Supplemental Appendices for

Sequential estimation of shape parameters in multivariate dynamic models

Dante Amengual
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<amengual@cemfi.es>

Gabriele Fiorentini
Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy
<fiorentini@ds.unifi.it>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

February 2012
Revised: December 2012

B General versions of the propositions

Here we extend the propositions in section 3 to those models in which reparametrisation 1 does not necessarily hold, but maintaining the assumption that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied. We also drop the ellipticity assumption when it is not essential for the statement of the proposition.

Proposition B1 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ *is i.i.d.* $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ *with bounded fourth moments, then* $\sqrt{T}(\tilde{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_0) \to N[0, \mathcal{F}(\boldsymbol{\phi}_0)]$*, where*

$$\mathcal{F}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0) + \mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0).$$

Proof. We can use standard arguments (see e.g. Newey and McFadden (1994)) to show that the sequential ML estimator of $\boldsymbol{\eta}$ is asymptotically equivalent to a MM estimator based on the linearised influence function

$$s_{\eta t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathcal{I}_{\boldsymbol{\theta}\eta}'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \mathbf{0}).$$

On this basis, the expression for $\mathcal{F}(\boldsymbol{\phi}_0)$ follows from the definitions of $\mathcal{B}(\boldsymbol{\phi}_0)$, $\mathcal{C}(\boldsymbol{\phi}_0)$ and $\mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)$ in Propositions 1 and 3 in Fiorentini and Sentana (2010), together with the martingale difference nature of $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ and $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$, and the fact that $E\left\{\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})\mathbf{e}_{rt}'(\boldsymbol{\phi})\right| \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\} = \mathbf{0}$. □

Proposition B2 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ *is i.i.d.* $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ *with bounded fourth moments, then* $\mathcal{I}^{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) \leq \mathcal{F}(\boldsymbol{\phi}_0)$*, with equality if and only if*

$$\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\left\{\mathcal{C}(\boldsymbol{\phi}_0) - \left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\right]^{-1}\right\}\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0) = \mathbf{0}.$$

Proof. A straightforward application of Theorem 5 in Pagan (1986) allows us to show that

$$\sqrt{T}(\tilde{\boldsymbol{\eta}}_T - \hat{\boldsymbol{\eta}}_T) \to N\left[0, \mathcal{Y}(\boldsymbol{\phi}_0)\right],$$

where

$$\mathcal{Y}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\left\{\mathcal{C}(\boldsymbol{\phi}_0) - \left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\right]^{-1}\right\}\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0).$$

Therefore, the sequential ML estimator will be asymptotically as efficient as the joint ML estimator if and only if $\mathcal{Y}(\boldsymbol{\phi}_0) = 0$. □

Proposition B3 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ *is i.i.d.* $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ *with bounded fourth moments, then the optimal sequential GMM estimator of* $\boldsymbol{\eta}$ *based on* $\mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$ *will be asymptotically equivalent to the optimal sequential GMM estimator based on* $\mathbf{n}_t^{\perp}(\tilde{\boldsymbol{\theta}}_T, \boldsymbol{\eta})$*, where*

$$\mathbf{n}_t^{\perp}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbf{n}_t(\boldsymbol{\theta}, \boldsymbol{\eta}) - \mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0}),$$

with

$$\mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0) = \lim_{T\to\infty} \frac{1}{T}\sum\nolimits_{t=1}^{T} E\left(-\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial \boldsymbol{\theta}'}\middle|\, \boldsymbol{\phi}_0\right),$$

are the residuals from the theoretical IV regression of $\mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta})$ *on* $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$ *using* $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ *as instruments.*

Proof. Under standard regularity conditions, we can use the expansion

$$\begin{aligned}\frac{1}{T}\sum\nolimits_{t=1}^{T} \mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta}_0) &= \frac{1}{T}\sum\nolimits_{t=1}^{T} \mathbf{n}_t(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) - \mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0)\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) + o_p(1)\\ &= [\mathbf{I}, -\mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)]\frac{1}{T}\sum\nolimits_{t=1}^{T}\left[\begin{array}{c}\mathbf{n}_t(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\\ \mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0;\mathbf{0})\end{array}\right] + o_p(1),\end{aligned}$$

to show that

$$\lim_{T\to\infty} V\left(\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} \mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta}_0)\middle|\, \boldsymbol{\phi}_0\right)$$

will be given by

$$\mathcal{E}_{\mathbf{n}} = [\mathbf{I}, -\mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)]\left(\begin{array}{cc}\mathcal{G}_{\mathbf{n}}(\boldsymbol{\phi}_0) & \mathcal{D}_{\mathbf{n}}(\boldsymbol{\phi}_0)\\ \mathcal{D}'_{\mathbf{n}}(\boldsymbol{\phi}_0) & \mathcal{B}(\boldsymbol{\phi}_0)\end{array}\right)\left(\begin{array}{c}\mathbf{I}\\ -\mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\end{array}\right),$$

where

$$\left(\begin{array}{cc}\mathcal{G}_{\mathbf{n}}(\boldsymbol{\phi}_0) & \mathcal{D}_{\mathbf{n}}(\boldsymbol{\phi}_0)\\ \mathcal{D}'_{\mathbf{n}}(\boldsymbol{\phi}_0) & \mathcal{B}(\boldsymbol{\phi}_0)\end{array}\right) = \lim_{T\to\infty} V\left(\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T}\left[\begin{array}{c}\mathbf{n}_t(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\\ \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0;\mathbf{0})\end{array}\right]\middle|\, \boldsymbol{\phi}_0\right).$$

Similarly, it is easy to see that under standard regularity conditions

$$\frac{1}{T}\sum\nolimits_{t=1}^{T} \mathbf{n}_t^{\perp}(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta}_0) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \mathbf{n}_t^{\perp}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) - \mathcal{N}_{\mathbf{n}^{\perp}}(\boldsymbol{\phi}_0)\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) + o_p(1),$$

where

$$\mathcal{N}_{\mathbf{n}^{\perp}}(\boldsymbol{\phi}_0) = \lim_{T\to\infty} \frac{1}{T}\sum\nolimits_{t=1}^{T} E\left(-\frac{\partial \mathbf{n}_t^{\perp}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial \boldsymbol{\theta}'}\middle|\, \boldsymbol{\phi}_0\right).$$

But since

$$\mathcal{N}_{\mathbf{n}^{\perp}}(\boldsymbol{\phi}_0) = \mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0) - \mathcal{N}_{\mathbf{n}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0) = \mathbf{0},$$

it immediately follows that

$$\lim_{T\to\infty} V\left(\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} \mathbf{n}_t^{\perp}(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta}_0)\middle|\, \boldsymbol{\phi}_0\right) = \mathcal{E}_{\mathbf{n}}(\boldsymbol{\phi}_0).$$

Finally, given that

$$\frac{\partial \mathbf{n}_t^{\perp}(\boldsymbol{\theta},\boldsymbol{\eta})}{\partial \boldsymbol{\eta}'} = \frac{\partial \mathbf{n}_t(\boldsymbol{\theta},\boldsymbol{\eta})}{\partial \boldsymbol{\eta}'},$$

it follows that the optimal sequential GMM estimators based on $\mathbf{n}_t(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})$ and $\mathbf{n}_t^{\perp}(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})$ will be asymptotically equivalent. □

Proposition B4 *Let $\mathcal{J}_M(\boldsymbol{\phi}_0)$ and $\mathcal{K}_M(\boldsymbol{\phi}_0)$ denote the asymptotic variances of the optimal sequential GMM estimators of $\boldsymbol{\eta}$ based on $\mathbf{p}'[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = \{p_2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],...,p_M[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}$ and $\ell_t'(\boldsymbol{\theta},\boldsymbol{\eta}) = [\ell_{2t}(\boldsymbol{\theta},\boldsymbol{\eta}),...,\ell_{Mt}(\boldsymbol{\theta},\boldsymbol{\eta})]$, respectively, which are the orthogonal polynomials and higher order moments of order 2 to M for $\varsigma_t(\boldsymbol{\theta}_0)$. If*

$$[\mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0) - \mathcal{N}_{\mathbf{p}}(\boldsymbol{\phi}_0)]\mathcal{A}^{-1}(\boldsymbol{\phi}_0) = \mathcal{D}_{\mathbf{q}}'(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0), \tag{B6}$$

where $\mathcal{N}_{\mathbf{o}}(\boldsymbol{\phi}_0) = cov\{\mathbf{o}[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0], \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)|\boldsymbol{\phi}_0\}$, $\mathcal{D}_{\mathbf{o}}(\boldsymbol{\phi}_0) = cov\{\mathbf{o}[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0], \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})|\boldsymbol{\phi}_0\}$, $\mathbf{o}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ are some generic influence functions and $\mathbf{q}'[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = \{q_2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],...,q_M[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}'$, with

$$q_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = \ell_{mt}(\boldsymbol{\theta},\boldsymbol{\eta}) - \sum_{j=2}^{m-1} \frac{cov\{\ell_{mt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0), p_j[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]|\boldsymbol{\phi}_0\}}{V\{p_j[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]|\boldsymbol{\phi}_0\}} q_j[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$$

for $j = 2,...,M$, then $\mathcal{J}_M(\boldsymbol{\phi}_0) \leq \mathcal{K}_M(\boldsymbol{\phi}_0)$, with equality if and only if $(\varsigma_t/N - 1)$ can be written as an exact linear combination of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})$, in which case (B6) necessarily holds.

Proof. The first thing to note is that the mapping from $\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})$ to $\mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ is bijective because the coefficients used to recursively construct $\mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ from $\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})$ are the same as the coefficients used to recursively construct $\mathbf{p}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ from $\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})$ and $p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ (see (C9)). Hence, the sequential GMM estimators of $\boldsymbol{\eta}$ based on $\mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ and $\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})$ will be asymptoticaly equivalent.

It is also straightforward to see that

$$cov\{q_2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\} = cov[\ell_{2t}(\boldsymbol{\theta},\boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = cov\{p_2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\}$$

because $cov[p_1(\boldsymbol{\theta},\boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] = 0$. We can then show by induction that

$$cov\{q_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\} = cov\{p_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\}$$

for all $m > 2$, so that

$$\mathcal{H}_{\mathbf{q}}(\boldsymbol{\phi}) = cov\{\mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\} = cov\{\mathbf{p}_t[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}], \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}\} = \mathcal{H}_{\mathbf{p}}(\boldsymbol{\phi}).$$

Therefore, we can ignore the Jacobians in comparing $\mathcal{J}_M(\boldsymbol{\phi}_0)$ with $\mathcal{K}_M(\boldsymbol{\phi}_0)$, focusing instead on the asymptotic variances of the relevant sample moments evaluated at $\tilde{\boldsymbol{\theta}}_T$, which we denote by $\mathcal{E}_{\mathbf{p}}(\boldsymbol{\phi})$ and $\mathcal{E}_{\mathbf{q}}(\boldsymbol{\phi})$. Following Proposition B3, we work with the modified influence functions

$$\begin{aligned}
\mathbf{q}^{\circ}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] &= \mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0}), \\
\mathbf{p}^{\circ}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] &= \mathbf{p}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] - \mathcal{N}_{\mathbf{p}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0}),
\end{aligned}$$

which lead to asymptotically equivalent estimators of $\boldsymbol{\eta}$ but are invariant to the sampling uncertainty surrounding $\tilde{\boldsymbol{\theta}}_T$. It is easy to see that the asymptotic variance of the sample average of $\mathbf{q}^\circ[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]$ is

$$\mathcal{E}_{\mathbf{q}}(\boldsymbol{\phi}_0)=\mathcal{G}_{\mathbf{q}}(\boldsymbol{\phi}_0)+\mathcal{N}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0)-2\mathcal{N}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}_{\mathbf{q}}(\boldsymbol{\phi}_0),$$

where $\mathcal{G}_{\mathbf{o}}(\boldsymbol{\phi}_0)$ denotes the asymptotic variance of the sample average of $\mathbf{o}[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]$, while the the asymptotic variance of the sample average of $\mathbf{p}^\circ[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]$ will be

$$\mathcal{E}_{\mathbf{p}}(\boldsymbol{\phi}_0)=\mathcal{G}_{\mathbf{p}}(\boldsymbol{\phi}_0)+\mathcal{N}'_{\mathbf{p}}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{N}_{\mathbf{p}}(\boldsymbol{\phi}_0).$$

For our purposes, though, it is more convenient to write $\mathbf{q}^\circ[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ as the sum of two asymptotically orthogonal components, namely:

$$\{\mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]-\mathcal{D}_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0})\}+\{\mathcal{D}_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)-\mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\}\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0}),$$

which yields

$$\mathcal{E}_{\mathbf{q}}(\boldsymbol{\phi}_0)=\mathcal{G}_{\mathbf{q}}(\boldsymbol{\phi}_0)-\mathcal{D}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}_{\mathbf{q}}(\boldsymbol{\phi}_0)$$
$$+[\mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0)-\mathcal{D}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)]'\mathcal{C}(\boldsymbol{\phi}_0)[\mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0)-\mathcal{D}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)].$$

In order to compare $\mathcal{G}_{\mathbf{p}}(\boldsymbol{\phi}_0)$ with $\mathcal{G}_{\mathbf{q}}(\boldsymbol{\phi}_0)-\mathcal{D}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}_{\mathbf{q}}(\boldsymbol{\phi}_0)$ we exploit the fact that $\mathbf{p}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]=\boldsymbol{\Upsilon}(\boldsymbol{\phi})\mathbf{n}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ while $\mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]-\mathcal{D}_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{\theta},\mathbf{0})=\boldsymbol{\Upsilon}(\boldsymbol{\phi})\mathbf{m}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$, where $\mathbf{n}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ and $\mathbf{m}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ are the residuals from the theoretical regressions of $\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})$ on $p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ and $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})$, respectively. Although the proof by induction of this statement is rather tedious, intuitively the reason is once again that the coefficients used to recursively construct $\mathbf{q}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ from $\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})$ are the same as the coefficients used to recursively construct $\mathbf{p}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ from $\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})$ and $p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$. It is then possible to prove that

$$V\{\mathbf{n}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}=V[\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{\phi}]-\frac{cov\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}cov'\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}}{\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})+\psi(\boldsymbol{\phi})}$$

while

$$V\{\mathbf{m}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}=V[\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{\phi}]-cov\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})|\boldsymbol{\phi}\}\mathcal{B}^{-1}(\boldsymbol{\phi})cov\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})|\boldsymbol{\phi}\}$$
$$=V[\ell_t(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{\phi}]-\frac{cov\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}cov'\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}}{\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})+\psi(\boldsymbol{\phi})}\frac{\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})}{\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})+\psi(\boldsymbol{\phi})},$$

with $\mathbf{g}(\boldsymbol{\phi})=cov\{p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\}\mathcal{B}^{-1}(\boldsymbol{\phi})$ and $\psi(\boldsymbol{\phi})=V\{p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}-\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}'(\boldsymbol{\phi})$, where we have used the fact that

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})=[\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})+\psi(\boldsymbol{\phi})]^{-1}\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]+\mathbf{w}_t,$$

with $\mathbf{w}_t$ orthogonal to $p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$, so that

$$\begin{aligned} cov\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})|\boldsymbol{\phi}\} &= cov\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),[\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})+\psi(\boldsymbol{\phi})]^{-1}\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\} \\ &= cov\{\ell_t(\boldsymbol{\theta},\boldsymbol{\eta}),p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}[\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathbf{g}(\boldsymbol{\phi})+\psi(\boldsymbol{\phi})]^{-1}\mathbf{g}'(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi}). \end{aligned}$$

As a result,

$$V\{\mathbf{m}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}\geq V\{\mathbf{n}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\},$$

with equality if and only if $\psi(\boldsymbol{\phi})=0$, in which case $(\varsigma_t/N-1)$ can be written as an exact linear combination of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})$.

Hence,

$$\begin{aligned} \mathcal{G}_{\mathbf{p}}(\boldsymbol{\phi}_0) &= \boldsymbol{\Upsilon}(\boldsymbol{\phi})V\{\mathbf{n}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}\boldsymbol{\Upsilon}'(\boldsymbol{\phi}) \\ &\leq \boldsymbol{\Upsilon}(\boldsymbol{\phi})V\{\mathbf{m}[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}\boldsymbol{\Upsilon}'(\boldsymbol{\phi})=\mathcal{G}_{\mathbf{q}}(\boldsymbol{\phi}_0)-\mathcal{D}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}_{\mathbf{q}}(\boldsymbol{\phi}_0). \end{aligned}$$

Therefore, given that

$$\mathcal{N}'_{\mathbf{p}}(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{N}_{\mathbf{p}}(\boldsymbol{\phi}_0)=[\mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0)-\mathcal{D}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)]'\mathcal{C}(\boldsymbol{\phi}_0)[\mathcal{N}_{\mathbf{q}}(\boldsymbol{\phi}_0)-\mathcal{D}'_{\mathbf{q}}(\boldsymbol{\phi}_0)\mathcal{B}^{-1}(\boldsymbol{\phi}_0)\mathcal{A}(\boldsymbol{\phi}_0)]$$

if condition (B6) holds, the main result in the proposition follows.

