Supplemental Appendices for

Specification tests for non-Gaussian maximum likelihood estimators

Gabriele Fiorentini

Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy

<gabriele.fiorentini@unifi.it>

Enrique Sentana

CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain

<sentana@cemfi.es>

August 2019

Revised: November 2020

B Auxiliary results

Lemma 1 *Let* $\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\mathbf{m}}'_T(\boldsymbol{\theta})\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\boldsymbol{\theta})$ *denote the GMM estimator of* $\boldsymbol{\theta}$ *over the parameter space* $\boldsymbol{\Theta}$ *based on the average influence functions* $\bar{\mathbf{m}}_T(\boldsymbol{\theta})$ *and weighting matrix* $\tilde{\mathcal{S}}_{mT}$*, and consider a homeomorphic and continuously differentiable transformation* $\boldsymbol{\pi}(.)$ *from the original parameters* $\boldsymbol{\theta}$ *to a new set of parameters* $\boldsymbol{\pi}$*, with* $\mathrm{rank}\left[\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right]$ *evaluated at* $\hat{\boldsymbol{\theta}}_T$ *equal to* $p = \dim(\boldsymbol{\theta})$*. If* $\hat{\boldsymbol{\theta}}_T \in int(\boldsymbol{\Theta})$*, then*

$$\begin{aligned}\hat{\boldsymbol{\theta}}_T &= \boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T),\\ \hat{\boldsymbol{\pi}}_T &= \boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_T),\end{aligned}$$

and

$$\bar{\mathbbm{m}}'_T(\hat{\boldsymbol{\pi}}_T)\tilde{\mathcal{S}}_{mT}\bar{\mathbbm{m}}_T(\hat{\boldsymbol{\pi}}_T) = \bar{\mathbf{m}}'_T(\hat{\boldsymbol{\theta}}_T)\tilde{\mathcal{S}}_{mT}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T),$$

where $\boldsymbol{\theta}(\boldsymbol{\pi})$ *is the inverse mapping such that* $\boldsymbol{\pi}[\boldsymbol{\theta}(\boldsymbol{\pi})] = \boldsymbol{\pi}$*,* $\bar{\mathbbm{m}}_T(\boldsymbol{\pi}) = \bar{\mathbf{m}}_T[\boldsymbol{\theta}(\boldsymbol{\pi})]$ *are the average influence functions written in terms of* $\boldsymbol{\pi}$*, and* $\hat{\boldsymbol{\pi}}_T = \arg\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \bar{\mathbbm{m}}'_T(\boldsymbol{\pi})\tilde{\mathcal{S}}_{mT}\bar{\mathbbm{m}}_T(\boldsymbol{\pi})$*.*

Proof. The interior solution assumption implies that the sample first-order condition characterising $\hat{\boldsymbol{\theta}}_T$ is

$$\frac{\partial\bar{\mathbf{m}}'_T(\hat{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{\mathbbm{m}T}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}, \tag{B1}$$

while the corresponding condition for $\hat{\boldsymbol{\pi}}_T$ will be

$$\frac{\partial\bar{\mathbbm{m}}'_T(\hat{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\tilde{\mathcal{S}}_{\mathbbm{m}T}\bar{\mathbbm{m}}_T(\hat{\boldsymbol{\pi}}_T) = \frac{\partial\boldsymbol{\theta}'(\hat{\boldsymbol{\pi}}_T)}{\partial\boldsymbol{\pi}}\frac{\partial\bar{\mathbf{m}}'_T[\boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)]}{\partial\boldsymbol{\theta}}\tilde{\mathcal{S}}_{\mathbbm{m}T}\bar{\mathbf{m}}_T[\boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)] = \mathbf{0} \tag{B2}$$

by the chain rule for derivatives. Given that $\mathrm{rank}\left[\partial\boldsymbol{\theta}'(\boldsymbol{\pi})/\partial\boldsymbol{\pi}\right]$ evaluated at $\boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_T)$ is p in view of our assumption on the rank of the direct Jacobian $\partial\boldsymbol{\pi}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ by virtue of the inverse mapping theorem, the above equations imply that $\hat{\boldsymbol{\theta}}_T = \boldsymbol{\theta}(\hat{\boldsymbol{\pi}}_T)$, whence the other two results trivially follow. □

This result confirms the numerical invariance of the GMM criterion to reparametrisations when the weighting matrix remains the same, a condition satisfied by the most popular choices, including the identity matrix, as well as the unconditional sample variance of the influence functions and its long-run counterpart when the initial estimators at which those matrices are evaluated satisfy $\boldsymbol{\pi}^i = \boldsymbol{\pi}(\boldsymbol{\theta}^i)$. Obviously, in exactly identified contexts, such as the one implicitly arising in maximum likelihood estimation, in which the usual sufficient identification condition $\mathrm{rank}\{E[\partial\mathbf{m}_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}']\} = p$ holds, the weighting matrix becomes irrelevant, at least in large samples, which allows us to replace the first order conditions (B1) and (B2) by $\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$, and $\bar{\mathbbm{m}}_T(\hat{\boldsymbol{\pi}}_T) = \mathbf{0}$, respectively. Aside from this change, the results of the lemma continue to hold.

Lemma 2 *Let* ς *denote a scalar random variable with continuously differentiable density function* $h(\varsigma;\boldsymbol{\eta})$ *over the possibly infinite domain* $[a,b]$*, and let* $m(\varsigma)$ *denote a continuously differentiable function over the same domain such that* $E\left[m(\varsigma)|\boldsymbol{\eta}\right] = k(\boldsymbol{\eta}) < \infty$*. Then*

$$E\left[\partial m(\varsigma)/\partial\varsigma|\,\boldsymbol{\eta}\right] = -E\left[m(\varsigma)\partial\ln h(\varsigma;\boldsymbol{\eta})/\partial\varsigma|\,\boldsymbol{\eta}\right],$$

as long as the required expectations are defined and bounded.

