i=0}^{\infty}\beta^i(vec\left(H_{j,t-1-i}\left(\gamma_j,\lambda\right)-\Gamma_j\right))'(H_{jt}^{-1/2})^{\otimes 2}vec(Z_{jt}Z'_{jt}-I_2)\right]}_{\partial l_{jt}(\gamma_j,\lambda)/\partial\beta}. \tag{B.27}$$

Next, we consider $\hat{\gamma}_j-\gamma_{j0}$. By following the same steps as PR (see the arguments leading to the final equation on p. 50 of PR), we obtain

$$\begin{aligned}
(\hat{\gamma}_j-\gamma_{j0}) &= \frac{1-\beta_0}{1-\alpha_0-\beta_0}\frac{1}{T}\sum_{t=1}^{T}\left(H_{jt}^{1/2}\right)^{\otimes 2}vec\left(Z_{jt}Z'_{jt}-I_2\right) \\
&\quad+\frac{1}{1-\alpha_0-\beta_0} \\
&\quad\times\left(\alpha_0\frac{1}{T}vec\left(X_{j0}X'_{j0}-X_{jT}X'_{jT}\right)+\beta_0\frac{1}{T}vec\left(H_{j0}-H_{jT}\right)\right).
\end{aligned} \tag{B.28}$$

We focus on the second term on the right-hand side of (B.28). By standard results,

$$\begin{aligned}
&\left|\left|\alpha_0 vec\left(X_{j0}X'_{j0}-X_{jT}X'_{jT}\right)+\beta_0 vec\left(H_{j0}-H_{jT}\right)\right|\right| \\
\leq\ &\alpha_0\left|\left|vec\left(X_{j0}X'_{j0}-X_{jT}X'_{jT}\right)\right|\right| \\
&+\beta_0\left|\left|vec\left(H_{j0}-H_{jT}\right)\right|\right| \\
\leq\ &\alpha_0\left|\left|vec\left(X_{j0}X'_{j0}\right)\right|\right|+\alpha_0\left|\left|vec\left(X_{jT}X'_{jT}\right)\right|\right| \\
&+\beta_0\left|\left|vec\left(H_{j0}\right)\right|\right|+\beta_0\left|\left|vec\left(H_{jT}\right)\right|\right| \\
=\ &\alpha_0\left|\left|X_{j0}\right|\right|^2+\alpha_0\left|\left|X_{jT}\right|\right|^2+\beta_0\left|\left|H_{j0}\right|\right|+\beta_0\left|\left|H_{jT}\right|\right|.
\end{aligned} \tag{B.29}$$

By Assumption 4.6 and Lemma B.5, we have $E[||X_{jt}||^2] < \infty$ and $E[||H_{jt}||] < \infty$, respectively, for all j and t. Moreover, by Assumption 4.2, α_0 and β_0 are both finite. Therefore, by (B.29) we obtain

$$E\left[||\alpha_0 vec(X_{j0}X_{j0}' - X_{jT}X_{jT}') + \beta_0 vec(H_{j0} - H_{jT})||\right] < \infty.$$

Then, by Chebychev's inequality, for any $\varepsilon > 0$

$$\begin{aligned} & P\left(\frac{1}{\sqrt{T}}\left|\left|\alpha_0 vec(X_{j0}X_{j0}' - X_{jT}X_{jT}') + \beta_0 vec\left(H_{j0} - H_{jT}\right)\right|\right| > \varepsilon\right) \\ \leq\ & \frac{E\left[\left|\left|\alpha_0 vec(X_{j0}X_{j0}' - X_{jT}X_{jT}') + \beta_0 vec\left(H_{j0} - H_{jT}\right)\right|\right|\right]}{T^{1/2}\varepsilon} \\ =\ & O\left(T^{-1/2}\right), \end{aligned}$$

as $T \to \infty$. Hence,

$$\frac{1}{\sqrt{T}}\left|\left|\alpha_0 vec(X_{j0}X_{j0}' - X_{jT}X_{jT}') + \beta_0 vec\left(H_{j0} - H_{jT}\right)\right|\right| = o_p(1),$$

and combining this result with (B.28) yields

$$(\hat{\gamma}_j - \gamma_{j0}) = \frac{1-\beta_0}{1-\alpha_0-\beta_0}\frac{1}{T}\sum_{t=1}^{T}(H_{jt}^{1/2})^{\otimes 2} vec(Z_{jt}Z_{jt}' - I_2) + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{B.30}$$

for all j as $T \to \infty$. Bringing (B.26), (B.27) and (B.30) together, we then have

$$\begin{bmatrix} \hat{\gamma} - \gamma_0 \\ \partial l_{NT}\left(\gamma_0, \lambda_0\right)/\partial\lambda \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_{Nt}\left(\gamma_0, \lambda_0\right) + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where

$$Y_{Nt}\left(\gamma_0, \lambda_0\right) = \begin{bmatrix} \frac{1-\beta_0}{1-\alpha_0-\beta_0}(H_{1t}^{1/2})^{\otimes 2} vec\left(Z_{1t}Z_{1t}' - I_2\right) \\ \vdots \\ \frac{1-\beta_0}{1-\alpha_0-\beta_0}(H_{Nt}^{1/2})^{\otimes 2} vec\left(Z_{Nt}Z_{Nt}' - I_2\right) \\ N^{-1}\sum_{j=1}^{N}\left[-\sum_{i=0}^{\infty}\beta_0^i vec(X_{j,t-1-i}X_{j,t-1-i}' - \Gamma_{j0})\right]'(H_{jt}^{-1/2})^{\otimes 2} vec(Z_{jt}Z_{jt}' - I_2) \\ N^{-1}\sum_{j=1}^{N}\left[-\sum_{i=0}^{\infty}\beta_0^i vec(H_{j,t-1-i}\left(\gamma_{j0}, \lambda_0\right) - \Gamma_{j0})\right]'(H_{jt}^{-1/2})^{\otimes 2} vec(Z_{jt}Z_{jt}' - I_2) \end{bmatrix}.$$

Notice that $Y_{Nt}(\gamma_0, \lambda_0)$ stated as such is equivalent to $[(Q_t^N)', (W_t^N)', (\tilde{W}_t^N)']'V_t^N$, where Q_t^N, W_t^N, $\tilde{W}_t^N$ and V_t^N are as defined in Section B.1. This yields the stated result. ■

Lemma B.10 *$E[||((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')'V_t^N||^2] < \infty$, where V_t^N, Q_t^N, W_t^N and $\tilde{W}_t^N$ are as defined in Section B.1.*

Proof of Lemma B.10. To keep the notation concise, in what follows we let $H_{jt} = H_{jt}(\gamma_{j0}, \lambda_0)$. This lemma is the scalar BEKK version of Lemma C.4 (which focusses on the non-scalar BEKK model of equation (14)). Consequently, a substantial portion of the arguments made in the Proof of Lemma C.4 can also be used here. In particular, by the same arguments as in the Proof of Lemma C.4, one can immediately show that for V_t^N, Q_t^N, W_t^N and $\tilde{W}_t^N$ as defined in Section B.1

$$\begin{aligned} E[||((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')'V_t^N||^2] &= E[(V_t^N)'(Q_t^N)'Q_t^N V_t^N] + E[(V_t^N)'(W_t^N)'W_t^N V_t^N] \\ &\quad + E[(V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N], \end{aligned} \tag{B.31}$$

and

$$E[(V_t^N)'(Q_t^N)'Q_t^N V_t^N] < \infty. \tag{B.32}$$

Moreover, letting

$$G_{j,t-1} = -\sum_{i=0}^{\infty} \beta_0^i vec(X_{j,t-1-i}X_{j,t-1-i}' - \Gamma_{j0}) \quad \text{and} \quad \tilde{G}_{j,t-1} = -\sum_{i=0}^{\infty} \beta_0^i vec(H_{j,t-1-i} - \Gamma_{j0}),$$

again by the same arguments that lead to (C.26) and (C.33) in the Proof of Lemma C.4, it is straightforward to obtain

$$E[(V_t^N)'(W_t^N)'W_t^N V_t^N] \le K\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\sqrt{E\left[||G_{j,t-1}||^2\right]}\sqrt{E\left[||G_{k,t-1}'||^2\right]}, \tag{B.33}$$

$$E[(V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N] \le K\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\sqrt{E\left[||\tilde{G}_{j,t-1}||^2\right]}\sqrt{E\left[||\tilde{G}_{k,t-1}'||^2\right]}. \tag{B.34}$$

Now, by Minkowski's inequality

$$\begin{aligned}\left\{E\left[||G_{j,t-1}||^2\right]\right\}^{1/2} &= \left\{E\left[\left|\left|\sum_{i=0}^{\infty}\beta_0^i vec(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0})\right|\right|^2\right]\right\}^{1/2} \\ &\leq \sum_{i=0}^{\infty}\left\{E\left[\left|\left|\beta_0^i vec(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0})\right|\right|^2\right]\right\}^{1/2} \\ &\leq \sum_{i=0}^{\infty}\beta_0^i\left\{E\left[\left|\left|vec(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0})\right|\right|^2\right]\right\}^{1/2}, \qquad \text{(B.35)}\end{aligned}$$

and by (A.2) and (A.3)

$$\begin{aligned}\left\{E\left[\left|\left|vec(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0})\right|\right|^2\right]\right\}^{1/2} &= \left\{E\left[\left|\left|X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right|\right|^2\right]\right\}^{1/2} \\ &\leq \left\{E\left[\left(\left|\left|X_{j,t-1-i}X'_{j,t-1-i}\right|\right|+||\Gamma_{j0}||\right)^2\right]\right\}^{1/2} \\ &= \left\{E\left[\left(||X_{j,t-1-i}||^2+||\Gamma_{j0}||\right)^2\right]\right\}^{1/2} \\ &= \left\{E\left[||X_{j,t-1-i}||^4+2\,||X_{j,t-1-i}||^2\,||\Gamma_{j0}||+||\Gamma_{j0}||^2\right]\right\}^{1/2} \\ &\leq K, \qquad \text{(B.36)}\end{aligned}$$

since $E[||X_{jt}||^4] < \infty$ by Assumption 4.6. Combining (B.35) with (B.36), we then obtain

$$\sqrt{E\left[||G_{j,t-1}||^2\right]} \leq \sum_{i=0}^{\infty} K\phi^i < \infty, \qquad \text{(B.37)}$$

since $0 < \beta_0 < 1$ by assumption. Noting that $E[||H_{jt}||^2] < \infty$ by Lemma B.5, it can similarly be proved that

$$\sqrt{E[||\tilde{G}_{j,t-1}||^2]} < \infty. \qquad \text{(B.38)}$$

Combining (B.33), (B.34), (B.37) and (B.38) yields

$$E\left[\left(V_t^N\right)'\left(W_t^N\right)'W_t^N V_t^N\right] < \infty \quad \text{and} \quad E\left[\left(V_t^N\right)'\left(\tilde{W}_t^N\right)'\tilde{W}_t^N V_t^N\right] < \infty. \qquad \text{(B.39)}$$

The stated result now follows from equations (B.31), (B.32) and (B.39). ■

Lemma B.11 *Let Ω_0 be defined as in Section B.1 and let $\hat{\gamma}_j$ be as defined in (13). Then, as $T \to \infty$ we have*

$$\sqrt{T}\begin{bmatrix} \hat{\gamma}-\gamma_0 \\ \partial l_{NT}(\gamma_0,\lambda_0)/\partial\lambda \end{bmatrix} \xrightarrow{d} N(0,\Omega_0).$$

Proof of Lemma B.11. Let Q_t^N, W_t^N, $\tilde{W}_t^N$ and V_t^N be as defined in Section B.1. Define $A_t^N = ((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')' V_t^N$. Notice that Q_t^N, W_t^N and $\tilde{W}_t^N$ are all $\mathcal{F}_{t-1}$-measurable (remember that $\mathcal{F}_t$ is the sigma algebra generated by the collection of all returns at and before time t). Moreover, V_t^N is independent of $\mathcal{F}_{t-1}$. Therefore, $(A_t^N, \mathcal{F}_t)$ yields an ergodic martingale difference sequence. Observe that by Lemma B.10 A_t^N is square-integrable. Then, by the same arguments as in the Proof of Lemma B.10 of PR an appropriate CLT exists and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} A_t^N \xrightarrow{d} N(0,\Omega_0) \qquad \text{as } T\to\infty, \tag{B.40}$$

where Ω_0 is as defined in Section B.1. Finally, invoking Lemma B.9 and using it together with (B.40) yields the stated result. ■

Lemma B.12 *Let $\lambda_1 = \alpha$ and $\lambda_2 = \beta$. Then, for $i = 1, 2$,*

$$\sqrt{T}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\left|\frac{\partial l_{jt}(\gamma_{j0},\lambda_0)}{\partial\lambda_i} - \frac{\partial l_{jt,h}(\gamma_{j0},\lambda_0)}{\partial\lambda_i}\right| = o_p(1), \qquad \textit{as } T\to\infty.$$