Finally, the proof that condition (B6) is satisfied when $\psi(\boldsymbol{\phi})=0$ involves the following steps. First, we can prove that in those circumstances

$$cov\{o[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\}=\frac{cov\{o[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}}{V\{p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]|\boldsymbol{\phi}\}}cov\{p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\}$$

for any influence function $o[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ which only depends on $\boldsymbol{\theta}$ through $\varsigma_t(\boldsymbol{\theta})$. Second, given that 0 is trivially orthogonal to both $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})$ and $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta})$, it is clear that the coefficients in the IV regression of $p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ on $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})$ using $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta})$ as instruments will coincide with the coefficients in the least squares projection of $p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ on $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})$ when $(\varsigma_t/N-1)$ can be written as an exact linear combination of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0})$, so that

$$cov\{p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{\phi}\}\mathcal{A}^{-1}(\boldsymbol{\phi})=cov\{p_1[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}],\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\}\mathcal{B}^{-1}(\boldsymbol{\phi}).$$

Finally, we can prove by induction that

$$E[\{q_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]-p_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{\phi}]\mathcal{A}^{-1}(\boldsymbol{\phi})=E\{q_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\}\mathcal{B}^{-1}(\boldsymbol{\phi})$$

when $\psi(\boldsymbol{\phi})=0$. □

Proposition B5 *If* $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ *is i.i.d.* $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ *with bounded fourth moments, then the efficient influence function is given by the efficient parametric score of* $\boldsymbol{\eta}$*:*

$$\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta}) = \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta},\boldsymbol{\eta}) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta}), \tag{B7}$$

which is the residual from the theoretical regression of $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$ *on* $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$*.*

Proof. The first thing to note is that

$$cov[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta},\boldsymbol{\eta}) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})] = \mathbf{0},$$

which means that

$$E\left[\frac{\partial \mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})}{\partial\boldsymbol{\theta}}\right] = \mathbf{0}$$

by virtue of the generalised information equality, which in turn implies that the asymptotic distribution of the sample average of $\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})$ will be invariant to parameter uncertainty in $\boldsymbol{\theta}$ (see Bontemps and Meddahi (2012) for further discussion of this point).

Following Newey and Powell (1998), if $\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})$ is efficient then it will satisfy

$$V\left[\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})\right] = -E\left[\frac{\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right].$$

But

$$V\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta},\boldsymbol{\eta}) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})\right] = \mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0),$$

which coincides with

$$-E\left[\frac{\mathbf{s}_{\boldsymbol{\eta}|\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right] = cov\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta},\boldsymbol{\eta}), \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta},\boldsymbol{\eta}) - \mathcal{I}'_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\eta})\right]. \quad \square$$

Proposition B6 *If* $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1},\boldsymbol{\phi}_0$ *is i.i.d.* $t(\mathbf{0},\mathbf{I}_N,\nu_0)$*, with* $\nu_0 > 8$*, then* $\sqrt{T}(\breve{\eta}_T - \eta_0) \to N\left[0,\mathcal{E}_\ell(\boldsymbol{\phi}_0)/\mathcal{H}^2(\boldsymbol{\phi}_0)\right]$ *and* $\sqrt{T}(\mathring{\eta}_T - \eta_0) \to N\left[0,\mathcal{E}_{\mathbf{p}}(\boldsymbol{\phi}_0)/\mathcal{H}^2(\boldsymbol{\phi}_0)\right]$*, where* $\breve{\eta}_T$ *and* $\mathring{\eta}_T$ *are the sequential MM estimators of* η *based on the square of* ς_t *and its second order polynomial, respectively, while*

$$\begin{aligned}
\mathcal{E}_\ell(\boldsymbol{\phi}_0) &= \mathcal{G}_\ell(\boldsymbol{\phi}_0) + \mathcal{N}'_\ell(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{N}_\ell(\boldsymbol{\phi}_0) - 2\mathcal{N}'_\ell(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathcal{D}_\ell(\boldsymbol{\phi}_0),\\
\mathcal{E}_p(\boldsymbol{\phi}_0) &= \mathcal{G}_p(\boldsymbol{\phi}_0) + \mathcal{N}'_p(\boldsymbol{\phi}_0)\mathcal{C}(\boldsymbol{\phi}_0)\mathcal{N}_p(\boldsymbol{\phi}_0),\\
\mathcal{D}_\ell(\boldsymbol{\phi}_0) &= cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,0),\ell_{2t}(\boldsymbol{\theta}_0,\eta_0)|\boldsymbol{\phi}_0] = \frac{4(\nu_0-2)(N+\nu_0-2)}{N(\nu_0-4)(\nu_0-6)}\mathbf{W}_s(\boldsymbol{\phi}_0),\\
\mathcal{G}_\ell(\boldsymbol{\phi}_0) &= V[\ell_{2t}(\boldsymbol{\theta}_0,\eta_0)|\boldsymbol{\phi}_0] = \frac{(\nu_0-2)^2}{(\nu_0-4)^2}\left[\frac{(N+6)(N+4)}{N(N+2)}\frac{(\nu_0-2)(\nu_0-4)}{(\nu_0-6)(\nu_0-8)} - 1\right],\\
\mathcal{G}_p(\boldsymbol{\phi}_0) &= V\{p_{2t}[\varsigma_t(\boldsymbol{\theta}_0),\eta_0]|\boldsymbol{\phi}_0\} = \mathcal{G}_\ell(\boldsymbol{\phi}_0) - \frac{8(\nu_0-2)^2(N+\nu_0-2)}{N(\nu_0-6)^2(\nu_0-4)},\\
\mathcal{N}_\ell(\boldsymbol{\phi}_0) &= cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\eta_0),\ell_{2t}(\boldsymbol{\theta}_0,\eta_0)|\boldsymbol{\phi}_0] = \frac{4(\nu_0-2)}{N(\nu_0-4)}\mathbf{W}_s(\boldsymbol{\phi}_0),\\
\mathcal{N}_p(\boldsymbol{\phi}_0) &= cov\{\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\eta_0),p_{2t}[\varsigma_t(\boldsymbol{\theta}_0),\eta_0]|\boldsymbol{\phi}_0\} = -\frac{8(\nu_0-2)}{N(\nu_0-4)(\nu_0-6)}\mathbf{W}_s(\boldsymbol{\phi}_0),\\
\mathcal{H}(\boldsymbol{\phi}_0) &= cov[s_{\eta t}(\boldsymbol{\theta}_0,\eta_0),\ell_{2t}(\boldsymbol{\theta}_0,\eta_0)|\boldsymbol{\phi}_0] = cov\{s_{\eta t}(\boldsymbol{\theta}_0,\eta_0),p_{2t}[\varsigma_t(\boldsymbol{\theta}_0),\eta_0]|\boldsymbol{\phi}_0\} = \frac{2\nu_0^2}{(\nu_0-4)^2}
\end{aligned}$$

and

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = \mathbf{Z}_d(\boldsymbol{\phi}_0)[\mathbf{0}', vec'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0][\mathbf{0}', vec'(\mathbf{I}_N)]'$$

$$= E\left\{ \frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]/\partial\boldsymbol{\theta}\cdot vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)] \right| \boldsymbol{\phi}_0 \Bigg\} = E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0] = -E\left\{ \partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \right| \boldsymbol{\phi}_0 \right\}. \quad \text{(B8)}$$

Proof. The linearised influence functions corresponding to $\breve{\eta}_T$ and $\mathring{\eta}_T$ are

$$\ell_{2t}(\boldsymbol{\theta}_0, \eta) - \mathcal{N}_\ell'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0),$$

and

$$p_2[\varsigma(\boldsymbol{\theta}_0), \eta] - \mathcal{N}_p'(\boldsymbol{\phi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0),$$

respectively, whence we can directly obtain the formulae for $\mathcal{E}_\ell(\boldsymbol{\phi}_0)$ and $\mathcal{E}_p(\boldsymbol{\phi}_0)$. Therefore, the only remaining task is to obtain closed-form expressions for the required moments. In this respect, we can use the law of iterated expectations to show that

$$\begin{aligned} cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0), \ell_{2t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] &= \mathbf{Z}_d(\boldsymbol{\phi}_0) \cdot E\{E[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, 0) \cdot \ell_{2t}(\boldsymbol{\theta}_0, \eta_0)|\varsigma_t; \boldsymbol{\phi}_0]|\boldsymbol{\phi}_0\} \\ &= \mathbf{W}_s(\boldsymbol{\phi}_0) E\left[\left(\frac{\varsigma_t}{N} - 1 \right) \ell_{2t}(\boldsymbol{\theta}_0, \eta_0) \right| \boldsymbol{\phi}_0 \Bigg] \end{aligned}$$

and

$$cov[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \eta_0), n_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\boldsymbol{\phi}_0] = \mathbf{Z}_d(\boldsymbol{\phi}_0) \cdot E\{E[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \eta_0) \cdot n_{\eta t}(\boldsymbol{\theta}_0, \eta_0)|\varsigma_t; \boldsymbol{\phi}_0]|\boldsymbol{\phi}_0\}$$
$$= \mathbf{W}_s(\boldsymbol{\phi}_0) E\left[\left(\frac{N + \nu_0}{\nu_0 - 2 + \varsigma_t} \frac{\varsigma_t}{N} - 1 \right) n_{\eta t}(\boldsymbol{\theta}_0, \eta_0) \right| \boldsymbol{\phi}_0 \Bigg].$$

Then, we can use the properties of the beta distribution to show that

$$\begin{aligned} E\left[\left(\frac{\varsigma_t^2}{N(N+2)} - \frac{\nu_0 - 2}{\nu_0 - 4} \right)^2 \right] &= \frac{(\nu_0 - 2)^2}{(\nu_0 - 4)^2} \left[\frac{(N+6)(N+4)}{N(N+2)} \frac{(\nu_0 - 2)(\nu_0 - 4)}{(\nu_0 - 6)(\nu_0 - 8)} - 1 \right], \\ E\left[\left(\frac{\varsigma_t}{N} - 1 \right) \left(\frac{\varsigma_t^2}{N(N+2)} - \frac{\nu_0 - 2}{\nu_0 - 4} \right) \right] &= \frac{4(\nu_0 - 2)(N + \nu_0 - 2)}{N(\nu_0 - 4)(\nu_0 - 6)}, \end{aligned}$$

and

$$E\left[\left(\frac{N + \nu_0}{\nu_0 - 2 + \varsigma_t} \frac{\varsigma_t}{N} - 1 \right) \left(\frac{\varsigma_t^2}{N(N+2)} - \frac{\nu_0 - 2}{\nu_0 - 4} \right) \right] = \frac{4(\nu_0 - 2)}{N(\nu_0 - 4)}.$$

On the other hand, since $p_2[\varsigma(\boldsymbol{\theta}_0), \eta_0]$ is the residual from the least squares projection of $\ell_{2t}(\boldsymbol{\theta}_0, \eta_0)$ on $\varsigma_t/N - 1$, we can obtain the relevant expressions for $p_2[\varsigma(\boldsymbol{\theta}_0), \eta_0]$ from those of $\ell_{2t}(\boldsymbol{\theta}_0, \eta_0)$ by using the fact that

$$E\left[\left(\frac{\varsigma_t}{N} - 1 \right)^2 \right] = \frac{2(N + \nu_0 - 2)}{N(\nu_0 - 4)}$$

and

$$E\left[\left(\frac{N + \nu_0}{\nu_0 - 2 + \varsigma_t} \frac{\varsigma_t}{N} - 1 \right) \left(\frac{\varsigma_t}{N} - 1 \right) \right] = \frac{2}{N}.$$

□

Proposition B7 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1};\varphi_0$, *is i.i.d.* $s(\mathbf{0},\mathbf{I}_N)$, *where* φ *includes* $\boldsymbol{\theta}$ *and the true shape parameters, but the spherical distribution assumed for estimation purposes does not necessarily nest the true density, then the asymptotic distribution of the sequential ML estimator of* $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\eta}}_T$, *will be given by*

$$\sqrt{T}(\tilde{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_\infty) \to N\left\{\mathbf{0}, \mathcal{H}_{rr}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathcal{E}_r(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathcal{H}_{rr}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\right\},$$

where $\boldsymbol{\phi}_\infty = (\boldsymbol{\theta}_0, \boldsymbol{\eta}'_\infty)'$, $\boldsymbol{\eta}_\infty$ *solves* $E[\,\mathbf{e}_{rt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_\infty)|\,\boldsymbol{\varphi}_0] = \mathbf{0}$, $\mathcal{H}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = -E[\,\partial\mathbf{e}_{rt}(\boldsymbol{\phi})/\partial\boldsymbol{\eta}'|\,\boldsymbol{\varphi}]$,

$$\begin{aligned}\mathcal{E}_r(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= [\mathcal{O}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]^{-1} \\ &\quad + [\mathcal{O}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]^{-1}\mathcal{O}'_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathcal{C}(\boldsymbol{\varphi}_0)\mathcal{O}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)[\mathcal{O}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]^{-1},\end{aligned}$$

$\mathcal{O}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = E[-\partial\mathbf{e}_{rt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_\infty)/\partial\boldsymbol{\theta}'|\,\boldsymbol{\varphi}]$ *and* $\mathcal{O}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) = V[\,\mathbf{e}_{rt}(\boldsymbol{\phi})|\,\boldsymbol{\varphi}]$.

Proof. To obtain the variance of the elliptically symmetric score of $\boldsymbol{\eta}$ under misspecification, we can follow exactly the same steps as in the proof of Proposition 10 in Fiorentini and Sentana (2010) by exploiting that $E[\,\mathbf{e}_{rt}(\boldsymbol{\phi})|\,\boldsymbol{\varphi}_0] = \mathbf{0}$ holds at the pseudo-true parameter values $\boldsymbol{\phi}_\infty = (\boldsymbol{\theta}_0, \boldsymbol{\eta}'_\infty)'$. Specifically, under standard regularity conditions

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\mathbf{e}_{rt}(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta}_\infty) = \frac{\sqrt{T}}{T}\sum_{t=1}^{T}\mathbf{e}_{rt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_\infty) + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\mathbf{e}_{rt}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_\infty)}{\partial\boldsymbol{\theta}'}\frac{\sqrt{T}}{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1).$$

Finally, the same steps used in the proof of Proposition 3 yield the expression for $\mathcal{E}_r(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$.□

C Orthogonal polynomials

The m^{th} orthogonal polynomial associated to a spherical distribution for $\boldsymbol{\varepsilon}_t^*$ is given by

$$p_m^s[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = \sum_{h=0}^{m} a_h^s(\boldsymbol{\eta})\varsigma_t^h(\boldsymbol{\theta}),$$

where $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$ and $\boldsymbol{\eta}$ are the shape parameters. The first two non-normalised polynomials are always $p_0[\varsigma_t(\boldsymbol{\theta})] = 1$ and

$$p_1[\varsigma_t(\boldsymbol{\theta})] = \frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1,$$

which do not depend on $\boldsymbol{\eta}$. Subsequent polynomials can be obtained by recursively regressing $\ell_{mt}(\boldsymbol{\theta},\boldsymbol{\eta})$ in (8) on $p_j[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ for $j = 0, 1, \ldots, m-1$. Specifically,

$$p_m[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = \ell_{mt}(\boldsymbol{\theta},\boldsymbol{\eta}) - \sum_{j=1}^{m-1}\frac{cov\{\ell_{mt}(\boldsymbol{\theta},\boldsymbol{\eta}), p_j[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}}{V\{p_j[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\}}p_j[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]. \quad \text{(C9)}$$

As a result, the polynomials have zero mean and are orthogonal to each other by construction, although not orthonormal unless we standardise them by their respective standard deviations.

Next, we present the coefficients for the second and third orthogonal polynomials of the distributions we use in Sections 3, 4 and 5 to illustrate our results.