Proof. If we differentiate

$$k(\boldsymbol{\eta}) = E\left[m(\varsigma)|\,\boldsymbol{\eta}\right] = \int_a^b m(\varsigma)h(\varsigma;\boldsymbol{\eta})d\varsigma$$

with respect to ς, we get

$$0 = \int_a^b \frac{\partial m(\varsigma)}{\partial\varsigma}h(\varsigma;\boldsymbol{\eta})d\varsigma + \int_a^b m(\varsigma)\frac{\partial h(\varsigma;\boldsymbol{\eta})}{\partial\varsigma}d\varsigma = \int_a^b \frac{\partial m(\varsigma)}{\partial\varsigma}h(\varsigma;\boldsymbol{\eta})d\varsigma + \int_a^b m(\varsigma)h(\varsigma;\boldsymbol{\eta})\frac{\partial \ln h(\varsigma;\boldsymbol{\eta})}{\partial\varsigma}d\varsigma,$$

as required. □

Lemma 3 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ with density function $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})$, where $\boldsymbol{\varrho} = \mathbf{0}$ denotes normality, then*

$$E\left\{\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}),\mathbf{e}'_{rt}(\boldsymbol{\theta},\boldsymbol{\varrho})\right]\middle|\, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} = [\mathcal{K}(\mathbf{0})\,|\mathbf{0}]. \tag{B3}$$

Proof. We can use the conditional analogue to the generalised information matrix equality (see e.g. Newey and McFadden (1994)) to show that

$$\begin{aligned} E\left\{\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})\left[\mathbf{s}'_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\varrho}),\mathbf{s}'_{\boldsymbol{\varrho}t}(\boldsymbol{\theta},\boldsymbol{\varrho})\right]\middle|\, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} &= -E\left\{\left[\frac{\partial\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})}{\partial\boldsymbol{\theta}'}\middle|\frac{\partial\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})}{\partial\boldsymbol{\varrho}'}\right]\middle|\, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} \\ &= -E\left\{[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta};\mathbf{0})|\mathbf{0}]\right|I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\} = [\mathcal{A}_t(\phi)|\mathbf{0}] \end{aligned}$$

irrespective of the conditional distribution of $\boldsymbol{\varepsilon}_t^*$, where we have used the fact that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$ does not vary with $\boldsymbol{\varrho}$ when regarded as the influence function for $\tilde{\boldsymbol{\theta}}_T$. Then, the required result follows from the martingale difference nature of both $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ and $\mathbf{e}_t(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)$. □

Lemma 4

$$\begin{aligned} &\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{K}_{NN}+\boldsymbol{\Upsilon} & \mathbf{E}_N\mathrm{M}_{sr} \\ \mathrm{M}'_{sr}\mathbf{E}_N & \mathcal{M}_{rr} \end{pmatrix}^{-1} \\ &\begin{pmatrix} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{E}_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}\begin{pmatrix} \mathrm{M}_{ss} & \mathrm{M}_{sr} \\ \mathrm{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{E}'_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}, \end{aligned} \tag{B4}$$

where $\mathcal{M}_{ss}$, $\mathcal{M}_{sr}$, $\mathcal{M}_{rr}$, $\boldsymbol{\Upsilon}$ and M_{sr} are defined in Proposition D2, and $\mathrm{M}_{ss} = (\mathbf{I}_N + \mathbf{E}'_N\boldsymbol{\Upsilon}\mathbf{E}_N)$ is a diagonal matrix of order N with typical element $\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$.

Proof. Using the partitioned inverse formula, we get

$$\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix}^{-1} = \begin{bmatrix} \mathcal{M}_{ss}^{-1} + \mathcal{M}_{ss}^{-1}\mathcal{M}_{sr}\mathcal{M}^{rr}\mathcal{M}'_{sr}\mathcal{M}_{ss}^{-1} & -\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathcal{M}'_{sr}\mathcal{M}_{ss}^{-1} & (\mathcal{M}_{rr} - \mathcal{M}'_{sr}\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr})^{-1} \end{bmatrix}.$$

Given that $\boldsymbol{\Upsilon}$ is diagonal, we can use Proposition 7 in Magnus and Sentana (2020), which yields

$$\begin{aligned} \mathcal{M}_{ss}^{-1} &= (\mathbf{K}_{NN}+\boldsymbol{\Upsilon})^{-1} = \boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N + \mathbf{E}_N(\mathbf{I}_N+\mathbf{E}'_N\boldsymbol{\Upsilon}\mathbf{E}_N)^{-1}\mathbf{E}'_N \\ &= \boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N + \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathbf{E}'_N. \end{aligned}$$

In turn, Theorem 7.4(i) in Magnus (1988) states that $\mathbf{K}_{NN}\mathbf{E}_N = \mathbf{E}_N$, which implies that $\mathcal{M}_{ss}\mathbf{E}_N = (\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\mathbf{E}_N = (\mathbf{I}_{N^2} + \boldsymbol{\Upsilon})\mathbf{E}_N = \mathbf{E}_N(\mathbf{I}_N + \mathbf{E}_N'\boldsymbol{\Upsilon}\mathbf{E}_N) = \mathbf{E}_N\mathrm{M}_{ss}$ by virtue of Proposition 3 in Magnus and Sentana (2020). Then, if we premultiply both sides by $\mathcal{M}_{ss}^{-1} = (\mathbf{K}_{NN} + \boldsymbol{\Upsilon})^{-1}$, we end up with $\mathbf{E}_N = \mathcal{M}_{ss}^{-1}\mathbf{E}_N\mathrm{M}_{ss}$, whence we finally obtain that $\mathcal{M}_{ss}^{-1}\mathbf{E}_N = \mathbf{E}_N\mathrm{M}_{ss}^{-1}$. Thus, $\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr} = \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$, where $\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$ is a diagonal matrix with typical element $\mathrm{M}_{sr}(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$. Therefore $\mathcal{M}_{sr}'\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr} = \mathrm{M}_{sr}'\mathbf{E}_N'\mathcal{M}_{ss}^{-1}\mathbf{E}_N\mathrm{M}_{sr} = \mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$ will be a diagonal $N \times N$ matrix with typical diagonal element $\mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$. In turn, this implies that $\mathcal{M}_{rr} - \mathcal{M}_{sr}'\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr} = \mathcal{M}_{rr} - \mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}$ is a diagonal matrix with typical element $\mathrm{M}_{rr}(\nu_i) - \mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$, so that $\mathcal{M}^{rr} = (\mathcal{M}_{rr} - \mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr})^{-1}$ is also diagonal. Moreover, $\mathcal{M}_{ss}^{-1}\mathcal{M}_{sr}\mathcal{M}^{rr} = \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$, where $\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}$ is once again diagonal with typical element $[\mathrm{M}_{sr}(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)]/[\mathrm{M}_{rr}(\nu_i)-\mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)/\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)]$.

