Dynamic score driven independent component analysis
Supplementary Appendix

Christian M. Hafner* Helmut Herwartz†

November 28, 2021

Appendices

C Proofs of the theorems

Let $\tilde{\delta}_t$ denote the process (2.3) where the starting value $\tilde{\delta}_0$ is drawn from the stationary distribution of $\tilde{\delta}_t$, and let $\tilde{L}(\theta)$ and $\tilde{l}_t(\theta)$ be defined accordingly. In practice these are not feasible and a fixed starting value δ_0 has to be used. One of the challenges is to show that this does not affect the asymptotic properties of the estimator. The true parameter is denoted θ_0, and we will use the notation $\delta_t = \delta_t(\theta)$ and $\delta_{0t} = \delta_t(\theta_0)$. Furthermore, let $C > 0$ denote a generic absolute constant that may take different values in different equations.

Lemma 1. *Under Assumptions (A1)-(A3), the model (2.1) is identifiable up to a permutation of the columns of $R(\delta_t)$. Under Assumptions (A1)-(A3) and (A5), it is*

*Corresponding author, Louvain Institute of Data Analysis and Modeling in Economics and Statistics (LIDAM), and ISBA, Université catholique de Louvain, Belgium. e-mail `christian.hafner@uclouvain.be`

†Department of Economics, University of Göttingen, Germany. e-mail `hherwartz@uni-goettingen.de`.

uniquely identifiable.

Proof: First, from (2.4) and (2.5), $\xi_{1t}(\theta)$ and $\xi_{2t}(\theta)$ are linear combinations of the true innovations, ξ_{1t} and ξ_{2t} with, by construction, identical second moment structure. The distributions of $\xi_t(\theta)$ and ξ_t would be identical under Gaussianity, which is excluded by Assumption (A3). As shown by Lancaster (1954), under the requirement of independence and "non-trivial" rotations, the only situation where non-identifiability occurs is the case of a normal distribution. As for 'trivial' rotations: Rotations by 180°, which correspond to the sign changes $\xi_t(\delta_{0t}-\pi) = -\xi_t$, are excluded by the restriction $|\delta_t - \delta_{0t}| < \pi$. Rotations by 90° correspond to a change in the order of the columns of $R(\delta_t)$: $\xi_{1t}(\delta_{0t} - \pi/2) = \xi_{2t}$ and $\xi_{2t}(\delta_{0t} - \pi/2) = \xi_{1t}$. This is not excluded by our restrictions and so $R(\delta_t)$ is identified only up to a change of columns under Assumptions (A1)-(A3). Adding Assumption (A5), the distributions of ξ_{1t} and ξ_{2t} are different, so that $R(\delta_t)$ is uniquely identified.

It remains to show that $\delta_t(\theta) = \delta_t(\theta_0)$ a.s. for all t implies that $\theta = \theta_0$. Let $\delta_t(\theta) = \omega + \psi(L)u_t$ and suppose that $\delta_t(\theta) = \delta_t(\theta_0)$ a.s. for all t. Then $\delta_t(\theta) - \delta_t(\theta_0) = 0 \Rightarrow \omega - \omega_0 = \left[\psi(L) - \psi_0(L)\right] u_t$ a.s. for all t. If $\psi(L) \neq \psi_0(L)$, the right hand side consists of some linear combination of u_{t-j}, $j > 0$, which has a nondegenerate distribution by Assumption (A2) and, hence, is not constant. Therefore, $\psi(L) = \psi_0(L)$ and $\omega = \omega_0$. $\square$

Proof of Theorem 1: We show strong consistency by contradiction. Suppose that $\hat{\theta}_T \not\to \theta_0$ a.s., so that for some arbitrary $\eta > 0$ the set $\Omega = \{\omega : \lim_{T\to\infty} \|\hat{\theta}_T - \theta_0\| \geq \eta\}$ has a positive probability. Since the set $\Lambda = \Theta \cap \{\theta : \|\theta - \theta_0\| \geq \eta\}$ is compact, we can find for each $\omega \in \Omega$ a convergent subsequence $\hat{\theta}_{T_i} \to \tilde{\theta} \in \Lambda$. From the definition of the QMLE,

$$\limsup_{i\to\infty} \frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\theta_0) \le \limsup_{i\to\infty}\sup_{\theta\in\Lambda} \frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\theta) = \limsup_{i\to\infty} \frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\hat{\theta}_{T_i}).$$

Lemma 3 implies that this also holds for the functions $\tilde{l}_t(\cdot)$, i.e.

$$\limsup_{i\to\infty} \frac{1}{T_i}\sum_{t=1}^{T_i} \tilde{l}_t(\theta_0) \le \limsup_{i\to\infty}\sup_{\theta\in\Lambda} \frac{1}{T_i}\sum_{t=1}^{T_i} \tilde{l}_t(\theta) = \limsup_{i\to\infty} \frac{1}{T_i}\sum_{t=1}^{T_i} \tilde{l}_t(\hat{\theta}_{T_i}).$$

From Lemmas 2(i) and 2(iii), $\mathrm{E}\tilde{l}_t(\theta_0) \le \mathrm{E}\sup_{\theta\in\Lambda} \tilde{l}_t(\theta)$, which contradicts Lemma 2(ii). Since $\eta > 0$ is arbitrary, the strong consistency follows. $\square$

Lemma 2: Under Assumptions (A1)-(A7),

i) $\frac{1}{T}\sum_{t=1}^{T} \tilde{l}_t(\theta_0) \to \mathrm{E}\tilde{l}_t(\theta_0),\ a.s.$

ii) $\mathrm{E}\tilde{l}_t(\theta_0) \ge \mathrm{E}\tilde{l}_t(\theta)$ with equality if and only if $\theta = \theta_0$.

iii) For any compact set $\Lambda \subseteq \Theta$, $\limsup_{T\to\infty}\sup_{\theta\in\Lambda} \frac{1}{T}\sum_{t=1}^{T} \tilde{l}_t(\theta) \le \mathrm{E}\sup_{\theta\in\Lambda} \tilde{l}_t(\theta)$ a.s.

Proof:

i) Since $\mathrm{E}\tilde{l}_t(\theta_0) = \mathrm{E}\sum_i \log g_i(\xi_{it})$ and $\mathrm{E}|\log g_i(\xi_{it})|$ is bounded by Assumption (A4), the desired result follows from the ergodic theorem.

ii) Note that from (2.7), $\xi_{1t}^2(\delta_t) = a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}$ and $\xi_{2t}^2(\delta_t) = (1-a_t)\xi_{1t}^2 + a_t\xi_{2t}^2 - b_t\xi_{1t}\xi_{2t}$, where $a_t = \cos^2(\Delta_t)$ and $b_t = \sin(2\Delta_t)$. Hence,

$$\begin{aligned}
\mathrm{E}\tilde{l}_t(\theta) - \mathrm{E}\tilde{l}_t(\theta_0) &= \frac{\nu_1+1}{2}\mathrm{E}\log(\nu_1 - 2 + \xi_{1t}^2) + \frac{\nu_2+1}{2}\mathrm{E}\log(\nu_2 - 2 + \xi_{2t}^2) \\
&- \frac{\nu_1+1}{2}\mathrm{E}\log(\nu_1 - 2 + a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}) \\
&- \frac{\nu_2+1}{2}\mathrm{E}\log(\nu_2 - 2 + (1-a_t)\xi_{1t}^2 + a_t\xi_{2t}^2 - b_t\xi_{1t}\xi_{2t})
\end{aligned}$$

It suffices to show that

$$\mathrm{E}\log(\nu_1 - 2 + \xi_{1t}^2) - \mathrm{E}\log(\nu_1 - 2 + a_t\xi_{it}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}) \le 0. \quad \text{(C.1)}$$

Conditioning on $\mathcal{F}_{t-1}$ and denoting $\mathrm{E}_t(\cdot) = \mathrm{E}[\cdot|\mathcal{F}_{t-1}]$, we have due to the law of iterated expectation

$$\mathrm{E}[\log(\nu_1 - 2 + \xi_{1t}^2) - \mathrm{E}_t\log(\nu_1 - 2 + a_t\xi_{it}^2 + (1-a_t)\xi_{2t}^2) + b_t\xi_{1t}\xi_{2t}] \le 0,$$

where $a_t \in [0,1)$ and $b_t \in (-1,1)$ are $\mathcal{F}_{t-1}$-measurable. Therefore, Assumption (A7) applies and (C.1) holds.