C.1 Orthogonal Laguerre polynomials for the standard normal distribution

The coefficients of the second order polynomial are

$$a_0^g = \frac{1}{2},$$
$$a_1^g = -\frac{1}{N}$$

and

$$a_2^g = \frac{1}{2N\,(N+2)};$$

while the third order polynomial coefficients become

$$a_0^g = \frac{1}{2},$$
$$a_1^g = -\frac{3}{2N},$$
$$a_2^g = \frac{3}{2N(N+2)}$$

and

$$a_3^g = -\frac{1}{2N\,(N+2)\,(N+4)}.$$

C.2 Orthogonal polynomials for the standardised Student t distribution

In this case the coefficients of the second order polynomial are

$$a_0^t(\eta) = \frac{N\,(N+2)\,(\nu-2)^3}{4\,(\nu-4)\,(\nu-6)},$$
$$a_1^t(\eta) = -\frac{(N+2)\,(\nu-2)}{2\,(\nu-6)}$$

and

$$a_2^t(\eta) = \frac{1}{4},$$

while the third order polynomial coefficients become

$$a_0^t(\eta) = -\frac{N\,(N+2)\,(N+4)\,(\nu-2)^2}{8\,(\nu-6)\,(\nu-8)\,(\nu-10)},$$
$$a_1^t(\eta) = \frac{(N+2)\,(N+4)\,(\nu-2)}{8\,(\nu-8)\,(\nu-10)},$$
$$a_2^t(\eta) = -\frac{(N+4)\,(\nu-2)}{8\,(\nu-10)},$$

and

$$a_3^t(\eta) = \frac{1}{24}.$$

C.3 Orthogonal polynomials for the standardised DSMN distribution

In this case the coefficients of the second order polynomial are

$$a_0^{ds}(\alpha,\varkappa) = \frac{N(N+2)}{8}\frac{1}{[\alpha(1-\varkappa)+\varkappa]^4}\left\{2(1-\alpha)\varkappa^4+(N+4)(1-\alpha)\alpha\varkappa^3 -2(N+2)(1-\alpha)\alpha\varkappa^2+(N+4)(1-\alpha)\alpha\varkappa+2\alpha^2\right\},$$

$$a_1^{ds}(\alpha,\varkappa) = -\frac{N+2}{8}\frac{1}{[\alpha(1-\varkappa)+\varkappa]^3}\left\{2(1-\alpha)(4+N\alpha)\varkappa^3 -N(1-\alpha)\alpha\varkappa^2-N(1-\alpha)\alpha\varkappa+\alpha\left[N(1-\alpha)+4\right]\right\},$$

and

$$a_2^{ds}(\alpha,\varkappa)=\frac{1}{8}\frac{1}{[\alpha(1-\varkappa)+\varkappa]^2}\left\{(1-\alpha)(2+N\alpha)\varkappa^2-2N(1-\alpha)\alpha\varkappa+\left[2+N(1-\alpha)\right]\alpha\right\}.$$

Similarly, the coefficients of the third order polynomial are

$$a_0^{ds}(\alpha,\varkappa)=-\frac{N(N+2)(N+4)}{192}\frac{1}{[(\alpha(1-\varkappa)+\varkappa]^6}$$
$$\times\left\{8(1-\alpha)^3\varkappa^9+(N+4)(N+6)(1-\alpha)^2\alpha\varkappa^8\right.$$
$$-4(N+2)(N+6)(1-\alpha)^2\alpha\varkappa^7+6(N+4)^2(1-\alpha)^2\alpha\varkappa^6$$
$$+(N+6)(1-\alpha)\alpha\left[(5N+24)\alpha-4(N+4)\right]\varkappa^5$$
$$-(N+6)(1-\alpha)\alpha\left[24\alpha+N(5\alpha-1)-8\right]\varkappa^4$$
$$+6(N+4)^2(1-\alpha)\alpha^2\varkappa^3-4(N+2)(N+6)(1-\alpha)\alpha^2\varkappa^2$$
$$\left.+(N+4)(N+6)(1-\alpha)\alpha^2\varkappa+8\alpha^3\right\},$$

$$a_1^{ds}(\alpha,\varkappa)=\frac{(N+2)(N+4)}{192}\frac{1}{[(\alpha(1-\varkappa)+\varkappa]^5}$$
$$\times\left\{(1-\alpha)^2\left[N(N+10)\alpha+24\right]\varkappa^8-2N(N+6)(1-\alpha)^2\alpha\varkappa^7\right.$$
$$-2N(N+8)(1-\alpha)^2\alpha\varkappa^6$$
$$+2(1-\alpha)\alpha\left[4N(N+7)-N(3N+14)\alpha+48\right]\varkappa^5$$
$$-(N+6)(7N+24)(1-\alpha)\alpha\varkappa^4$$
$$+2(1-\alpha)\alpha\left[N(3\alpha N+N+14(\alpha+1))+48\right]\varkappa^3$$
$$-2N(N+8)(1-\alpha)\alpha^2\varkappa^2-2N(N+6)(1-\alpha)\alpha^2\varkappa$$
$$\left.+\left[24-N(N+10)(\alpha-1)\right]\alpha^2\right\},$$

$$
\begin{aligned}
a_2^{ds}(\alpha,\varkappa) \;=\; & -\frac{N+4}{192}\frac{1}{[(\alpha(1-\varkappa)+\varkappa]^4} \\
& \times\left\{2(1-\alpha)^2\left[N(N+8)\alpha+12\right]\varkappa^7 - N(7N+38)(1-\alpha)^2\alpha\varkappa^6\right. \\
& +(1-\alpha)\alpha\left[8(N+2)(N+3)-N(7N+26)\alpha\right]\varkappa^5 \\
& -2(N+2)(N+6)(1-\alpha)\alpha\varkappa^4 - 2(N+2)(N+6)(1-\alpha)\alpha\varkappa^3 \\
& +(1-\alpha)\alpha\left[N(7\alpha N+N+26\alpha+14)+48\right]\varkappa^2 \\
& \left.-N(7N+38)(1-\alpha)\alpha^2\varkappa+2\left[12+N(N+8)(1-\alpha)\right]\alpha^2\right\}
\end{aligned}
$$

and

$$
\begin{gathered}
a_3^{ds}(\alpha,\varkappa) = \frac{1}{192}\frac{1}{[(\alpha(1-\varkappa)+\varkappa]^3} \\
\times\left\{(1/\alpha)^2\left[N(N+6)\alpha+8\right]\varkappa^6 - 4N(N+4)(1-\alpha)^2\alpha\varkappa^5\right. \\
+(1-\alpha)\alpha\left[6(N+2)^2 - N(5N+14)\alpha\right]\varkappa^4 \\
-4(N+2)(N+4)(1-\alpha)\alpha\varkappa^3 \\
+(1-\alpha)\alpha\left[N(5\alpha N+N+14\alpha+10)+24\right]\varkappa^2 \\
\left.-4N(N+4)(1-\alpha)\alpha^2\varkappa+\left[8-N(N+6)(\alpha-1)\right]\alpha^2\right\}.
\end{gathered}
$$

C.4 Orthogonal polynomials for the standardised 3^{rd}-order PE distribution

The coefficients of the second order polynomial are

$$
\begin{gathered}
a_0^{pe}(c_2,c_3) = \frac{1}{4}\left(-8c_2^2+2N(N+8)c_2+N\left(N(N+2)-12c_3\right)\right), \\
a_1^{pe}(c_2,c_3) = -\frac{1}{2}\left[N(N+2)+(N+6)c_2-3c_3\right],
\end{gathered}
$$

and

$$
a_2^{pe}(c_2,c_3) = \frac{1}{4}\left(N+2c_2\right).
$$

In turn, the coefficients of the third order one are

$$
\begin{gathered}
a_0^{pe}(c_2,c_3) = -\frac{1}{24}\left\{(N+2)^2(N+4)N^3 - 4\left(N^3+32N+192\right)c_2^3\right. \\
+6c_2^2\left[N(N+4)(N(N+2)+48)+4((N-4)N+48)c_3\right] \\
-12c_3\left[(N+2)(7N+48)N^2+3(N(N+6)+24)c_3N-72c_3^2\right] \\
-6c_2N^2(N+2)(N+4)(N+12) \\
\left.+24c_2c_3\left[N(N(N+14)+120)+6(N-12)c_3\right]\right\},
\end{gathered}
$$

$$a_1^{pe}(c_2,c_3) = \frac{1}{8}\left\{N^2(N+4)(N+2)^2 - 12\left[3(N(N+10)+32)+4c_2\right]c_3^2\right.$$

$$+2(N+4)c_2\left[N(N+2)(3N+32)+c_2\left(N(3N+14)-2Nc_2+96\right)\right]$$

$$\left.-8\left[3N(N+2)(3N+20)+2c_2\left(N(N+14)-Nc_2+96\right)\right]c_3\right\},$$

$$a_2^{pe}(c_2,c_3) = \frac{1}{8}\left\{(N+2)(N+4)N^2 - 36(N+8)c_3^2\right.$$

$$+2(N+4)c_2\left[N(3N+26)+(3N-2c_2+8)\,c_2\right]$$

$$\left.-4\left[15N(N+6)+2\left(N-c_2+12\right)c_2\right]c_3\right\},$$

and

$$a_3^{pe}(c_2,c_3) = \frac{1}{24}\left[N^2(N+2) - 4c_2^3 + 6Nc_2^2 + 6N(N+6)c_2 - 12c_3\left(4N+3c_3\right)\right].$$

D Auxiliary results

D.1 Positivity of Laguerre expansions

To identify the region in the parameter space for which $P_J(\varsigma) = 1 + \sum_{j=2}^{J} c_j \cdot p_j(\varsigma, N) \geq 0$ it is convenient to reparametrise $P_J(\varsigma)$ as $\mathring{P}_J(c_2, c_3, \mathbf{t})$, with $\mathbf{t} = (\varsigma, c_4, ..., c_J)$. For each value of $\mathbf{t} \in \mathbb{R}^{J-2}$, the equation $\mathring{P}_J(c_2, c_3, \mathbf{t}) = 0$ defines a straight line in the $\mathbf{t}$-hyperplane. To determine the set of $\boldsymbol{\eta}$'s as a function of $\mathbf{t}$ such that $\mathring{P}_J(c_2, c_3, \mathbf{t})$ remains zero for small variations of $\mathbf{t}$, we should also impose $\partial\mathring{P}_J(c_2, c_3, \mathbf{t})/\partial\mathbf{t} = 0$. Finally, once this bound is found, we need to determine the subregion in which $P_J(\varsigma) \geq 0$ for $\varsigma \geq 0$.

D.1.1 Second order expansion

In this simple case the positivity region corresponds to those values of c_2 for which the polynomial $1 + c_2 \cdot p_2(t, N)$ is positive. Since the vertex of this quadratic function occurs at $t = N + 2 > 0$, positivity requires that its roots are either complex or double, which holds for $0 \leq c_2 \leq N$.

D.1.2 Third order expansion

For a given ς, the 3^{rd} order polynomial frontier that guarantees positivity must satisfy the following two equations in two unknowns

$$\begin{cases} 1 + c_2 \cdot p_2(t,N) + c_3 \cdot p_3(t,N) = 0 \\ c_2 \cdot \partial p_2(t,N)/\partial t - c_3 \cdot \partial p_3(t,N)/\partial t = 0 \end{cases}$$

whose solution is

$$c_2(t) = \frac{8 + 6N + N^2 - 8t - 2Nt + t^2}{8A(N,t)} \quad \text{and } c_3(t) = \frac{N + 2 - t}{2A(N,t)}$$

with

$$\begin{aligned} A(N,t) &= \frac{N^3 t + N t^3 - 5N^2}{24} + \frac{t^3 - N^3}{12} - \frac{N^4 + t^4}{96} \\ &\quad + \frac{Nt - 2N}{3} + \frac{N^2 t - N t^2 - t^2}{4} - \frac{N^2 t^2}{16}. \end{aligned}$$

The solid (dashed) black line in Figure 1 represents the frontier defined by positive (negative) values of ς. Notice that if we imposed the above conditions for all $\varsigma \in \mathbb{R}$, then $c_3 = 0$ and $0 \leq c_2 \leq N$. Such a frontier, however, is overly restrictive because it does not take into account the non-negativity of ς. In this sense, the blue line represents the tangent of $P_3(\varsigma)$ at $\varsigma = 0$ while the red line is the tangent of $P_3(\varsigma)$ when $\varsigma \to +\infty$. The grey area, therefore, defines the admissible set in the (c_2, c_3) space. Focusing on $\varsigma \in \mathbb{R}_+$ only allows for a larger range of (c_2, c_3) with $c_3 < 0$, which is given by the difference between the dashed black line and the blue one.

D.2 Higher order moments

The higher order moment parameter of spherical random variables defined in (E10) for the four distributions that we use to illustrate our results are:

(a) Student t distribution with $\nu = 1/\eta$ degrees of freedom:

$$1 + \tau_m^t(\eta) = (1 - 2\eta)^{m-1} \prod_{j=2}^{m} \frac{1}{(1 - 2j\eta)} \quad \text{when} \quad \eta < (2m)^{-1}.$$

(b) Discrete scale mixture of normals distribution with mixing probability α and variance ratio $\varkappa$:

$$1 + \tau_m^{ds}(\alpha, \varkappa) = \frac{\alpha + (1 - \alpha)\,\varkappa^m}{[\alpha + (1 - \alpha)\varkappa]^m}.$$

(c) 3^{rd}-order polynomial expansion distribution with parameters c_2 and c_3:

$$1 + \tau_m^{pe}(\alpha, \varkappa) = 1 + \frac{2m(m-1)}{N(N+2)} c_2 - \frac{4m\,[2 + m(m-3)]}{N(N+2)(N+4)} c_3.$$

Derivation of the results:

(a) If ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is a Gamma variate with mean ν and variance 2ν, with ζ_t and ξ_t mutually independent, then the uncentred moments of integer order r of $(\nu/N) \times (\zeta_t/\xi_t)$ are given by

$$E\left[\left(\frac{\zeta_t/N}{\xi_t/\nu}\right)^r\right] = \left(\frac{\nu}{N}\right)^r \frac{r - 1 + N/2}{-1 + \nu/2} \frac{r - 2 + N/2}{-2 + \nu/2} \times \cdots \times \frac{1 + N/2}{-(r-1) + \nu/2} \frac{N/2}{-r + \nu/2}$$

(Mood, Graybill and Boes, 1974). Given that $\varsigma_t = (\nu-2)\zeta_t/\xi_t$, it is straightforward to see that

$$E\left\{\left[(\nu-2)\frac{\zeta_t}{\xi_t}\right]^m\right\} = \frac{N}{2}\left[\frac{2(\nu-2)}{\nu}\right]^{m-1}\prod_{j=2}^{m}\frac{(N/2+j-1)\nu}{\nu-2j}$$

from where the result follows directly.

(b) When $\boldsymbol{\varepsilon}_t^*$ is distributed as a DSMN, ς_t is a two-component scale mixture of χ_N^2's, so that conditioning on the mixing variate s,

$$E[\varsigma_t^m|s=1] = \left[\frac{1}{\alpha+(1-\alpha)\varkappa}\right]^m E(\zeta_t^m) \text{ and } E[\varsigma_t^m|s=0] = \left[\frac{\varkappa}{\alpha+(1-\alpha)\varkappa}\right]^m E(\zeta_t^m)$$

where ζ_t is a χ_N^2 variate. Then, the required expression follows directly from the law of iterated expectations.

(c) Since $E[\varsigma_t^m p_{N/2-1,j}(\varsigma_t)|\mathbf{0}] = 0$ for $m < j$, we only need to compute $E[\varsigma_t^m p_{N/2-1,j}(\varsigma_t)|\mathbf{0}]$ for $m \geq j$, which can be written in terms of the higher order moments of the Gaussian distribution. For the 2^{nd}-order Laguerre polynomial we have

$$\begin{aligned} E[\varsigma_t^m p_{N/2-1,2}(\varsigma_t)|\mathbf{0}] &= \frac{1}{2}E[\varsigma_t^m|\mathbf{0}] - \frac{1}{N}E[\varsigma_t^{m+1}|\mathbf{0}] + \frac{1}{2N\,(N+2)}E[\varsigma_t^{m+2}|\mathbf{0}] \\ &= \left[\frac{1}{2} - \frac{2(N/2+m+1)}{N} + \frac{4(N/2+m+1)(N/2+m+2)}{2N\,(N+2)}\right]E[\varsigma_t^m|\mathbf{0}] \\ &= \frac{2m(m-1)}{N(N+2)}E[\varsigma_t^m|\mathbf{0}]. \end{aligned}$$

The same procedure applied to the 3^{rd}-order Laguerre polynomial yields the required result.

D.3 Moment generating functions

Not surprisingly, the moment generating function of a spherical random variable $\boldsymbol{\varepsilon}_t^*$ depends only on ς. Although it cannot be defined for the Student t distribution, it takes the following forms for the remaining distributions that we consider:

(a) Discrete scale mixture of normals distribution with mixing probability α and variance ratio $\varkappa$:

$$\Upsilon_{ds}(t|\alpha,\varkappa) \equiv E[e^{t\varsigma_t}|(\alpha,\varkappa)'] = \alpha\left[1-\frac{2t}{\alpha+(1-\alpha)\varkappa}\right]^{-N/2} + (1-\alpha)\left[1-\frac{2\varkappa t}{\alpha+(1-\alpha)\varkappa}\right]^{-N/2}.$$

(b) 3^{rd}-order polynomial expansion with parameters c_2 and c_3:

$$\Upsilon_{pe}^{J=3}(t|c_2,c_3) \equiv E[e^{t\varsigma_t}|(c_2,c_3)'] = (1-2t)^{-N/2}\left[1+\frac{2t^2}{(1-2t)^2}c_2 - \frac{4t^3}{(1-2t)^3}c_3\right].$$

Derivation of the results:

(a) Since ς_t is a two-component scale mixture of $\chi_N^{2\prime}s$, we can compute $E[e^{t\varsigma_t}|\alpha,\varkappa,s]$ for $s=1$ and $s=0$ by exploiting the fact that the relevant conditional distribution are Gamma with shape parameter $N/2$ and scale parameters

$$\frac{2}{\alpha+(1-\alpha)\varkappa} \quad \text{and} \quad \frac{2\varkappa}{\alpha+(1-\alpha)\varkappa}$$

respectively. Finally, the law of iterated expectation yields the desired result.

(b) The moment generating function of the polynomial expansion distribution can be easily obtained by applying Lemma 1 in Amengual and Sentana (2011). For the 2^{nd}-order Laguerre polynomial we have

$$\begin{aligned} E[e^{t\varsigma_t}p_{N/2-1,2}(\varsigma_t)|\mathbf{0}] &= \frac{1}{2}E[e^{t\varsigma_t}|\mathbf{0}]-\frac{1}{N}E[\varsigma_t e^{t\varsigma_t}|\mathbf{0}]+\frac{1}{2N\,(N+2)}E[\varsigma_t^2 e^{t\varsigma_t}|\mathbf{0}] \\ &= \frac{1}{2}\left(\frac{1}{1-2t}\right)^{N/2}-\left(\frac{1}{1-2t}\right)^{N/2+1}+\frac{1}{2}\left(\frac{1}{1-2t}\right)^{N/2+2} \\ &= (1-2t)^{-N/2}\left[\frac{(1-2t)^2-2(1-2t)+1}{2(1-2t)^2}\right] \\ &= (1-2t)^{-N/2}\,\frac{2t^2}{(1-2t)^2}. \end{aligned}$$

The same procedure applied to the 3^{rd}-order Laguerre polynomial yields the required result.