Proof of Lemma B.12. To keep the notation concise, in what follows we will omit the arguments of a function whenever it is evaluated at the true parameter values; e.g. $l_{jt} = l_{jt}(\gamma_{j0},\lambda_0)$, $H_{jt} = H_{jt}(\gamma_{j0},\lambda_0)$ etc. First, notice that inequality (B.32) of Hafner and Preminger (2009) can be modified to accommodate our two-dimensional setting with the indices (j,t). Then, for some r such that $0 < r < 1$,

$$\begin{aligned} & E\left[\left|\frac{\partial l_{jt}}{\partial\lambda_i} - \frac{\partial l_{jt,h}}{\partial\lambda_i}\right|^r\right] \\ \le\ & KE\left[\left(K + K\|X_{jt}\|^{2r}\right)\left\|\frac{\partial H_{jt}}{\partial\lambda_i} - \frac{\partial H_{jt,h}}{\partial\lambda_i}\right\|^r\right] \end{aligned}$$

$$+KE\left[\left(K+K\left|\left|X_{jt}\right|\right|^{2r}\right)\times\left|\left|\frac{\partial H_{jt,h}}{\partial\lambda_i}H_{jt,h}^{-1}\right|\right|^{r}\times\left|\left|H_{jt}-H_{jt,h}\right|\right|^{r}\right]$$
$$+KE\left[\left|\left|X_{jt}\right|\right|^{2r}\left|\left|H_{jt}-H_{jt,h}\right|\right|^{r}\times\left|\left|\frac{\partial H_{jt,h}}{\partial\lambda_i}H_{jt,h}^{-1}\right|\right|^{r}\right]. \tag{B.41}$$

We note that $\sup_{\theta_j}||H_{jt,h}^{-1}(\gamma_j,\lambda)||<\infty$ and $E[||X_{jt}||^6]<\infty$ by (B.9) and Assumption 4.6, respectively. Then, by (B.41) and Hölder's inequality it can be shown that if

$$E\left[||H_{jt}-H_{jt,h}||\right]=O\left(\phi^t\right), \tag{B.42}$$
$$E\left[\left|\left|\frac{\partial H_{jt}}{\partial\lambda_i}-\frac{\partial H_{jt,h}}{\partial\lambda_i}\right|\right|\right]=O\left(t\phi^t\right), \tag{B.43}$$
$$E\left[\left|\left|\frac{\partial H_{jt,h}}{\partial\lambda_i}\right|\right|\right]<\infty. \tag{B.44}$$

then

$$E\left[\left|\frac{\partial l_{jt}}{\partial\lambda_i}-\frac{\partial l_{jt,h}}{\partial\lambda_i}\right|^{1/4}\right]=O\left(t\phi^t\right), \tag{B.45}$$

We start with (B.42): by Lemma B.5

$$E\left[\sup_{\theta_j}||H_{jt}(\gamma_j,\lambda)||\right]<\infty\quad\text{and}\quad E\left[\sup_{\theta_j}\left|\left|\frac{\partial H_{jt}(\gamma_j,\lambda)}{\partial\theta_i}\right|\right|\right]<\infty, \tag{B.46}$$

for all j and i, where θ_i is as defined in Lemma B.5. Next, by a recursion argument

$$\begin{aligned}H_{jt}(\gamma_j,\lambda)-H_{jt,h}(\gamma_j,\lambda) &= \beta\left(H_{j,t-1}(\gamma_j,\lambda)-H_{j,t-1,h}(\gamma_j,\lambda)\right)\\ &= \beta^t\left(H_{j0}(\gamma_j,\lambda)-h_j\right).\end{aligned} \tag{B.47}$$

Then,

$$\begin{aligned}E\left[\sup_{\theta_j}||H_{jt}(\gamma_j,\lambda)-H_{jt,h}(\gamma_j,\lambda)||\right] &= E\left[\sup_{\theta_j}\left|\left|\beta^t\left(H_{j0}(\gamma_j,\lambda)-h_j\right)\right|\right|\right]\\ &= O\left(\phi^t\right),\end{aligned} \tag{B.48}$$

which follows from (B.46), (B.47), Assumption 4.2 and the fact that h_j is a constant. This proves (B.42). Next, we consider (B.43):

$$
\begin{aligned}
E\left[\sup_{\theta_j}\left\|\frac{\partial H_{jt}(\gamma_j,\lambda)}{\partial\lambda_i}-\frac{\partial H_{jt,h}(\gamma_j,\lambda)}{\partial\lambda_i}\right\|\right] &= E\left[\sup_{\theta_j}\left\|\frac{\partial}{\partial\lambda_i}\left[\beta^t\left(H_{j0}(\gamma_j,\lambda)-h_j\right)\right]\right\|\right] \\
&= O\left(t\phi^t\right), \qquad \text{(B.49)}
\end{aligned}
$$

where the first equality follows from (B.47) whereas the second equality is due to (B.46); see also the argument leading to equation (B.48) in PR. Hence, (B.43) holds. Finally, (B.44) holds by Lemma B.5. Hence, the conditions (B.42)-(B.44) are satisfied and (B.45) holds.

Now, notice that $\lim_{T\to\infty}\sum_{t=1}^{T} O\left(t\phi^t\right) < \infty$ since ϕ^t converges to 0 exponentially whereas t diverges to ∞ linearly. Then, following PR, we use the generalised Chebychev's inequality, the c_r inequality (see White (2001), p.35) and (B.45) to obtain that for any $\varepsilon > 0$,

$$
\begin{aligned}
& P\left(\sqrt{T}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\left|\frac{\partial l_{jt}}{\partial\lambda_i}-\frac{\partial l_{jt,h}}{\partial\lambda_i}\right| > \varepsilon\right) \\
\leq\ & \frac{1}{\varepsilon^{1/4}N^{1/4}T^{1/8}}E\left[\left(\sum_{j=1}^{N}\sum_{t=1}^{T}\left|\frac{\partial l_{jt}}{\partial\lambda_i}-\frac{\partial l_{jt,h}}{\partial\lambda_i}\right|\right)^{1/4}\right] \\
\leq\ & \frac{1}{\varepsilon^{1/4}N^{1/4}T^{1/8}}\sum_{j=1}^{N}\sum_{t=1}^{T}E\left[\left|\frac{\partial l_{jt}}{\partial\lambda_i}-\frac{\partial l_{jt,h}}{\partial\lambda_i}\right|^{1/4}\right] \\
\leq\ & \frac{1}{\varepsilon^{1/4}N^{1/4}T^{1/8}}\sum_{j=1}^{N}\sum_{t=1}^{T}O\left(t\phi^t\right) \\
\leq\ & \frac{1}{\varepsilon^{1/4}N^{1/4}T^{1/8}}\sum_{j=1}^{N}K,
\end{aligned}
$$

which is $O(T^{-1/8})$ since N is fixed. The desired result follows. ■

Lemma B.13 *Let θ_i be the i^{th} entry of θ, where $i = 1, ..., 4N+2$. Then,*

$$
\sup_{\theta}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\left|\frac{\partial^2 l_{jt}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}}-\frac{\partial^2 l_{jt,h}(\gamma_j,\lambda)}{\partial\theta_i\partial\theta_{i'}}\right| = o_{a.s}(1), \qquad as\ T\to\infty,
$$

for all i and i'.

Proof of Lemma B.13. The proof follows the same ideas as the Proof of Lemma B.12

above. Modifying the inequality (B.36) of Hafner and Preminger (2009) for our particular case with both cross-section and time indices, and using Hölder's inequality, we observe that if

$$E\left[||X_{jt}||^2\right] < \infty, \tag{B.50}$$

$$E\left[\sup_{\theta_j} ||vec\left(H_{jt}\left(\gamma_j, \lambda\right)\right) - vec\left(H_{jt,h}\left(\gamma_j, \lambda\right)\right)||\right] = O\left(\phi^t\right), \tag{B.51}$$

$$E\left[\sup_{\theta_j} \left|\left|\frac{\partial vec\left(H_{jt}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i} - \frac{\partial vec\left(H_{jt,h}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i}\right|\right|\right] = O\left(t\phi^t\right), \tag{B.52}$$

$$E\left[\sup_{\theta_j} \left|\left|\frac{\partial^2 vec\left(H_{jt}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i \partial \theta_{i'}} - \frac{\partial^2 vec\left(H_{jt,h}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i \partial \theta_{i'}}\right|\right|\right] = O\left(t^2\phi^t\right), \tag{B.53}$$

$$E\left[\sup_{\theta_j} \left|\left|\frac{\partial vec\left(H_{jt}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i}\right|\right|\right] < \infty, \tag{B.54}$$

$$E\left[\sup_{\theta_j} \left|\left|\frac{\partial vec\left(H_{jt,h}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i}\right|\right|\right] < \infty, \tag{B.55}$$

$$E\left[\sup_{\theta_j} \left|\left|\frac{\partial^2 vec\left(H_{jt,h}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i \partial \theta_{i'}}\right|\right|\right] < \infty, \tag{B.56}$$

then

$$E\left[\sup_{\theta_j} \left|\frac{\partial^2 l_{jt}\left(\gamma_j, \lambda\right)}{\partial \theta_i \partial \theta_{i'}} - \frac{\partial^2 l_{jt,h}\left(\gamma_j, \lambda\right)}{\partial \theta_i \partial \theta_{i'}}\right|^{1/4}\right] = O\left(t^2\phi^t\right). \tag{B.57}$$

We now verify (B.50)-(B.56). First, $E[||X_{jt}||^2] < \infty$ by Assumption 4.6. From (B.48) we already know that (B.51) holds. By (B.49) we also know that (B.52) holds when the derivative is taken with respect to α or β. However, from (B.47) it is obvious that, given Lemma B.5, we can obtain the same result for the derivative with respect to γ_j, as well. Notice that (B.52) also holds if θ_i is not one of the parameters contained in $(\gamma_j', \lambda')'$, as in that case the derivatives are identically equal to zero. Hence, (B.52) holds. Next, we consider (B.53). Using (B.47), the triangle inequality and Lemma B.5

$$E\left[\sup_{\theta_j} \left|\left|\frac{\partial^2 vec\left(H_{jt}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i \partial \theta_{i'}} - \frac{\partial^2 vec\left(H_{jt,h}\left(\gamma_j, \lambda\right)\right)}{\partial \theta_i \partial \theta_{i'}}\right|\right|\right]$$

$$
\begin{aligned}
&= E\left[\sup_{\theta_j}\left|\left|\frac{\partial^2}{\partial\theta_i\partial\theta_{i'}}\beta^t vec\left[\left(H_{j0}\left(\gamma_j,\lambda\right)-h_j\right)\right]\right|\right|\right] \\
&\leq E\left[\sup_{\theta_j}\left|\left|\frac{\partial^2}{\partial\theta_i\partial\theta_{i'}}\beta^t vec H_{j0}\left(\gamma_j,\lambda\right)\right|\right|\right] + E\left[\sup_{\theta_j}\left|\left|\frac{\partial^2}{\partial\theta_i\partial\theta_{i'}}\beta^t vec\left(h_j\right)\right|\right|\right] \\
&= O\left(t^2\phi^t\right);
\end{aligned}
$$

see also the arguments on page 54 of PR who obtain the same result in their equation (B.52). The bounds in (B.54), (B.55) and (B.56) follow from Lemma B.5. Hence, (B.57) holds. Now, by the triangle and c_r inequalities, and (B.57), we have

$$
\begin{aligned}
E\left[\sup_{\theta}\left|\frac{\partial^2 l_{Nt}\left(\gamma,\lambda\right)}{\partial\theta_i\partial\theta_{i'}}-\frac{\partial^2 l_{Nt,h}\left(\gamma,\lambda\right)}{\partial\theta_i\partial\theta_{i'}}\right|^{1/4}\right] &\leq \frac{1}{N}\sum_{j=1}^{N}E\left[\sup_{\theta_j}\left|\frac{\partial^2 l_{jt}\left(\gamma_j,\lambda\right)}{\partial\theta_i\partial\theta_{i'}}-\frac{\partial^2 l_{jt,h}\left(\gamma_j,\lambda\right)}{\partial\theta_i\partial\theta_{i'}}\right|^{1/4}\right] \\
&= O\left(t^2\phi^t\right),
\end{aligned}
$$

since N is fixed. Notice that this is akin to equation (B.51) of PR with $r = 1/4$, except that here we have $l_{Nt}\left(\gamma,\lambda\right)$ and $l_{Nt,h}\left(\gamma,\lambda\right)$, rather than $l_{jt}\left(\gamma_j,\lambda\right)$ and $l_{jt,h}\left(\gamma_j,\lambda\right)$, due to the averaging across pairs. This averaging does not affect the asymptotic arguments since N is fixed. Then, by the same arguments as in the last part of the Proof of Lemma B.11 of PR, the stated result follows. ■

C Non-scalar BEKK

In this section we provide the consistency and asymptotic normality of the composite likelihood estimator when the pairwise likelihood functions are based on the non-scalar BEKK model given in (14). The only required change is that Assumptions 4.2 and 4.4 be replaced by Assumptions C.1 and C.2 below, respectively. The latter is a simple modification of Assumption 4.4 to accommodate the change in the number of parameters.

In what follows, for a $(K \times K)$ matrix Z with eigenvalues $\xi_1, ..., \xi_K$, we define $\rho(Z) = \max_{1 \leq k \leq K} |\xi_k|$.

Assumption C.1 *The parameter matrices A and B are such that $\rho(A^{\otimes 2} + B^{\otimes 2}) < 1$.*

Assumption C.2 *For every pair j, $\gamma_j \in \Theta_\gamma$ and $\lambda \in \Theta_\lambda$ where Θ_γ and Θ_λ are compact subsets of $\mathbb{R}^4$ and $\mathbb{R}^8$, respectively.*

Theorem C.1 follows, which we provide below. In Section C.1 we define some additional notation, while consistency and asymptotic normality are proved in Sections C.2 and C.3. Required lemmas are presented and proved in Section C.4.

Theorem C.1 *Suppose Assumptions 4.1, 4.3, 4.5–4.7, C.1 and C.2 hold. Let $\hat{\theta} = (\hat{\gamma}_1', ..., \hat{\gamma}_N', \hat{\lambda}')'$ be the composite likelihood estimator as defined in (13), with $H_{jt}(\gamma_j, \lambda)$ given by the non-scalar BEKK model in (14). Then, as $T \to \infty$ we have $\hat{\theta} - \theta_0 \xrightarrow{a.s.} 0$ and*

$$\sqrt{T} \begin{bmatrix} \hat{\gamma} - \gamma_0 \\ \hat{\lambda} - \lambda_0 \end{bmatrix} \xrightarrow{d} \begin{bmatrix} I_{4N} & 0_{4N \times 8} \\ -J_N^{-1}(\theta_0) K_N(\theta_0) & -J_N^{-1}(\theta_0) \end{bmatrix} N\left(0, \Omega_0\right),$$

where $J_N(\theta_0)$, $K_N(\theta_0)$ and Ω_0 are as defined in Section C.1.