iii) For any compact subset $\Lambda \subseteq \Theta$ we have that

$$\limsup_{T\to\infty} \sup_{\bar\theta\in\Lambda} \frac{1}{T}\sum_{t=1}^{T} l_t(\bar\theta) \le \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sup_{\bar\theta\in\Lambda} l_t(\bar\theta).$$

Now from Lemma 2(i) we have that $\mathrm{E}\sup_{\bar\theta\in\Lambda}\tilde{l}_t(\bar\theta) \in \mathbb{R}\cup\{+\infty\}$ and since $\{\sup_{\bar\theta\in\Lambda} l_t(\bar\theta)\}$ is a stationary and ergodic process we have that (see e.g. Pfanzagl (1969))

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sup_{\bar\theta\in N} l_t(\bar\theta) \le \mathrm{E}\sup_{\bar\theta\in N} l_t(\bar\theta),$$

and the desired result follows. □

Lemma 3: Under Assumptions (A1)-(A11) we have that

$$\limsup_{T\to\infty} \sup_{\theta\in\Theta} \left| \tfrac{1}{T}\sum_{t=1}^{T} l_t(\theta) - \tfrac{1}{T}\sum_{t=1}^{T} \tilde{l}_t(\theta) \right| = 0 \;\; a.s.$$

Proof: By Cesaro's mean theorem it is sufficient to check that $\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)|$ tends to zero a.s. If $\mathrm{E}\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)|^r$ is bounded by a summable sequence in t

for some $r > 0$, by the generalized Chebyshev inequality, $\sum_{t=1}^{\infty} \mathrm{P} \sup_{\theta\in\Theta} |l_t(\theta) - \tilde{l}_t(\theta)| > \eta < \infty$, for all $\eta > 0$. Then the desired result follows by the Borel-Cantelli lemma.

Let $a_t = \cos^2(\delta_t - \delta_{0t})$, $\tilde{a}_t = \cos^2(\tilde{\delta}_t - \tilde{\delta}_{0t})$, $b_t = \sin(2\Delta_t)$ and $\tilde{b}_t = \sin(2\tilde{\Delta}_t)$. Furthermore, let $\tilde{\xi}_{it} = \xi_{it}(\tilde{\delta}_{0t})$ and note that $\xi_t - \tilde{\xi}_t = (R'(\delta_{0t})R(\tilde{\delta}_{0t}) - I_2)\tilde{\xi}_t$, where $R'(\delta_{0t})R(\tilde{\delta}_{0t})$ is a matrix of bounded and continuously differentiable functions, so that there exists a constant $C > 0$ such that

$$\|\xi_t - \tilde{\xi}_t\| \leq C|\delta_{0t} - \tilde{\delta}_{0t}|\|\tilde{\xi}_t\|. \tag{C.2}$$

Then,

$$\begin{aligned}
l_t(\theta) - \tilde{l}_t(\theta) &= \frac{\nu_1+1}{2}\log(\nu_1 - 2 + \tilde{a}_t\tilde{\xi}_{1t}^2 + (1-\tilde{a}_t)\tilde{\xi}_{2t}^2) + b_t\tilde{\xi}_{1t}\tilde{\xi}_{2t} \\
&+ \frac{\nu_2+1}{2}\log(\nu_2 - 2 + (1-\tilde{a}_t)\tilde{\xi}_{1t}^2 + \tilde{a}_t\tilde{\xi}_{2t}^2) - b_t\tilde{\xi}_{1t}\tilde{\xi}_{2t} \\
&- \frac{\nu_1+1}{2}\log(\nu_1 - 2 + a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2) + b_t\xi_{1t}\xi_{2t} \\
&- \frac{\nu_2+1}{2}\log(\nu_2 - 2 + (1-a_t)\xi_{1t}^2 + a_t\xi_{2t}^2) - b_t\xi_{1t}\xi_{2t}
\end{aligned}$$

It suffices to show that, for some $r > 0$ and $\rho \in (0,1)$,

$$\mathrm{E}\sup_{\theta\in\Theta}\left|\log(h(\delta_t)) - \log(h(\tilde{\delta}_t))\right|^r = O(\rho^t) \tag{C.3}$$

where $h(\delta_t) := \nu_1 - 2 + a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}$. Note that $a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t} \geq 0$ a.s., and that by Assumption (A6), $h(\delta_t)$ is bounded below by an absolute constant $\epsilon > 0$. Moreover, a_t and b_t are bounded, continuously differentiable functions of $\Delta_t = \delta_t - \delta_{0t}$. Hence, $\log(h(\delta_t))$ is a Lipschitz-continuous function, since it has a bounded first derivative. An application of the mean value theorem implies that there exists a constant $C > 0$ such that

$$\left|\log(h(\delta_t)) - \log(h(\tilde{\delta}_t))\right| \leq C\left|h(\delta_t) - h(\tilde{\delta}_t)\right| \tag{C.4}$$

Because a_t and b_t are bounded functions, we further have

$$\left|h(\delta_t) - h(\tilde{\delta}_t)\right| \leq C\|\xi_t - \tilde{\xi}_t\| \leq C|\delta_{0t} - \tilde{\delta}_{0t}|\|\tilde{\xi}_t\| \tag{C.5}$$

using (C.2). Therefore,

$$\sup_{\theta\in\Theta}\left|\log(h(\delta_t)) - \log(h(\tilde{\delta}_t))\right|^r \leq C\|\tilde{\xi}_t\|^r|\delta_{0t} - \tilde{\delta}_{0t}|^r \tag{C.6}$$

For $0 < r < 1$, $\mathrm{E}[\|\tilde{\xi}_t\|^r] \leq 2^r$ by Assumption (A1) and Jensen's inequality.

To establish (C.3) it remains to show that $\mathrm{E}|\delta_{0t} - \tilde{\delta}_{0t}|^r = O(\rho^t)$. By the definition of δ_t in (2.3), we have

$$\delta_t - \tilde{\delta}_t = \beta(\delta_{t-1} - \tilde{\delta}_{t-1}) + \kappa(u_{t-1} - \tilde{u}_{t-1}) \tag{C.7}$$

By the mean value theorem, $u_t - \tilde{u}_t = \frac{\partial u_t}{\partial \delta_t}(\bar{\delta}_t)(\delta_t - \tilde{\delta}_t)$ where $\bar{\delta}_t$ is on the line segment joining δ_t and $\tilde{\delta}_t$, so that

$$\delta_t - \tilde{\delta}_t = \prod_{i=1}^{t-1} x_{t-i}(\bar{\delta}_{t-i})(\delta_0 - \tilde{\delta}_0),$$

where $x_t = \beta + \kappa\partial u_t/\partial\delta_t(\bar{\delta}_t)$. Note that $|x_t|$ is a stationary ergodic sequence of non-negative random variables with $\mathrm{E}\log|x_t| < 0$, which is implied by Assumption (A11). Thus, by Lemma 2.4 of Straumann and Mikosch (2006), $\prod_{i=1}^{t-1}|x_{t-i}(\bar{\delta}_{t-i})| \to_{e.a.s.} 0$, meaning that there exists a $\gamma > 1$ s.t. $\gamma^t\prod_{i=1}^{t-1}|x_{t-i}(\bar{\delta}_{t-i})| \to_{a.s.} 0$. This implies that $\mathrm{E}\prod_{i=1}^{t-1}|x_{t-i}(\bar{\delta}_{t-i})| = O(\rho^t)$, and

$$\mathrm{E}\left|\delta_t - \tilde{\delta}_t\right|^r = \mathrm{E}\prod_{i=1}^{t-1}\left|x_{t-i}(\bar{\delta}_{t-i})\right|^r \mathrm{E}|\delta_0 - \tilde{\delta}_0|^r = O(\rho^t) \tag{C.8}$$

for some $r > 0$. Hence, (C.3) follows by (C.8) and (C.6). □

Proof of Theorem 2: Let $N(\theta_0)$ be an arbitrarily small compact set around θ_0. By strong consistency, we have that for sufficiently large T, $\hat{\theta}_T \in$ int $N(\theta_0)$ a.s., hence the mean-value expansion of the score vector around θ_0 gives

$$
\begin{aligned}
0 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t(\hat{\theta}_T)}{\partial\theta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}\right) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} \qquad \text{(C.9)} \\
&+ \left[\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} + J\right) + \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\tilde{\theta}_T)}{\partial\theta\partial\theta'} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}\right) - J\right]\sqrt{T}(\hat{\theta}_T - \theta_0)
\end{aligned}
$$

where $\tilde{\theta}_T$ is on the line segment between $\hat{\theta}_T$ and θ_0, and recall that $J = -\mathrm{E}\left(\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}\right)$.