If we put all these pieces together, we end up with

$$\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}_{sr}' & \mathcal{M}_{rr} \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{M}_{ss}^{-1} + \mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & -\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & \mathcal{M}^{rr} \end{pmatrix}$$

$$= \left\{ \begin{matrix} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}_N'(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}_N' + \mathbf{E}_N(\mathrm{M}_{ss}^{-1} + \mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1})\mathbf{E}_N' & -\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1}\mathbf{E}_N' & \mathcal{M}^{rr} \end{matrix} \right\}$$

$$= \begin{pmatrix} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}_N'(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}_N' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{E}_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathrm{M}_{ss}^{-1} + \mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1} & -\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}_{sr}'\mathrm{M}_{ss}^{-1} & \mathcal{M}^{rr} \end{pmatrix} \begin{pmatrix} \mathbf{E}_N' & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{\Delta}_N[\boldsymbol{\Delta}_N'(\mathbf{K}_{NN} + \boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}_N' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{E}_N & 0 \\ 0 & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathrm{M}_{ss} & \mathrm{M}_{sr} \\ \mathrm{M}_{sr}' & \mathcal{M}_{rr} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{E}_N' & 0 \\ 0 & \mathbf{I}_N \end{pmatrix},$$

as claimed. □

Proposition B1 *If model (18) with cross-sectionally independent symmetric structural shocks generates a covariance stationary process, then:*

1. *Its information matrix is block diagonal between $(\boldsymbol{\tau}', \mathbf{a}')'$ and $(\mathbf{c}', \boldsymbol{\varrho}')'$*
2. *The asymptotic covariance matrix of the restricted and unrestricted ML estimators of $(\boldsymbol{\tau}', \mathbf{a}')'$ will be given by*

$$\begin{bmatrix} 1 & \boldsymbol{\mu}' & \dots & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') & \dots & (\boldsymbol{\Gamma}(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}'(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') & \dots & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') \end{bmatrix}^{-1} \otimes \mathbf{C}\mathcal{M}_{ll}^{-1}\mathbf{C}',$$

   where $\boldsymbol{\Gamma}(p)$ is the p^{th} autocovariance matrix of $\mathbf{y}_t$ and $\mathcal{M}_{ll}$ is defined in Proposition D2.
3. *The asymptotic covariance matrices of the restricted and unrestricted ML estimators of $\mathbf{c}$ and $\boldsymbol{\varrho}$ are given by*

$$(\mathbf{I}_N \otimes \mathbf{C})\mathcal{M}_{ss}^{-1}(\mathbf{I}_N \otimes \mathbf{C}') \text{ and}$$

$$\begin{bmatrix} (\mathbf{I}_N \otimes \mathbf{C}'^{-1})\mathcal{M}_{ss}(\mathbf{I}_N \otimes \mathbf{C}^{-1}) & (\mathbf{I}_N \otimes \mathbf{C}'^{-1})\mathcal{M}_{sr} \\ \mathcal{M}_{sr}'(\mathbf{I}_N \otimes \mathbf{C}^{-1}) & \mathcal{M}_{rr} \end{bmatrix}^{-1},$$

   respectively, where $\mathcal{M}_{ss}$, $\mathcal{M}_{sr}$ and $\mathcal{M}_{rr}$ are also defined in Proposition D2 and the rank of the difference between the asymptotic variances of these two estimators of $\mathbf{c}$ is N.

Proof. Given the mapping between the structural and reduced form parameters, the contribution to the conditional log-likelihood function from observation t $(t = 1, \ldots, T)$ will be

$$l_t(\mathbf{y}_t; \boldsymbol{\phi}) = -\ln|\mathbf{C}| + l[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1] + \ldots + l[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N],$$

where $l[\varepsilon^*_{it}(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]$ is the univariate log-likelihood function for the i^{th} structural shock $\varepsilon^*_{it}(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}) = \mathbf{C}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = (\mathbf{y}_t - \boldsymbol{\tau} - \boldsymbol{\Phi}_1\mathbf{y}_{t-1} - \ldots - \boldsymbol{\Phi}_p\mathbf{y}_{t-p})$. To compute the gradient and information matrix, we rely on the expressions in Supplemental Appendix D.3 because the assumed multivariate distribution for $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta})$ is not elliptically symmetric despite the marginal distributions of its components being symmetric. Given that the conditional mean vector and covariance matrix of (18) are given by

$$\begin{aligned} \boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\tau} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p}, \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{C}\mathbf{C}', \end{aligned}$$

respectively, straightforward algebra shows that

$$\begin{aligned} \mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \frac{\partial \boldsymbol{\mu}'_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2 \times N} \end{pmatrix} \mathbf{C}^{-1\prime}, \\ \mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} [\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = \begin{pmatrix} \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{N^2 \times N^2} \\ \vdots \\ \mathbf{0}_{N^2 \times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}), \end{aligned}$$

which means that the conditional mean and variance parameters are variation free. This fact, combined with the symmetry of the Student t and the formulas in Proposition D2, immediately implies that the information matrix will be block diagonal. Specifically, the block of the information matrix corresponding to the $N + pN^2$ conditional mean parameters $(\boldsymbol{\tau}, \mathbf{a})$ will be

$$E[\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ll}\mathbf{Z}'_{lt}(\boldsymbol{\theta})] = E\begin{bmatrix} 1 & \mathbf{y}'_{t-1} & \cdots & \mathbf{y}'_{t-p} \\ \mathbf{y}_{t-1} & \mathbf{y}_{t-1}\mathbf{y}'_{t-1} & \cdots & \mathbf{y}_{t-1}\mathbf{y}'_{t-p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t-p} & \mathbf{y}_{t-p}\mathbf{y}'_{t-1} & \cdots & \mathbf{y}_{t-p}\mathbf{y}'_{t-p} \end{bmatrix} \otimes \mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1} \tag{B5}$$

$$= \begin{bmatrix} 1 & \boldsymbol{\mu}' & \cdots & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\boldsymbol{\Gamma}(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}'(p-1) + \boldsymbol{\mu}\boldsymbol{\mu}') & \cdots & (\boldsymbol{\Gamma}(0) + \boldsymbol{\mu}\boldsymbol{\mu}') \end{bmatrix} \otimes \mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1}. \tag{B6}$$

In turn, the (conditional) information matrix for the unrestricted ML estimators of the N^2 structural shock coefficients $\mathbf{c}$ and the N shape parameters $\boldsymbol{\varrho}$ will be:

$$\begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix}.$$

In this respect, we can use the results in Proposition D2 to prove that

$$\begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix} = \begin{pmatrix} \mathbf{K}_{NN}+\boldsymbol{\Upsilon} & \mathbf{E}_N\mathrm{M}_{sr} \\ \mathrm{M}'_{sr}\mathbf{E}_N & \mathcal{M}_{rr} \end{pmatrix}.$$