D.4 Marginal and conditional distributions required for VaR and CoVaR calculations

Theorem 2.6 in Fang, Kotz and Ng (1990) characterises the marginal distribution of a partition of $\boldsymbol{\varepsilon}_t^*$ into n components. In particular, if we split $\boldsymbol{\varepsilon}_t^*$ into its first n elements, $\boldsymbol{\varepsilon}_{1t}^*$, and the remaining $N-n$ ones, $\boldsymbol{\varepsilon}_{2t}^*$ say, this theorem implies that

$$\begin{bmatrix} \boldsymbol{\varepsilon}_{1t}^* \\ \boldsymbol{\varepsilon}_{2t}^* \end{bmatrix} = \begin{bmatrix} e_t d_t \mathbf{u}_{1t} \\ e_t(1-d_t)\mathbf{u}_{2t} \end{bmatrix},$$

where e_t is the generating variate, $d_t \sim Beta[n/2,(N-n)/2]$ and $\mathbf{u}_{1t}$ and $\mathbf{u}_{2t}$ are two independent vectors which are uniformly distributed on the unit sphere surface in $\mathbb{R}^n$ and $\mathbb{R}^{N-n}$, respectively.

D.4.1 Marginal densities and CDFs of $z_{1t}=[\varepsilon_{1t}^*(\boldsymbol{\theta})]^2$

In the particular case of univariate marginals, it is easy to obtain the marginal probability density function of ε_{it}^{*2} (see Mood, Graybill, and Boes, 1974) by computing

$$h_1^s(z,\boldsymbol{\eta})=\frac{\sqrt{\pi}\Gamma(N/2-1/2)}{\Gamma(N/2)}\int_0^\infty h^s\left(\frac{z}{y},\boldsymbol{\eta},N\right)\times\frac{(1-y)^{\frac{N}{2}-\frac{3}{2}}}{\sqrt{y}}dy.$$

>From here, we can easily obtain $f_1^s(\varepsilon,\boldsymbol{\eta})$ using the change of variable formula as $h_1^s(\varepsilon^2,\boldsymbol{\eta})\cdot|z|$.

Student t For the N-variate standardised Student t distribution with $\nu = 1/\eta$ degrees of freedom, the univariate marginal probability density function of $z = \varepsilon^2$ is

$$h_1^t(z,\eta) = \frac{\Gamma[1/2(1+\eta^{-1})]}{\sqrt{\pi}\Gamma[(2\eta)^{-1}]} \frac{1}{\sqrt{z(z+\eta^{-1}-2)}} \left[\frac{1-2\eta}{1-2\eta+\eta z}\right]^{\frac{1}{2\eta}},$$

while its cumulative distribution function is

$$H_1^t(z,\eta) = \frac{\Gamma[1/2(1+\eta^{-1})]}{\sqrt{\pi}\Gamma[(2\eta)^{-1}]} \cdot i \cdot Beta\left(-\frac{\eta z}{1-2\eta}, \frac{1}{2}, \frac{\eta-1}{2\eta}\right),$$

where $i = \sqrt{-1}$ and $Beta(z,a,b)$ is the incomplete beta function, defined by

$$Beta\,(z,a,b) = \int_0^z u^{a-1}(1-u)^{b-1}du.$$

Discrete scale mixture of normals For the N-variate standardised DSMN distribution with mixing probability α and variance ratio $\varkappa$:

$$h_1^{ds}(z,\alpha,\varkappa) = \frac{1}{\sqrt{z}\sqrt{2\pi}} \left\{\alpha\sqrt{\frac{1}{\varpi}}\exp\left(-\frac{1}{2\varpi}z\right) + \left(\frac{1-\alpha}{\varkappa}\right)\sqrt{\frac{1}{\varpi\varkappa}}\exp\left(-\frac{1}{2\varpi\varkappa}z\right)\right\},$$

where $\varpi = [\alpha + (1-\alpha)\varkappa]^{-1}$, while its cumulative distribution function is

$$H_1^{ds}(z,\alpha,\varkappa) = (1-\alpha)\,\mathrm{erf}\left(\frac{\sqrt{z}}{\sqrt{2\varpi\varkappa}}\right) + \alpha\,\mathrm{erf}\left(\frac{\sqrt{z}}{\sqrt{2\varpi}}\right),$$

where $\mathrm{erf}\,(x)$ is the standard "error function" defined by $\mathrm{erf}\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$.

3^{rd}-order polynomial expansion For the N-variate standardised 3^{rd}-order PE with parameters c_2 and c_3

$$h_1^{pe}(z,c_2,c_3) = \left\{1 + \frac{[z(z-6)+3]}{2N(N+2)}c_2 - \frac{[z(45+z(z-15))-15]}{2N(N+2)(N+4)}c_3\right\} \frac{1}{\sqrt{z}\sqrt{2\pi}}\exp\left(-\frac{z}{2}\right),$$

while its cumulative distribution function is

$$H_1^{pe}(z,c_2,c_3) = \frac{[15+(z-10)z]\,c_3 - (N+4)(z-3)c_2}{N(N+2)(N+4)\sqrt{2\pi}}\sqrt{z}\exp\left(-\frac{z}{2}\right) + \mathrm{erf}\left(\frac{\sqrt{z}}{\sqrt{2}}\right).$$

D.4.2 Cumulative density functions of conditionals $(\varepsilon_{1t}^*(\boldsymbol{\theta})|\varepsilon_{2t}^*(\boldsymbol{\theta}))$

Using again Theorem 2.6 in Fang, Kotz and Ng (1990) we can obtain the marginal bivariate distribution $f_{1,2}^s(\varepsilon_1,\varepsilon_2,\boldsymbol{\eta})$, which together with $f_1^s(\varepsilon,\boldsymbol{\eta}) = h_1^s(\varepsilon^2,\boldsymbol{\eta})\cdot|z|$, allow us to obtain the conditional pdfs. In this way,

Student t

$$F^{t}_{1|2}(\varepsilon_1,\varepsilon_2,\eta)=\frac{1}{2}\left\{1+\frac{\Gamma[1+(2\eta)^{-1}]}{\sqrt{\pi}\Gamma[(1+\eta)(2\eta)^{-1}]}Beta\left(-\frac{\eta\varepsilon_1^2}{1+(\varepsilon_2^2-2)\eta},\frac{1}{2},-\frac{1}{2\eta}\right)\right\}.$$

Discrete scale mixture of normals

$$\begin{aligned}F^{ds}_{1|2}(\varepsilon_1,\varepsilon_2,\alpha,\varkappa) &= \left\{(1-\alpha)\exp\left(\frac{2\alpha+\varkappa}{2}\varepsilon_2^2\right)\sqrt{\frac{1}{\varpi\varkappa}}+\alpha\exp\left(\frac{\alpha+\varkappa+\alpha\varkappa^2}{2\varkappa}\varepsilon_2^2\right)\sqrt{\frac{1}{\varpi}}\right\}^{-1}\\ &\quad \frac{1}{2\sqrt{\varpi\varkappa}}\left\{(1-\alpha)\exp\left(\frac{2\alpha+\varkappa}{2}\varepsilon_2^2\right)\left[1+\text{erf}\left(\sqrt{\frac{1}{2\varpi\varkappa}}\varepsilon_1\right)\right]\right.\\ &\quad \left.+\alpha\exp\left(\frac{\alpha+\varkappa+\alpha\varkappa^2}{2\varkappa}\varepsilon_2^2\right)\sqrt{\varkappa}\left[1+\text{erf}\left(\sqrt{\frac{1}{2\varpi}}\varepsilon_1\right)\right]\right\},\end{aligned}$$

where $\varpi=[\alpha+(1-\alpha)\varkappa]^{-1}$.

3^{rd}-order polynomial expansion

$$\begin{aligned}F^{pe}_{1|2}(\varepsilon_1,\varepsilon_2,c_2,c_3,N) &= \frac{1}{2}+\frac{1}{2\sqrt{\pi}}\left[2+\frac{\varepsilon_2^4-6\varepsilon_2^2+3}{N(N+2)}c_2-\frac{\varepsilon_2^6-15\varepsilon_2^4+45\varepsilon_2^2-15}{N(N+2)(N+4)}c_3\right]^{-1}\\ &\quad\times\exp\left(-\frac{\varepsilon_1^2}{2}\right)\left\{\exp\left(-\frac{\varepsilon_1^2}{2}\right)\sqrt{\pi}\,\text{erf}\left(\frac{\varepsilon_1}{\sqrt{2}}\right)\right.\\ &\quad\times\left(2+\frac{\varepsilon_2^4-6\varepsilon_2^2+3}{N(N+2)}c_2-\frac{\varepsilon_6^6-15\varepsilon_2^4+45\varepsilon_2^2-15}{N(N+2)(N+4)}c_3\right)\\ &\quad\left.+\sqrt{2}\varepsilon_1\left(\frac{\left(\varepsilon_1^4-13\varepsilon_1^2+3\varepsilon_2^4+3\left(\varepsilon_1^2-9\right)\varepsilon_2^2+33\right)}{N(N+2)(N+4)}c_3-\frac{\varepsilon_1^2+2\varepsilon_2^2-5}{N(N+2)}c_2\right)\right\}.\end{aligned}$$

D.5 Standard errors for parametric VaR and CoVaR

Given that $q_1(\lambda,\boldsymbol{\eta})$ satisfies

$$\lambda=F\left[q_1(\lambda,\boldsymbol{\eta}),\boldsymbol{\eta}\right]=\int_0^{q_1(\lambda,\boldsymbol{\eta})}f_1\left(\varepsilon_{1t}^*;\boldsymbol{\eta}\right)d\varepsilon_{1t}^*,$$

if we differentiate this expression with respect to $\boldsymbol{\eta}$ we obtain

$$0=f_1\left[q_1(\lambda,\boldsymbol{\eta});\boldsymbol{\eta}\right]\frac{\partial q_1(\lambda,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}+\int_0^{q_1(\lambda,\boldsymbol{\eta})}\frac{\partial f_1\left(\varepsilon_{1t}^*;\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}}d\varepsilon_{1t}^*,$$

whence

$$\frac{\partial q_1(\lambda,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}=-\frac{1}{f_1\left[q_1(\lambda,\boldsymbol{\eta});\boldsymbol{\eta}\right]}\int_0^{q_1(\lambda,\boldsymbol{\eta})}\frac{\partial f_1\left(\varepsilon_{1t}^*;\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}}d\varepsilon_{1t}^*.$$

To relate this expression to the asymptotic variances of the non-parametric quantile estimators, it is convenient to write

$$\begin{aligned}\int_0^{q_1(\lambda,\boldsymbol{\eta})}\frac{\partial f_1\left(\varepsilon_{1t}^*;\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}}d\varepsilon_{1t}^* &= \int_0^{q_1(\lambda,\boldsymbol{\eta})}\frac{\partial\ln f_1\left(\varepsilon_{1t}^*;\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}}f_1\left(\varepsilon_{1t}^*;\boldsymbol{\eta}\right)d\varepsilon_{1t}^*\\ &= \Pr\left[\varepsilon_{1t}^*\leq q_1(\lambda,\boldsymbol{\eta})\right]E\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\varepsilon_{1t}^*\leq q_1(\lambda,\boldsymbol{\eta})\right.\right],\end{aligned}$$

where, importantly, the distribution used to compute the foregoing expectation is the same as the distribution used for estimation purposes. Hence, we will have that

$$V[q_1(\lambda,\hat{\boldsymbol{\eta}}_T)] = \frac{\lambda^2}{f_1^2\left[q_1\left(\lambda,\boldsymbol{\eta}\right);\boldsymbol{\eta}_0\right]} E\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\varepsilon_{1t}^*\leq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}_0\right.\right] V\left[\hat{\boldsymbol{\eta}}\left|\boldsymbol{\eta}_0\right.\right] E\left[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\varepsilon_{1t}^*\leq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}_0\right.\right].$$

Further, given that

$$0 = E\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\boldsymbol{\eta}_0\right.\right] = \lambda E\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\varepsilon_{1t}^*\leq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}_0\right.\right] + (1-\lambda) E\left[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\varepsilon_{1t}^*\geq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}_0\right.\right],$$

we can finally write

$$V[q_1(\lambda,\hat{\boldsymbol{\eta}}_T)] = \frac{\lambda(1-\lambda)}{f^2\left[q_1\left(\lambda,\boldsymbol{\eta}\right);\boldsymbol{\eta}_0\right]} E\left[\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\varepsilon_{1t}^*\leq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}_0\right.\right] V\left[\hat{\boldsymbol{\eta}}\left|\boldsymbol{\eta}_0\right.\right] E\left[\mathbf{s}'_{\boldsymbol{\eta}t}(\boldsymbol{\phi})\left|\varepsilon_{1t}^*\geq q_1\left(\lambda,\boldsymbol{\eta}\right),\boldsymbol{\eta}_0\right.\right].$$

Let $f_{1,2}$ denote the joint bivariate distribution of ε_{1t}^* and ε_{2t}^*. By definition, we know that $q_{2|1}\left(\lambda_2,\lambda_1,\boldsymbol{\eta}\right)$ satisfies

$$\begin{aligned}
\lambda_{2|1} &= \int_{-\infty}^{q_1(\lambda_1,\boldsymbol{\eta})} f_1(\varepsilon_{1t}^*;\boldsymbol{\eta}) \left(\int_{-\infty}^{q_{2|1}(\lambda_2,\lambda_1,\boldsymbol{\eta})} f_{2|1}(\varepsilon_{2t}^*,\varepsilon_{1t}^*;\boldsymbol{\eta}) d\varepsilon_{2t}^*\right) d\varepsilon_{1t}^* \\
&= \int_{-\infty}^{q_1(\lambda_1,\boldsymbol{\eta})} \int_{-\infty}^{q_{2|1}(\lambda_2,\lambda_1,\boldsymbol{\eta})} f_{1,2}(\varepsilon_{1t}^*,\varepsilon_{2t}^*;\boldsymbol{\eta}) d\varepsilon_{2t}^* d\varepsilon_{1t}^* \\
&= \int_{-\infty}^{0} \int_{-\infty}^{0} f_{1,2}(\varepsilon_{1t}^*+q_1(\lambda_1,\boldsymbol{\eta}),\varepsilon_{2t}^*+q_{2|1}\left(\lambda_2,\lambda_1,\boldsymbol{\eta}\right);\boldsymbol{\eta}) d\varepsilon_{2t}^* d\varepsilon_{1t}^*,
\end{aligned}$$

where we have achieved constant limits of integration in the last expression by means of the change of variable

$$u(\varepsilon_{1t}^*,\varepsilon_{2t}^*) = \varepsilon_{1t}^* + q_1(\lambda_1,\boldsymbol{\eta}), \text{ and } v(\varepsilon_{1t}^*,\varepsilon_{2t}^*) = \varepsilon_{2t}^* + q_{2|1}\left(\lambda_2,\lambda_1,\boldsymbol{\eta}\right),$$

whose Jacobian is 1. Differentiating the previous expression with respect to $\boldsymbol{\eta}$ yields

$$\begin{aligned}
0 =& \int_{-\infty}^{0}\int_{-\infty}^{0} \frac{f_{1,2}(\varepsilon_{1t}^*+q_1(\lambda_1,\boldsymbol{\eta}),\varepsilon_{2t}^*+q_{2|1}\left(\lambda_2,\lambda_1,\boldsymbol{\eta}\right);\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} d\varepsilon_{2t}^* d\varepsilon_{1t}^* \\
&+\frac{\partial q_1(\lambda_1,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} \int_{-\infty}^{0}\int_{-\infty}^{0} \frac{\partial f_{1,2}(\varepsilon_{1t}^*+q_1(\lambda_1,\boldsymbol{\eta}),\varepsilon_{2t}^*+q_{2|1}\left(\lambda_2,\lambda_1,\boldsymbol{\eta}\right);\boldsymbol{\eta})}{\partial\varepsilon_{1t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^* \\
&+\frac{\partial q_{2|1}\left(\lambda_2,\lambda_1,\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}} \int_{-\infty}^{0}\int_{-\infty}^{0} \frac{\partial f_{1,2}(\varepsilon_{1t}^*+q_1(\lambda_1,\boldsymbol{\eta}),\varepsilon_{2t}^*+q_{2|1}\left(\lambda_2,\lambda_1,\boldsymbol{\eta}\right);\boldsymbol{\eta})}{\partial\varepsilon_{2t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^*.
\end{aligned}$$

Finally, undoing the change of variable we obtain

$$\begin{aligned}
\frac{\partial q_{2|1}(\lambda_2,\lambda_1,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} =& -\left(\int_{-\infty}^{q_1(\lambda_1,\boldsymbol{\eta})}\int_{-\infty}^{q_{2|1}(\lambda_2,\lambda_1,\boldsymbol{\eta})} \frac{\partial f_{1,2}(\varepsilon_{1t}^*,\varepsilon_{2t}^*;\boldsymbol{\eta})}{\partial\varepsilon_{2t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^*\right)^{-1} \\
&\times\left\{\int_{-\infty}^{q_1(\lambda_1,\boldsymbol{\eta})}\int_{-\infty}^{q_{2|1}(\lambda_2,\lambda_1,\boldsymbol{\eta})} \frac{\partial f_{1,2}(\varepsilon_{1t}^*,\varepsilon_{2t}^*;\boldsymbol{\eta})}{\partial\varepsilon_{1t}^*} d\varepsilon_{2t}^* d\varepsilon_{1t}^*\right. \\
&\times\frac{1}{f_1\left[q_1(\lambda_1,\boldsymbol{\eta});\boldsymbol{\eta}\right]} \int_{-\infty}^{q_1(\lambda_1,\boldsymbol{\eta})} \frac{\partial f_1(\varepsilon_{1t}^*;\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} d\varepsilon_{1t}^* \\
&\left.-\int_{-\infty}^{q_1(\lambda_1,\boldsymbol{\eta})}\int_{-\infty}^{q_{2|1}(\lambda_2,\lambda_1,\boldsymbol{\eta})} \frac{f_{1,2}(\varepsilon_{1t}^*,\varepsilon_{2t}^*;\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} d\varepsilon_{2t}^* d\varepsilon_{1t}^*\right\}.
\end{aligned}$$

E Inference with elliptical innovations

Some useful distribution results

It is easy to combine the representation of elliptical distributions with the higher order moments of a multivariate normal vector in Balestra and Holly (1990) to prove that the third and fourth moments of a spherically symmetric distribution with $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$ are given by

$$E(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ) = \mathbf{0},$$

and

$$E(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime}) = E[vec(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime}) vec'(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^\circ)] = (\kappa_0 + 1)[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec\,(\mathbf{I}_N)\, vec'\,(\mathbf{I}_N)],$$

respectively. In this regard, note that since $E(e_t^4) \geq E^2(e_t^2) = N^2$ by the Cauchy-Schwarz inequality, with equality if and only if $e_t = \sqrt{N}$ so that $\boldsymbol{\varepsilon}_t^\circ$ is proportional to $\mathbf{u}_t$, then $\kappa_0 \geq -2/(N+2)$, the minimum value being achieved in the uniformly distributed case. For example, $\kappa_0 = 2/(\nu_0 - 4)$ in the Student t case with $\nu_0 > 4$, and $\kappa_0 = 0$ under normality.