The first term in (C.9) converges to zero in probability by Lemma 7. Next, we show that the second term, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}$, obeys a CLT. As $\partial \tilde{l}_t(\theta_0)/\partial\theta$ is the product of an i.i.d. term, u_t, and $\partial\tilde{\delta}_t(\theta_0)/\partial\theta$, it is strictly stationary and ergodic if $\partial\tilde{\delta}_t(\theta_0)/\partial\theta$ is strictly stationary and ergodic, for which Assumption (A11) is sufficient since $\mathrm{E}(x_t^2) < 1$ implies $\mathrm{E}\log|x_t| < 0$ by Jensen's inequality. We have that $\mathrm{E}\left[\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}|\mathcal{F}_{t-1}\right] = 0$, where $\mathcal{F}_t$ is the σ-field generated by the current and past values of e_t, i.e. $\mathcal{F}_t \equiv \sigma(e_t, e_{t-1}, \ldots)$. Thus, the score is also a martingale difference sequence. Lemma 4 shows that $\mathrm{E}\left\|\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}'\right\|$ is finite, so that we can apply the CLT of Scott (1973) and the Cramer-Wold device to establish the asymptotic normality of the score function.

Next, we show that the first term inside the square parenthesis in (C.9) converges a.s. to zero. The second derivative is given by

$$
\frac{\partial^2 \tilde{l}_t}{\partial\theta\partial\theta'} = \frac{\partial^2\delta_t}{\partial\theta\partial\theta'}u_t + \frac{\partial u_t}{\partial\delta_t}\frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}.
$$

The score u_t and its derivative, $\frac{\partial u_t}{\partial\delta_t}$, are independent of δ_t and its derivatives. Furthermore, $\frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}$ is stationary and ergodic by Corollary 6 of Harvey (2013), and $\frac{\partial^2\delta_t}{\partial\theta\partial\theta'}$ is stationary and ergodic by Corollary 7 of Harvey (2013). This implies by the ergodic theorem that

$$
\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} \rightarrow_{a.s.} \mathrm{E}(u_t)\mathrm{E}\frac{\partial^2\delta_t}{\partial\theta\partial\theta'} + \mathrm{E}\frac{\partial u_t}{\partial\delta_t}\mathrm{E}\frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'} \qquad \text{(C.10)}
$$

The expectation $\mathrm{E}\frac{\partial^2 \delta_t}{\partial\theta\partial\theta'}$ is finite if $\mathrm{E}[\frac{\partial^2 u_t}{\partial \delta_t^2}] < \infty$, which holds by Lemma 6. Hence, the first term in (C.10) is zero because $\mathrm{E}(u_t) = 0$. The second term is equal to $-J$, which shows that the first term inside the square parenthesis in (C.9) converges a.s. to zero. In addition, it follows by standard arguments that $-\mathrm{E}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}$ is positive definite.

Lemmas 8 and 9 imply that the second term inside the square parenthesis in (C.9) is $o_{a.s.}(1)$. Since we have shown that the second term in (C.9) obeys a CLT, asymptotic normality follows from Slutsky's theorem. Finally, following Theorem 1 and Appendix A of Harvey (2013) to calculate the expectation of the outer product of $\partial\delta_t/\partial\theta$, we obtain the expression for the asymptotic covariance matrix in (3.8). $\square$

Lemma 4: Under Assumptions (A1)-(A11), $\mathrm{E}\left\|\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}'\right\| < \infty$.

Proof: Note that

$$\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta'} = u_t^2 \frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta}\frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta'}$$

As u_t is i.i.d. and $\partial\tilde{\delta}_t/\partial\theta$ is $\mathcal{F}_{t-1}$-measurable, by the law of iterated expectations,

$$\mathrm{E}\left\|\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta'}\right\| = \mathrm{E}(u_t^2)\mathrm{E}\left\|\frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta}\frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta'}\right\| \tag{C.11}$$

Let $\bar{\nu} = \max(\nu_1, \nu_2) < \infty$ and $\underline{\nu} = \min(\nu_1, \nu_2)$. By Assumption (A6), $\underline{\nu} > 2$. By the definition of u_t in (3.2),

$$u_t^2 \leq (\underline{\nu} - 2)^{-2}(\bar{\nu} + 1)^2 \xi_{1t}^2 \xi_{2t}^2, \quad a.s.$$

and hence, using Assumption (A1),

$$E(u_t^2) \leq (\underline{\nu} - 2)^{-2}(\bar{\nu} + 1)^2 < \infty.$$

Furthermore,

$$\mathrm{E}\left\|\frac{\partial\tilde{\delta}_t(\theta_0)}{\partial\theta}\frac{\partial\tilde{\delta}_t(\theta_0)}{\partial\theta'}\right\| \leq \mathrm{E}\left\|\frac{\partial\tilde{\delta}_t(\theta_0)}{\partial\theta}\right\|^2 < \infty$$

where the first inequality is an application of the Cauchy-Schwarz inequality, and the second follows by Lemma 10 of Harvey (2013) and Assumption (A11). □

Lemma 5: Under Assumption (A1), $\mathrm{E}\left|\frac{\partial u_t}{\partial\delta_t}\right| < \infty$.

Proof: Write $\frac{\partial u_t}{\partial\delta_t}$ in (3.3) as $\frac{\partial u_t}{\partial\delta_t} = (\nu_2+1)I_1 + (\nu_1+1)I_2$, where

$$I_1 = \frac{(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{2t}^4+\xi_{1t}^2\xi_{2t}^2}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} \tag{C.12}$$

$$I_2 = \frac{(\nu_1-2)(\xi_{1t}^2-\xi_{2t}^2)+\xi_{1t}^4+\xi_{1t}^2\xi_{2t}^2}{\left\{\nu_1-2+\xi_{1t}(\delta_t)^2\right\}^2} \tag{C.13}$$

By the triangle inequality $\mathrm{E}\left|\frac{\partial u_t}{\partial\delta_t}\right| \leq (\nu_2+1)\mathrm{E}|I_1| + (\nu_1+1)\mathrm{E}|I_2|$. It suffices to consider the first term, I_1. By symmetry, the result then applies also to the second term. Note that

$$\begin{aligned}\frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{2t}^4+\xi_{1t}^2\xi_{2t}^2\right|}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} &\leq \frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right|}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} + \frac{\xi_{2t}^4}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} \\ &\leq \frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right|}{\left\{\nu_2-2\right\}^2} + 1 \ a.s.\end{aligned} \tag{C.14}$$

and due to the finite second moments of ξ_t by Assumption (A1),

$$\mathrm{E}|I_1| \leq \mathrm{E}\frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right|}{(\nu_2-2)^2} + 1 \tag{C.15}$$

$$\leq \mathrm{E}\frac{\left|\xi_{2t}^2-\xi_{1t}^2\right|}{\nu_2-2} + \mathrm{E}\frac{\xi_{1t}^2\xi_{2t}^2}{(\nu_2-2)^2} + 1 \tag{C.16}$$

$$\leq \frac{2}{\nu_2-2} + \frac{1}{(\nu_2-2)^2} + 1 < \infty \tag{C.17}$$

□

Lemma 6: Under Assumption (A9), $\mathrm{E}\left|\frac{\partial^2 u_t}{\partial \delta_t^2}\right| < \infty$.