Hence, the information matrix will be

$$\begin{aligned} & \begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathcal{M}_{ss} & \mathcal{M}_{sr} \\ \mathcal{M}'_{sr} & \mathcal{M}_{rr} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix} \\ = & \begin{bmatrix} (\mathbf{I}_N\otimes\mathbf{C}'^{-1})(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})(\mathbf{I}_N\otimes\mathbf{C}^{-1}) & (\mathbf{I}_N\otimes\mathbf{C}'^{-1})\mathbf{E}_N\mathrm{M}_{sr} \\ \mathrm{M}'_{sr}\mathbf{E}_N(\mathbf{I}_N\otimes\mathbf{C}^{-1}) & \mathcal{M}_{rr} \end{bmatrix}. \end{aligned}$$

If we then use the expressions in Lemma 4, we can easily show that the inverse of the information matrix will be

$$\begin{bmatrix} (\mathbf{I}_N\otimes\mathbf{C})\{\boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N+\mathbf{E}_N\mathrm{M}^{ss}\mathbf{E}'_N\}(\mathbf{I}_N\otimes\mathbf{C}) & -(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr} \\ -\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{C}) & \mathcal{M}^{rr} \end{bmatrix},$$

where $\mathrm{M}^{ss}=\mathrm{M}_{ss}^{-1}+\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}$.

In contrast, if we assume that the shape parameters are fixed at their true values, the asymptotic covariance matrix of the restricted ML estimators of $\mathbf{c}$ will be

$$\begin{aligned} (\mathbf{I}_N\otimes\mathbf{C})\mathcal{M}_{ss}^{-1}(\mathbf{I}_N\otimes\mathbf{C}') &= (\mathbf{I}_N\otimes\mathbf{C})\boldsymbol{\Delta}_N[\boldsymbol{\Delta}'_N(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N]^{-1}\boldsymbol{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{C}') \\ &\quad +(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{C}'). \end{aligned}$$

Therefore, the efficiency loss from simultaneously estimating the N shape parameters $\boldsymbol{\varrho}$ will be

$$(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{C}'),$$

which has rank N rather than N^2 because $\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}\mathrm{M}'_{sr}\mathrm{M}_{ss}^{-1}\mathbf{E}'_N$ is a diagonal matrix of rank N in which the non-zero diagonal elements are

$$\frac{1}{\psi_i^2}\frac{\mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)}{\mathrm{M}_{ss}^2(\boldsymbol{\varrho}_i)}\left[\mathrm{M}_{rr}(\boldsymbol{\varrho}_i)-\frac{\mathrm{M}_{sr}^2(\boldsymbol{\varrho}_i)}{\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)}\right]^{-1}.$$

Finally, note that since the ranks of $(\mathbf{I}_N\otimes\mathbf{C}'^{-1})$ and $\mathcal{M}_{sr}=\mathbf{E}_N\mathrm{M}_{sr}$ are N^2 and N, respectively, Sylvester's rank inequality implies that

$$rank[(\mathbf{I}_N\otimes\mathbf{C})\mathbf{E}_N\mathrm{M}_{ss}^{-1}\mathrm{M}_{sr}\mathcal{M}^{rr}]=N,$$

so that Holly's (1982) condition holds. □

Proposition B2 *If model (18) with cross-sectionally independent symmetric structural shocks generates a covariance stationary process, then the asymptotic covariance matrix of the Gaussian PML estimators is block diagonal between* $(\boldsymbol{\tau}',\mathbf{a}')'$ *and* $\boldsymbol{\sigma}$*, with the first block given by*

$$\begin{bmatrix} 1 & \boldsymbol{\mu}' & \dots & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}(0)+\boldsymbol{\mu}\boldsymbol{\mu}') & \dots & (\boldsymbol{\Gamma}(p-1)+\boldsymbol{\mu}\boldsymbol{\mu}') \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu} & (\boldsymbol{\Gamma}'(p-1)+\boldsymbol{\mu}\boldsymbol{\mu}') & \dots & (\boldsymbol{\Gamma}(0)+\boldsymbol{\mu}\boldsymbol{\mu}') \end{bmatrix}^{-1} \otimes \boldsymbol{\Sigma}$$

and the second block by

$$[\mathbf{D}_N'(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{D}_N]^{-1}\mathbf{D}_N'(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1})\mathcal{K}(\mathbf{C}^{-\prime}\otimes\mathbf{C}^{-1\prime})\mathbf{D}_N[\mathbf{D}_N'(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{D}_N]^{-1},$$

where $\mathcal{K}=E[vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^*-\mathbf{I}_N)vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^*-\mathbf{I}_N)']$ *is the* $N^2\times N^2$ *matrix of fourth-order moments of the structural shocks.*

Proof. The information matrix equality implies that the expected value of the (minus) Hessian of the Gaussian pseudo log-likelihood usually coincides with the value of the true information matrix under normality. Therefore, we could exploit the fact that $\mathcal{M}_{ll}=\mathbf{I}_N$ and $\mathbf{C}^{-1\prime}\mathcal{M}_{ll}\mathbf{C}^{-1}=\mathbf{\Sigma}^{-1}$ under normality to simplify the expressions we have already derived for $\boldsymbol{\tau}$ and $\mathbf{a}$ in Proposition B1. However, the situation is slightly more complicated for $\boldsymbol{\sigma}$ because the number of parameters that can be identified by the Gaussian and non-Gaussian PMLs is different. For that reason, we use the expressions in Proposition C2 to prove that the bottom block of the (minus) expected value of the Hessian will be given by

$$\mathcal{A}_{\boldsymbol{\sigma}\boldsymbol{\sigma}}=\frac{1}{4}\mathbf{D}_N'(\mathbf{\Sigma}^{-\frac{1}{2}}\otimes\mathbf{\Sigma}^{-\frac{1}{2}})(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{\Sigma}^{-\frac{1}{2}\prime}\otimes\mathbf{\Sigma}^{-\frac{1}{2}\prime})\mathbf{D}_N=\frac{1}{2}\mathbf{D}_N'(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{D}_N$$

regardless of the choice of square root matrix in view of the properties of the duplication and commutation matrix in Magnus and Neudecker (2919).