An alternative characterisation can be based on the higher order moment parameter of spherical random variables introduced by Berkane and Bentler (1986), $\tau_m(\boldsymbol{\eta})$, which Maruyama and Seo (2003) relate to higher order moments as

$$E[\varsigma_t^m|\boldsymbol{\eta}] = [1 + \tau_m(\boldsymbol{\eta})]E[\varsigma_t^m|\mathbf{0}] \text{ where } E[\varsigma_t^m|\mathbf{0}] = 2^m \prod_{j=1}^{m}(N/2 + j - 1). \tag{E10}$$

For the elliptical examples mentioned above, we derive expressions for $\tau_m(\boldsymbol{\eta})$ in Appendix D.2. A noteworthy property of these examples is that their moments are always bounded, with the exception of the Student t. Appendix D.3 contains the moment generating functions for the Kotz, the DSMN and the 3^{rd}-order PE.

E.1 The log-likelihood function, its score and information matrix

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta})'$ denote the $p + q$ parameters of interest, which we assume variation free. Ignoring initial conditions, the log-likelihood function of a sample of size T for those values of $\boldsymbol{\theta}$ for which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ has full rank will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\,[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$, where $d_t(\boldsymbol{\theta}) = -1/2 \ln |\boldsymbol{\Sigma}_t(\boldsymbol{\theta})|$ corresponds to the Jacobian, $c(\boldsymbol{\eta})$ to the constant of integration of the assumed density, and $g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]$ to its kernel, where $\varsigma_t(\boldsymbol{\theta})=\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})=\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta})$.

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $s_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. Then, it is straightforward to show that if $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, $c(\boldsymbol{\eta})$ and $g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]$ are differentiable

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}+\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varsigma}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\left[\mathbf{Z}_{lt}(\boldsymbol{\theta}),\mathbf{Z}_{st}(\boldsymbol{\theta})\right]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}),\\
\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= \partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta}+\partial g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}=\mathbf{e}_{rt}(\boldsymbol{\phi}),
\end{aligned}\qquad\text{(E11)}$$

where

$$\begin{aligned}
\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N)\\
\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\},\\
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial\boldsymbol{\mu}_t^{\prime}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}),\\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial vec^{\prime}\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})],\\
\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), &\text{(E12)}\\
\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right\}, &\text{(E13)}
\end{aligned}$$

$\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^{\prime}$ and $\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}^{\prime}$ depend on the particular specification adopted, and

$$\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]=-2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma \qquad\text{(E14)}$$

can be understood as a damping factor that reflects the kurtosis of the specific distribution assumed for estimation purposes (see Appendix E.3.1 for further details). But since $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ is equal to 1 under Gaussianity, it is straightforward to check that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992), in which case:

$$\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})=\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right]=\left\{\begin{array}{c}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right]\end{array}\right\}.$$

Given correct specification, the results in Crowder (1976) imply that $\mathbf{e}_t(\boldsymbol{\phi})=[\mathbf{e}_{dt}^{\prime}(\boldsymbol{\phi}),\mathbf{e}_{rt}(\boldsymbol{\phi})]^{\prime}$ evaluated at $\boldsymbol{\phi}_0$ follows a vector martingale difference, and therefore, the same is true of the score vector $\mathbf{s}_t(\boldsymbol{\phi})$. His results also imply that, under suitable regularity conditions, the asymptotic distribution of the feasible, joint ML estimator will be $\sqrt{T}(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}_0)\rightarrow N\left[\mathbf{0},\mathcal{I}^{-1}(\boldsymbol{\phi}_0)\right]$, where $\mathcal{I}(\boldsymbol{\phi}_0)=E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$,

$$\begin{aligned}
\mathcal{I}_t(\boldsymbol{\phi}) &= V\left[\mathbf{s}_t(\boldsymbol{\phi})|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right]=\mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\phi})\mathbf{Z}_t^{\prime}(\boldsymbol{\theta})=-E\left[\mathbf{h}_t(\boldsymbol{\phi})|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}\right],\\
\mathbf{Z}_t(\boldsymbol{\theta}) &= \left(\begin{array}{cc}\mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{array}\right)=\left(\begin{array}{ccc}\mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{array}\right),
\end{aligned}$$

$\mathbf{h}_t(\boldsymbol{\phi})$ denotes the Hessian function $\partial \mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$ and $\mathcal{M}(\boldsymbol{\phi}) = V\left[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right]$.

The following result, which reproduces Proposition 2 in Fiorentini and Sentana (2010), contains the required expressions to compute the information matrix of the ML estimators:

Proposition E1 *If* $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}$ *is i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ *with density* $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$, *then*

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathbf{0} & \mathcal{M}_{sr}'(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix},$$
$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = V[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N,$$
$$\mathcal{M}_{ss}(\boldsymbol{\eta}) = V[\mathbf{e}_{st}(\boldsymbol{\phi})|\,\boldsymbol{\phi}] = \mathrm{M}_{ss}(\boldsymbol{\eta})\,(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + [\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N),$$
$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\,\boldsymbol{\phi}] = -E\left\{\partial\mathbf{e}_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\eta}'|\,\boldsymbol{\phi}\right\} = vec(\mathbf{I}_N)\mathrm{M}_{sr}(\boldsymbol{\eta}),$$
$$\mathcal{M}_{rr}(\boldsymbol{\eta}) = V[\,\mathbf{e}_{rt}(\boldsymbol{\phi})|\,\boldsymbol{\phi}] = -E[\partial\mathbf{e}_{rt}(\boldsymbol{\phi})/\partial\boldsymbol{\eta}'|\,\boldsymbol{\phi}],$$
$$\mathrm{M}_{ll}(\boldsymbol{\eta}) = E\left\{\delta^2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\theta})}{N}\middle|\,\boldsymbol{\phi}\right\} = E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t(\boldsymbol{\theta})}{N} + \delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\middle|\,\boldsymbol{\phi}\right\},$$
$$\mathrm{M}_{ss}(\boldsymbol{\eta}) = \frac{N}{N+2}\left[1 + V\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t}{N}\middle|\,\boldsymbol{\phi}\right\}\right] = E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)}\middle|\,\boldsymbol{\phi}\right\} + 1,$$
$$\mathrm{M}_{sr}(\boldsymbol{\eta}) = E\left[\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\theta})}{N} - 1\right\}\mathbf{e}_{rt}'(\boldsymbol{\phi})\middle|\,\boldsymbol{\phi}\right] = -E\left\{\frac{\varsigma_t(\boldsymbol{\theta})}{N}\frac{\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\boldsymbol{\eta}'}\middle|\,\boldsymbol{\phi}\right\}.$$

Fiorentini, Sentana and Calzolari (2003) provide the relevant expressions for the multivariate standardised Student t, while the expressions for the Kotz distribution and the DSMN are given in Amengual and Sentana (2010).[16]

E.2 Gaussian pseudo maximum likelihood estimators of $\boldsymbol{\theta}$

If the interest of the researcher lied exclusively in $\boldsymbol{\theta}$, which are the parameters characterising the conditional mean and variance functions, then one attractive possibility would be to estimate a restricted version of the model in which $\boldsymbol{\eta}$ is set to zero. Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \mathbf{0})$ denote such a PML estimator of $\boldsymbol{\theta}$. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the conditional distribution of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is not Gaussian, provided that it has bounded fourth moments. The proof is based on the fact that in those circumstances, the pseudo log-likelihood score, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})$, is a vector martingale difference sequence when evaluated at $\boldsymbol{\theta}_0$, a property that inherits from $\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})$. The asymptotic distribution of the PML estimator of $\boldsymbol{\theta}$ is stated in the following result, which reproduces Proposition 3.2 in Fiorentini and Sentana (2010):

[16] The expression for $\mathrm{M}_{ss}(\kappa)$ for the Kotz distribution in Amengual and Sentana (2010) contains a typo. The correct value is $(N\kappa + 2)/[(N+2)\kappa + 2]$.

Proposition E2 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ *is i.i.d.* $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ *with* $\kappa_0<\infty$, *and the regularity conditions in Bollerslev and Wooldridge (1992) are satisfied, then* $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\to N[\mathbf{0},\mathcal{C}(\boldsymbol{\phi}_0)]$, *where*

$$\begin{gathered}\mathcal{C}(\boldsymbol{\phi})=\mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi}),\\ \mathcal{A}(\boldsymbol{\phi})=-E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]=E\left[\mathcal{A}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],\\ \mathcal{A}_t(\boldsymbol{\phi})=-E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta};\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}]=\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(0)\mathbf{Z}'_{dt}(\boldsymbol{\theta}),\\ \mathcal{B}(\boldsymbol{\phi})=V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]=E\left[\mathcal{B}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],\\ \mathcal{B}_t(\boldsymbol{\phi})=V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta};\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}]=\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\kappa)\mathbf{Z}'_{dt}(\boldsymbol{\theta}),\\ \textit{and }\mathcal{K}(\kappa)=V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}]=\begin{bmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & (\kappa+1)\left(\mathbf{I}_{N^2}+\mathbf{K}_{NN}\right)+\kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\end{bmatrix},\end{gathered}$$

which only depends on $\boldsymbol{\eta}$ *through the population coefficient of multivariate excess kurtosis.*

But if κ_0 is infinite then $\mathcal{B}(\boldsymbol{\phi}_0)$ will be unbounded, and the asymptotic distribution of some or all the elements of $\tilde{\boldsymbol{\theta}}_T$ will be non-standard, unlike that of $\hat{\boldsymbol{\theta}}_T$ (see Hall and Yao (2003)).

E.3 Computational details

E.3.1 Scores and first order conditions

The damping factor (E14) reduces to

$$\delta^t[\varsigma_t(\boldsymbol{\theta}),\eta]=(N\eta+1)/[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})]$$

for the Student t,

$$\delta^{ds}[\varsigma_t(\boldsymbol{\theta}),\alpha,\varkappa]=[\alpha+(1-\alpha)\varkappa]\cdot\frac{\alpha+(1-\alpha)\varkappa^{-(N/2+1)}\exp\left[-\frac{[\alpha+(1-\alpha)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]}{\alpha+(1-\alpha)\varkappa^{-N/2}\exp\left[-\frac{[\alpha+(1-\alpha)\varkappa](1-\varkappa)}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right]}$$

for the DSMN, and

$$\delta^{pe}[\varsigma_t(\boldsymbol{\theta}),c_2,c_3]=1-\frac{\sum_{j=1}^J c_j p_{N/2,j}[\varsigma_t(\boldsymbol{\theta})]}{1+\sum_{j=1}^J c_j p_{N/2-1,j}[\varsigma_t(\boldsymbol{\theta})]}$$

for the PE.

As for $\mathbf{e}_{rt}(\boldsymbol{\theta},\boldsymbol{\eta})$, Fiorentini Sentana and Calzolari (2003) show that in the multivariate Student t case it becomes

$$\begin{aligned}s^t_{\eta t}(\boldsymbol{\theta},\eta)\;&=\;\frac{N}{2\eta(1-2\eta)}-\frac{1}{2\eta^2}\left[\psi\left(\frac{N\eta+1}{2\eta}\right)-\psi\left(\frac{1}{2\eta}\right)\right]\\ &-\frac{N\eta+1}{2\eta(1-2\eta)}\frac{\varsigma_t(\boldsymbol{\theta})}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})}+\frac{1}{2\eta^2}\ln\left(1+\frac{\eta}{1-2\eta}\varsigma_t(\boldsymbol{\theta})\right).\end{aligned}$$

For the multivariate discrete scale mixture of normals, we can use (E11) to write the score with respect to the mixing parameter α as

$$s^{ds}_{\alpha}(\boldsymbol{\theta},\alpha,\varkappa)=\frac{N}{2}\frac{1-\varkappa}{[\alpha+\varkappa(1-\alpha)]}+\frac{\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\alpha},$$

where

$$\begin{aligned}\frac{\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial \alpha} &= \frac{1}{\exp\left(g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\right)}\left\{\exp\left(-\frac{1}{2\varpi}\varsigma_t(\boldsymbol{\theta})\right)-\varkappa^{-N/2}\exp\left(-\frac{1}{2\varpi\varkappa}\varsigma_t(\boldsymbol{\theta})\right)\right\}\\ &\quad -\frac{1}{\exp\left(g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\right)}\frac{1-\varkappa}{[\alpha+\varkappa(1-\alpha)]^2}\frac{1}{2\varpi^2}\varsigma_t(\boldsymbol{\theta})\\ &\quad \times\alpha\exp\left(-\frac{1}{2\varpi}\varsigma_t(\boldsymbol{\theta})\right)+(1-\alpha)\varkappa^{-N/2-1}\exp\left(-\frac{1}{2\varpi\varkappa}\varsigma_t(\boldsymbol{\theta})\right),\end{aligned}$$

and the score with respect to the relative scale parameter $\varkappa$ as

$$s^{ds}_{\varkappa}(\boldsymbol{\theta},\alpha,\varkappa)=\frac{N}{2}\frac{1-\alpha}{[\alpha+\varkappa(1-\alpha)]}+\frac{\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial \varkappa},$$

where

$$\begin{aligned}\frac{\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial \varkappa} &= \frac{1}{\exp\left(g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\right)}\left\{-\frac{N}{2}(1-\alpha)\varkappa^{-N/2-1}\exp\left(-\frac{1}{2\varpi\varkappa}\varsigma_t(\boldsymbol{\theta})\right)\right.\\ &\quad -\varsigma_t(\boldsymbol{\theta})\frac{1-\alpha}{2}\alpha\exp\left(-\frac{1}{2\varpi}\varsigma_t(\boldsymbol{\theta})\right)\\ &\quad \left.+(1-\alpha)\alpha\frac{\varsigma_t(\boldsymbol{\theta})}{2\varkappa^2}\varkappa^{-N/2}\exp\left(-\frac{1}{2\varpi\varkappa}\varsigma_t(\boldsymbol{\theta})\right)\right\}.\end{aligned}$$

Finally, the scores of the 3^{rd} order PE distribution with respect to c_2 and c_3 will be

$$s^{pe}_{c_2t}(\boldsymbol{\theta},c_2,c_3)=\frac{p_{N/2,2}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}{1+\sum_{j=1}^J c_j p_{N/2-1,j}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}$$

and

$$s^{pe}_{c_3t}(\tilde{\boldsymbol{\theta}}_T,c_2,c_3)=\frac{p_{N/2,3}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}{1+\sum_{j=1}^J c_j p_{N/2-1,j}[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)]}.$$

We can then use $\tilde{\boldsymbol{\theta}}_T$ to obtain a sequential ML estimator of $\boldsymbol{\eta}$ as $\tilde{\boldsymbol{\eta}}_T=\arg\max_\eta L_T(\tilde{\boldsymbol{\theta}}_T,\boldsymbol{\eta})$, possibly subject to some inequality constraints on $\boldsymbol{\eta}$. For example, in the Student t case $\tilde{\eta}_T$ will be characterised by the first-order Kuhn-Tucker (KT) conditions

$$\bar{s}_{\eta T}(\tilde{\boldsymbol{\theta}}_T,\tilde{\eta}_T)+\tilde{\upsilon}_T=0;\quad \tilde{\eta}_T\geq 0;\quad \tilde{\upsilon}_T\geq 0;\quad \tilde{\upsilon}_T\cdot\tilde{\eta}_T=0,$$

where $\bar{s}_{\eta T}(\boldsymbol{\theta},\eta)$ is the sample mean of $s_{\eta t}(\boldsymbol{\theta},\eta)$, and $\tilde{\upsilon}_T$ the KT multiplier associated to the constraint $\eta\geq 0$.