Proof: We have

$$\frac{\partial^2 u_t}{\partial \delta_t^2} = (\nu_2+1)(T_{11}+T_{12}) + (\nu_1+1)(T_{21}+T_{22}),$$

where

$$T_{11} = \frac{-2\xi_{2t}\xi_{1t}\left(\xi_{1t}^2+\xi_{2t}^2+2(\nu_2-2)\right)}{\left(\nu_2-2+\xi_{2t}^2\right)^2} \tag{C.18}$$

$$T_{12} = \frac{-4\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{2t}^4+\xi_{1t}^2\xi_{2t}^2\right)}{\left(\nu_2-2+\xi_{2t}^2\right)^3} \tag{C.19}$$

$$T_{21} = \frac{2\xi_{2t}\xi_{1t}\left(\xi_{1t}^2+\xi_{2t}^2+2(\nu_1-2)\right)}{\left(\nu_1-2+\xi_{1t}^2\right)^2} \tag{C.20}$$

$$T_{22} = \frac{4\xi_{1t}\xi_{2t}\left((\nu_1-2)(\xi_{1t}^2-\xi_{2t}^2)+\xi_{1t}^4+\xi_{1t}^2\xi_{2t}^2\right)}{\left(\nu_2-2+\xi_{1t}^2\right)^3} \tag{C.21}$$

By the c_r-inequality we have $\mathrm{E}\left|\frac{\partial^2 u_t}{\partial \delta_t^2}\right| \leq (\nu_2+1)(\mathrm{E}|T_{11}|+\mathrm{E}|T_{12}|)+(\nu_1+1)(\mathrm{E}|T_{21}|+\mathrm{E}|T_{22}|)$. It suffices to show that $\mathrm{E}|T_{11}|+\mathrm{E}|T_{12}|<\infty$. By symmetry, the result then also applies to $\mathrm{E}|T_{21}|+\mathrm{E}|T_{22}|$. Note that $|T_{11}| \leq |2(\nu_2-2)^{-1}\xi_{2t}\xi_{1t}\left(\xi_{1t}^2+\xi_{2t}^2+2(\nu_2-2)\right)$ a.s., and hence, as a direct consequence of Assumption (A9), $\mathrm{E}|T_{11}|<\infty$. The term $|T_{12}|$ can be bounded almost surely by

$$\begin{aligned}
|T_{12}| &\leq \frac{4|\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right)|}{\left(\nu_2-2+\xi_{2t}^2\right)^3} + \frac{4|\xi_{1t}\xi_{2t}^5|}{\left(\nu_2-2+\xi_{2t}^2\right)^3} \\
&\leq 4(\nu_2-2)^{-3}|\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right)| + \frac{4(\nu_2-2)^{-1}|\xi_{1t}\xi_{2t}^5|}{\left(\nu_2-2+\xi_{2t}^2\right)^2} \\
&\leq 4(\nu_2-2)^{-3}|\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right)| + 4(\nu_2-2)^{-1}|\xi_{1t}\xi_{2t}|, \ \ a.s.
\end{aligned}$$

and by Assumption (A9) both terms on the right hand side have finite expectation, and therefore $\mathrm{E}|T_{12}|<\infty$. □

Lemma 7: Under Assumptions (A1)-(A11), $T^{-1/2}\sum_{t=1}^{T}\frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} = o_{\mathrm{P}}(1)$.

Proof: By the generalized Chebyshev inequality and the c_r inequality we have for $\varepsilon > 0$ and $r > 0$,

$$\mathrm{P}\left(\left\| T^{-1/2}\sum_{t=1}^{T}\frac{\partial l_t(\theta_0)}{\partial\theta_i} - \frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_i}\right\|^r \geq \varepsilon\right) \leq T^{-r/2}\sum_{t=1}^{T}\mathrm{E}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right\|^r.$$

Hence, it is sufficient to show that for some $r > 0$, there exists a $\rho \in (0,1)$ such that,

$$\mathrm{E}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right\|^r = O(t\rho^t). \tag{C.22}$$

Given the score function in (3.6), we have that (the argument θ_0 is omitted for simplicity),

$$\frac{\partial l_t}{\partial\theta} - \frac{\partial\tilde{l}_t}{\partial\theta} = (u_t - \tilde{u}_t)\frac{\partial\delta_t}{\partial\theta} + \tilde{u}_t\left(\frac{\partial\delta_t}{\partial\theta} - \frac{\partial\tilde{\delta}_t}{\partial\theta}\right),$$

and by the c_r inequality and the independence of u_t and $\frac{\partial\delta_t}{\partial\theta}$,

$$\mathrm{E}\left\|\frac{\partial l_t}{\partial\theta} - \frac{\partial\tilde{l}_t}{\partial\theta}\right\|^r \leq c_r\mathrm{E}|u_t - \tilde{u}_t|^r\mathrm{E}\|\frac{\partial\delta_t}{\partial\theta}\|^r + c_r\mathrm{E}|\tilde{u}_t|^r\mathrm{E}\|\frac{\partial\delta_t}{\partial\theta} - \frac{\partial\tilde{\delta}_t}{\partial\theta}\|^r$$

Now, $\mathrm{E}|\tilde{u}_t|^r < \infty$ by Assumption (A2), and $\mathrm{E}\|\frac{\partial\delta_t}{\partial\theta}\|^r < \infty$ for $0 < r \leq 2$ by Lemma 10 of Harvey (2013). To analyze $u_t - \tilde{u}_t$ it suffices to consider

$$\frac{\xi_{1t}(\delta_t)\xi_{2t}(\delta_t)}{\nu_2 - 2 + \xi_{2t}(\delta_t)^2} - \frac{\xi_{1t}(\tilde{\delta}_t)\xi_{2t}(\tilde{\delta}_t)}{\nu_2 - 2 + \xi_{2t}(\tilde{\delta}_t)^2}.$$

The function $f(x,y) = xy/(\nu_2 - 2 + y^2)$ is Lipschitz continuous as it is both Lipschitz continuous relative to x and to y. Therefore, there exists $C > 0$ such that

$$f(\xi_{1t},\xi_{2t}) - f(\tilde{\xi}_{1t},\tilde{\xi}_{2t}) \leq C\|\xi_t - \tilde{\xi}_t\|$$

and by (C.2), $\|\xi_t - \tilde{\xi}_t\| \leq C|\delta_{0t} - \tilde{\delta}_{0t}|\|\tilde{\xi}_t\|$. By symmetry, the same applies to the term

$$\frac{\xi_{1t}(\delta_t)\xi_{2t}(\delta_t)}{\nu_1 - 2 + \xi_{1t}(\delta_t)^2} - \frac{\xi_{1t}(\tilde{\delta}_t)\xi_{2t}(\tilde{\delta}_t)}{\nu_1 - 2 + \xi_{1t}(\tilde{\delta}_t)^2}.$$

Hence, $\mathrm{E}|u_t - \tilde{u}_t|^r \leq C\mathrm{E}|\delta_t - \tilde{\delta}_t|^r \mathrm{E}\|\tilde{\xi}_t\|^r = O(\rho^t)$, using (C.8) and $\mathrm{E}\|\tilde{\xi}_t\|^r < \infty$.

Furthermore, note that

$$\begin{aligned}\frac{\partial \delta_t}{\partial \theta} - \frac{\partial \tilde{\delta}_t}{\partial \theta} &= A(z_{t-1} - \tilde{z}_{t-1}) + \frac{\partial \tilde{\delta}_{t-1}}{\partial \theta}(x_{t-1} - \tilde{x}_{t-1}) + x_{t-1}\left(\frac{\partial \delta_{t-1}}{\partial \theta} - \frac{\partial \tilde{\delta}_{t-1}}{\partial \theta}\right) \\ &= \sum_{j=1}^{t}\prod_{i=1}^{j-1} x_{t-i}\left\{A(z_{t-j} - \tilde{z}_{t-j}) + \frac{\partial \tilde{\delta}_{t-j}}{\partial \theta}(x_{t-j} - \tilde{x}_{t-j})\right\} + \prod_{i=1}^{t} x_{t-i}\frac{\partial \tilde{\delta}_0}{\partial \theta}\end{aligned}$$

with the convention $\prod_{i=1}^{0} = 1$. Now, $\mathrm{E}|z_t - \tilde{z}_t|^r = O(\rho^t)$ follows from the above results for $\mathrm{E}|u_t - \tilde{u}_t|^r$ and $\mathrm{E}|\delta_t - \tilde{\delta}_t|^r$. Then, $x_t - \tilde{x}_t = \kappa(\partial u_t/\partial \delta_t - \partial \tilde{u}_t/\partial \delta_t)$, and the fact that $\mathrm{E}|\partial u_t/\partial \delta_t - \partial \tilde{u}_t/\partial \delta_t|^r = O(\rho^t)$ follows by Lipschitz continuity of $\partial u_t/\partial \delta_t$ and the same arguments used to establish $\mathrm{E}|u_t - \tilde{u}_t|^r = O(\rho^t)$ above. Due to the independence of $\{x_t\}$, we obtain

$$\mathrm{E}|\frac{\partial \delta_t}{\partial \theta} - \frac{\partial \tilde{\delta}_t}{\partial \theta}|^r \leq C\sum_{j=1}^{t}\prod_{i=1}^{j-1}\mathrm{E}|x_{t-i}|^r \mathrm{E}|\varpi_{t-j}|^r + C\prod_{i=1}^{t}\mathrm{E}|x_{t-i}|^r \mathrm{E}|\frac{\partial \tilde{\delta}_0}{\partial \theta}|^r$$