As for the matrix $\mathcal{B}$, which contains the asymptotic variance of the Gaussian scores, the symmetry of the marginal distributions of the structural shocks together with the cross-sectional independence across shocks imply that we will also obtain a block diagonal expression with the same block for the conditional mean parameters as $\mathcal{A}$. In contrast, the block for the conditional variance parameters $\boldsymbol{\sigma}$ will be different. To obtain it, we can use the expressions in Proposition C2 with $\mathbf{C}$ playing the role of $\mathbf{\Sigma}^{\frac{1}{2}}$ to exploit the cross-sectional independence of the structural shocks, which leads to

$$\mathcal{B}_{\boldsymbol{\sigma}\boldsymbol{\sigma}}=\frac{1}{4}\mathbf{D}_N'(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1})\mathcal{K}(\mathbf{C}^{-\prime}\otimes\mathbf{C}^{-1\prime})\mathbf{D}_N,$$

where $\mathcal{K}$ is equal to $\mathbf{K}_{NN}$ plus a a block diagonal matrix in which each of the N blocks is diagonal of size $N\times N$ with the following structure:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & \ddots & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \kappa_{ii}(\boldsymbol{\varrho}_i) & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In the Student t case, $\kappa_{ii}(\boldsymbol{\varrho}_i)=(\nu_i+2)/(\nu_i-4)$. □

Proposition B3 *If model (18) with cross-sectionally independent symmetric structural shocks generates a covariance stationary process, then the scores and information matrix of* $\boldsymbol{\sigma}_L$ *and* $\boldsymbol{\omega}$ *are given by*

$$\begin{bmatrix} s_{\boldsymbol{\sigma}_L}(\boldsymbol{\theta};\boldsymbol{\varrho})\\ s_{\boldsymbol{\omega}}(\boldsymbol{\theta};\boldsymbol{\varrho})\end{bmatrix}=\begin{bmatrix}\mathbf{L}_N(\mathbf{I}_N\otimes\mathbf{\Sigma}_L^{-1\prime})(\mathbf{Q}\otimes\mathbf{Q})\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\mathbf{Q})\end{bmatrix}\mathbf{e}_{st}(\boldsymbol{\phi})$$

and

$$\left[\begin{array}{c}\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime})(\mathbf{Q}\otimes\mathbf{Q})\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\mathbf{Q})\end{array}\right]\mathcal{M}_{ss}\left[\begin{array}{cc}(\mathbf{Q}'\otimes\mathbf{Q}')(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1})\mathbf{L}_N & (\mathbf{I}_N\otimes\mathbf{Q})\cdot\partial vec(\mathbf{Q})/\partial\boldsymbol{\omega}'\end{array}\right].$$

Proof. As in Proposition 14, the proof builds up on Proposition B1. Specifically, given that $vec(\mathbf{C})=(\mathbf{Q}'\otimes\mathbf{I}_N)vec(\boldsymbol{\Sigma}_L)=(\mathbf{Q}'\otimes\mathbf{I}_N)\mathbf{L}_N'vech(\boldsymbol{\Sigma}_L)$, straightforward algebra shows that

$$\frac{\partial\mathbf{c}}{\partial\boldsymbol{\sigma}_L'}=(\mathbf{Q}'\otimes\mathbf{I}_N)\mathbf{L}_N'.$$

Similarly, given that we can also write $vec(\mathbf{C})=(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L)vec(\mathbf{Q})$, we will have that

$$\frac{\partial\mathbf{c}}{\partial\boldsymbol{\omega}'}=(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L)\frac{\partial vec(\mathbf{Q})}{\partial\boldsymbol{\omega}'},$$

where $\partial vec(\mathbf{Q})/\partial\boldsymbol{\omega}'$ depends on the particular parametrisation of orthogonal matrices chosen (see Magnus, Pijls and Sentana (2020)). Given that

$$s_{\mathbf{c}}(\boldsymbol{\theta};\boldsymbol{\varrho})=(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{e}_{st}(\boldsymbol{\phi}),$$

this direct approach allows us to obtain the scores for $\boldsymbol{\sigma}_L$ and $\boldsymbol{\omega}$ as

$$\left[\begin{array}{c}s_{\boldsymbol{\sigma}_L}(\boldsymbol{\theta};\boldsymbol{\varrho})\\ s_{\boldsymbol{\omega}}(\boldsymbol{\theta};\boldsymbol{\varrho})\end{array}\right]=\left(\begin{array}{c}\partial\mathbf{c}'/\partial\boldsymbol{\sigma}_L\\ \partial\mathbf{c}'/\partial\boldsymbol{\omega}\end{array}\right)s_{\mathbf{c}}(\boldsymbol{\theta};\boldsymbol{\varrho})=\left[\begin{array}{c}\mathbf{L}_N(\mathbf{Q}\otimes\mathbf{I}_N)\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L')\end{array}\right]s_{\mathbf{c}}(\boldsymbol{\theta};\boldsymbol{\varrho}).$$

But since $\mathbf{C}=\boldsymbol{\Sigma}_L\mathbf{Q}$ so $\mathbf{C}^{-1}=\mathbf{Q}'\boldsymbol{\Sigma}_L^{-1}$ and $\mathbf{C}^{-1\prime}=\boldsymbol{\Sigma}_L^{-1\prime}\mathbf{Q}$, we have that

$$\begin{aligned}\left[\begin{array}{c}\mathbf{L}_N(\mathbf{Q}\otimes\mathbf{I}_N)\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L')\end{array}\right](\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) &= \left[\begin{array}{c}\mathbf{L}_N(\mathbf{Q}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime}\mathbf{Q})\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L')(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime}\mathbf{Q})\end{array}\right]\\ &= \left[\begin{array}{c}\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime})(\mathbf{Q}\otimes\mathbf{Q})\\ \partial vec'(\mathbf{Q})/\partial\boldsymbol{\omega}\cdot(\mathbf{I}_N\otimes\mathbf{Q})\end{array}\right],\end{aligned}$$

whence the expression for the scores and information matrix immediately follows. The dependence of the scores $s_{\boldsymbol{\sigma}_L}(\boldsymbol{\theta};\boldsymbol{\varrho})$ on $\mathbf{Q}$ simply reflects the fact that we have defined $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})=\mathbf{C}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ in terms of the true underlying independent shocks. We explain how to compute $\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_L^{-1\prime})$ efficiently at the end of Appendix D.1. □