C.1 Definitions

Throughout Section C we use

$$H_{jt}(\gamma_j, \lambda) = \Gamma_j - A\Gamma_j A' - B\Gamma_j B' + AX_{j,t-1}X_{j,t-1}'A' + BH_{j,t-1}(\gamma_j, \lambda)B', \quad \text{(C.1)}$$

$$H_{jt,h}(\gamma_j, \lambda) = \Gamma_j - A\Gamma_j A' - B\Gamma_j B' + AX_{j,t-1}X_{j,t-1}'A' + BH_{j,t-1,h}(\gamma_j, \lambda)B',$$

where $\gamma_j = vec(\Gamma_j)$, $\lambda = (vec(A)', vec(B)')'$. As in Section B, $H_{j0,h}(\gamma_j, \lambda) = h_j > 0$, where h_j is the starting value.

As for the remaining definitions: We first note that the definitions for $J_N(\theta_0)$ and $K_N(\theta_0)$ made in (B.6) are still valid, except that $l_{jt}(\gamma_j, \lambda)$ is now based on the non-scalar BEKK model in (C.1). The definition of the asymptotic variance matrix also remains the same as in Section B.1:

$$\Omega_0 = E\left[((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')' V_t^N (V_t^N)' ((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')\right];$$

however, the definitions of some of the variables V_t^N, W_t^N, $\tilde{W}_t^N$ and Q_t^N are now different. To start with, they are now based on $H_{jt}(\gamma_j, \lambda)$ as defined in (C.1). Moreover, although Q_t^N is still a $(4N \times 4N)$ block diagonal matrix with the j^{th} diagonal block given by $Q_{jt} = D(H_{jt}^{1/2})^{\otimes 2}$, we now have $D = (I_4 - A_0^{\otimes 2} - B_0^{\otimes 2})^{-1}(I_4 - B_0^{\otimes 2})$. The definitions of W_{jt} and $\tilde{W}_{jt}$ are also different now due to the presence of matrix valued parameters A and B :

$$W_{jt} = \left[-\sum_{i=0}^{\infty} \left(B_0^{\otimes 2}\right)^i M_{j,t-1-i}(\gamma_{j0}, \lambda_0)\right]' \left(H_{jt}^{-1/2}(\gamma_{j0}, \lambda_0)\right)^{\otimes 2}, \tag{C.2}$$

$$\tilde{W}_{jt} = \left[-\sum_{i=0}^{\infty} \left(B_0^{\otimes 2}\right)^i \tilde{M}_{j,t-1-i}(\gamma_{j0}, \lambda_0)\right]' \left(H_{jt}^{-1/2}(\gamma_{j0}, \lambda_0)\right)^{\otimes 2}, \tag{C.3}$$

where

$$\begin{aligned} M_{jt}(\gamma_j, \lambda) &= \{[A(X_{jt}X_{jt}' - \Gamma_j)] \otimes I_2\} + \{I_2 \otimes [A(X_{jt}X_{jt}' - \Gamma_j)]\}K_{2,2}, \\ \tilde{M}_{jt}(\gamma_j, \lambda) &= \{[B(H_{jt}(\gamma_j, \lambda) - \Gamma_j)] \otimes I_2\} + \{I_2 \otimes [B(H_{jt}(\gamma_j, \lambda) - \Gamma_j)]\} K_{2,2}. \end{aligned}$$

Here, $K_{2,2}$ is the (4×4) commutation matrix. In particular, for any (2×2) matrix M, we have $K_{2,2}vec(M) = vec(M')$. Notice that $M_{jt}(\gamma_j, \lambda)$ and $\tilde{M}_{jt}(\gamma_j, \lambda)$ as defined here are analogous to the definitions in equations (B.28) and (B.29) in PR.

The definitions of V_t^N, W_t^N and $\tilde{W}_t^N$ remain otherwise the same as in Section B.1. In particular, $V_{jt} = vec(Z_{jt}Z_{jt}' - I_2)$, $V_t^N = (V_{1t}', ..., V_{Nt}')'$, $W_t^N = N^{-1}(W_{1t}, W_{2t}, ..., W_{Nt})$ and $\tilde{W}_t^N = N^{-1}(\tilde{W}_{1t}, \tilde{W}_{2t}, ..., \tilde{W}_{Nt})$ for W_{jt} and $\tilde{W}_{jt}$ as defined in equations (C.2) and (C.3).

C.2 Proof of consistency

First, notice that under the maintained assumptions all results of PR hold for every individual pair $j = 1, ..., N$. In particular for each j we have

$$||\hat{\gamma}_j - \gamma_{j0}|| = o_{a.s.}(1) \quad \text{as } T \to \infty, \tag{C.4}$$

$$\sup_{\lambda} |l_{jT}(\gamma_{j0}, \lambda) - l_{jT,h}(\hat{\gamma}_j, \lambda)| = o_{a.s.}(1) \quad \text{as } T \to \infty, \tag{C.5}$$

$$\sup_{\theta_j} |l_{jT}(\gamma_j, \lambda) - E[l_{jt}(\gamma_j, \lambda)]| = o_{a.s.}(1) \quad \text{as } T \to \infty, \tag{C.6}$$

$$E[l_{jt}(\gamma_{j0}, \lambda)] > E[l_{jt}(\gamma_{j0}, \lambda_0)] \quad \text{if} \quad \lambda \neq \lambda_0. \tag{C.7}$$

The consistency result of (C.4) follows from equation (A.6) of PR, whereas (C.5)-(C.7) are due to Lemmas B.1-B.3 of PR. By using the same arguments as in Section B.2 above, since N is fixed (C.5)-(C.7) are sufficient to establish

$$\sup_{\lambda} |l_{NT}(\gamma_0, \lambda) - l_{NT,h}(\hat{\gamma}, \lambda)| = o_{a.s.}(1) \quad \text{as } T \to \infty,$$

$$\sup_{\theta} |l_{NT}(\gamma, \lambda) - E[l_{Nt}(\gamma, \lambda)]| = o_{a.s.}(1) \quad \text{as } T \to \infty,$$

$$E[l_{Nt}(\gamma_0, \lambda)] > E[l_{Nt}(\gamma_0, \lambda_0)] \quad \text{for any } \lambda \neq \lambda_0.$$

Then, the remaining arguments in Section B.2 above hold and one obtains $\hat{\lambda} - \lambda_0 = o_{a.s.}(1)$ as $T \to \infty$. Together with (C.4) this establishes

$$\hat{\theta} - \theta_0 = o_{a.s.}(1) \quad \text{as} \quad T \to \infty,$$

as desired.

C.3 Proof of asymptotic normality

Starting with the same expansion as in (B.5), one can use Lemmas C.1, C.2, C.5 and C.6 to proceed in exactly the same way as in Section B.3 and prove the asymptotic normality result of Theorem C.1.

C.4 Lemmas

Remark 2 *Throughout Section C.4 we assume that Assumptions 4.1, 4.3, 4.5–4.7, C.1 and C.2 hold.*

Lemma C.1 *Let θ_i be the i^{th} entry of θ, where $i = 1, ..., 4N + 8$. Then, as $T \to \infty$*

$$\sup_{\theta} \left| \frac{\partial^2 l_{NT}(\gamma, \lambda)}{\partial\theta_i \partial\theta_{i'}} - E\left[\frac{\partial^2 l_{Nt}(\gamma, \lambda)}{\partial\theta_i \partial\theta_{i'}}\right]\right| = o_{a.s.}(1) \qquad \text{for all } i \text{ and } i'.$$

Proof of Lemma C.1. Under the maintained assumptions we can invoke Lemma B.6 of PR to obtain

$$\sup_{\theta_j} \left| \frac{\partial^2 l_{jT}(\gamma_j, \lambda)}{\partial\theta_i \partial\theta_{i'}} - E\left[\frac{\partial^2 l_{jT}(\gamma_j, \lambda)}{\partial\theta_i \partial\theta_{i'}}\right]\right| = o_{a.s.}(1), \qquad \text{as } T \to \infty$$

for each pair $j = 1, ..., N$ (note that when θ_i and/or $\theta_{i'}$ is not one of the parameters contained in (γ_j', λ'), the above result holds trivially since $\partial^2 l_{jT}(\gamma_j, \lambda) / \partial\theta_i \partial\theta_{i'} = 0$). Since N is fixed, this result and the triangle inequality are sufficient to obtain the statement of Lemma C.1. ■

Lemma C.2 *$E[\partial^2 l_{Nt}(\gamma_0, \lambda_0) / \partial\lambda\partial\lambda']$ is non-singular.*

Proof of Lemma C.2. This is a straightforward extension of Lemma B.7 of PR to the specific case of composite likelihood function. Consequently, the proof is almost identical to the Proof of Lemma B.7 of PR.

To keep the notation concise, in what follows we use $H_{jt} = H_{jt}(\gamma_{j0}, \lambda_0)$. Let λ_i be the i^{th} entry of λ, where $i = 1, ..., 8$. Remembering that $E\left[X_{jt}X_{jt}' | \mathcal{F}_{j,t-1}\right] = H_{jt}$, and modifying equation (B.22) of PR by adding the pair index j, we have

$$E\left[\frac{\partial^2 l_{jt}(\gamma_{j0}, \lambda_0)}{\partial\lambda_i \partial\lambda_{i'}}\middle| \mathcal{F}_{j,t-1}\right] = tr\left(H_{jt}^{-1}\frac{\partial H_{jt}}{\partial\lambda_{i'}} H_{jt}^{-1}\frac{\partial H_{jt}}{\partial\lambda_i}\right). \tag{C.8}$$

Next, we define $h_{jt,i} = (H_{jt}^{-1/2})^{\otimes 2} k_{jt,i}$ and $k_{jt,i} = vec(\partial H_{jt} / \partial\lambda_i)$. Then, (C.8) can be written as

$$\begin{aligned}
& E\left[\left.\frac{\partial^2 l_{jt}\left(\gamma_{j0},\lambda_0\right)}{\partial\lambda_i\partial\lambda_{i'}}\right|\mathcal{F}_{j,t-1}\right] \\
= \; & tr\left(H_{jt}^{-1}\frac{\partial H_{jt}}{\partial\lambda_{i'}}H_{jt}^{-1}\frac{\partial H_{jt}}{\partial\lambda_i}\right) \\
= \; & \left[vec\left(\frac{\partial H_{jt}}{\partial\lambda_i}\right)\right]'\left[H_{jt}^{-1}\otimes H_{jt}^{-1}\right]vec\left(\frac{\partial H_{jt}}{\partial\lambda_{i'}}\right) \\
= \; & \left[vec\left(\frac{\partial H_{jt}}{\partial\lambda_i}\right)\right]'(H_{jt}^{-1/2})^{\otimes 2}\times(H_{jt}^{-1/2})^{\otimes 2}vec\left(\frac{\partial H_{jt}}{\partial\lambda_{i'}}\right) \\
= \; & h'_{jt,i}h_{jt,i'},
\end{aligned} \tag{C.9}$$

where the second equality follows from result 7.2(11) of Lütkepohl (1996). Then, by (C.9) we obtain

$$\begin{aligned}
E\left[\frac{\partial^2 l_{Nt}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i\partial\lambda_{i'}}\right] &= \frac{1}{N}\sum_{j=1}^{N}E\left[\frac{\partial^2 l_{jt}\left(\gamma_{j0},\lambda_0\right)}{\partial\lambda_i\partial\lambda_{i'}}\right] \\
&= \frac{1}{N}\sum_{j=1}^{N}E\left\{E\left[\left.\frac{\partial^2 l_{jt}\left(\gamma_{j0},\lambda_0\right)}{\partial\lambda_i\partial\lambda_{i'}}\right|\mathcal{F}_{j,t-1}\right]\right\} \\
&= \frac{1}{N}\sum_{j=1}^{N}E\left[h'_{jt,i}h_{jt,i'}\right].
\end{aligned}$$