where $\varpi_t = A(z_t - \tilde{z}_t) + \frac{\partial \tilde{\delta}_t}{\partial \theta}(x_t - \tilde{x}_t)$ with $\mathrm{E}|\varpi_t|^r = O(\rho^t)$. Using Assumption (A11) and Jensen's inequality, $\mathrm{E}|x_{t-i}|^r \leq b < 1$, for $0 < r \leq 2$. Hence,

$$\mathrm{E}\|\frac{\partial \delta_t}{\partial \theta} - \frac{\partial \tilde{\delta}_t}{\partial \theta}\|^r \leq C\sum_{j=1}^{t} b^{j-1}\mathrm{E}|\varpi_{t-j}|^r + O(b^t) = O(tb^{j-1})O(\rho^{t-j}) + O(b^t) = O(t\bar{\rho}^t)$$

where $\bar{\rho} = \max(b, \rho)$. This implies that

$$\mathrm{E}\left\|\frac{\partial l_t}{\partial \theta} - \frac{\partial \tilde{l}_t}{\partial \theta}\right\|^r = O(t\bar{\rho}^t),$$

which is sufficient to establish the result.

□

Lemma 8: Under Assumptions (A1)-(A11), there exists a neighborhood $N(\theta_0) \subseteq \Theta$ around θ_0 such that

$$\sup_{\theta \in N(\theta_0)}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| = o_{a.s.}(1)$$

Proof: We have

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'} &= (\tilde{u}_t - u_t)\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} + u_t\left(\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'}\right) \\
&+ \left(\frac{\partial\tilde{u}_t}{\partial\delta_t} - \frac{\partial u_t}{\partial\delta_t}\right)\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} + \frac{\partial u_t}{\partial\delta_t}\left(\frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{\delta}_t}{\partial\theta'} - \frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}\right)
\end{aligned}$$

In the proof of Lemma 7 we have shown that $\mathrm{E}\sup_{\theta\in N(\theta_0)}|\tilde{u}_t - u_t|^r = O(\rho^t)$ and $\mathrm{E}\sup_{\theta\in N(\theta_0)}|\frac{\partial\tilde{u}_t}{\partial\delta_t} - \frac{\partial u_t}{\partial\delta_t}|^r = O(\rho^t)$, while $\mathrm{E}\sup_{\theta\in N(\theta_0)}\|\frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{\delta}_t}{\partial\theta'} - \frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}\|^r = O(t\rho^t)$. It remains to consider the second term.

$$\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} = \frac{\partial A z_{t-1}}{\partial\theta'} + \frac{\partial\delta_{t-1}}{\partial\theta}\frac{\partial x_{t-1}}{\partial\theta'} + x_{t-1}\frac{\partial^2\tilde{\delta}_{t-1}}{\partial\theta\partial\theta'}$$

Following the same arguments as in the proof of Lemma 7, we can show that

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial A z_{t-1}}{\partial\theta'} - \frac{\partial A z_{t-1}}{\partial\theta'}\right\|^r = O(t\rho^t)$$

and

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial\delta_t}{\partial\theta}\frac{\partial x_t}{\partial\theta'} - \frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{x}_t}{\partial\theta'}\right\|^r = O(t\rho^t)$$

Let $\zeta_t = \frac{\partial A z_t}{\partial\theta'} - \frac{\partial A z_t}{\partial\theta'} + \frac{\partial\delta_t}{\partial\theta}\frac{\partial x_t}{\partial\theta'} - \frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{x}_t}{\partial\theta'} + \frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'}(\tilde{x}_t - x_t)$. Then,

$$\begin{aligned}
\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'} &= \zeta_{t-1} + x_{t-1}\left(\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'}\right) && \text{(C.23)} \\
&= \sum_{j=1}^{t}\prod_{i=1}^{j-1} x_{t-i}\zeta_{t-j} + \prod_{i=1}^{t} x_{t-i}\frac{\partial^2\tilde{\delta}_0^*}{\partial\theta\partial\theta'} && \text{(C.24)}
\end{aligned}$$

and since $\mathrm{E}\sup_{\theta\in N(\theta_0)}|\zeta_t|^r = O(t\rho^t)$ and $\mathrm{E}\sup_{\theta\in N(\theta_0)}\prod_{i=1}^{j-1}|x_{t-i}|^r = O(b^{j-1})$, we have

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'}\right\| = O(tb^{j-1})O(t\rho^{t-j}) + O(b^t) = O(t^2\bar{\rho}^t),$$

This implies that

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| = O(t^2\bar{\rho}^t),$$

which is a summable sequence in t. Hence, the result follows by an application of the generalized Chebychev inequality and the Borel-Cantelli lemma.

$\square$

Lemma 9: Under Assumptions (A1–(A11),

$$\left\| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \tilde{l}_t(\tilde{\theta}_T)}{\partial\theta\partial\theta'} - \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} \right\| \to 0 \;\; a.s.$$

where $\tilde{\theta}_T$ is on the line segment between $\hat{\theta}_T$ and θ_0.

Proof: Applying the mean-value theorem for the second order derivative function about θ_0, gives

$$\left\| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \tilde{l}_t(\tilde{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} \right\| \le \sup_{\theta\in N(\theta_0)} \left\{ \left\| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial\theta'} \left(\frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'} \right) \right\| \cdot \left\| \tilde{\theta}_T - \theta_0 \right\| \right\} \tag{C.25}$$

where $N(\theta_0) \subseteq \Theta$ is a small neighborhood around θ_0.

It suffices to check that

$$\mathrm{E} \sup_{\theta\in N(\theta_0)} \left\| \frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'} \right\| < \infty, \tag{C.26}$$

where, almost surely,

$$\begin{aligned} \frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'} &= u_t \frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{\delta}_t(\theta)}{\partial\theta\partial\theta'} + \frac{\partial u_t}{\partial\delta_t} \mathrm{vec}\left(\frac{\partial^2 \tilde{\delta}_t(\theta)}{\partial\theta\partial\theta'} \right) \frac{\partial\delta_t}{\partial\theta'} \\ &+ \frac{\partial^2 u_t}{\partial\delta_t^2} \mathrm{vec}\left(\frac{\partial\tilde{\delta}_t}{\partial\theta} \frac{\partial\tilde{\delta}_t}{\partial\theta'} \right) \frac{\partial\tilde{\delta}_t}{\partial\theta'} + \frac{\partial u_t}{\partial\delta_t} (I_m + K_{mm}) (\frac{\partial\delta_t}{\partial\theta} \otimes I_m) \frac{\partial^2 \delta_t}{\partial\theta\partial\theta'} \end{aligned}$$

and where K_{mm} is the commutation matrix and $m = \dim(\theta)$.

The second, third and fourth terms can be uniformly bounded similar to Lemma 7. For the first term, note that

$$\frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{\delta}_t(\theta)}{\partial\theta\partial\theta'} = \frac{\partial \mathrm{vec}}{\partial\theta'} \left(\frac{\partial A z_{t-1}}{\partial\theta'} + \frac{\partial\delta_{t-1}}{\partial\theta} \frac{\partial x_{t-1}}{\partial\theta'} + x_{t-1} \frac{\partial^2 \tilde{\delta}_{t-1}}{\partial\theta\partial\theta'} \right)$$