To obtain the asymptotic variances of $\boldsymbol{\sigma}_L$, we can alternatively use the following two-step procedure. First, we go from the structural loading matrix $\mathbf{C}$ to $\boldsymbol{\Sigma}$. Given that $d\boldsymbol{\Sigma}=(d\mathbf{C})\mathbf{C}'+\mathbf{C}(d\mathbf{C}')$, it immediately follows that

$$\begin{aligned}dvec(\boldsymbol{\Sigma}) &= (\mathbf{C}\otimes\mathbf{I}_N)dvec(\mathbf{C})+(\mathbf{I}_N\otimes\mathbf{C})dvec(\mathbf{C}')\\ &= (\mathbf{C}\otimes\mathbf{I}_N)dvec(\mathbf{C})+(\mathbf{I}_N\otimes\mathbf{C})\mathbf{K}_{NN}dvec(\mathbf{C})=(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{C}\otimes\mathbf{I}_N)dvec(\mathbf{C}),\end{aligned}$$

so that

$$\frac{\partial\boldsymbol{\sigma}}{\partial\mathbf{c}'}=\mathbf{D}_N^+(\mathbf{I}_{N^2}+\mathbf{K}_{NN})(\mathbf{C}\otimes\mathbf{I}_N),$$

where $\mathbf{D}_N^+$ is the Moore-Penrose inverse of the duplication matrix (see Magnus, 1988). Using this Jacobian, the delta method allows us to obtain the asymptotic covariance matrix of the restricted and unrestricted MLEs of the reduced form parameters $\boldsymbol{\sigma}$, but not their scores because $\text{rank}(\partial\boldsymbol{\sigma}/\partial\mathbf{c}')=N(N+1)/2$, so we cannot invert it. Then, we can go from $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}_L$ by exploiting expression (E13) in Appendix D.1.

Lemma 5

$$[(\mathbf{\Psi}\otimes\mathbf{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\mathbf{\Delta}_N \vdots \mathbf{E}_N\mathbf{\Psi}^{-1}]^{-1} = \left\{ \begin{array}{c} \mathbf{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})] \\ \mathbf{\Psi}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{J}) \end{array} \right\}.$$

Proof. Let us look at the four blocks of

$$\left\{ \begin{array}{c} \mathbf{\Delta}'_N(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})[\mathbf{I}_{N^2}-\mathbf{E}_N\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{J})(\mathbf{\Psi}^{-1}\otimes\mathbf{\Psi})] \\ \mathbf{\Psi}\mathbf{E}'_N(\mathbf{I}_N\otimes\mathbf{J}) \end{array} \right\} [\ (\mathbf{\Psi}\otimes\mathbf{\Psi}^{-1})(\mathbf{I}_N\otimes\mathbf{J}^{-1})\mathbf{\Delta}_N \quad \mathbf{E}_N\mathbf{\Psi}^{-1}\ ].$$

The northwestern block is

$$\begin{aligned} & \mathbf{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi}^{-1} \otimes \mathbf{\Psi})(\mathbf{\Psi} \otimes \mathbf{\Psi}^{-1})(\mathbf{I}_N \otimes \mathbf{J}^{-1})\mathbf{\Delta}_N \\ & -\mathbf{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi}^{-1} \otimes \mathbf{\Psi})\mathbf{E}_N\mathbf{E}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi}^{-1} \otimes \mathbf{\Psi})(\mathbf{\Psi} \otimes \mathbf{\Psi}^{-1})(\mathbf{I}_N \otimes \mathbf{J}^{-1})\mathbf{\Delta}_N \\ = \ & \mathbf{\Delta}'_N\mathbf{\Delta}_N - \mathbf{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi}^{-1} \otimes \mathbf{\Psi})\mathbf{E}_N\mathbf{E}'_N\mathbf{\Delta}_N = \mathbf{I}_{N(N-1)} \end{aligned}$$

by virtue of Proposition 4 in Magnus and Sentana (2020). Similarly, the northeastern block is

$$\begin{aligned} & \mathbf{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi}^{-1} \otimes \mathbf{\Psi})\mathbf{E}_N\mathbf{\Psi}^{-1} - \mathbf{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi}^{-1} \otimes \mathbf{\Psi})\mathbf{E}_N\mathbf{E}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi}^{-1} \otimes \mathbf{\Psi})\mathbf{E}_N\mathbf{\Psi}^{-1} \\ = \ & \mathbf{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N\mathbf{\Psi}^{-1} - \mathbf{\Delta}'_N(\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N\mathbf{E}'_N(\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N\mathbf{\Psi}^{-1} = \mathbf{0} \end{aligned}$$

thanks to Propositions 2 and 3 in Magnus and Sentana (2020), together with the fact that the diagonal elements of $\mathbf{J}$ are normalised to 1. The same propositions also imply that the southwestern block will be

$$\mathbf{\Psi}\mathbf{E}'_N(\mathbf{I}_N \otimes \mathbf{J})(\mathbf{\Psi} \otimes \mathbf{\Psi}^{-1})(\mathbf{I}_N \otimes \mathbf{J}^{-1})\mathbf{\Delta}_N = \mathbf{\Psi}\mathbf{E}'_N\mathbf{\Delta}_N = \mathbf{0},$$

while the sourtheastern one

$$\mathbf{\Psi}\mathbf{E}'_N(\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N\mathbf{\Psi}^{-1} = \mathbf{\Psi}(\mathbf{I}_N \odot \mathbf{J})\mathbf{\Psi}^{-1} = \mathbf{I}_N,$$

as claimed. □

C The special case of spherical distributions

C.1 Some useful distribution results

A spherically symmetric random vector of dimension N, $\boldsymbol{\varepsilon}_t^\bullet$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^\bullet = e_t\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and e_t is a non-negative random variable independent of $\mathbf{u}_t$, whose distribution determines the distribution of $\boldsymbol{\varepsilon}_t^\bullet$. The variables e_t and $\mathbf{u}_t$ are referred to as the generating variate and the uniform base of the spherical distribution. Assuming that $E(e_t^2) < \infty$, we can standardise $\boldsymbol{\varepsilon}_t^\bullet$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^\bullet) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\bullet) = \mathbf{I}_N$. Specifically, if $\boldsymbol{\varepsilon}_t^\bullet$ is distributed as a standardised multivariate Student t random vector of dimension N with ν_0 degrees of freedom, then $e_t = \sqrt{(\nu_0 - 2)\zeta_t/\xi_t}$, where ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is an independent Gamma variate with mean $\nu_0 > 2$ and variance

$2\nu_0$. If we further assume that $E(e_t^4) < \infty$, then the coefficient of multivariate excess kurtosis κ_0, which is given by $E(e_t^4)/[N(N+2)]-1$, will also be bounded. For instance, $\kappa_0 = 2/(\nu_0-4)$ in the Student t case with $\nu_0 > 4$, and $\kappa_0 = 0$ under normality. In this respect, 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 $\boldsymbol{\varepsilon}_t^\bullet$ is proportional to $\mathbf{u}_t$, then $\kappa_0 \geq -2/(N+2)$, the minimum value being achieved in the uniformly distributed case.