Next, letting $\mathcal{H}_{jt}=(H_{jt}^{-1/2})^{\otimes 2}$, $h_{jt}=(h_{jt,1},h_{jt,2},...,h_{jt,8})$ and $k_{jt}=(k_{jt,1},k_{jt,2},...,k_{jt,8})$ we have $h_{jt}=\mathcal{H}_{jt}k_{jt}$. Notice that since in our particular case where the composite likelihood is based on the non-scalar BEKK model in (14), λ is (8×1). Consequently, h_{jt} and k_{jt} are (1×8). Now, for $E\left[\partial^2 l_{Nt}\left(\gamma_0,\lambda_0\right)/\partial\lambda\partial\lambda'\right]$ to be singular we must have some non-zero (8×1) vector $c=(c_1,...,c_8)'$, such that

$$\begin{aligned}
c'E\left[\frac{\partial^2 l_{Nt}\left(\gamma_0,\lambda_0\right)}{\partial\lambda\partial\lambda'}\right]c &= \frac{1}{N}\sum_{j=1}^{N}c'E\left[\frac{\partial^2 l_{jt}\left(\gamma_{j0},\lambda_0\right)}{\partial\lambda\partial\lambda'}\right]c \\
&= \frac{1}{N}\sum_{j=1}^{N}c'E\left[(h_{jt,1},...,h_{jt,8})'(h_{jt,1},...,h_{jt,8})\right]c \\
&= \frac{1}{N}\sum_{j=1}^{N}E\left[c'h'_{jt}h_{jt}c\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{j=1}^{N} E\left[c'k'_{jt}\mathcal{H}^2_{jt}k_{jt}c\right] \\
&= 0, \qquad \text{(C.10)}
\end{aligned}
$$

which is possible only if

$$
k_{jt}c = \sum_{i=1}^{8} c_i \frac{\partial}{\partial \lambda_i}\left[vec\left(H_{jt}(\gamma_j,\lambda)\right]\right|_{\theta_j=\theta_{j0}} = 0, \qquad \text{(C.11)}
$$

for all j and t. Using notation similar to that of the Proof of Lemma B.7 of PR, we define

$$
\begin{aligned}
\omega_j &= \left(I_4 - A^{\otimes 2} - B^{\otimes 2}\right)\gamma_j, \qquad \omega_{j0} = \left(I_4 - A_0^{\otimes 2} - B_0^{\otimes 2}\right)\gamma_{j0}, \\
\tilde{\omega}_{j0} &= \left.\sum_{i=1}^{8} c_i \frac{\partial}{\partial \lambda_i}\omega_j\right|_{\theta_j=\theta_{j0}}, \qquad \tilde{A}_0 = \left.\sum_{i=1}^{8} c_i \frac{\partial}{\partial \lambda_i}A^{\otimes 2}\right|_{\theta_j=\theta_{j0}}, \qquad \tilde{B}_0 = \left.\sum_{i=1}^{8} c_i \frac{\partial}{\partial \lambda_i}B^{\otimes 2}\right|_{\theta_j=\theta_{j0}}.
\end{aligned}
$$

Now, if (C.11) holds, then

$$
\begin{aligned}
0 &= \left.\sum_{i=1}^{8} c_i \frac{\partial}{\partial \lambda_i}\left\{\omega_j + A^{\otimes 2} vec\left(X_{j,t-1}X'_{j,t-1}\right) + B^{\otimes 2} vec\left[H_{j,t-1}\left(\gamma_j,\lambda\right)\right]\right\}\right|_{\theta_j=\theta_{j0}} \\
&= \tilde{\omega}_{j0} + \tilde{A}_0 vec\left(X_{j,t-1}X'_{j,t-1}\right) + \tilde{B}_0 vec(H_{j,t-1}) \\
&\quad + B_0^{\otimes 2}\sum_{i=1}^{8} c_i \left(\left.\frac{\partial}{\partial \lambda_i} vec\left[H_{j,t-1}\left(\gamma_j,\lambda\right)\right]\right|_{\theta_j=\theta_{j0}}\right) \\
&= \tilde{\omega}_{j0} + \tilde{A}_0 vec\left(X_{j,t-1}X'_{j,t-1}\right) + \tilde{B}_0 vec(H_{j,t-1}), \qquad \text{(C.12)}
\end{aligned}
$$

where the last line follows from $k_{j,t-1}c = 0$ by (C.11). Subtracting (C.12), which is equal to zero, from $vec\left(H_{jt}\right)$ yields another expression for $vec\left(H_{jt}\right)$, given by

$$
\begin{aligned}
vec(H_{jt}) &= \omega_{j0} + A_0^{\otimes 2} vec\left(X_{j,t-1}X'_{j,t-1}\right) + B_0^{\otimes 2} vec(H_{j,t-1}) \\
&\quad - \tilde{\omega}_{j0} - \tilde{A}_0 vec\left(X_{j,t-1}X'_{j,t-1}\right) - \tilde{B}_0 vec(H_{j,t-1}) \\
&= (\omega_{j0} - \tilde{\omega}_{j0}) + (A_0^{\otimes 2} - \tilde{A}_0) vec\left(X_{j,t-1}X'_{j,t-1}\right) + (B_0^{\otimes 2} - \tilde{B}_0) vec(H_{j,t-1}),
\end{aligned}
$$

in terms of the new parameters $(\omega_{j0} - \tilde{\omega}_{j0})$, $(A_0^{\otimes 2} - \tilde{A}_0)$ and $(B_0^{\otimes 2} - \tilde{B}_0)$. Since $c_1, ..., c_8$ cannot be equal to zero, these parameters are different from $(\omega_{j0}, A_0^{\otimes 2}, B_0^{\otimes 2})$. By the same argument as in the Proof of Lemma B.7 of PR, this violates the condition that $vec\left(H_{jt}\right)$

has a unique representation. This holds for all j and t, independent of N and T. Therefore, Assumption 4.5 is violated. This means that (C.11) and, therefore, (C.10) cannot hold, implying that $E\left[\partial^2 l_{Nt}\left(\gamma_0, \lambda_0\right)/\partial\lambda\partial\lambda'\right]$ is non-singular. ■

Lemma C.3 *Let Q_t^N, W_t^N, $\tilde{W}_t^N$ and V_t^N be as defined in Section C.1. Then, for $\hat{\gamma}_j$ as defined in (13), we have*

$$\begin{bmatrix} \hat{\gamma} - \gamma_0 \\ \partial l_{NT}\left(\gamma_0, \lambda_0\right)/\partial\lambda \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} Q_t^N \\ W_t^N \\ \tilde{W}_t^N \end{bmatrix} V_t^N + o_p\left(\frac{1}{\sqrt{T}}\right) \quad \text{as } T \to \infty.$$

Proof of Lemma C.3. To keep the notation concise, in what follows we use $H_{jt} = H_{jt}\left(\gamma_{j0}, \lambda_0\right)$. Under the maintained assumptions we can invoke Lemma B.8 of PR to obtain

$$\begin{bmatrix} \hat{\gamma}_j - \gamma_{j0} \\ \partial l_{jT}\left(\gamma_{j0}, \lambda_0\right)/\partial\lambda \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_{jt}\left(\gamma_{j0}, \lambda_0\right) + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{C.13}$$

for each pair j as $T \to \infty$, where

$$Y_{jt}\left(\gamma_{j0}, \lambda_0\right) = \begin{bmatrix} D(H_{jt}^{1/2})^{\otimes 2} vec\left(Z_{jt}Z_{jt}' - I_2\right) \\ \left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{j,t-1-i}\left(\gamma_{j0}, \lambda_0\right)\right]' (H_{jt}^{-1/2})^{\otimes 2} vec\left(Z_{jt}Z_{jt}' - I_2\right) \\ \left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i \tilde{M}_{j,t-1-i}\left(\gamma_{j0}, \lambda_0\right)\right]' (H_{jt}^{-1/2})^{\otimes 2} vec\left(Z_{jt}Z_{jt}' - I_2\right) \end{bmatrix}$$

and $M_{jt}\left(\gamma_j, \lambda\right)$ and $\tilde{M}_{jt}\left(\gamma_j, \lambda\right)$ are as defined in Section C.1. Next, let

$$Y_{Nt}\left(\gamma_0, \lambda_0\right) = \begin{bmatrix} D(H_{1t}^{1/2})^{\otimes 2} vec\left(Z_{1t}Z_{1t}' - I_2\right) \\ \vdots \\ D(H_{Nt}^{1/2})^{\otimes 2} vec\left(Z_{Nt}Z_{Nt}' - I_2\right) \\ N^{-1}\sum_{j=1}^{N}\left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{j,t-1-i}\left(\gamma_{j0}, \lambda_0\right)\right]' (H_{jt}^{-1/2})^{\otimes 2} vec\left(Z_{jt}Z_{jt}' - I_2\right) \\ N^{-1}\sum_{j=1}^{N}\left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i \tilde{M}_{j,t-1-i}\left(\gamma_{j0}, \lambda_0\right)\right]' (H_{jt}^{-1/2})^{\otimes 2} vec\left(Z_{jt}Z_{jt}' - I_2\right) \end{bmatrix},$$

and notice that $Y_{Nt}\left(\gamma_0, \lambda_0\right) = [(Q_t^N)', (W_t^N)', (\tilde{W}_t^N)']' V_t^N$ for Q_t^N, W_t^N, $\tilde{W}_t^N$ and V_t^N as defined in Section C.1. Then, since N is fixed, by the result in (C.13) we have

$$\begin{bmatrix} \hat{\gamma} - \gamma_0 \\ \partial l_{NT}(\gamma_0, \lambda_0) / \partial \lambda \end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} Q_t^N \\ W_t^N \\ \tilde{W}_t^N \end{bmatrix} V_t^N + o_p\left(\frac{1}{\sqrt{T}}\right),$$

as $T \to \infty$, as stated. ■

Lemma C.4 *Let V_t^N, Q_t^N, W_t^N and $\tilde{W}_t^N$ be as defined in Section C.1. Then, we have $E[||((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')' V_t^N||^2] < \infty$.*

Proof of Lemma C.4. To keep the notation concise, in what follows we use $H_{jt} = H_{jt}(\gamma_{j0}, \lambda_0)$. We first note that under the maintained assumptions we have

$$\sup_{\theta_j} ||H_{jt}^{-1/2}(\gamma_j, \lambda)|| = \sup_{\theta_j} \sqrt{tr[H_{jt}^{-1}(\gamma_j, \lambda)]} \leq \sqrt{\sup_{\theta_j} tr[(\Gamma_j - A\Gamma_j A' - B\Gamma_j B')^{-1}]} \leq K, \tag{C.14}$$

where the first inequality follows from (A.5) and the last inequality follows from Assumption C.2; see also the second-to-last equation on page 42 of PR. Moreover, we also have

$$E[||H_{jt}||^2] < \infty \tag{C.15}$$

by the same arguments as those leading to equation (B.20) of PR.

Next, we derive some simple results which will be used later. First,

$$\begin{aligned}
\left|\left|H_{jt}^{-1/2} \otimes H_{jt}^{-1/2}\right|\right| &= \sqrt{tr\left((H_{jt}^{-1/2} \otimes H_{jt}^{-1/2})(H_{jt}^{-1/2} \otimes H_{jt}^{-1/2})\right)} \\
&= \sqrt{tr\left((H_{jt}^{-1} \otimes H_{jt}^{-1})\right)} \\
&= \sqrt{tr\left(H_{jt}^{-1}\right) tr\left(H_{jt}^{-1}\right)} \\
&= \left|tr\left(H_{jt}^{-1}\right)\right| \\
&= \left|tr\left(H_{jt}^{-1} I_2\right)\right| \\
&\leq \left|\left|H_{jt}^{-1}\right|\right| \, ||I_2|| \\
&= \sqrt{2} \left|\left|H_{jt}^{-1}\right|\right|
\end{aligned}$$

$$\leq \quad K, \tag{C.16}$$

since $||H_{jt}^{-1}|| \leq ||H_{jt}^{-1/2}||^2$, and $||H_{jt}^{-1/2}|| < \infty$ by (C.14). Note that the second equality in (C.16) follows from (A.8), while the third equality is due to (A.7). Second, for two $(m \times 1)$ matrices A and B,

$$\begin{aligned} ||AB'|| &= \sqrt{tr\left(AB'BA'\right)} \\ &= \sqrt{tr\left(A'AB'B\right)} \\ &\leq \sqrt{||A'A||\,||B'B||} \\ &\leq \sqrt{||A||^2\,||B||^2} \\ &= ||A||\,||B||\,, \end{aligned} \tag{C.17}$$

where the first inequality follows from (A.6), while the second inequality is due to (A.3). Third,

$$\begin{aligned} & E\left[\left|\left|vec\left(Z_{jt}Z_{jt}' - I_2\right) vec\left(Z_{jt}Z_{jt}' - I_2\right)'\right|\right|\right] \\ = \; & E\left[\left|\left|vec\left(Z_{jt}Z_{jt}' - I_2\right)\right|\right|^2\right] \\ = \; & E\left[\left|\left|\left(Z_{jt}Z_{jt}' - I_2\right)\right|\right|^2\right] \\ \leq \; & E\left[\left(\left|\left|Z_{jt}Z_{jt}'\right|\right| + ||I_2||\right)^2\right] \\ = \; & E\left[\left(||Z_{jt}||^2 + ||I_2||\right)^2\right] \\ = \; & E\left[||Z_{jt}||^4 + 2\sqrt{2}\,||Z_{jt}||^2 + 2\right] \end{aligned}$$

where we have used (A.2), (A.3) and the triangle inequality. Notice that,

$$E[||Z_{jt}||^4] = E[||H_{jt}^{-1/2}X_{jt}||^4] \leq E[||H_{jt}^{-1/2}||^4||X_{jt}||^4] \leq KE[||X_{jt}||^4] < \infty, \tag{C.18}$$

since $||H_{jt}^{-1/2}|| < \infty$ by (C.14) and $E[||X_{jt}||^4] < \infty$ by Assumption 4.6. Also, the first inequality in (C.18) follows from (A.4). It then follows that

$$E[||vec(Z_{jt}Z_{jt}' - I_2)vec(Z_{jt}Z_{jt}' - I_2)'||] \leq K. \tag{C.19}$$

Finally, for any $j, k = 1, ..., N$, by Hölder's inequality and similar arguments as before

$$E[||Z_{kt}Z_{kt}' - I_2|| \times ||Z_{jt}Z_{jt}' - I_2||] \leq \sqrt{E[||Z_{kt}Z_{kt}' - I_2||^2]}\sqrt{E[||Z_{jt}Z_{jt}' - I_2||^2]} < \infty.$$

We now focus on the main proof. We have

$$\begin{aligned}
& \left|\left|\left((Q_t^N)', (W_t^N)', \left(\tilde{W}_t^N\right)'\right)' V_t^N\right|\right|^2 \\
\leq \; & \left|tr[(V_t^N)'(Q_t^N)'Q_t^N V_t^N]\right| + \left|tr[(V_t^N)'(W_t^N)'W_t^N V_t^N]\right| + \left|tr[(V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N]\right| \\
= \; & (V_t^N)'(Q_t^N)'Q_t^N V_t^N + (V_t^N)'(W_t^N)'W_t^N V_t^N + (V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N. \qquad \text{(C.20)}
\end{aligned}$$