A typical element of this matrix, denoting the i-th row of A as A_i, is given by

$$
\begin{aligned}
\frac{\partial^3 \tilde{\delta}_t(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} &= \frac{\partial A_i}{\partial\theta_j}\frac{\partial z_{t-1}}{\partial\theta_k} + \frac{\partial A_i}{\partial\theta_k}\frac{\partial z_{t-1}}{\partial\theta_j} + A_i \frac{\partial^2 z_{t-1}}{\partial\theta_j \partial\theta_k} \\
&+ \frac{\partial^2 \delta_{t-1}}{\partial\theta_i \partial\theta_k}\frac{\partial x_{t-1}}{\partial\theta_j} + \frac{\partial \delta_{t-1}}{\partial\theta_i}\frac{\partial^2 x_{t-1}}{\partial\theta_j \partial\theta_k} + \frac{\partial x_{t-1}}{\partial\theta_k}\frac{\partial^2 \delta_{t-1}}{\partial\theta_i \partial\theta_j} + x_{t-1}\frac{\partial^3 \tilde{\delta}_{t-1}(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \\
&= \sum_{j=1}^{\infty} \vartheta_{t-j} \prod_{i=1}^{j-1} x_{t-i}
\end{aligned}
$$

where $\vartheta_t = \frac{\partial A_i}{\partial\theta_j}\frac{\partial z_t}{\partial\theta_k} + \frac{\partial A_i}{\partial\theta_k}\frac{\partial z_t}{\partial\theta_j} + A_i \frac{\partial^2 z_t}{\partial\theta_j \partial\theta_k} + \frac{\partial^2 \delta_t}{\partial\theta_i \partial\theta_k}\frac{\partial x_t}{\partial\theta_j} + \frac{\partial \delta_t}{\partial\theta_i}\frac{\partial^2 x_t}{\partial\theta_j \partial\theta_k} + \frac{\partial x_t}{\partial\theta_k}\frac{\partial^2 \delta_t}{\partial\theta_i \partial\theta_j}$. Each term of ϑ_t can be uniformly bounded in a neighborhood of θ_0 and a repeated application of the c_r inequality yields $\mathrm{E}\sup_{\theta \in N(\theta_0)} |\vartheta_t| < \infty$. Since $\mathrm{E}\prod_{i=1}^{j} |x_{t-i}| = O(\rho^j)$ as in the proof of Lemma 3, the infinite sum in (C.27) is convergent and we obtain

$$
\mathrm{E} \sup_{\theta \in N(\theta_0)} \left| \frac{\partial^3 \tilde{\delta}_t(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < \infty
$$

which implies (C.26). Finally, since by the ergodic theorem, given (C.26),

$$
\limsup_{T\to\infty} \left\| \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta \in N(\theta_0)} \frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta \partial\theta'} \right\| = \mathrm{E} \left\| \sup_{\theta \in N(\theta_0)} \frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta \partial\theta'} \right\| < \infty
$$

the desired result follows because $\hat{\theta}_T \to \theta_0$ a.s. by Theorem 1.

$\square$

D Monte Carlo simulation

In this section we provide insights into the finite sample first and second order behaviour of the PML estimator in score-driven rotation models. We first outline the simulation design and subsequently discuss simulation outcomes. The section concludes with some recommendations that one might draw from the Monte-Carlo analysis for empirical work.

D.1 Simulation design

The purpose of the simulation study is to uncover the scope of the PML estimator provided in Section 3 for evaluating the dynamic model in (2.1) and (2.3) and, hence, its potential to extract the underlying independent components from orthogonalized data. The data generating processes are specified with two alternative choices of β, i.e. $\beta = 0.9, 0.95$, and three alternative choices of κ, i.e. $\kappa = 0.1, 0.05, 0.025$. For all data generating processes we set $\omega = 1.30$. The alternative choices of β and κ are combined to obtain overall six parameterizations of the underlying data generating process ($\theta = (\omega, \beta, \kappa)'$).

Throughout we generate data from marginal standardized Student-t random variables with $\nu_1 = 5$ and $\nu_2 = 6$ degrees of freedom. For PML estimation we distinguish two scenarios. First, we assume the correct degrees of freedom ($\nu_i^{(0)} = \nu_i$, $i = 1, 2$). In the second scenario, the interest turns to the effects of performing PML estimation under a false assumption on the degrees of freedom ($\nu_1^{(0)} \neq \nu_1, \nu_2^{(0)} \neq \nu_2$). Specifically, we set $\nu_1^{(0)} = 7$ and $\nu_2^{(0)} = 8$ (instead of $\nu_1^{(0)} = 5$ and $\nu_2^{(0)} = 6$). Alternative sample sizes are $T = 1000, 2500, 5000$ and $T = 10000$. For each experiment we draw $T + 250$ sample observations, and disregard the first 250 observations to immunize the outcomes of the Monte-Carlo analysis against initial effects.

For the numerical log-likelihood maximization we use the initial parameters as

$\theta^{(0)} = (1.63, 0.98, 0.01)'$ in the case of $T = 10000$. For analysing shorter samples of length $T = 1000, 2500$ and $T = 5000$ we try four alternative initializations $\theta^{(0)}$ and select the particular choice which obtains the largest log-likelihood statistics (the alternative choices are (1.95,0.98,0.01)', (1.43,0.95,0.05)', (1.17,0.90,0.10)' and (0.98,0.70,0.15)'). We use the BHHH algorithm of Berndt et al. (1974) for log-likelihood maximization which has the convenient feature that its implementation only requires first order derivatives. In a stylized form the BHHH algorithm delivers updates of the estimated parameter vector as

$$\hat{\theta}^{(s)} = \hat{\theta}^{(s-1)} + \phi_s \left(\sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} \Big|_{\theta=\hat{\theta}^{(s-1)}} \right)^{-1} \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \Big|_{\theta=\hat{\theta}^{(s-1)}},$$

where s indexes the iteration sequence of the optimizer and $\phi_s > 0$ is used to modify the step length at each iteration. Specifically, we employ alternative choices $\phi_s \in \{1.0, 0.5, 0.25, 0.10, 0.05\}$ to detect most effective updates of the parameter vector at each iteration step.

Each Monte-Carlo experiment is performed with $R = 1000$ replications. To characterise the performance of PML estimation we document the following summary statistics for the typical elements of $\hat{\theta}$ denoted $\hat{\theta}_i$:

- Root mean squared error: $\text{RMSE}(\hat{\theta}_i) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_{r,i} - \theta_i)^2}$, where $\hat{\theta}_{r,i}$ is the estimator of the i-th component of θ obtained from the r-th Monte-Carlo replication.

- Bias: $\text{BIAS}(\hat{\theta}_i) = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_{r,i} - \theta_i$

- Standard deviation: $\text{SD}(\hat{\theta}_i) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_{r,i} - \bar{\theta}_i)^2}$, where $\bar{\theta}_i = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_{r,i}$

- Approximation based on $\widehat{I}^{-1}$: $\widehat{\text{SD}}_I(\hat{\theta}_i) = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\mathcal{S}_{ii}^{(r)}}$, with $\mathcal{S}^{(r)} = \widehat{I}_r^{-1}/T$

- Approximation based on $\widehat{V}$: $\widehat{\text{SD}}_V(\hat{\theta}_i) = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\mathcal{S}_{ii}^{(r)}}$, with $\mathcal{S}^{(r)} = \widehat{V}_r/T$

D.2 Simulation results

Simulation results for DGPs parameterized with $\beta = 0.90$ and $\beta = 0.95$ are documented in Tables D.1 and D.2, respectively. The left (right) hand side panels of the tables show simulation outcomes for experiments with $\nu_i^{(0)} = \nu_i$, $i = 1, 2$ ($\nu_1^{(0)} \neq \nu_1$, $\nu_2^{(0)} \neq \nu_2$), i.e., PML estimation under knowledge (misspecification) of the actual underlying distributional models. The Monte-Carlo analysis allows for a couple of insights with respect to the first and second order properties of the nonlinear PML estimator:

1. Confirming our results of Theorem 1, all model parameters are estimated consistently. For instance, regarding the DGP with $\beta = 0.9$, $\kappa = 0.05$ the documented RMSE statistics shrink for all parameters with increasing sample size. Focussing on estimates for the response parameter κ, RMSE (absolute BIAS) statistics shrink from 10.68E-02 (1.66E-02) in case of $T = 1000$ to 1.71E-02 (0.17E-02) in case of $T = 10000$ (left hand side panels of Table D.1). With $T = 2500$ sample observations, these performance statistics are 4.99E-02 (RMSE) and 1.13E-02 (absolute BIAS). Since BIAS estimates vanish with increasing sample size, the RMSE statistics approach the documented standard errors of the estimators.

2. As it is implied by the documented BIAS statistics, it holds for all alternative DGPs that the PML estimator results in an underestimation of ω and β, and an overestimation of κ.

3. The consistency of the PML estimator holds irrespective of assuming correct or false degrees of freedom for the underlying distributions of independent innovations. However, assuming false degrees of freedom ($\nu_1^{(0)} = 7$ and $\nu_2^{(0)} = 8$) comes with a loss in estimator precision. For instance, under an assumption

of false degrees of freedom, the relative efficiency in terms of the RMSE of estimating κ is about 69% in the case of $T = 2500$ and about 64% for $T = 10000$ ($\beta = 0.9$, $\kappa = 0.05$). Hence, as an implication for empirical practice, it is worth to strive for a most suitable distribution, since it promises efficiency gains of parameter estimates.