Fiorentini, Sentana and Calzolari (2003) show that in the multivariate Student t case $s_{\eta t}(\boldsymbol{\theta},0)$ is proportional to the second generalised Laguerre polynomial. Similarly, Amengual and Sentana (2011) show that this is also the case for the score of the scale parameter of a DSMN. Therefore,

$$s^t_{\eta t}(\boldsymbol{\theta},0)=s^{ds}_{\varkappa t}(\boldsymbol{\theta},0)=p^g_{N/2-1,2}[\varsigma_t(\boldsymbol{\theta})].$$

Amengual and Sentana (2011) also provide the corresponding expressions for the α-component of $e_{rt}^{DSMN}(\boldsymbol{\theta},0)$ in the case of "outliers", which is given by

$$\lim_{\alpha\to 0^+} s_{\alpha t}^{ds}(\boldsymbol{\theta},\alpha,\varkappa) = \varkappa^{N/2}\exp\left(\frac{1-\varkappa}{2}\varsigma_t(\boldsymbol{\theta})\right) - 1 - \frac{1-\varkappa}{2\varkappa}(\varsigma_t(\boldsymbol{\theta})-N).$$

In contrast, in the case of "inliers" it will be given by

$$\lim_{\alpha\to 1^-} s_{\alpha t}^{ds}(\boldsymbol{\theta},\alpha,\varkappa) = 1 - \varkappa^{-N/2}\exp\left(\frac{\varkappa-1}{2\varkappa}\varsigma_t(\boldsymbol{\theta})\right) - \frac{1-\varkappa}{2}(\varsigma_t(\boldsymbol{\theta})-N).$$

As for the polynomial expansion, we saw in Appendix D.1 that the shape parameters are also inequality constrained. Not surprisingly, Amengual and Sentana (2011) also show that $\mathbf{e}_{rt}^{pe}(\boldsymbol{\theta},\mathbf{0}) = \{p_{N/2-1,2}^{g}[\varsigma_t(\boldsymbol{\theta})], p_{N/2-1,3}^{g}[\varsigma_t(\boldsymbol{\theta})]\}'$.

E.3.2 Numerical issues

Random number generation We sample Student t and DSMN exploiting the decomposition presented in section 2.1. Specifically, we simulate standardised versions of all these distributions by appropriately mixing a N-dimensional spherical normal vector with a univariate gamma random variable, and, in the case of DSMN, a draw from a scalar uniform, which we obtain from the NAG Fortran 77 Mark 19 library routines G05DDF, G05FFF and G05CAF, respectively (see Numerical Algorithm Group (2001) for details). To draw innovations from a PE, we use a modification of the inversion method. Specifically, we first compute the square Euclidean norm of the N-dimensional spherical normal vector, ζ say, which is distributed as a χ^2 with N degrees of freedom. We then use the G05NCF routine to find the solution to the equation $F(\varsigma, c_2, c_3, N) = F_{\chi^2_N}(\zeta)$, where

$$\begin{aligned} F(\varsigma, c_2, c_3, N) &= 1 - \frac{\Gamma(N/2,\varsigma/2)}{\Gamma(N/2)} - c_2 \times \frac{\varsigma^{N/2}e^{-\varsigma/2}}{2^{N/2+2}\Gamma(N/2+2)}(\varsigma - 2 - N) \\ &\quad + c_3 \times \frac{\varsigma^{N/2}e^{-\varsigma/2}}{2^{N/2+3}\Gamma(N/2+3)}\left[\varsigma^2 - 2d(N+4) + (N+2)(N+4)\right], \end{aligned}$$

with $\varsigma = \zeta$ as starting value. In this way, we make sure that the three distributions that we simulate share the random draws from the underlying $N \times 1$ uniform vector, which minimises Monte Carlo variability.

Estimation strategy Our estimation procedure employs the following numerical strategy. First, we estimate the conditional mean and variance parameters $\boldsymbol{\theta}$ under normality with a scoring algorithm that combines the E04LBF routine with the analytical expressions for the score and the $\mathcal{A}(\boldsymbol{\phi}_0)$ matrix in Proposition E2. Then, we compute consistent estimators of $\boldsymbol{\eta}$ using the expressions in the next subsection, which we use as initial values for the optimisation procedure that obtains the sequential ML estimator $\tilde{\boldsymbol{\eta}}_T$ with the E04JYF routine. This estimator is then used as initial value for the efficient sequential MM estimator, which is obtained with the C05NCF routine. Since our model admits reparametrisation (1), we use expression (9) with M_{ss} and M_{sr} computed either analytically, or by Monte Carlo integration or quadrature. We use again $\tilde{\boldsymbol{\eta}}_T$ as initial value for the sequential GMM estimators based on orthogonal polynomials using the E04JYF routine. As for the joint ML estimators, we employ the E04JBF routine with numerical derivatives starting from the PML estimators of $\boldsymbol{\theta}$ and the sequential ML estimators of $\boldsymbol{\eta}$.

Initial consistent estimators of shape parameters

Student *t* The initial value of η is the moment estimator proposed by Fiorentini, Sentana and Calzolari (2003):

$$\eta_{init} = \max\left[\frac{\hat{\tau}_2^t}{(4\hat{\tau}_2^t+2)}, 0\right] \quad \text{where} \quad \hat{\tau}_2^t = \frac{\frac{1}{T}\sum_{t=1}^{T}\varsigma_t^2(\tilde{\boldsymbol{\theta}}_T)}{(N+2)N} - 1.$$

Discrete scale mixture of normals The initial values for α and $\varkappa$ are obtained by running a standard EM algorithm that does not impose $E[\varsigma_t(\tilde{\boldsymbol{\theta}}_T)] = N$.

3rd-order polynomial expansion The initial values for c_2 and c_3 are moment estimators obtained as

$$c_{2:init} = \frac{\sum_{t=1}^{T}\varsigma_t^2(\tilde{\boldsymbol{\theta}}_T)}{4T} - \frac{N(N+2)}{4}$$

and

$$c_{3:init} = \frac{N+4}{2}c_{2:init} + \frac{N(N+2)(N+4)}{24} - \frac{\sum_{t=1}^{T}\varsigma_t^3(\tilde{\boldsymbol{\theta}}_T)}{24T}.$$

F Supplemental tables and figures

Table F1: Maximum likelihood estimates of mean and variance parameters

Asset	Parameter	Gaussian	Student t	DSMN	PE
EMU commercial bank index					
	μ_M	0.136	0.202	0.187	0.168
	σ_M^2	11.290	9.089	8.595	10.848
	γ_M	0.152	0.106	0.109	0.132
	β_M	0.834	0.882	0.880	0.854
	ψ_M	−0.040	−0.026	−0.028	−0.038
Deutsche Bank					
	a_{DEU}	−0.070	−0.037	−0.0375	−0.0520
	b_{DEU}	1.197	1.147	1.1529	1.1835
	ω_{DEU}	7.559	6.551	6.800	8.910
	γ_{DEU}	0.089	0.061	0.063	0.072
	β_{DEU}	0.894	0.930	0.928	0.917
	ψ_{DEU}	0.000	−0.022	−0.026	−0.010
BNP Paribas					
	a_{BNP}	0.061	−0.004	0.016	0.025
	b_{BNP}	1.193	1.194	1.185	1.186
	ω_{BNP}	10.150	13.400	12.814	14.649
	γ_{BNP}	0.113	0.096	0.098	0.104
	β_{BNP}	0.870	0.899	0.898	0.889
	ψ_{BNP}	0.039	−0.028	−0.021	0.024
Banco Santander					
	a_{SAN}	0.066	0.092	0.074	0.072
	b_{SAN}	1.049	1.033	1.031	1.047
	ω_{SAN}	24.367	27.659	29.889	38.566
	γ_{SAN}	0.153	0.077	0.077	0.115
	β_{SAN}	0.839	0.921	0.921	0.881
	ψ_{SAN}	0.074	0.007	0.020	0.033
Unicredit Group					
	a_{UNI}	−0.109	−0.152	−0.134	−0.124
	b_{UNI}	1.036	1.049	1.054	1.045
	ω_{UNI}	51.005	42.404	40.377	53.926
	γ_{UNI}	0.092	0.071	0.073	0.081
	β_{UNI}	0.906	0.927	0.926	0.917
	ψ_{UNI}	−0.190	−0.148	−0.138	−0.169

Notes: The balanced panel includes 984 weekly observations from mid October 1993 to the end of August 2012. Excess returns are computed by subtracting the continuously compounded rate of return on the one-week Eurocurrency rate in DM/Euros applicable over the relevant week. ML and SML denote joint and sequential maximum likelihood estimator, respectively. We consider a generalised version of (10) in which we allow both systematic and idiosyncratic variances to evolve over time as GQARCH(1,1) processes i.e. $\sigma^2_{Mt} = \sigma^2_M + \gamma_M(\varepsilon^2_{Mt-1} - \sigma^2_M) + \psi_M\varepsilon_{Mt-1} + \beta_M(\sigma^2_{Mt-1} - \sigma^2_M)$ for the variance of the bank index and $\omega_{it} = \omega_i + \gamma_i(\varepsilon^2_{it-1} - \omega_i) + \psi_i\varepsilon_{it-1} + \beta_i(\omega_{it-1} - \omega_i)$ for the idiosyncratic variance of bank i.

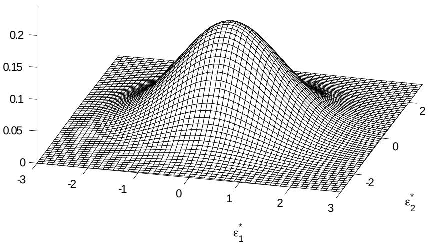
Table F2**:** Maximum likelihood estimates of shape parameters

	Student		DSMN		PE	
	ML	SML	ML	SML	ML	SML
Student t						
η	0.154	0.148				
DSMN						
α			0.159	0.173		
$\varkappa$			0.260	0.275		
PE						
c_2					2.412	2.262
c_3					−0.708	−0.619
VaR and CoVaR quantities						
VaR (1%)	2.549	2.538	2.563	2.556	2.538	2.524
CoVaR (5%)	2.121	2.094	2.171	2.141	2.031	2.006

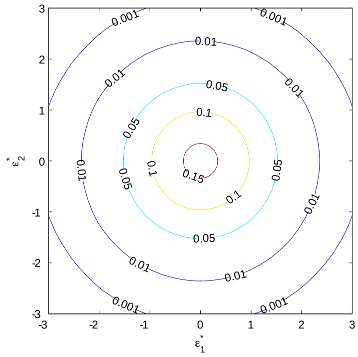
Notes: The balanced panel includes 984 weekly observations from mid October 1993 to the end of August 2012. For model specification see Section 6. Excess returns are computed by subtracting the continuously compounded rate of return on the one-week Eurocurrency rate in DM/Euros applicable over the relevant week. ML and SML denote joint and sequential maximum likelihood estimator, respectively. For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. For DSMN innovations, αdenotes the mixing probability and $\varkappa$ is the variance ratio of the two components. In turn, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2-1$ in the case of PE innovations.

Figure F1: Densities and contours of bivariate elliptical distributions

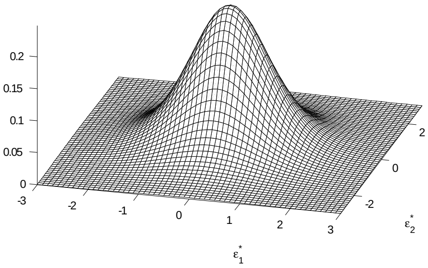
(a) Standardised bivariate normal density



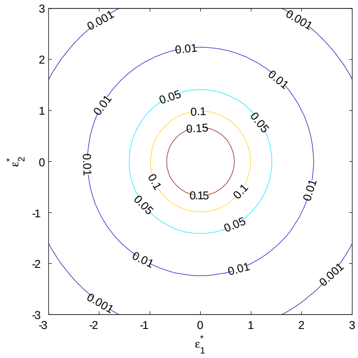
(b) Contours of a standardised bivariate normal density



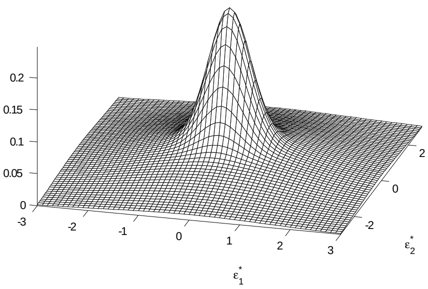
(c) Standardised bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)



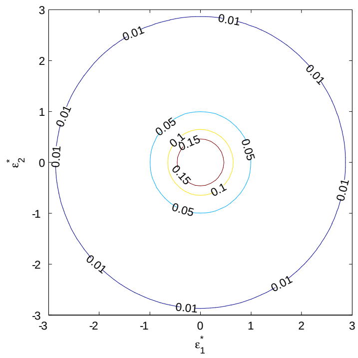
(d) Contours of a standardised bivariate Student t density with 8 degrees of freedom ($\eta = 0.125$)



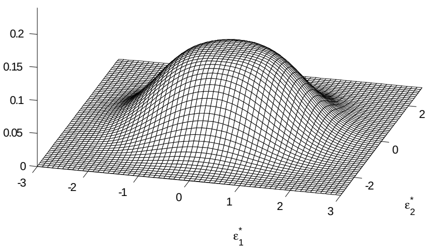
(e) Standardised bivariate DSMN density with multivariate excess kurtosis $\kappa = 0.125$ ($\alpha = 0.5$)



(f) Contours of a standardised bivariate DSMN density with multivariate excess kurtosis $\kappa = 0.125$ ($\alpha = 0.5$)



(g) Standardised bivariate 3^{rd}-order PE with parameters $c_2 = 0$ and $c_3 = -0.2$



(h) Contours of a standardised 3^{rd}-order PE with parameters $c_2 = 0$ and $c_3 = -0.2$

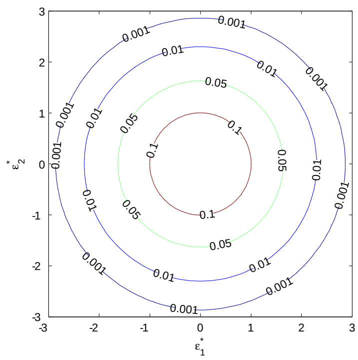
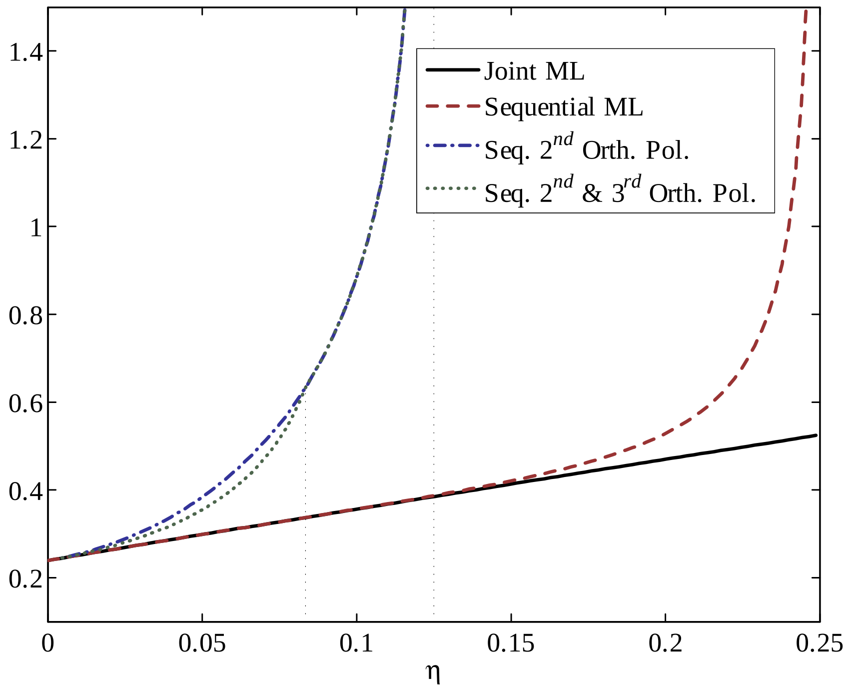
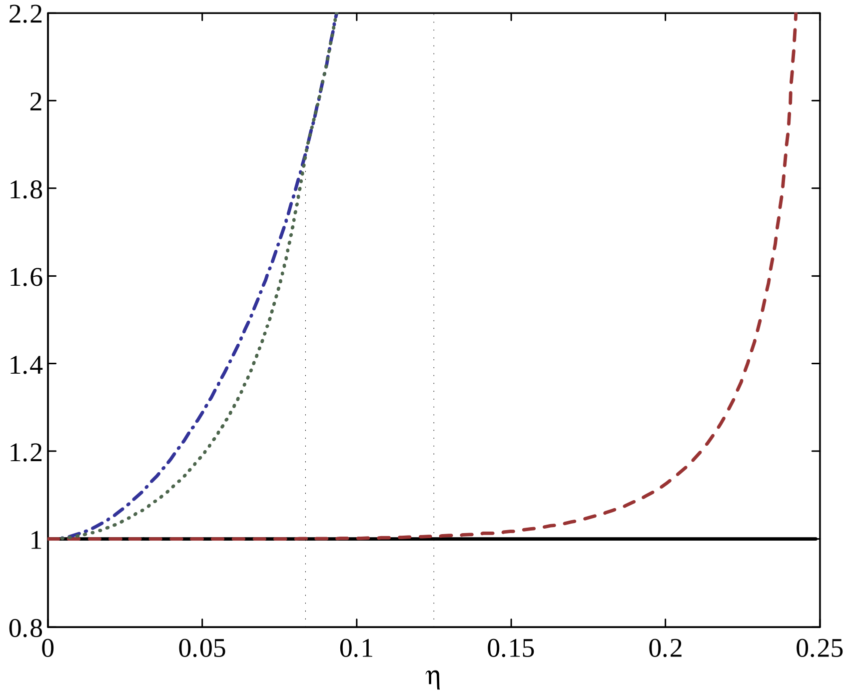