We consider the expectation of each term in (C.20) individually. First, using the definitions in Section C.1 we have

$$\begin{aligned}
& E[(V_t^N)'(Q_t^N)'Q_t^N V_t^N] \\
= \; & \sum_{j=1}^N E[vec(Z_{jt}Z_{jt}' - I_2)'(H_{jt}^{1/2})^{\otimes 2} D' D (H_{jt}^{1/2})^{\otimes 2} vec(Z_{jt}Z_{jt}' - I_2)]. \qquad \text{(C.21)}
\end{aligned}$$

Now,

$$\begin{aligned}
& E\left[vec(Z_{jt}Z_{jt}' - I_2)'(H_{jt}^{1/2})^{\otimes 2} D' D (H_{jt}^{1/2})^{\otimes 2} vec(Z_{jt}Z_{jt}' - I_2)\right] \\
= \; & E\left\{tr\left[vec\left(Z_{jt}Z_{jt}' - I_2\right)' (H_{jt}^{1/2})^{\otimes 2} D' D (H_{jt}^{1/2})^{\otimes 2} vec\left(Z_{jt}Z_{jt}' - I_2\right)\right]\right\} \\
\leq \; & E\left\{\left|tr\left[(H_{jt}^{1/2})^{\otimes 2} D' D (H_{jt}^{1/2})^{\otimes 2} vec\left(Z_{jt}Z_{jt}' - I_2\right) vec\left(Z_{jt}Z_{jt}' - I_2\right)'\right]\right|\right\} \\
\leq \; & E\left\{\left|\left|(H_{jt}^{1/2})^{\otimes 2} D' D (H_{jt}^{1/2})^{\otimes 2}\right|\right| \times \left|\left|vec\left(Z_{jt}Z_{jt}' - I_2\right) vec\left(Z_{jt}Z_{jt}' - I_2\right)'\right|\right|\right\} \\
= \; & ||D||^2 \, E[||(H_{jt}^{1/2})^{\otimes 2}||^2] \times E\left[\left|\left|vec\left(Z_{jt}Z_{jt}' - I_2\right) vec\left(Z_{jt}Z_{jt}' - I_2\right)'\right|\right|\right] \\
= \; & ||D||^2 \, E\left[\left|tr[(H_{jt}^{1/2})^{\otimes 2}(H_{jt}^{1/2})^{\otimes 2}]\right|\right] \times K \\
= \; & K \, ||D||^2 \, E\left[\left|tr\left[(H_{jt})^{\otimes 2}\right]\right|\right] \\
= \; & K \, ||D||^2 \, E\left[\left(tr\left(H_{jt}\right)\right)^2\right] \\
\leq \; & K \, ||D||^2 \, 2E\left[||H_{jt}||^2\right] \\
< \; & \infty \qquad \text{(C.22)}
\end{aligned}$$

where we use (A.6) to obtain the second inequality, (A.4) to obtain the second equality, (C.19) to obtain the third equality, (A.7) to obtain the fifth equality, and (C.15) to obtain the last inequality. By (C.21) and (C.22), and since N is fixed, we finally obtain

$$E\left[\left(V_t^N\right)'\left(Q_t^N\right)' Q_t^N V_t^N\right] < \infty. \tag{C.23}$$

Next, we focus on the expectations of the remaining two terms in (C.20). First, letting

$$\begin{aligned}
G_{j,t-1} &= -\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right), \\
\tilde{G}_{j,t-1} &= -\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i \tilde{M}_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right),
\end{aligned}$$

we notice that

$$\begin{aligned}
E\left[||W_{jt}'W_{kt}||\right] &= E\left[||(H_{jt}^{-1/2}\otimes H_{jt}^{-1/2})'G_{j,t-1}G_{k,t-1}'(H_{kt}^{-1/2}\otimes H_{kt}^{-1/2})||\right] \\
&\leq E\left[||(H_{jt}^{-1/2})^{\otimes 2}|| \times ||G_{j,t-1}|| \times \left|\left|G_{k,t-1}'\right|\right| \times ||(H_{jt}^{-1/2})^{\otimes 2}||\right] \\
&\leq K\sqrt{E\left[||G_{j,t-1}||^2\right]}\sqrt{E\left[||G_{k,t-1}'||^2\right]},
\end{aligned} \tag{C.24}$$

and similarly,

$$E\left[||\tilde{W}_{jt}'\tilde{W}_{kt}||\right] \leq K\sqrt{E\left[||\tilde{G}_{j,t-1}||^2\right]}\sqrt{E\left[||\tilde{G}_{k,t-1}'||^2\right]}, \tag{C.25}$$

where we have used (A.4), (C.16) and Hölder's inequality. Now,

$$\begin{aligned}
& E\left[\left(V_t^N\right)'\left(W_t^N\right)' W_t^N V_t^N\right] \\
=\ & E\left(tr\left\{\begin{bmatrix} vec\left(Z_{1t}Z_{1t}'-I_2\right) \\ \vdots \\ vec\left(Z_{Nt}Z_{Nt}'-I_2\right)\end{bmatrix}'\left(W_t^N\right)' W_t^N \begin{bmatrix} vec\left(Z_{1t}Z_{1t}'-I_2\right) \\ \vdots \\ vec\left(Z_{Nt}Z_{Nt}'-I_2\right)\end{bmatrix}\right\}\right) \\
=\ & \frac{1}{N^2}E\left(tr\left\{\sum_{j=1}^{N}\sum_{k=1}^{N}\left[vec\left(Z_{jt}Z_{jt}'-I_2\right)\right]' W_{jt}'W_{kt} vec\left(Z_{kt}Z_{kt}'-I_2\right)\right\}\right) \\
\leq\ & \frac{1}{N^2}E\left(\sum_{j=1}^{N}\sum_{k=1}^{N}\left|tr\left\{W_{jt}'W_{kt} vec\left(Z_{kt}Z_{kt}'-I_2\right)\left[vec\left(Z_{jt}Z_{jt}'-I_2\right)\right]'\right\}\right|\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} E\left[\left|\left|W_{jt}'W_{kt}\right|\right| \times \left|\left|vec\left(Z_{kt}Z_{kt}'-I_2\right)\left[vec\left(Z_{jt}Z_{jt}'-I_2\right)\right]'\right|\right|\right] \\
&\leq \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} E\left[\left|\left|W_{jt}'W_{kt}\right|\right|\right] \times E\left[\left|\left|vec\left(Z_{kt}Z_{kt}'-I_2\right)\left[vec\left(Z_{jt}Z_{jt}'-I_2\right)\right]'\right|\right|\right] \\
&\leq \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} E\left[\left|\left|W_{jt}'W_{kt}\right|\right|\right] E\left[\left|\left|Z_{kt}Z_{kt}'-I_2\right|\right| \times \left|\left|Z_{jt}Z_{jt}'-I_2\right|\right|\right] \\
&\leq K\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} E\left[\left|\left|W_{jt}'W_{kt}\right|\right|\right] \\
&\leq K\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} \sqrt{E\left[||G_{j,t-1}||^2\right]}\sqrt{E\left[||G_{k,t-1}'||^2\right]} \qquad \text{(C.26)}
\end{aligned}
$$

where the second inequality follows from (A.6), the third inequality follows from $\left|\left|W_{jt}'W_{kt}\right|\right|$ being measurable with respect to $\mathcal{F}_{t-1}$, the fourth inequality follows from (A.2) and (C.17), the second-to-last inequality follows from (C.19) and the final inequality follows from (C.24). Next, we investigate $\sqrt{E[||G_{j,t-1}||^2]}$. Using Minkowski's inequality and the already utilised matrix algebra results,

$$
\begin{aligned}
&\sqrt{E[||G_{j,t-1}||^2]} \\
&= \left\{E\left[\left|\left|\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)\right|\right|^2\right]\right\}^{1/2} \\
&\leq \left\{E\left[\left(\sum_{i=0}^{\infty}||(B_0^{\otimes 2})^i M_{j,t-1-i}(\gamma_{j0},\lambda_0)||\right)^2\right]\right\}^{1/2} \\
&\leq \sum_{i=0}^{\infty}\left\{E\left[\left|\left|(B_0^{\otimes 2})^i M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)\right|\right|^2\right]\right\}^{1/2} \\
&\leq \sum_{i=0}^{\infty}\left\{E\left[\left|tr\left\{\left[(B_0^{\otimes 2})^i M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)\right]'\left[(B_0^{\otimes 2})^i M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)\right]\right\}\right|\right]\right\}^{1/2} \\
&\leq \sum_{i=0}^{\infty}\left\{E\left[\left|tr\left\{\left[(B_0^{\otimes 2})^i\right]'\left(B_0^{\otimes 2}\right)^i M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)'\right\}\right|\right]\right\}^{1/2} \\
&\leq \sum_{i=0}^{\infty}\left\{E\left[\left|\left|\left[(B_0^{\otimes 2})^i\right]'(B_0^{\otimes 2})^i\right|\right| \times \left|\left|M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)M_{j,t-1-i}\left(\gamma_{j0},\lambda_0\right)'\right|\right|\right]\right\}^{1/2}
\end{aligned}
$$

$$= \sum_{i=0}^{\infty} \left|\left|(B_0^{\otimes 2})^i\right|\right| \left\{E\left[||M_{j,t-1-i}(\gamma_{j0},\lambda_0)||^2\right]\right\}^{1/2}. \tag{C.27}$$

Next,

$$\begin{aligned}
&\left\{E\left[||M_{j,t-1-i}(\gamma_{j0},\lambda_0)||^2\right]\right\}^{1/2} \\
= &\left\{E\left[\left|\left|\left\{\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right\}+\left\{I_2\otimes\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\right\}K_{2,2}\right|\right|^2\right]\right\}^{1/2} \\
\leq &\left\{E\left[\left(\left|\left|\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right|\right|+\left|\left|\left\{I_2\otimes\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\right\}K_{2,2}\right|\right|\right)^2\right]\right\}^{1/2} \\
\leq &\left\{E\left[\left|\left|\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right|\right|^2\right]\right\}^{1/2} \\
&+\left\{E\left[\left|\left|\left\{I_2\otimes\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\right\}K_{2,2}\right|\right|^2\right]\right\}^{1/2},
\end{aligned} \tag{C.28}$$

and

$$\begin{aligned}
&E\left[\left|\left|\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right|\right|^2\right] \\
= &E\left[\left|tr\left(\left\{\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right\}'\left\{\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right\}\right)\right|\right] \\
= &E\left[\left|tr\left(\left\{\left[\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)A'_0\right]\otimes I_2\right\}\left\{\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right\}\right)\right|\right] \\
= &E\left[\left|tr\left\{\left[\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)A'_0A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\otimes I_2\right\}\right|\right] \\
= &E\left[\left|tr\left[\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)A'_0A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]tr\left(I_2\right)\right|\right] \\
= &2E\left[\left|tr\left[\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)A'_0A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\right|\right] \\
\leq &2E\left[\left|\left|\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right|\right|^2\times||A_0||^2\right] \\
= &2\,||A_0||^2\,E\left[\left|\left|\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right|\right|^2\right] \\
\leq &2\,||A_0||^2\,E\left[\left(\left|\left|X_{j,t-1-i}X'_{j,t-1-i}\right|\right|+||\Gamma_{j0}||\right)^2\right] \\
= &2\,||A_0||^2\,E\left[\left(||X_{j,t-1-i}||^2+||\Gamma_{j0}||\right)^2\right] \\
\leq &K
\end{aligned} \tag{C.29}$$

since $E[||X_{jt}||^4]<\infty$ by Assumption 4.6. By a similar reasoning, one can also prove that

$$E\left[\left|\left|\left\{I_2\otimes\left[A_0\left(X_{j,t-1-i}X'_{j,t-1-i}-\Gamma_{j0}\right)\right]\right\}K_{2,2}\right|\right|^2\right]\leq K. \tag{C.30}$$

Combining (C.28), (C.29) and (C.30) with (C.27) we obtain

$$\sqrt{E\left[||G_{j,t-1}||^2\right]} \leq \sum_{i=0}^{\infty} \left|\left| \left(B_0^{\otimes 2}\right)^i \right|\right| K \leq \sum_{i=0}^{\infty} K\phi^i < \infty, \tag{C.31}$$

where we used equation (B.15) of PR. Hence, by (C.26) and (C.31) we finally obtain

$$E\left[\left(V_t^N\right)' \left(W_t^N\right)' W_t^N V_t^N\right] < \infty. \tag{C.32}$$

Finally, we consider $E[(V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N]$. Notice that the only difference between $E[(V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N]$ and $E[(V_t^N)'(W_t^N)'W_t^N V_t^N]$ is that the former is based on $\tilde{M}_{jt}(\gamma_{j0}, \lambda_0)$ instead of $M_{jt}(\gamma_{j0}, \lambda_0)$. Hence, it is straightforward to modify the arguments leading to (C.26) and obtain

$$\begin{aligned} E[(V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N] &\leq K\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} E\left[\left|\left|\tilde{W}_{jt}'\tilde{W}_{kt}\right|\right|\right] \\ &\leq K\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} \sqrt{E\left[||\tilde{G}_{j,t-1}||^2\right]}\sqrt{E\left[||\tilde{G}_{k,t-1}'||^2\right]}, \end{aligned} \tag{C.33}$$

where the last line follows from (C.25). Similarly, (C.27) can be modified to yield

$$\sqrt{E[||\tilde{G}_{j,t-1}||^2]} \leq \sum_{i=0}^{\infty} ||(B_0^{\otimes 2})^i|| \{E[||\tilde{M}_{j,t-1-i}(\gamma_{j0}, \lambda_0)||^2]\}^{1/2}. \tag{C.34}$$