Then, it is easy to combine the representation of spherical distributions above 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 $V(\boldsymbol{\varepsilon}_t^\bullet) = \mathbf{I}_N$ are given by

$$E(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime} \otimes \boldsymbol{\varepsilon}_t^\bullet) = \mathbf{0}, \tag{C1}$$

$$E(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime} \otimes \boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime}) = E[vec(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^{\bullet\prime}) vec'(\boldsymbol{\varepsilon}_t^\bullet \boldsymbol{\varepsilon}_t^\bullet)] = (\kappa_0+1)[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\,(\mathbf{I}_N)\,vec'\,(\mathbf{I}_N)], \tag{C2}$$

where $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see e.g. Magnus and Neudecker (2019)).

C.2 Likelihood, score and Hessian for spherically symmetric distributions

Let $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$ denote the assumed conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and the shape parameters, where $c(\boldsymbol{\eta})$ corresponds to the constant of integration, $g(\varsigma_t, \boldsymbol{\eta})$ to its kernel and $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. 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 $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, where $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$ is the Jacobian, $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ 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 $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. 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, then

$$\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + \partial g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \tag{C3}$$

while

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial\varsigma}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})]\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{array}\right] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \tag{C4}$$

where

$$\begin{aligned}
\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N), \\
\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\}, && \text{(C5)} \\
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}), && \text{(C6)} \\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})], && \text{(C7)} \\
\mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), && \text{(C8)} \\
\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N\right\}, && \text{(C9)}
\end{aligned}$$

and

$$\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma \tag{C10}$$

is a damping factor that reflects the tail-thickness of the distribution assumed for estimation purposes. Importantly, while both $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ and $\mathbf{e}_{dt}(\boldsymbol{\phi})$ depend on the specific choice of square root matrix $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ does not, a property that inherits from $l_t(\boldsymbol{\phi})$. As we shall see in Supplemental Appendix D, this result is not generally true for non-spherical distributions.

Obviously, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})$ reduces to the multivariate normal expression in Bollerslev and Wooldridge (1992), in which case:

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

Assuming further twice differentiability of the different functions involved, we will have that the Hessian function $\mathbf{h}_t(\boldsymbol{\phi}) = \partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$ will be

$$\begin{aligned}
\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{(\partial\varsigma)^2}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma}\frac{\partial^2\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, && \text{(C11)}\\
\mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) &= \partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varsigma\partial\boldsymbol{\eta}', && \text{(C12)}\\
\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) &= \partial^2 c(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}' + \partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}',
\end{aligned}$$

where

$$\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 2\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{Z}_{st}'(\boldsymbol{\theta}) - \frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\otimes\mathbf{I}_p\right\}\partial vec\left\{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\right\}/\partial\boldsymbol{\theta}', \tag{C13}$$

$$\begin{aligned}
\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' &= 2\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + 8\mathbf{Z}_{st}(\boldsymbol{\theta})[\mathbf{I}_N\otimes\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})]\mathbf{Z}_{st}'(\boldsymbol{\theta}) + 4\mathbf{Z}_{lt}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\\
&+4\mathbf{Z}_{st}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{lt}'(\boldsymbol{\theta}) - 2[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_p]\partial vec[\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}]\partial\boldsymbol{\theta}'\\
&-\{vec'[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]\otimes\mathbf{I}_p\}\partial vec\{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\}/\partial\boldsymbol{\theta}'.
\end{aligned}$$

Note that $\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ and $\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ depend on the dynamic model specification, while $\partial^2 g(\varsigma,\eta)/(\partial\varsigma)^2$, $\partial^2 g(\varsigma,\eta)/\partial\varsigma\partial\boldsymbol{\eta}'$ and $\partial g(\varsigma,\eta)/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ depend on the specific spherical distribution assumed for estimation purposes (see Fiorentini, Sentana and Calzolari (2003) for expressions for $\delta(\varsigma_t,\boldsymbol{\eta})$, $c(\boldsymbol{\eta})$, $g(\varsigma_t,\boldsymbol{\eta})$ and its derivatives in the multivariate Student t case, Amengual and Sentana (2010) for the Kotz distribution and discrete scale mixture of normals, and Amengual, Fiorentini and Sentana (2013) for polynomial expansions).

C.3 Asymptotic distribution

Given correct specification, the results in Crowder (1976) imply that $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}_{dt}'(\boldsymbol{\phi}), \mathbf{e}_{rt}(\boldsymbol{\phi})]'$ 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 joint ML estimator will be $\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \to N\left[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)\right]$, where $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$,

$$\begin{aligned}
\mathcal{I}_t(\boldsymbol{\phi}) &= V\left[\mathbf{s}_t(\boldsymbol{\phi})|I_{t-1};\boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\phi})\mathbf{Z}_t'(\boldsymbol{\theta}) = -E\left[\mathbf{h}_t(\boldsymbol{\phi})|I_{t-1};\boldsymbol{\phi}\right],\\
\mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix}, && \text{(C14)}
\end{aligned}$$

and $\mathcal{M}(\phi) = V\left[\mathbf{e}_t(\phi)|\phi\right]$. In particular, Crowder (1976) requires: (i) ϕ_0 is locally identified and belongs to the interior of the admissible parameter space, which is a compact subset of $\mathbb{R}^{p+q}$; (ii) the Hessian matrix is non-singular and continuous throughout some neighbourhood of ϕ_0; (iii) there is uniform convergence to the integrals involved in the computation of the mean vector and covariance matrix of $\mathbf{s}_t(\phi)$; and (iv) $-E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\phi)\right]T^{-1}\sum_t \mathbf{h}_t(\phi) \xrightarrow{p} \mathbf{I}_{p+q}$, where $E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\phi)\right]$ is positive definite on a neighbourhood of ϕ_0.

As for $\tilde{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}})$, assuming that $\bar{\boldsymbol{\eta}}$ coincides with the true value of this parameter vector, the same arguments imply that $\sqrt{T}[\tilde{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}) - \boldsymbol{\theta}_0] \to N\left[\mathbf{0}, \mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}(\phi_0)\right]$, where $\mathcal{I}_{\boldsymbol{\theta\theta}}(\phi_0)$ is the relevant block of the information matrix.