Figure F2: Asymptotic efficiency of Student t estimators

Asymptotic standard errors of η estimators



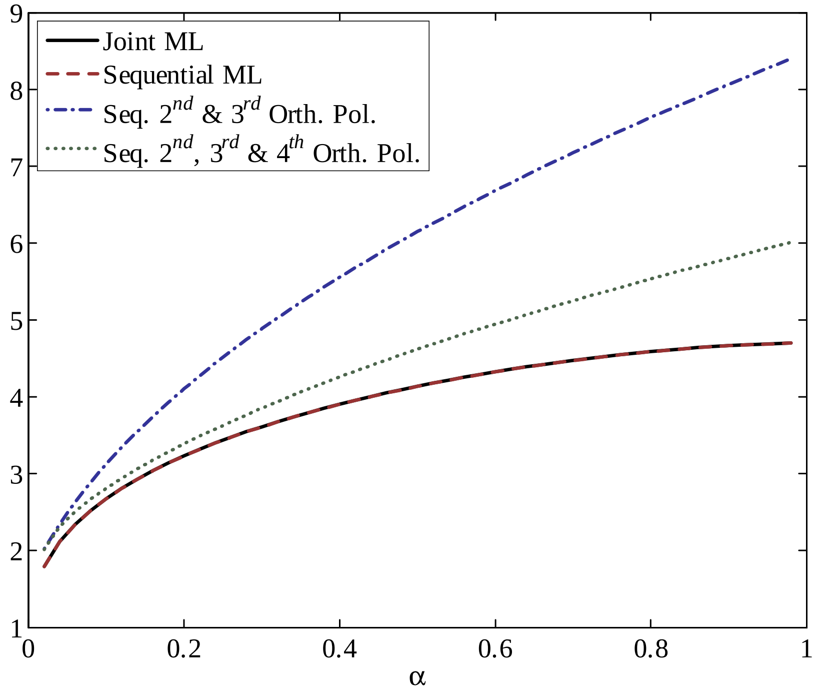
Relative efficiency of η estimators (with respect to Joint ML)



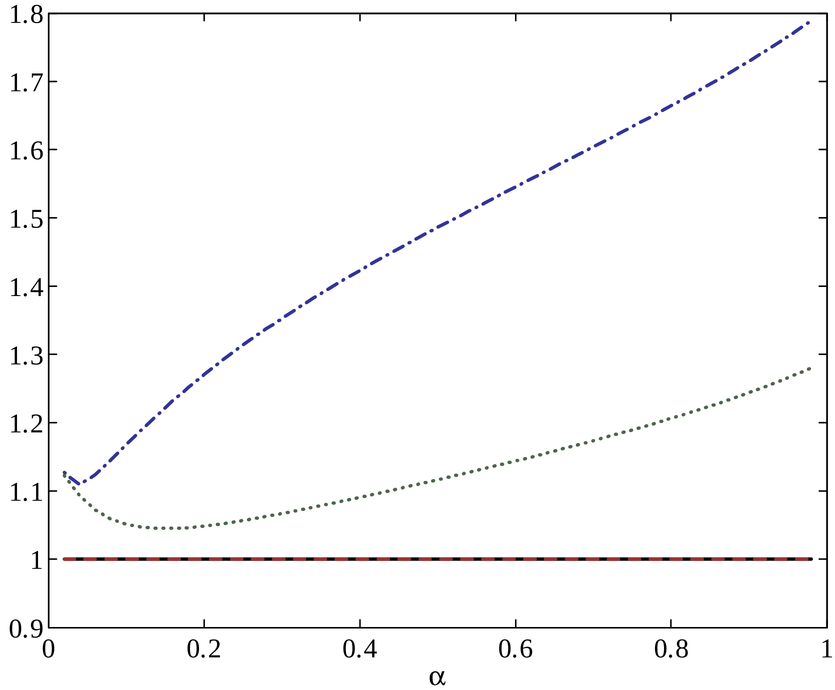
Notes: $N = 5$. For Student t innovations with ν degrees of freedom, $\eta = 1/\nu$. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F3: (a) Asymptotic efficiency of DSMN estimators ($\varkappa = 0.5$)

Asymptotic standard errors of α estimators



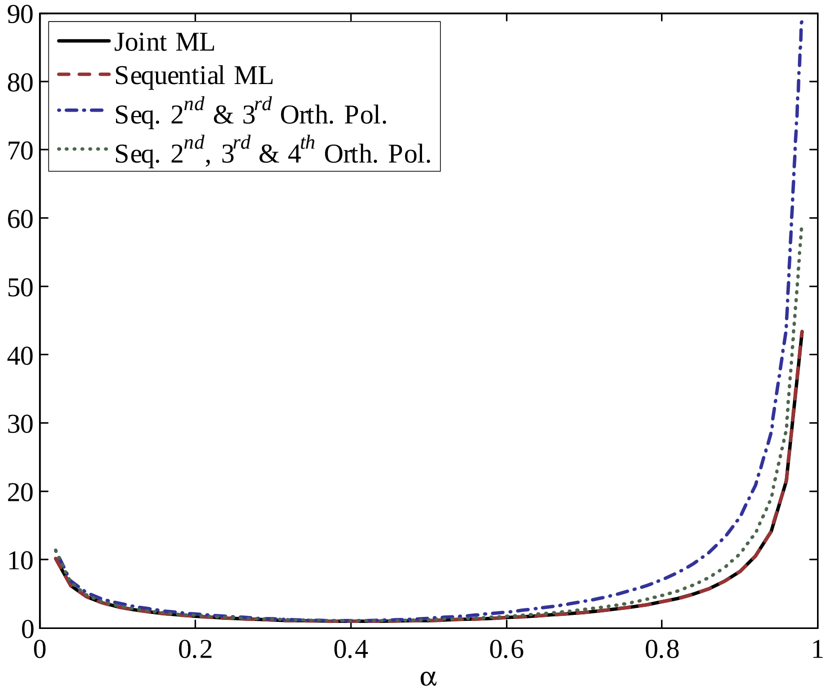
Relative efficiency of α estimators (with respect to Joint ML)



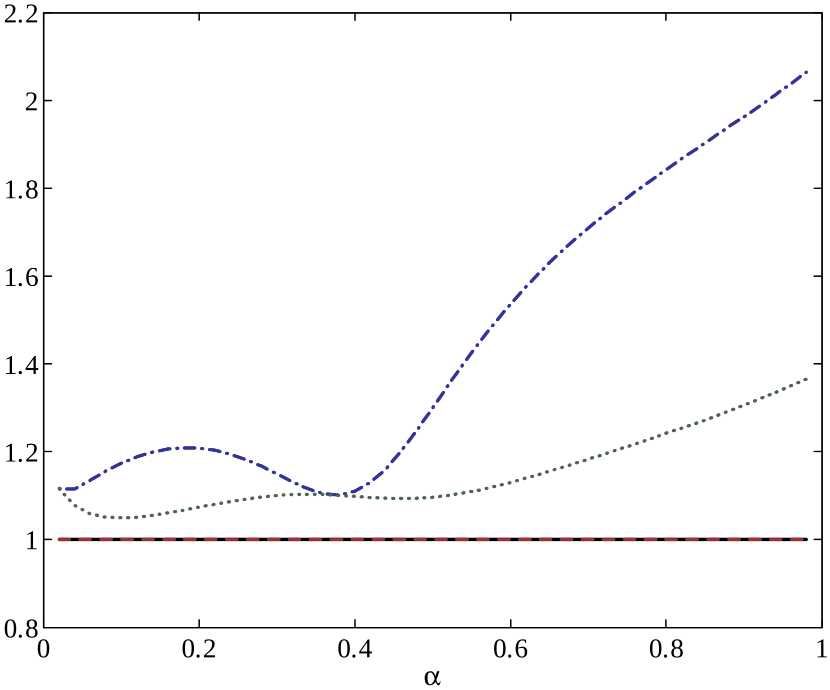
Notes: $N = 5$ and $\varkappa = 0.5$. For DSMN innovations, α denotes the mixing probability and $\varkappa$ is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F3: (b) Asymptotic efficiency of DSMN estimators ($\varkappa = 0.5$)

Asymptotic standard errors of $\varkappa$ estimators



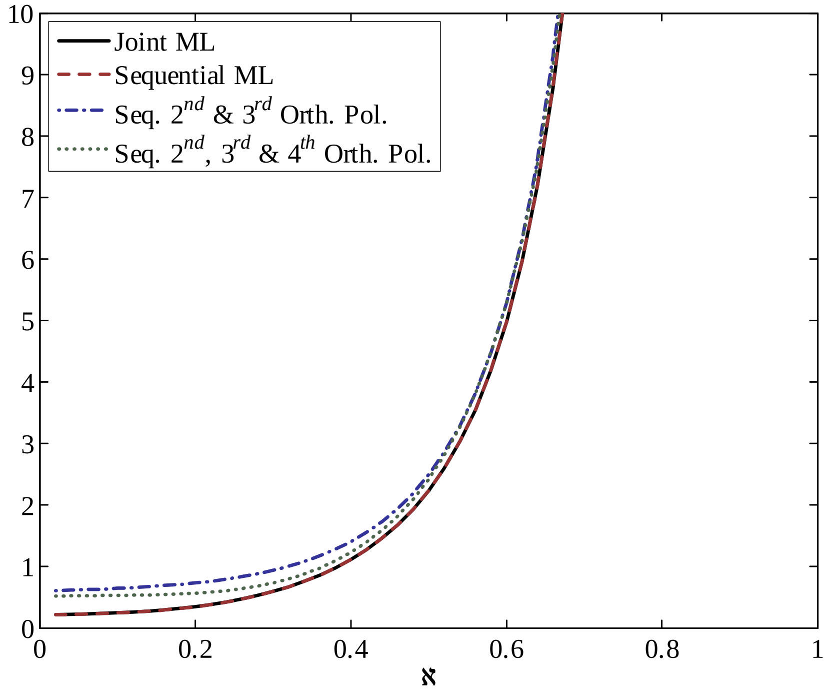
Relative efficiency of $\varkappa$ estimators (with respect to Joint ML)



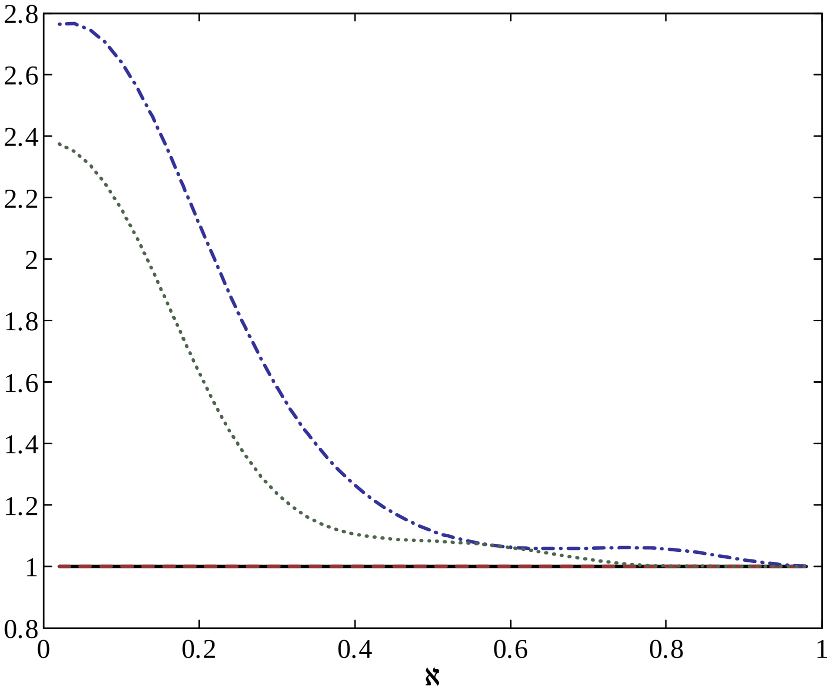
Notes: $N = 5$ and $\varkappa = 0.5$. For DSMN innovations, α denotes the mixing probability and $\varkappa$ is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F3: (c) Asymptotic efficiency of DSMN estimators ($\alpha = 0.05$)

Asymptotic standard errors of α estimators



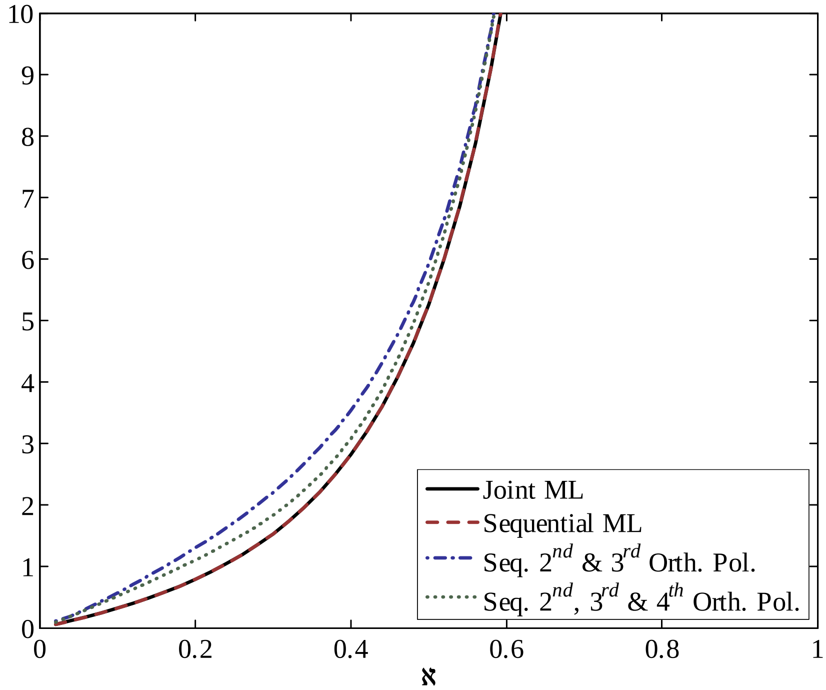
Relative efficiency of α estimators (with respect to Joint ML)



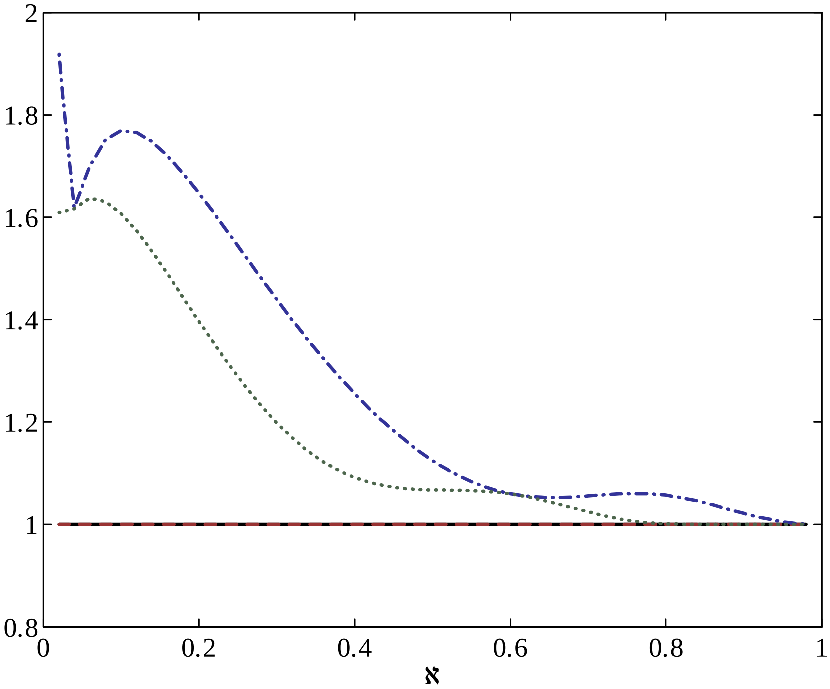
Notes: $N = 5$ and $\alpha = 0.05$. For DSMN innovations, α denotes the mixing probability and $\varkappa$ is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F3: (d) Asymptotic efficiency of DSMN estimators ($\alpha = 0.05$)

Asymptotic standard errors of $\varkappa$ estimators



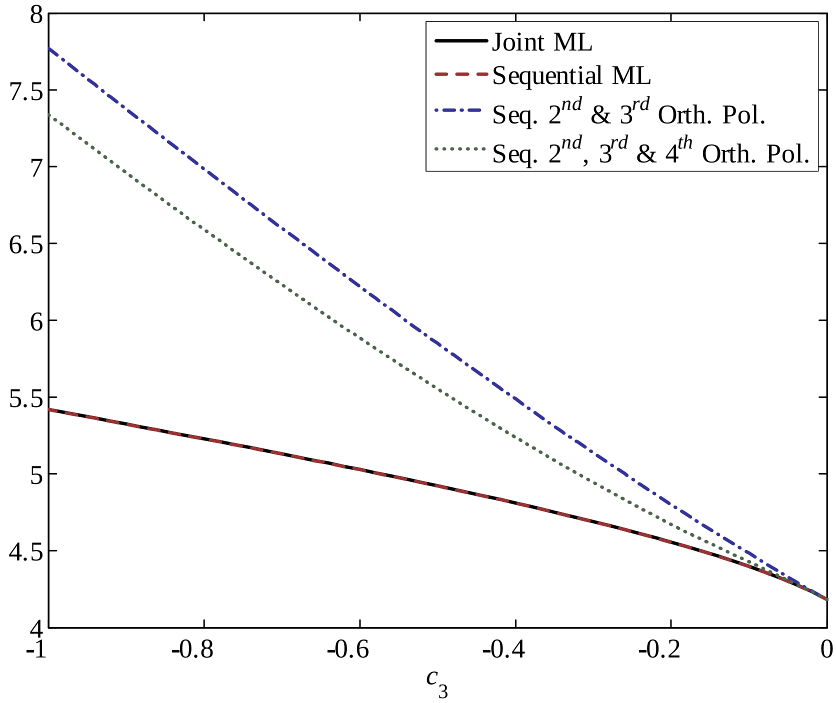
Relative efficiency of $\varkappa$ estimators (with respect to Joint ML)