Now,

$$\begin{aligned} \left\{E\left[\left|\left|\tilde{M}_{j,t-1-i}\left(\gamma_{j0}, \lambda_0\right)\right|\right|^2\right]\right\}^{1/2} &\leq \left\{E\left[||\{[B_0\left(H_{jt} - \Gamma_{j0}\right)] \otimes I_2\}||^2\right]\right\}^{1/2} \\ &\quad + \left\{E\left[||\{I_2 \otimes [B_0\left(H_{jt} - \Gamma_{j0}\right)]\} K_{2,2}||^2\right]\right\}^{1/2} \end{aligned}$$

and notice that,

$$\begin{aligned} E\left[||\{[B_0\left(H_{jt} - \Gamma_{j0}\right)] \otimes I_2\}||^2\right] &= E\left[\left|tr\left(\{[B_0\left(H_{jt} - \Gamma_{j0}\right)] \otimes I_2\}' \{[B_0\left(H_{jt} - \Gamma_{j0}\right)] \otimes I_2\}\right)\right|\right] \\ &= E\left[\left|tr\left(\{[\left(H_{jt} - \Gamma_{j0}\right) B_0'B_0\left(H_{jt} - \Gamma_{j0}\right)] \otimes I_2\}\right)\right|\right] \\ &\leq 2E\left[\left|tr\left\{[\left(H_{jt} - \Gamma_{j0}\right) B_0'B_0\left(H_{jt} - \Gamma_{j0}\right)]\right\}\right|\right] \end{aligned}$$

$$
\begin{aligned}
&= 2E\left[|tr\left\{\left[(H_{jt}-\Gamma_{j0})\,(H_{jt}-\Gamma_{j0})\,B_0'B_0\right]\right\}|\right] \\
&\leq 2E\left[||H_{jt}-\Gamma_{j0}||^2\,||B_0||^2\right] \\
&= 2\,||B_0||^2\,E\left[(||H_{jt}||+||\Gamma_{j0}||)^2\right] \\
&< \infty,
\end{aligned}
$$

since $E[||H_{jt}||^2] < \infty$ by (C.15). $E[||\{I_2 \otimes [B_0(H_{jt}-\Gamma_{j0})]\}K_{2,2}||^2]$ can be bounded in a similar way. Hence $\{E[||\tilde{M}_{j,t-1-i}(\gamma_{j0},\lambda_0)||^2]\}^{1/2} < \infty$ and together with (C.34) this implies that

$$
\sqrt{E[||\tilde{G}_{j,t-1}||^2]} \leq \sum_{i=0}^{\infty} ||(B_0^{\otimes 2})^i||\{E[||\tilde{M}_{j,t-1-i}(\gamma_{j0},\lambda_0)||^2]\}^{1/2} < \infty, \tag{C.35}
$$

using the same arguments as those leading to (C.31). Consequently, combining (C.33) and (C.35) we have

$$
E\left[(V_t^N)'(\tilde{W}_t^N)'\tilde{W}_t^N V_t^N\right] < \infty. \tag{C.36}
$$

Finally, taking the expectation of (C.20) and combining this with (C.23), (C.32) and (C.36) yields the stated result. ■

Lemma C.5 *Let Ω_0 be as defined in Section C.1. Then, for $\hat{\gamma}_j$ as defined in (13), we have*

$$
\sqrt{T}\begin{bmatrix} \hat{\gamma}-\gamma_0 \\ \partial l_{NT}(\gamma_0,\lambda_0)/\partial\lambda \end{bmatrix} \xrightarrow{d} N(0,\Omega_0) \qquad \text{as } T\to\infty.
$$

Proof of Lemma C.5. Let V_t^N, Q_t^N, W_t^N and $\tilde{W}_t^N$ be as defined in Section C.1. Notice that V_t^N is independent of $\mathcal{F}_{t-1}$, the information set at time $t-1$. Also, $((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')'$is measureable with respect to $\mathcal{F}_{t-1}$. Therefore, $(((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')'V_t^N, \mathcal{F}_t)$, $t = 1, ..., T$, is an ergodic martingale difference sequence for any N. Moreover, under the maintained assumptions $((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')'V_t^N$ is square integrable by Lemma C.4. Hence, by the same arguments as in the Proof of Lemma B.10 of PR a CLT applies and

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}((Q_t^N)', (W_t^N)', (\tilde{W}_t^N)')'V_t^N \xrightarrow{d} N(0,\Omega_0), \quad \text{as } T\to\infty. \tag{C.37}
$$

Lemma C.3 and (C.37) together yield the desired result. ■

Lemma C.6 *Let λ_i be the i^{th} entry of λ, where $i = 1, ..., 8$. Also, let θ_i be the i^{th} entry of θ, where $i = 1, ..., 4N + 8$. Then,*

$$\left| \sqrt{T} \left(\frac{\partial l_{NT}(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial l_{NT,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} \right) \right| \xrightarrow{p} 0 \quad \text{for } i = 1, ..., 8,$$
$$\sup_{\theta} \left| \frac{\partial^2 l_{NT}(\gamma, \lambda)}{\partial \theta_i \partial \theta_{i'}} - \frac{\partial^2 l_{NT,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_{i'}} \right| \xrightarrow{a.s.} 0 \quad \text{for } i, i' = 1, ..., 4N + 8,$$

as $T \to \infty$.

Proof of Lemma C.6. Under the maintained assumptions we can invoke Lemma B.11 of PR to obtain

$$\left| \sqrt{T} \left(\frac{\partial l_{jT}(\gamma_{j0}, \lambda_0)}{\partial \lambda_i} - \frac{\partial l_{jT,h}(\gamma_{j0}, \lambda_0)}{\partial \lambda_i} \right) \right| = o_p(1), \tag{C.38}$$
$$\sup_{\theta_j} \left| \frac{\partial^2 l_{jT}(\gamma_j, \lambda)}{\partial \theta_i \partial \theta_{i'}} - \frac{\partial^2 l_{jT,h}(\gamma_j, \lambda)}{\partial \theta_i \partial \theta_{i'}} \right| = o_{a.s.}(1), \tag{C.39}$$

for each pair $j = 1, ..., N$ as $T \to \infty$. Notice that (C.39) also holds when θ_i and/or $\theta_{i'}$ is not any of the parameters contained in $(\gamma_j', \lambda')'$, since in that case the derivatives are both identically equal to zero. Hence, given that N is fixed, the desired results follow by (C.38), (C.39) and the triangle inequality. ■

D Additional material

D.1 Additional material for Section 4.2

In this part we provide heuristic proofs for the consistency and asymptotic normality of the composite likelihood estimator discussed in Section 4.2. In what follows, in addition to the definitions made in Section 4.2 we also use $\hat{\theta} = (\hat{\theta}'_1, ..., \hat{\theta}'_N)'$, $\theta_0 = (\theta'_{10}, ..., \theta'_{N0})'$.

D.1.1 Consistency

Our argument follows along the same lines of Section 3.2 of Aielli (2013). In particular, we assume that

$$\sup_{\theta_j,\phi} \left|\left|\hat{S}_j(\theta_j, \phi) - S_{j0}(\theta_j, \phi)\right|\right| \xrightarrow{p} 0, \tag{D.1}$$

$$\sup_{\theta_j,S_j,\phi} |l_{jT}(\theta_j, S_j, \phi) - E[l_{jT}(\theta_j, S_j, \phi)]| \xrightarrow{p} 0, \tag{D.2}$$

for all j as $T \to \infty$. We also assume that it has already been established that

$$\hat{\theta}_j \xrightarrow{p} \theta_{j0}, \tag{D.3}$$

for all j as $T \to \infty$.

We start with the consistency of $\hat{\phi}$. First, (D.1) and (D.3) can be used to obtain

$$\sup_{\phi} ||\hat{S}_j(\hat{\theta}_j, \phi) - S_{j0}(\theta_{j0}, \phi)|| \xrightarrow{p} 0, \tag{D.4}$$

for all j as $T \to \infty$ (see, e.g., Theorem 3.7 of White (1994)). By (D.4) and (D.2) we have

$$\sup_{\phi} \left|\left|\hat{s}(\hat{\theta}, \phi) - s(\theta_0, \phi)\right|\right| \xrightarrow{p} 0, \tag{D.5}$$

$$\sup_{\theta,s,\phi} |l_{NT}(\theta, s, \phi) - E\left[l_{NT}(\theta, s, \phi)\right]| \xrightarrow{p} 0, \tag{D.6}$$

respectively. Then (again by Theorem 3.7 of White (1994) or a similar result), one can use

(D.3), (D.5) and (D.6) together to obtain

$$\sup_{\phi} \left| l_{NT}(\hat{\theta}, \hat{s}(\hat{\theta}, \phi), \phi) - E\left[l_{NT}(\theta_0, s(\theta_0, \phi), \phi)\right] \right| \xrightarrow{p} 0. \tag{D.7}$$

Remember that $\hat{\phi} = \arg\max_{\phi} l_{NT}(\hat{\theta}, \hat{s}(\hat{\theta}, \phi), \phi)$. Hence, if one can show that $E[l_{NT}(\theta_0, s(\theta_0, \phi), \phi)]$ is uniquely maximised at ϕ_0, then under certain conditions (D.7) would be sufficient to yield $\hat{\phi} \xrightarrow{p} \phi_0$ (e.g., Theorem 3.4 of White (1994)).

Finally, $\hat{\phi} \xrightarrow{p} \phi$, (D.1) and (D.3) can together be used to obtain $\hat{s}_j(\hat{\theta}_j, \hat{\phi}) \xrightarrow{p} s_j(\theta_{j0}, \phi_0)$ for all j as $T \to \infty$ - for a typical example of such a result see Lemma A.1 of Wooldridge (1994).

D.1.2 Inference

In discussing inference, in addition to the notation defined in Section 4.2 , we use the following notation: first, $\hat{s}_j(\theta_j, \phi)$, $s_j(\theta_j, \phi)$ and s_j are vectorised versions of $\hat{S}_j(\theta_j, \phi)$, $S_j(\theta_j, \phi)$ and S_j, where S_j is some (2×2) matrix. We also let $\theta = (\theta_1', ..., \theta_N')'$, and define $\hat{\theta}$ and θ_0 similarly. Moreover, $\hat{s}(\theta, \phi) = (\hat{s}_1(\theta_1, \phi)', ..., \hat{s}_N(\theta_N, \phi)')'$, and $s(\theta, \phi)$, and s are defined similarly. Next, we define the population estimating equation. First, let

$$\mathcal{M}(\theta) = (M_1(\theta_1)', ..., M_N(\theta_N)')',$$

where

$$M_j(\theta_j) = (d\dot{l}_{j_1T}(\eta_{j_1})/d\eta_{j_1}', d\dot{l}_{j_2T}(\eta_{j_2})/d\eta_{j_2}')'.$$

Under standard conditions, $E[\mathcal{M}(\theta_0)] = 0_{6N}$. Second, we define,

$$\mathcal{G}(\theta, s, \phi) = (G_1(\theta_1, s_1, \phi)', ..., G_N(\theta_N, s_N, \phi)')',$$

where

$$G_j(\theta_j, s_j, \phi) = \hat{s}_j(\theta_j, \phi) - s_j.$$

Notice that under stationarity $E[\hat{s}_j(\theta_j,\phi)] = s_j(\theta_j,\phi)$ for any (θ_j,ϕ) and j; therefore, $E[\mathcal{G}(\theta_0, s(\theta_0,\phi_0),\phi_0)] = 0_{4N}$. Third, we define

$$l_{NT}(\theta, s(\theta,\phi),\phi) = \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} l_{jt}(\theta_j, S_j(\theta_j,\phi),\phi).$$

From the identification condition for ϕ_0, it follows that $E[dl_{NT}(\theta_0, s(\theta_0,\phi_0),\phi_0)/d\phi] = 0_2$. Then, the population estimating equation is given by

$$g_{NT}(\theta, s,\phi) = (\mathcal{M}(\theta)', \mathcal{G}(\theta,s,\phi)', dl_{NT}(\theta, s(\theta,\phi),\phi)/d\phi')',$$

where $E[g_{NT}(\theta_0, s(\theta_0,\phi_0),\phi_0)] = 0_{10N+2}$.

Our argument is based on a series of mean-value expansions. In what follows, the "$\sim$" sign denotes a mean value; e.g. $\tilde{\phi}$ is a mean value between $\hat{\phi}$ and ϕ_0. Moreover, $\mathcal{R}$ is a generic remainder term, the exact value of which may differ from line to line. Also, to keep the notation concise we use the following shorthand notation for derivatives: $d_\phi = d/d\phi$, $d_{\phi\phi'} = d^2/d\phi d\phi'$, $d_{\phi\theta_j'} = d^2/d\phi d\theta_j'$ etc. Finally, we define $l_{NT}(\theta,\hat{s}(\theta,\phi),\phi) = (NT)^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T} l_{jt}(\theta_j, \hat{S}_j(\theta_j,\phi),\phi)$.