4. Approximating estimation uncertainty by the asymptotic result in (3.8) suffers from strong biases in small samples. Under the assumption of a correctly specified distributional model ($I = J$) both covariance estimators $\widehat{I}^{-1}$ and $\widehat{V}$ can be expected to quantify estimation uncertainty reliably in sufficiently large samples. As it turns out, however, both estimators tend to underestimate the empirical variance of $\hat{\kappa}$ and overestimate the variances of $\hat{\omega}$ and $\hat{\beta}$, irrespective of the true response parameter and the actual sample size. The estimation biases for all these second order characteristics are sizeable in small samples ($T = 1000, 2500$) and vanish only slowly. For example, in the model with $\beta = 0.9$, $\kappa = 0.05$ and in case of $T = 2500$, the estimate of the standard deviation of $\hat{\kappa}$ using $\widehat{I}^{-1}$ falls short of the true empirical standard deviation by about 30%, while that of $\hat{\beta}$ exceeds the benchmark by about 3%.

5. Using the robust covariance estimator $\widehat{V}$ if the correct distributional model is employed for PML estimation ($\nu_i^{(0)} = \nu_i$, $i = 1, 2$) results in a loss of estimator precision in comparison with using $\widehat{I}^{-1}$. For instance, in the scenario described before ($T = 2500, \beta = 0.9$, $\kappa = 0.05$), the assessment of estimation uncertainty by means of $\widehat{V}$ yields standard error approximations which fall short of the empirical standard error by 47% (SD of $\hat{\kappa}$) and 15% (SD of $\hat{\beta}$). In the large sample case of $T = 10000$, the simulation outcomes for both variance estimators approach each other. Moreover, these performance measures come

close to the empirical standard errors of the estimators.

6. For scenarios of misspecified distributional models ($\nu_1^{(0)} = 7$ and $\nu_2^{(0)} = 8$), the large sample case of $T = 10000$ is not sufficient to indicate an overall lead of the robust covariance estimator $\widehat{V}$ over the nonrobust estimator $\widehat{I}^{-1}$. For instance, for the DGP with $\beta = 0.9$ and $\kappa = 0.05$, the Monte-Carlo standard deviation of $\hat{\beta}$ (scaled by 100) is 9.16, while the approximations obtained from $\widehat{I}^{-1}$ and $\widehat{V}$ are, respectively, 7.90 and 5.77.

7. The documented results on finite sample biases of PML covariance estimation indicate that one should be careful when interpreting commonly provided inferential outcomes (e.g. t-ratios). For instance, sample-based assessments of I^{-1} or V might be employed to construct confidence intervals for estimators $\hat{\kappa}$. It might be of interest to obtain the frequencies at which such confidence intervals fail to include the true parameter κ, corresponding to the size of an equivalent Wald test of $H_0 : \kappa = \kappa_0$. One can also obtain the empirical non-coverage of $\kappa = 0$, which would correspond to the power under classical conditions. In our case, however, the case $\kappa = 0$ is excluded by Assumption (A8), i.e., κ is at the boundary of the admissible parameter space. Hence, there are unidentified parameters under the null hypothesis, and the asymptotic theory developed in Section 3 does not apply for this case. To formally test the hypothesis $H_0 : \kappa = 0$, i.e. constancy of the rotation angle δ_t, we recommend Lagrange Multiplier (LM) tests, e.g. Portmanteau-type tests for serial correlation of the scores, which we will use in the empirical illustration in Section 4, see e.g. Chapter 2.5 of Harvey (2013).

   Table D.3 shows for the case $\beta = 0.9$, $\kappa = 0.05$ empirical frequencies of non-coverage for such confidence intervals with nominal 99% and 95% coverage.

The results confirm that analytical approximations of estimation uncertainty for $\hat{\kappa}$ are too small, such that the constructed confidence intervals fail to achieve their nominal coverage throughout. In the large sample case of $T = 10000$, the empirical type-I error probabilities still exceed their nominal counterparts, where the empirical sizes are 9.7 and 5.2 for nominal significance levels of 5% and 1%, respectively. One might interpret the results documented for a nominal level of 99% to imitate a size adjusted test for the null hypothesis of insignificance of $\hat{\kappa}$ with 5% nominal significance. With this interpretation of 99% confidence intervals, PML inference shows considerable (size adjusted) power against a model with static rotation ($\kappa = 0$). Defined in this sense, (size adjusted) power increases from 18.4 ($T = 1000$) to 87.1 ($T = 10000$) percent.

D.3 Recommendations for empirical practice

The simulation results indicate that the parameters governing the dynamic rotation patterns can be retrieved consistently from the data. The scales of estimation errors shrink quickly in case of realistic sample sizes covering 4 or 10 years of daily data ($T = 1000, 2500$). To guard the empirical analysis against (inefficient) estimates derived from local extremes of the pseudo likelihood, it is worth to initialize non-linear optimizations by means of alternative parameter selections. The precision of parameter estimation benefits from conditioning PML estimation on the correct distributional assumptions ($\nu_i^{(0)} = \nu_i$, $i = 1, 2$). Hence, for empirical practice it is reasonable (i) to estimate the degrees of freedom (ν_i) and dynamic parameters in θ jointly or (ii) to condition the PML estimation of the dynamic parameters in θ on a variety of alternative assumptions with regard to the degrees of freedom parameters ($\nu_i^{(0)}$, $i = 1, 2$). Owing to computational complexity the advice in (ii) seems preferable to the suggestion in (i).

The approximation of estimation uncertainty by means of $\widehat{I}^{-1}$ and $\widehat{V}$ suffers from sizeable biases in small samples and converges only slowly to the true second order features of the estimators. In light of (i) the convenient first order properties of the PML estimator and (ii) its asymptotic normality shown in Theorem 2, bootstrap approaches may be suggested as an alternative for inferential analysis in finite samples. The underlying model assumption of independence largely facilitates the implementation of bootstrap schemes by means of independent drawings with replacement. For the empirical application in Section 4 we propose a specific bootstrap scheme to complement analytical covariance estimates.

κ	T	θ_i	RMSE	BIAS	SD	$\widehat{\text{SD}}_I$	$\widehat{\text{SD}}_V$	RMSE	BIAS	SD	$\widehat{\text{SD}}_I$	$\widehat{\text{SD}}_V$
.100	1000	ω	6.97	-0.48	6.95	46.73	16.33	8.45	-0.10	8.45	13.49	7.18
		β	32.35	-10.16	30.72	109.00	123.51	36.70	-11.80	34.75	28.67	12.68
		κ	9.04	0.69	9.01	5.20	2.38	12.29	4.10	11.59	9.64	10.12
	2500	ω	3.90	-0.17	3.90	10.12	2.61	4.03	-0.14	4.03	4.41	2.52
		β	14.72	-3.09	14.39	7.05	5.39	14.98	-3.23	14.63	10.30	6.08
		κ	3.76	0.53	3.72	2.87	1.99	6.28	3.66	5.10	4.98	2.42
	5000	ω	2.58	-0.10	2.58	2.57	3.35	2.61	-0.10	2.61	3.04	1.90
		β	5.13	-0.95	5.04	3.64	2.98	5.01	-0.96	4.92	4.69	2.85
		κ	2.30	0.16	2.30	1.88	1.43	4.29	2.96	3.11	3.17	1.78
	10000	ω	1.77	-0.11	1.77	1.65	1.48	1.84	-0.12	1.83	2.09	1.25
		β	2.29	-0.32	2.27	2.05	1.65	2.48	-0.40	2.44	2.65	1.61
		κ	1.39	0.02	1.39	1.25	1.05	3.36	2.72	1.97	2.08	1.18
.050	1000	ω	6.57	-0.62	6.54	59.25	15.51	6.62	-0.28	6.61	79.87	19.26
		β	55.73	-25.44	49.59	40.88	29.70	56.39	-25.82	50.13	551.26	434.31
		κ	10.68	1.66	10.55	5.91	3.22	14.15	3.54	13.70	11.16	4.54
	2500	ω	3.83	-0.22	3.83	16.76	16.46	3.80	-0.20	3.80	18.60	3.98
		β	34.02	-12.16	31.77	32.75	27.09	38.38	-13.64	35.87	41.80	23.62
		κ	4.99	1.13	4.86	3.38	2.62	7.26	2.84	6.68	5.93	3.34
	5000	ω	2.56	-0.12	2.56	2.98	3.45	2.58	-0.13	2.57	3.87	2.40
		β	19.90	-5.08	19.24	12.88	11.63	24.14	-6.54	23.24	19.04	12.74
		κ	2.86	0.47	2.82	2.23	1.94	4.23	2.03	3.71	3.82	2.47
	10000	ω	1.77	-0.07	1.76	1.73	1.68	1.77	-0.08	1.77	2.19	1.70
		β	7.98	-1.68	7.80	5.93	5.62	9.37	-1.96	9.16	7.90	5.77
		κ	1.71	0.17	1.70	1.47	1.36	2.68	1.58	2.16	2.44	1.72
.025	1000	ω	6.75	-0.18	6.75	54.37	22.47	6.63	-0.34	6.62	370.82	126.34
		β	62.52	-29.66	55.03	296.95	250.95	65.85	-31.46	57.84	4293.2	3278.3
		κ	11.13	1.64	11.01	5.94	2.61	15.08	2.08	14.93	11.19	3.64
	2500	ω	3.65	-0.32	3.63	48.65	15.90	3.68	-0.23	3.67	102.13	13.77
		β	55.13	-24.71	49.28	906.42	748.17	62.10	-28.47	55.20	105.56	46.73
		κ	6.00	1.37	5.85	3.54	2.80	8.21	2.27	7.89	6.22	3.40
	5000	ω	2.47	-0.06	2.47	14.74	3.03	5.54	-0.05	5.54	141.14	45.68
		β	42.99	-16.12	39.86	183.17	201.70	47.97	-18.32	44.34	41.65	27.24
		κ	3.43	0.72	3.36	2.40	2.07	4.72	1.52	4.47	4.07	3.01
	10000	ω	1.71	-0.07	1.70	18.77	1.90	1.72	-0.06	1.72	3.70	4.99
		β	25.56	-7.47	24.45	26.49	25.71	30.73	-9.48	29.24	29.21	21.10
		κ	1.95	0.36	1.91	1.59	1.47	2.83	1.18	2.57	2.66	1.88