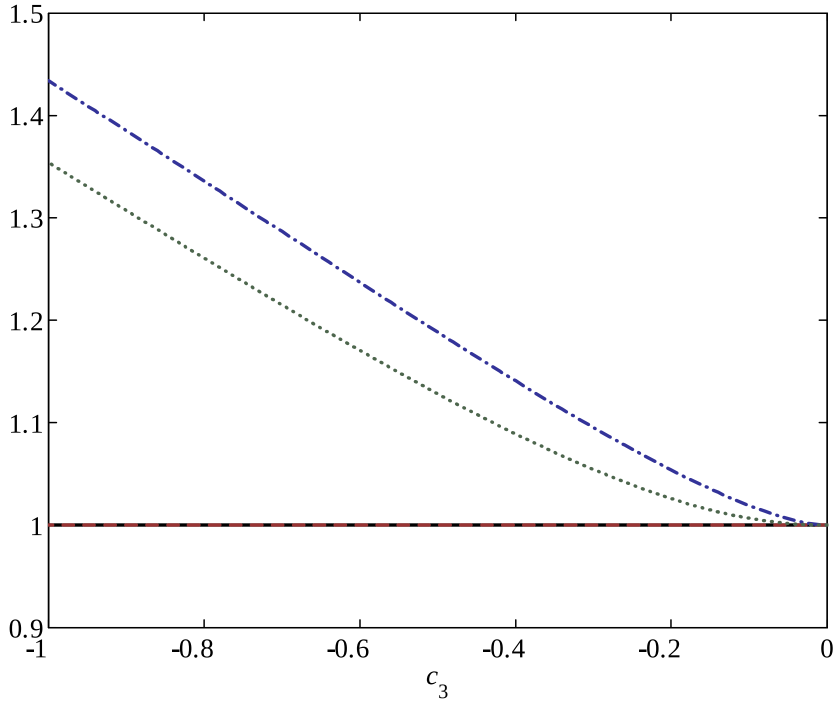
Notes: $N = 5$ and $\alpha = 0.05$. For DSMN innovations, α denotes the mixing probability and $\varkappa$ is the variance ratio of the two components. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F4: (a) Asymptotic efficiency of PE estimators ($c_2 = 0$)

Asymptotic standard errors of c_2 estimators



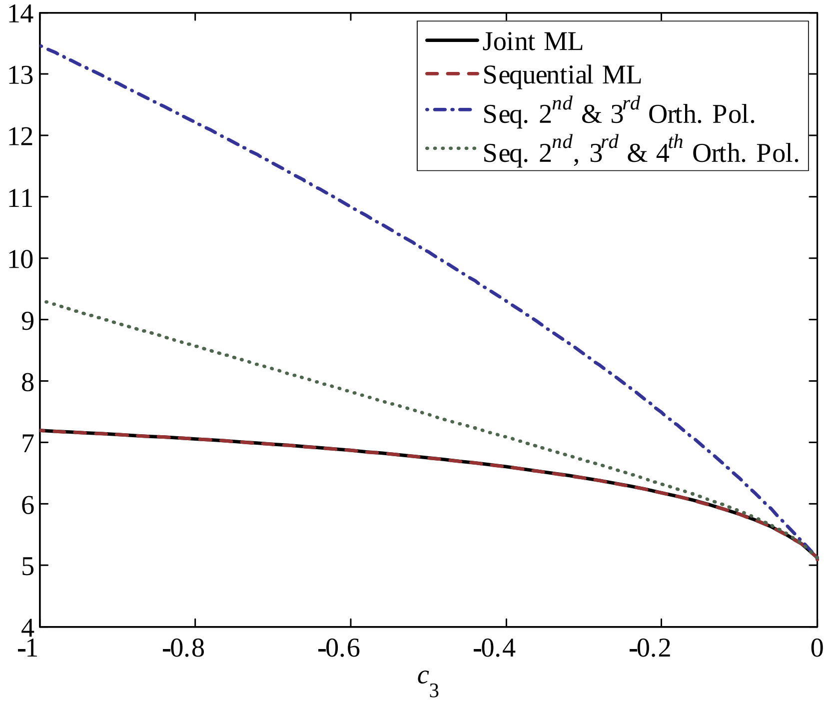
Relative efficiency of c_2 estimators (with respect to Joint ML)



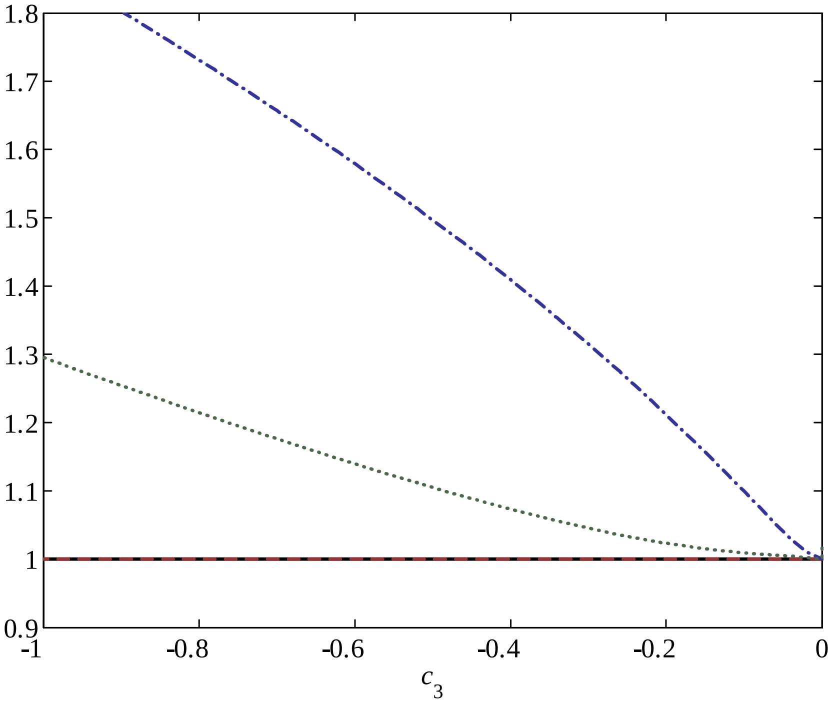
Notes: $N = 5$ and $c_2 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2-1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F4: (b) Asymptotic efficiency of PE estimators ($c_2 = 0$)

Asymptotic standard errors of c_3 estimators



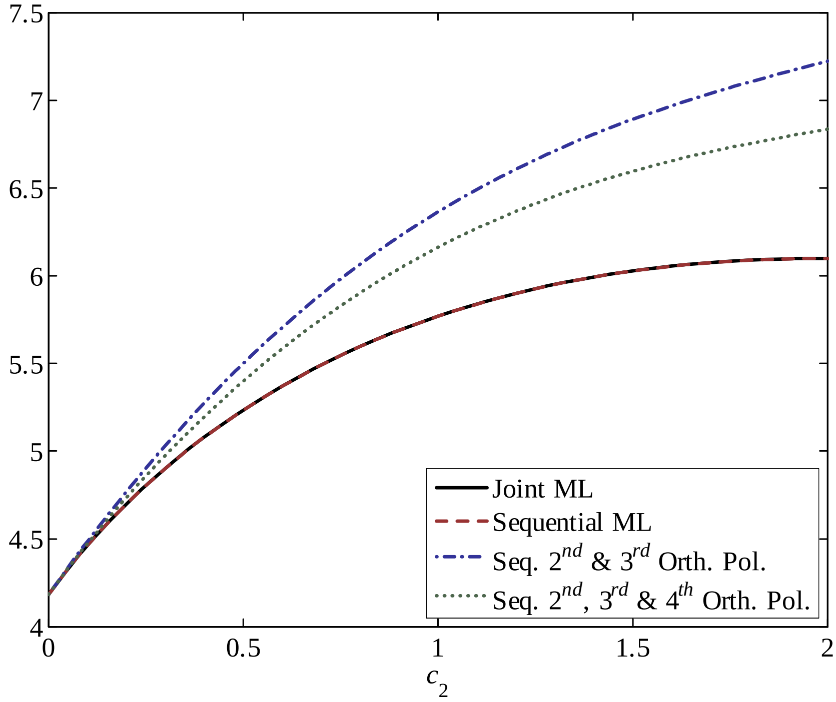
Relative efficiency of c_3 estimators (with respect to Joint ML)



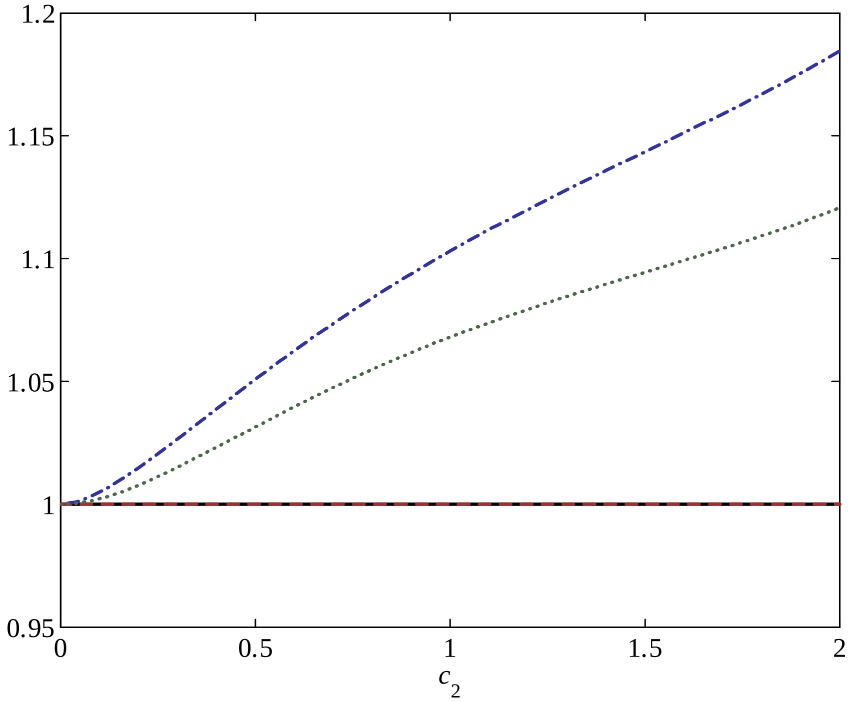
Notes: $N = 5$ and $c_2 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2-1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F4: (c) Asymptotic efficiency of PE estimators ($c_3 = 0$)

Asymptotic standard errors of c_2 estimators



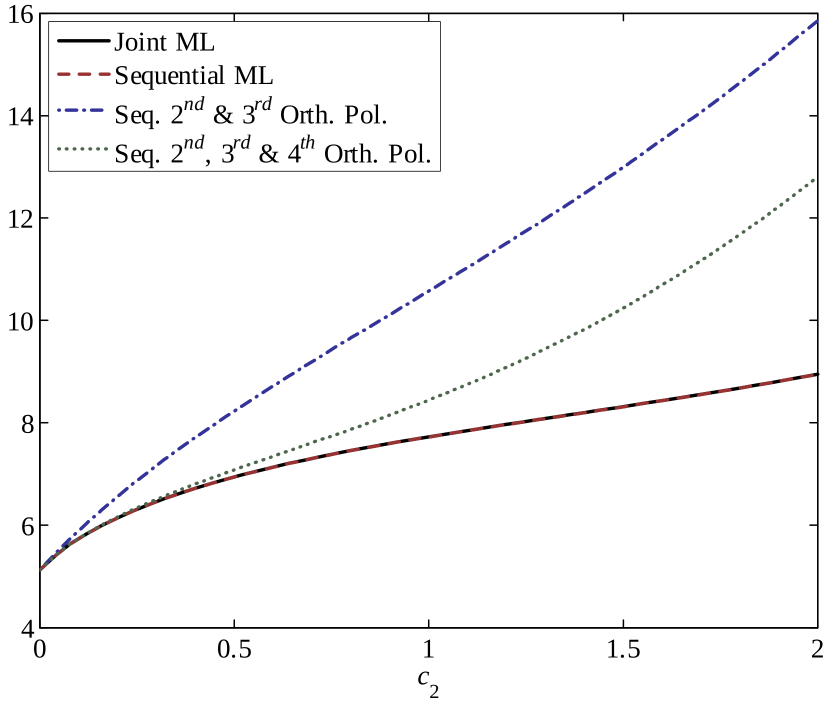
Relative efficiency of c_2 estimators (with respect to Joint ML)



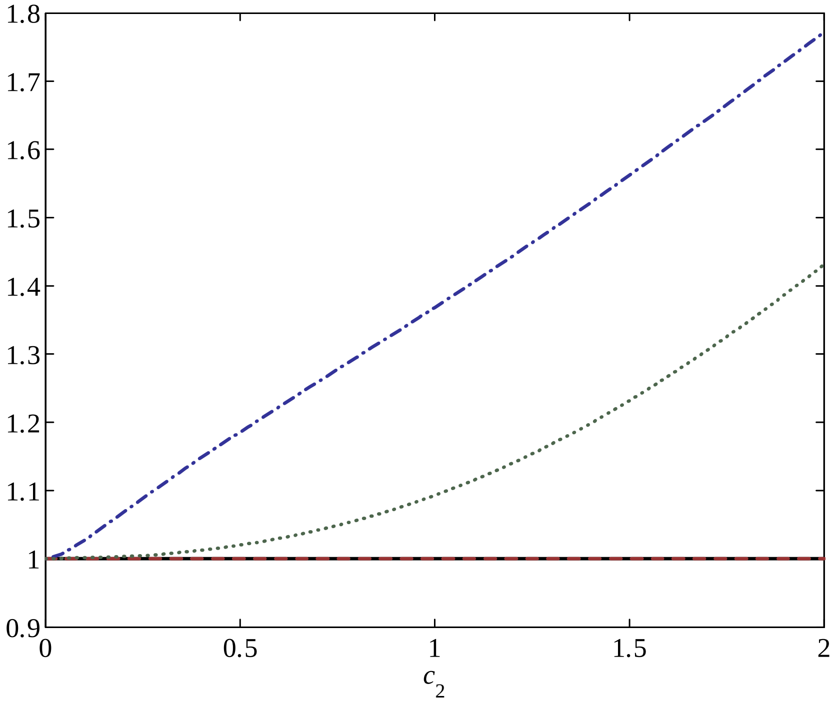
Notes: $N = 5$ and $c_3 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2-1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

Figure F4: (d) Asymptotic efficiency of PE estimators ($c_3 = 0$)

Asymptotic standard errors of c_3 estimators



Relative efficiency of c_3 estimators (with respect to Joint ML)



Notes: $N = 5$ and $c_3 = 0$. For PE innovations, c_2 and c_3 denote the coefficients associated to the 2^{nd} and 3^{rd} Laguerre polynomials with parameter $N/2-1$, respectively. Expressions for the asymptotic variances of the different estimators are given in Section 3.

References

Amengual, D. and Sentana, E. (2010): "A comparison of mean-variance efficiency tests", *Journal of Econometrics* 154, 16-34.

Amengual, D. and Sentana, E. (2011): "Inference in multivariate dynamic models with elliptical innovations", mimeo, CEMFI.

Balestra, P. and Holly, A. (1990): "A general Kronecker formula for the moments of the multivariate normal distribution", DEEP Cahier 9002, University of Lausanne.

Berkane, M. and Bentler, P.M. (1986): "Moments of elliptically distributed random variates", *Statistics and Probability Letters* 4, 333-335.

Bollerslev, T. and Wooldridge, J. M. (1992): "Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews* 11, 143-172.

Crowder, M.J. (1976): "Maximum likelihood estimation for dependent observations", *Journal of the Royal Statistical Society* B, 38, 45-53.

Fang, K.T., Kotz, S. and Ng, K.W. (1990): *Symmetric multivariate and related distributions*, Chapman and Hall.

Fiorentini, G. and Sentana, E. (2010): "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models", mimeo, CEMFI.

Hall, P. and Yao, Q. (2003): "Inference in ARCH and GARCH models with heavy-tailed errors", *Econometrica* 71, 285-317.

Maruyama, Y. and Seo, T. (2003): "Estimation of moment parameter in elliptical distributions", *Journal of the Japan Statistical Society* 33, 215-229.

Mood, A.M., Graybill, F.A., and Boes, D.C. (1974): *Introduction to the theory of Statistics*, (3rd ed.), McGraw Hill.

NAG (2001): *NAG Fortran 77 Library Mark 19 Reference Manual.*

Newey, W.K. (1984): "A method of moments interpretation of sequential estimators", *Economics Letters* 14, 201-206.

Newey, W.K. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047-1070.

Newey, W.K. and McFadden, D.L. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

Newey, W.K. and Powell, J.L. (1998): "Two-step estimation, optimal moment conditions, and sample selection models", mimeo, MIT.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics* 30, 415-443.