We start with the expansion of $d_\phi l_{NT}(\hat{\theta},\hat{s}(\hat{\theta},\hat{\phi}),\hat{\phi})$ about $\hat{\phi} = \phi_0$:

$$\begin{aligned} d_\phi l_{NT}(\hat{\theta},\hat{s}(\hat{\theta},\hat{\phi}),\hat{\phi}) &= d_\phi l_{NT}(\hat{\theta},\hat{s}(\hat{\theta},\phi_0),\phi_0) + [d_{\phi\phi'} l_{NT}(\hat{\theta},\hat{s}(\hat{\theta},\tilde{\phi}),\tilde{\phi})](\hat{\phi}-\phi_0) \\ &= d_\phi l_{NT}(\hat{\theta},\hat{s}(\hat{\theta},\phi_0),\phi_0) \\ &\quad + E[d_{\phi\phi'} l_{NT}(\theta_0, s\left(\theta_0,\phi_0\right),\phi_0)](\hat{\phi}-\phi_0) + \mathcal{R}, \end{aligned} \tag{D.8}$$

where, remembering that $\hat{\theta} \xrightarrow{p} \theta_0$, $\hat{\phi} \xrightarrow{p} \phi_0$ and $\hat{s}(\hat{\theta},\hat{\phi}) \xrightarrow{p} s(\theta_0,\phi_0)$ as obtained in Section D.1.1, we assume that

$$d_{\phi\phi'} l_{NT}(\hat{\theta},\hat{s}(\hat{\theta},\tilde{\phi}),\tilde{\phi}) - E[d_{\phi\phi'} l_{NT}(\theta_0, s\left(\theta_0,\phi_0\right),\phi_0)] \xrightarrow{p} 0_{2\times 2}.$$

In the following expansions we will implicitly assume that similar convergence results hold for the terms evaluated at parameter estimates and/or mean values. Next, by similar ideas, expanding $d_\phi l_{jT}(\hat{\theta}_j, \hat{S}_j(\hat{\theta}_j, \phi_0), \phi_0)$ around $\hat{\theta}_j = \theta_{j0}$ yields

$$\begin{aligned} d_\phi l_{jT}(\hat{\theta}_j, \hat{S}_j(\hat{\theta}_j, \phi_0), \phi_0) &= d_\phi l_{jT}(\theta_{j0}, \hat{S}_j(\theta_{j0}, \phi_0), \phi_0) + d_{\phi\theta_j'} l_{jT}(\tilde{\theta}_j, \hat{S}_j(\tilde{\theta}_j, \phi_0), \phi_0)(\hat{\theta}_j - \theta_{j0}), \\ &= d_\phi l_{jT}(\theta_{j0}, \hat{S}_j(\theta_{j0}, \phi_0), \phi_0) \\ &\quad + E\left[d_{\phi\theta_j'} l_{jT}(\theta_{j0}, S_j(\theta_{j0}, \phi_0), \phi_0)\right](\hat{\theta}_j - \theta_{j0}) + \mathcal{R}. \end{aligned} \tag{D.9}$$

Next, we expand $d_\phi l_{jT}(\theta_{j0}, \hat{S}_j(\theta_{j0}, \phi_0), \phi_0)$ around $\hat{s}_j(\theta_{j0}, \phi_0) = s_j(\theta_{j0}, \phi_0)$ and obtain,

$$\begin{aligned} d_\phi l_{jT}(\theta_{j0}, \hat{S}_j(\theta_{j0}, \phi_0), \phi_0) &= d_\phi l_{jT}(\theta_{j0}, S_j(\theta_{j0}, \phi_0), \phi_0) \\ &\quad + d_{\phi s_j'} l_{jT}(\theta_{j0}, \tilde{S}_j(\theta_{j0}, \phi_0), \phi_0)[\hat{s}_j(\theta_{j0}, \phi_0) - s_j(\theta_{j0}, \phi_0)] \\ &= d_\phi l_{jT}(\theta_{j0}, S_j(\theta_{j0}, \phi_0), \phi_0) \\ &\quad + E\left[d_{\phi s_j'} l_{jT}(\theta_{j0}, S_j(\theta_{j0}, \phi_0), \phi_0)\right] G_j(\theta_{j0}, s_j(\theta_{j0}, \phi_0), \phi_0) \\ &\quad + \mathcal{R}, \end{aligned} \tag{D.10}$$

where in obtaining the second equality we used $G_j(\theta_{j0}, s_j(\theta_{j0}, \phi_0), \phi_0) = \hat{s}_j(\theta_{j0}, \phi_0) - s_j(\theta_{j0}, \phi_0)$. Next, expanding $d_{\eta_{j_1}} \dot{l}_{j_1T}(\hat{\eta}_{j_1})$ around $\hat{\eta}_{j_1} = \eta_{j_10}$, by similar arguments as before we have

$$d_{\eta_{j_1}} \dot{l}_{j_1T}(\hat{\eta}_{j_1}) = d_{\eta_{j_1}} \dot{l}_{j_1T}(\eta_{j_10}) + E[d_{\eta_{j_1}\eta_{j_1}'} \dot{l}_{j_1T}(\eta_{j_10})](\hat{\eta}_{j_1} - \eta_{j_10}) + \mathcal{R},$$

and similarly for $d_{\eta_{j_2}} \dot{l}_{j_2T}(\hat{\eta}_{j_2})$. Then, remembering that $d_{\eta_{j_1}} \dot{l}_{j_1T}(\hat{\eta}_{j_1}) = d_{\eta_{j_2}} \dot{l}_{j_2T}(\hat{\eta}_{j_2}) = 0_3$ and $\theta_j = (\eta_{j_1}', \eta_{j_2}')'$, we have

$$(\hat{\theta}_j - \theta_{j0}) = \begin{bmatrix} \left\{E[-d_{\eta_{j_1}\eta_{j_1}'} \dot{l}_{j_1T}(\eta_{j_10})]\right\}^{-1} d_{\eta_{j_1}} \dot{l}_{j_1T}(\eta_{j_10}) \\ \left\{E[-d_{\eta_{j_2}\eta_{j_2}'} \dot{l}_{j_2T}(\eta_{j_20})]\right\}^{-1} d_{\eta_{j_2}} \dot{l}_{j_2T}(\eta_{j_20}) \end{bmatrix} + \mathcal{R}. \tag{D.11}$$

Bringing (D.8)-(D.11) together and remembering that $d_\phi l_{NT}(\hat{\theta}, \hat{s}(\hat{\theta}, \hat{\phi}), \hat{\phi}) = 0_2$, by definition, leads to

$$0_2 = d_\phi l_{NT}(\theta_0, s(\theta_0, \phi_0), \phi_0)$$

$$+\frac{1}{N}\sum_{j=1}^{N}\underbrace{E\left[d_{\phi s_j'}l_{jT}(\theta_{j0},S_j(\theta_{j0},\phi_0),\phi_0)\right]G_j(\theta_{j0},s_j(\theta_{j0},\phi_0),\phi_0)}_{A_{jT}}$$

$$-\frac{1}{N}\sum_{j=1}^{N}\underbrace{E\left[d_{\phi\theta_j'}l_{jT}(\theta_{j0},S_j(\theta_{j0},\phi_0),\phi_0)\right]\begin{bmatrix}\left\{E[d_{\eta_{j_1}\eta_{j_1}'}\dot{l}_{j_1T}(\eta_{j_10})]\right\}^{-1}d_{\eta_{j_1}}\dot{l}_{j_1T}(\eta_{j_10})\\ \left\{E[d_{\eta_{j_2}\eta_{j_2}'}\dot{l}_{j_2T}(\eta_{j_20})]\right\}^{-1}d_{\eta_{j_2}}\dot{l}_{j_2T}(\eta_{j_20})\end{bmatrix}}_{B_{jT}}$$

$$+E[d_{\phi\phi'}l_{NT}(\theta_0,s\left(\theta_0,\phi_0\right),\phi_0)](\hat{\phi}-\phi_0)+\mathcal{R}. \tag{D.12}$$

Notice that,

$$\frac{1}{N}\sum_{j=1}^{N}A_{jT}=E[d_{\phi s'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)]\mathcal{G}(\theta_0,s(\theta_0,\phi_0),\phi_0), \tag{D.13}$$

and

$$\frac{1}{N}\sum_{j=1}^{N}B_{jT}=E[d_{\phi\theta'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)]\{E[d_{\theta'}\mathcal{M}(\theta_0)]\}^{-1}\mathcal{M}(\theta_0). \tag{D.14}$$

Then, solving (D.12) for $(\hat{\phi}-\phi_0)$, and using (D.13) and (D.14) yields

$$\begin{aligned}\hat{\phi}-\phi_0 &= -\{E[d_{\phi\phi'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)]\}^{-1}d_{\phi}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)\\ &\quad -\{E[d_{\phi\phi'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)]\}^{-1}E[d_{\phi s'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)]\mathcal{G}(\theta_0,s(\theta_0,\phi_0),\phi_0)\\ &\quad +\{E[d_{\phi\phi'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)]\}^{-1}E[d_{\phi\theta'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)]\{E[d_{\theta'}\mathcal{M}(\theta_0)]\}^{-1}\mathcal{M}(\theta_0)\\ &\quad +\mathcal{R}\\ &= J_{NT}g_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)+\mathcal{R}, \end{aligned} \tag{D.15}$$

where

$$J_{NT}=\left[J_{NT;\phi\phi'}^{-1}J_{NT;\phi\theta'}\{E[d_{\theta'}\mathcal{M}(\theta_0)]\}^{-1}\quad -J_{NT;\phi\phi'}^{-1}J_{NT;\phi s'}\quad -J_{NT;\phi\phi'}^{-1}\right],$$

and

$$J_{NT;\phi\phi'} = E[d_{\phi\phi'}l_{NT}(\theta_0,s(\theta_0,\phi_0),\phi_0)],$$

$$J_{NT;\phi s'} = E[d_{\phi s'} l_{NT}(\theta_0, s(\theta_0, \phi_0), \phi_0)],$$
$$J_{NT;\phi\theta'} = E[d_{\phi\theta'} l_{NT}(\theta_0, s(\theta_0, \phi_0), \phi_0)].$$

Now, letting

$$\Sigma_N = \lim_{T\to\infty} Var(\sqrt{T} g_{NT}(\theta_0, s(\theta_0, \phi_0), \phi_0)), \tag{D.16}$$
$$J_N = \lim_{T\to\infty} J_{NT}, \tag{D.17}$$

if (i) the remainder $\mathcal{R}$ in expansion (D.15) is $o_p(T^{-1/2})$, and (ii) $g_{NT}(\theta_0, s(\theta_0, \phi_0), \phi_0) \xrightarrow{d} N(0_{10N\times 2}, \Sigma_N)$, then (D.15) yields,

$$\sqrt{T}(\hat{\phi} - \phi_0) \xrightarrow{d} N(0_2, J_N \Sigma_N J_N') \quad \text{as } T \to \infty,$$

as desired. The final asymptotic expansion in (D.15) is a combination of a multitude of expansions. As such, its remainder has a complicated form and obtaining its rate will require substantial work. As for the second requirement, since $E[g_{NT}(\theta_0, s(\theta_0, \phi_0), \phi_0)] = 0_{10N+2}$ it is reasonable to expect that this term satisfies a central limit theorem. However, proving that such a result exists will, again, be subject of substantial work. We note that these comments are not peculiar to the composite likelihood method - similar conditions would also be required for obtaining the asymptotic distribution for the maximum likelihood estimator of the DCC/cDCC model. The complications arise not due to the estimation method employed, but rather due to the underlying model itself (DCC/cDCC).

D.2 Additional material for Section 5.2

In this section we provide the additional Figure D.1, which presents the results of the efficiency analysis of Section 5.2 for the BEKK model.

D.3 Additional material for Section 5.4

Figures D.2 and D.3 replicate Figure 2 in Section 5.4 in the main paper by reducing the sample size from 2000 to 500 (Figure D.2) and increasing the cross-sectional dimension to 200 (Figure D.3). Reducing the sample size increases the bias in the estimates of β when

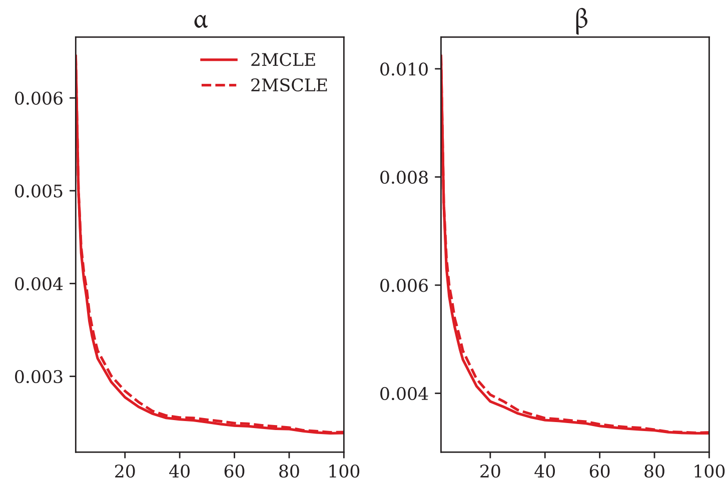

Figure D.1: Standard deviation for the CL estimator based on all pairs (2MCLE) and on a subset of pairs (2MSCLE), as L varies from 2 to 100. Calculated from simulated data for the scalar BEKK model with $\alpha = .05$, $\beta = .93$.

the number of moments is 6 or 8. When the number of moments is only 4, decreasing the sample size reduces the bias, suggesting that the asymptotic bias term may dominate the distribution when out theory does not apply. Figure D.4 produces similar Q-Q plots for parameters estimated using 2MLE. While these parameters appear to lie along the 45-degree line, they are severely biased. The bias is reflected in the y-axis values which range from -6 to -26 for α and 0 to 13 for β, depending on the model configuration.