Table D.1: Simulation results for DGPs with $\omega = 1.3$ and $\beta = 0.90$. The true parameter κ is given in the first column. PML estimation results are shown for presuming correct ($\nu_i^{(0)} = \nu_i$, left hand side panels) and false degrees of freedom ($\nu_i^{(0)} \neq \nu_1$, right hand side). Documented loss measures include root mean squared errors (RMSE), bias (BIAS) and standard deviation estimates (SD). Further performance statistics are PML approximations of SD based on estimates $\widehat{I}^{-1}$ and $\widehat{V}$ of the asymptotic covariance matrix. All summary statistics are multiplied with 100.

κ	T	θ_i	RMSE	BIAS	SD	$\widehat{\text{SD}}_I$	$\widehat{\text{SD}}_V$	RMSE	BIAS	SD	$\widehat{\text{SD}}_I$	$\widehat{\text{SD}}_V$
.100	1000	ω	11.44	-0.08	11.44	37.85	17.08	9.53	-0.04	9.53	22.83	44.18
		β	29.12	-7.93	28.02	910.49	713.45	31.22	-8.61	30.00	22.20	9.79
		κ	6.45	0.62	6.42	4.10	1.83	9.89	3.70	9.18	7.46	9.18
	2500	ω	4.53	-0.19	4.52	3.93	2.03	4.65	-0.16	4.64	5.04	0.58
		β	4.25	-0.95	4.14	2.61	1.99	4.79	-1.04	4.67	3.95	2.29
		κ	2.55	0.23	2.53	1.99	1.12	4.79	3.12	3.63	3.43	1.14
	5000	ω	2.95	-0.08	2.95	2.66	1.31	3.11	-0.08	3.11	3.37	0.76
		β	1.52	-0.32	1.48	1.23	0.79	1.56	-0.31	1.53	1.61	0.69
		κ	1.48	0.04	1.48	1.28	0.67	3.47	2.79	2.06	2.12	0.69
	10000	ω	1.96	-0.06	1.96	1.80	0.91	2.12	-0.11	2.11	2.25	0.56
		β	0.84	-0.13	0.83	0.73	0.41	0.89	-0.12	0.88	0.93	0.34
		κ	0.94	0.04	0.93	0.83	0.44	2.96	2.63	1.36	1.36	0.46
.050	1000	ω	11.02	- 0.09	11.02	477.52	115.01	10.76	-0.13	10.76	135.59	35.57
		β	47.50	-20.56	42.82	45.65	33.67	50.68	-22.29	45.51	4243.3	1847.8
		κ	9.65	2.09	9.42	5.32	2.71	13.85	4.26	13.17	9.86	3.30
	2500	ω	4.26	-0.22	4.26	9.58	4.16	4.28	-0.24	4.28	9.57	5.14
		β	23.34	-6.57	22.39	13.88	13.15	26.03	-7.22	25.01	17.89	11.61
		κ	4.07	0.91	3.97	2.64	2.13	5.72	2.48	5.16	4.51	2.27
	5000	ω	2.85	-0.09	2.85	2.78	2.54	2.85	-0.10	2.85	3.53	2.48
		β	11.16	-2.01	10.97	5.45	4.96	11.91	-2.14	11.72	7.98	5.55
		κ	2.02	0.24	2.00	1.63	1.44	3.08	1.70	2.57	2.73	1.80
	10000	ω	1.95	-0.06	1.94	1.92	1.78	1.96	-0.06	1.96	2.42	1.82
		β	2.74	-0.56	2.68	2.21	2.06	3.00	-0.62	2.93	2.91	2.08
		κ	1.15	0.07	1.15	1.06	0.94	2.10	1.43	1.54	1.75	1.24
.025	1000	ω	7.98	-0.19	7.98	70.32	27.12	6.70	-0.42	6.68	66.86	18.18
		β	67.20	-35.00	57.37	148.03	89.57	69.18	-35.98	59.09	66.44	52.41
		κ	10.73	1.99	10.55	5.75	2.88	14.88	1.95	14.75	10.84	3.50
	2500	ω	3.87	-0.28	3.86	33.10	15.90	3.96	-0.24	3.96	64.40	8.06
		β	48.57	-21.26	43.67	45.78	39.04	52.42	-22.88	47.17	49.82	31.83
		κ	5.41	1.42	5.22	3.23	2.58	7.69	2.53	7.26	5.59	3.01
	5000	ω	2.63	-0.09	2.62	64.50	102.22	2.64	-0.11	2.64	8.20	6.52
		β	33.78	-10.90	31.98	108.06	97.22	37.98	-12.80	35.76	33.68	25.11
		κ	2.83	0.66	2.75	1.97	1.73	3.84	1.42	3.56	3.34	2.21
	10000	ω	1.82	-0.06	1.82	1.84	1.82	1.83	-0.07	1.83	2.32	1.80
		β	16.25	-3.85	15.79	9.54	9.06	20.36	-4.88	19.77	12.81	9.33
		κ	1.52	0.28	1.49	1.23	1.16	2.28	1.06	2.02	2.06	1.48

Table D.2: Simulation results for DGPs with $\omega = 1.3$ and $\beta = 0.95$. For further comments see Table D.1.

	95% coverage				99% coverage			
T	1000	2500	5000	10000	1000	2500	5000	10000
$\kappa = 0.05$	26.8	18.2	12.9	9.7	20.6	12.7	8.4	5.2
$\kappa = 0$	30.5	46.7	73.4	97.2	18.7	24.3	48.2	87.1

Table D.3: Non-coverage frequencies for $\kappa = 0.05$ and $\kappa = 0$, of confidence intervals for $\hat{\kappa}$ that are determined by means of standard errors implied by $\widehat{I}^{-1}$. The true model parameters are $\kappa = 0.05$ and $\beta = 0.9$.

References

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974). "Estimation and inference in nonlinear structural models". *Annals of Economic and Social Measurement* 3/4, pp. 653–665.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Vol. 52. Cambridge University Press.

Lancaster, H. O. (1954). "Traces and cumulants of quadratic forms in normal variables". *Journal of the Royal Statistical Society. Series B (Methodological)* 16.2, pp. 247–254.

Pfanzagl, J. (1969). "On the measurability and consistency of minimum contrast estimates". *Metrika: International Journal for Theoretical and Applied Statistics* 14.1, pp. 249–272.

Scott, D. J. (1973). "Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach". *Advances in Applied Probability* 5.1, pp. 119–137.

Straumann, D. and T. Mikosch (Oct. 2006). "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach". *Annals of Statistics* 34.5, pp. 2449–2495.