D.4 Additional material for Section 5.5

To measure the effects of changing L and T on the conditioning numbers of the 2MLE and 2MCLE estimators in Section 5.5, we consider a simple regression specification,

$$y_{LT,r} = \beta_0 + \beta_1 \ln L + \beta_2 \ln T + \eta_{LT,r},$$

where $y_{LT,r}$ is one of the two measures, $\bar{u}_{LT}$ or $\bar{u}_{c,LT}$ and r is the replication index. The idea here is that β_1 and β_2 are measures of the exponents of L and T, in the sense that $y_{LT,r} = O(\ln L^{\beta_1} T^{\beta_2})$. The estimates are $\hat{\beta}_1 = 1.0046$ and $\hat{\beta}_2 = -0.8383$ for 2MLE, and $\hat{\beta}_1 = 0.0487$ and $\hat{\beta}_2 = -0.1448$ for 2MCLE. Hence, in line with our earlier observation,

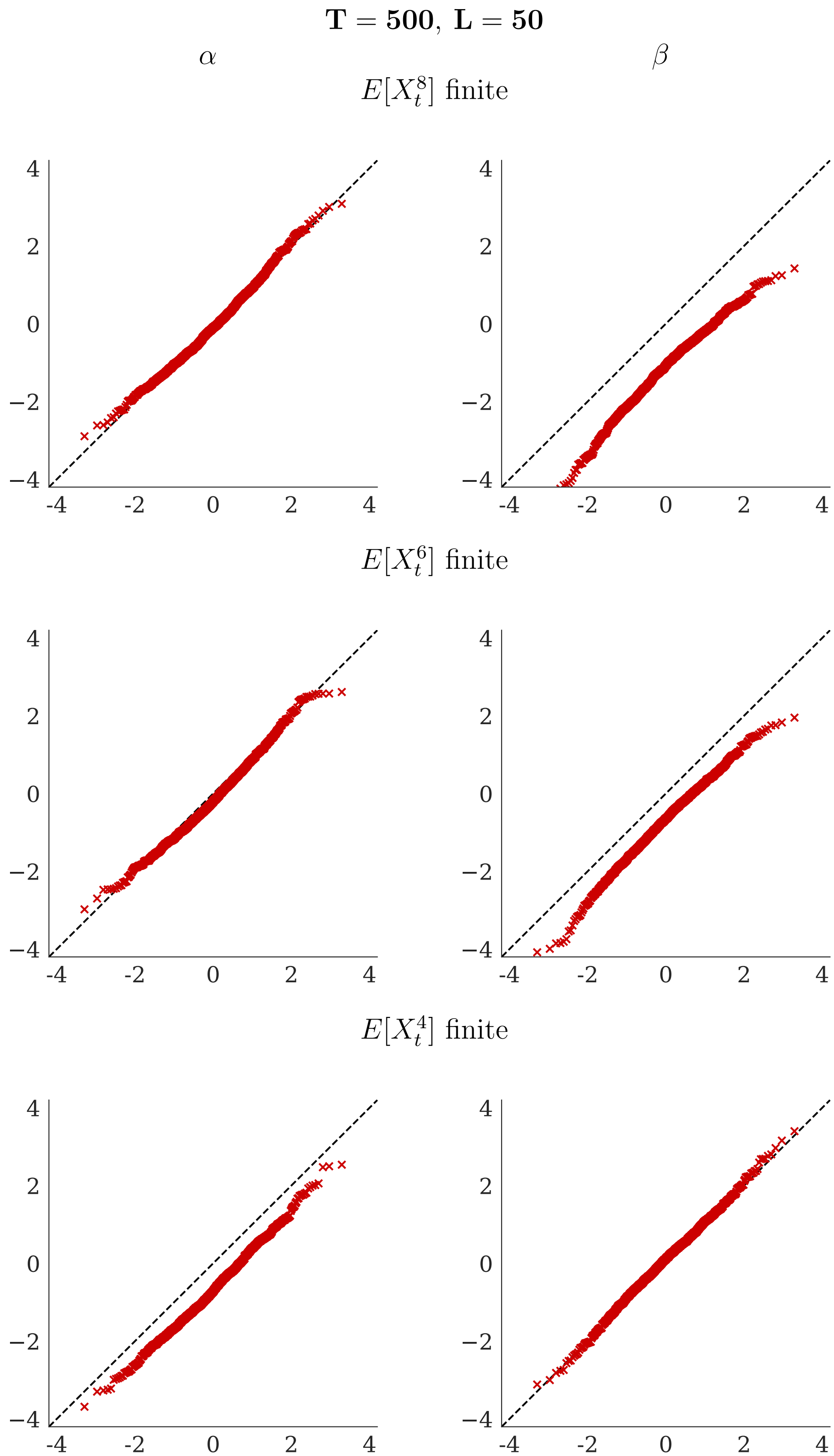

Figure D.2: Q-Q plots of estimates of α and β from scalar BEKK models parameterised to have 8, 6 and 4 finite moments. The normalised parameter errors are plotted along the y-axis. All estimates were produced from models with $L = 50$ and $T = 500$ using 2MCLE.

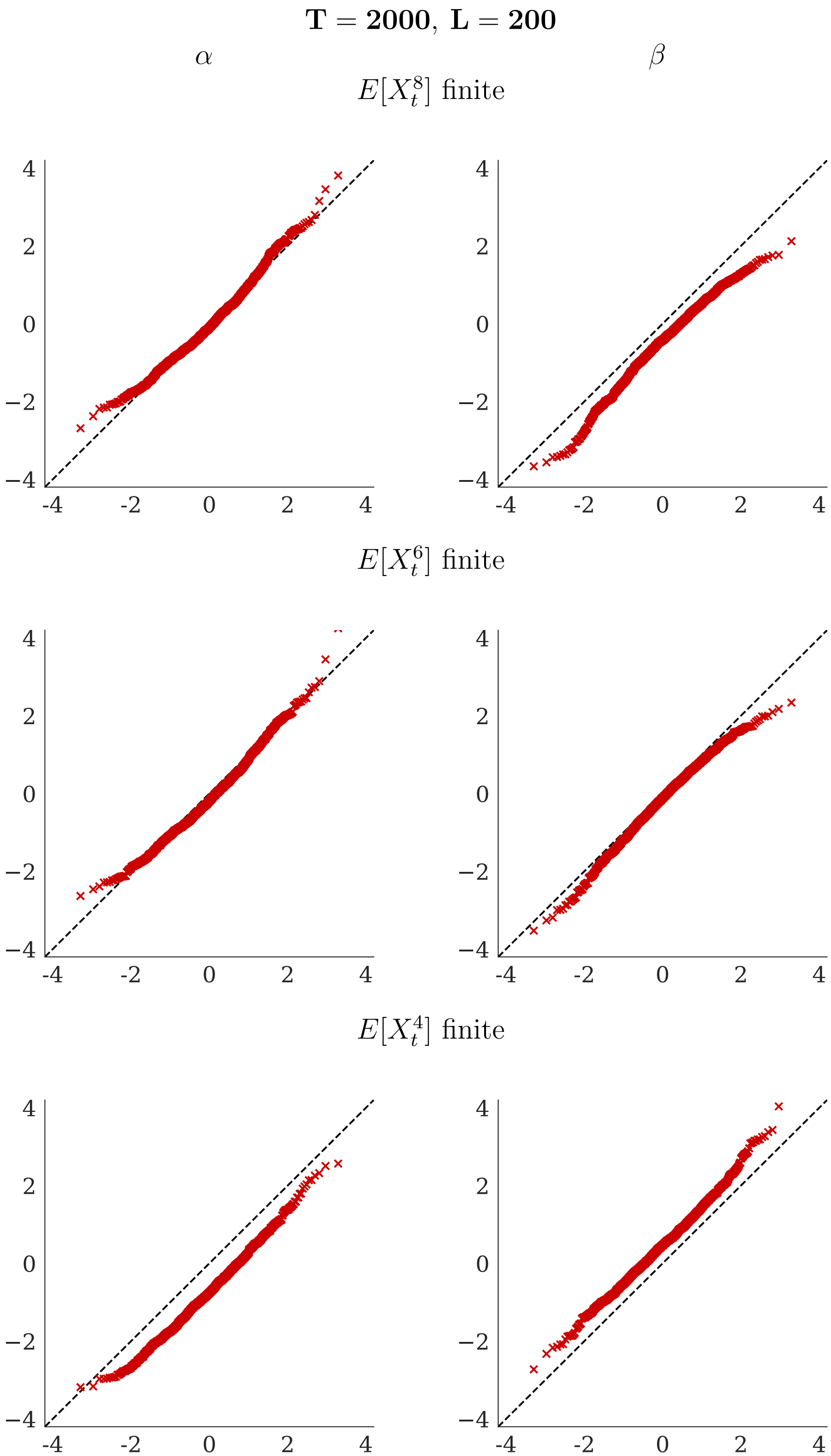

Figure D.3: Q-Q plots of estimates of α and β from scalar BEKK models parameterised to have 8, 6 and 4 finite moments. The normalised parameter errors are plotted along the y-axis. All estimates were produced from models with $L = 200$ and $T = 2000$ using 2MCLE.

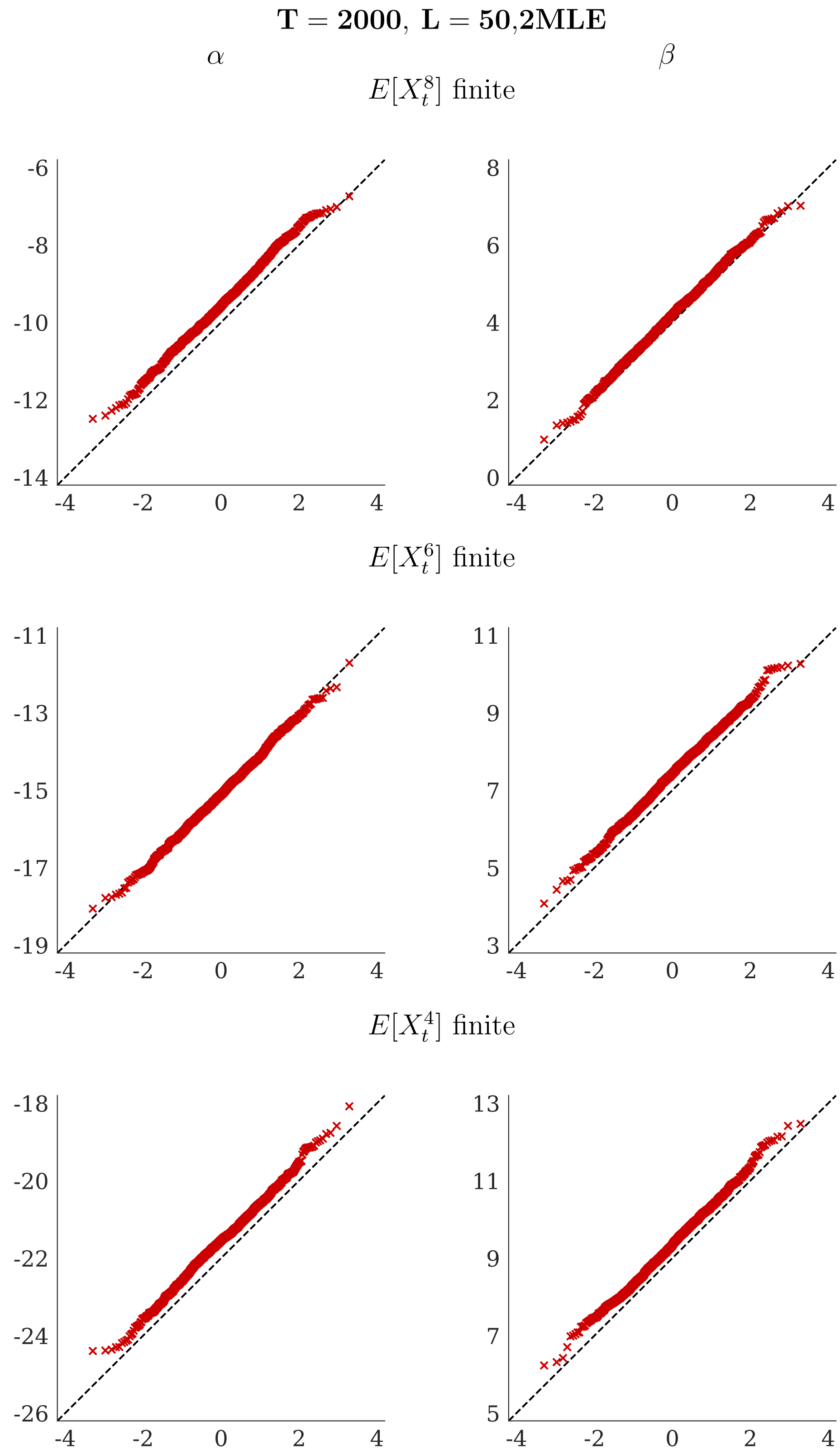

Figure D.4: Q-Q plots of estimates of α and β from scalar BEKK models parameterised to have 8, 6 and 4 finite moments. The normalised parameter errors are plotted along the y-axis. All estimates were produced from models with $L = 50$ and $T = 2000$ using 2MLE.

for both methods, larger values of L increase the estimation error while larger values of T decrease the error. The long-run estimator is an important source of noise in these problems, and an ill-conditioned target influences the precision of the dynamic parameters. In extreme cases, when the cross-section size is larger than the time-series dimension, the 2MLE estimator is not feasible. This is reflected in $\hat{\beta}_1$ and $\hat{\beta}_2$: for 2MLE the effect of L is of a greater magnitude than T, whereas it is the other way around for 2MCLE. These indicate that diagonally increasing both the cross-section size and the sample size will not ruin the 2MCLE estimator by creating large errors in the fitted values of H_{jt}. On the other hand, for the 2MLE, growing the cross-section size along with the sample length will produce large errors. This is consistent with our Monte Carlo findings.

References

Aielli, G. P. (2013). Consistent estimation of large scale dynamic conditional correlations. *Journal of Business and Economic Statistics 31*, 282–299.

Hafner, C. M. and A. Preminger (2009). Asymptotic theory for a factor GARCH model. *Econometric Theory 25*, 336–363.

Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: Wiley.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians* (2 ed.). Orlando: Academic Press.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2639–2738. North-Holland.