The next proposition, which originally appeared as Proposition 1 in Fiorentini and Sentana (2007), generalises Propositions 3 in Lange, Little and Taylor (1989), 1 in Fiorentini, Sentana and Calzolari (2003) and 5.2 in Hafner and Rombouts (2007), providing detailed expressions for $\mathcal{M}(\phi)$ in models with non-zero conditional means:

Proposition C1 *If $\boldsymbol{\varepsilon}^*_t|I_{t-1};\phi$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ with density* $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$*, then*

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathbf{0} & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix}, \tag{C15}$$

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = \mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N, \tag{C16}$$

$$\mathcal{M}_{ss}(\boldsymbol{\eta}) = \mathrm{M}_{ss}(\boldsymbol{\eta})\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right) + [\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N), \tag{C17}$$

$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = vec(\mathbf{I}_N)\mathrm{M}_{sr}(\boldsymbol{\eta}), \tag{C18}$$

$$\mathrm{M}_{ll}(\boldsymbol{\eta}) = E\left[\delta^2(\varsigma_t, \boldsymbol{\eta})\frac{\varsigma_t}{N}\bigg|\boldsymbol{\eta}\right] = E\left[\frac{2\partial\delta(\varsigma_t, \boldsymbol{\eta})}{\partial\varsigma}\frac{\varsigma_t}{N} + \delta(\varsigma_t, \boldsymbol{\eta})\bigg|\boldsymbol{\eta}\right],$$

$$\mathrm{M}_{ss}(\boldsymbol{\eta}) = \frac{N}{N+2}\left\{1 + V\left[\delta(\varsigma_t, \boldsymbol{\eta})\frac{\varsigma_t}{N}\bigg|\boldsymbol{\eta}\right]\right\} = \frac{N}{N+2}E\left[\frac{2\partial\delta(\varsigma_t, \boldsymbol{\eta})}{\partial\varsigma}\left(\frac{\varsigma_t}{N}\right)^2\bigg|\boldsymbol{\eta}\right] + 1,$$

$$\mathrm{M}_{sr}(\boldsymbol{\eta}) = E\left\{\left[\delta(\varsigma_t, \boldsymbol{\eta})\frac{\varsigma_t}{N} - 1\right]\mathbf{e}'_{rt}(\phi)\bigg|\phi\right\} = -E\left[\frac{\varsigma_t}{N}\frac{\partial\delta(\varsigma_t, \boldsymbol{\eta})}{\partial\boldsymbol{\eta}'}\bigg|\boldsymbol{\eta}\right].$$

Proof. For our purposes it is convenient to rewrite $\mathbf{e}_{dt}(\phi_0)$ as

$$\begin{aligned} \mathbf{e}_{lt}(\phi_0) &= \delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0) = \delta(\varsigma_t, \boldsymbol{\eta}_0)\sqrt{\varsigma_t}\mathbf{u}_t, \\ \mathbf{e}_{st}(\phi_0) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0) - \mathbf{I}_N\right\} = vec\left[\delta(\varsigma_t, \boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}'_t - \mathbf{I}_N\right], \end{aligned}$$

where ς_t and $\mathbf{u}_t$ are mutually independent for any standardised spherical distribution, with $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}'_t) = N^{-1}\mathbf{I}_N$, $E(\varsigma_t) = N$ and $E(\varsigma_t^2) = N(N+2)(\kappa_0 + 1)$. Importantly, we only need to compute unconditional moments because ς_t and $\mathbf{u}_t$ are independent of $\mathbf{z}_t$ and I_{t-1} by assumption. Then, it easy to see that

$$E[\mathbf{e}_{lt}(\phi)|\phi] = E[\delta(\varsigma_t, \boldsymbol{\eta})\sqrt{\varsigma_t}|\boldsymbol{\eta}] \cdot E(\mathbf{u}_t) = \mathbf{0},$$

and that

$$E[\mathbf{e}_{st}(\phi)|\phi] = vec\left\{E\left[\delta(\varsigma_t, \boldsymbol{\eta}_0)\varsigma_t|\boldsymbol{\eta}\right] \cdot E(\mathbf{u}_t\mathbf{u}'_t) - \mathbf{I}_N\right\} = vec(\mathbf{I}_N)\left\{E\left[\delta(\varsigma_t, \boldsymbol{\eta}_0)(\varsigma_t/N)|\boldsymbol{\eta}\right] - 1\right\}.$$

In this context, we can use expression (2.21) in Fang, Kotz and Ng (1990) to write the density function of ς_t as

$$h(\varsigma_t;\boldsymbol{\eta}) = \frac{\pi^{N/2}}{\Gamma(N/2)}\varsigma_t^{N/2-1}\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})], \tag{C19}$$

whence

$$[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1] = -\frac{2}{N}\left[1+\varsigma_t\cdot\partial\ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma\right]. \tag{C20}$$

On this basis, we can use Lemma 2 in Supplemental Appendix B to show that $E(\varsigma_t)=N<\infty$ implies

$$E\left[\varsigma_t\cdot\partial\ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma|\boldsymbol{\eta}\right] = -E\left[1\right] = -1,$$

which in turn implies that

$$E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1|\boldsymbol{\eta}\right]=0 \tag{C21}$$

in view of (C20). Consequently, $E[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}]=\mathbf{0}$, as required.

Similarly, we can also show that

$$\begin{aligned} E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{lt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left\{\delta^2(\varsigma_t,\boldsymbol{\eta})\varsigma_t\mathbf{u}_t\mathbf{u}_t'|\boldsymbol{\eta}\right\} = \mathbf{I}_N\cdot E[\delta^2(\varsigma_t,\boldsymbol{\eta}_0)(\varsigma_t/N)|\boldsymbol{\eta}], \\ E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{st}'(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left\{\delta(\varsigma_t,\boldsymbol{\eta})\sqrt{\varsigma_t}\mathbf{u}_t vec'\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N\right]|\boldsymbol{\eta}\right\}=\mathbf{0} \end{aligned}$$

by virtue of (C1), and

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_0)\mathbf{e}_{st}'(\boldsymbol{\phi}_0)|\boldsymbol{\phi}] &= E\left\{vec\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N\right]vec'\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N\right]|\boldsymbol{\eta}\right\} \\ &= E\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t|\boldsymbol{\eta}\right]^2\frac{1}{N(N+2)}[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec\left(\mathbf{I}_N\right)vec'\left(\mathbf{I}_N\right)] \\ &\quad -2E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)|\boldsymbol{\eta}\right]vec\left(\mathbf{I}_N\right)vec'\left(\mathbf{I}_N\right)+vec\left(\mathbf{I}_N\right)vec'\left(\mathbf{I}_N\right) \\ &= \frac{N}{(N+2)}E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)|\boldsymbol{\eta}\right]^2(\mathbf{I}_{N^2}+\mathbf{K}_{NN}) \\ &\quad +\left\{\frac{N}{(N+2)}E\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)|\boldsymbol{\eta}\right]^2-1\right\}vec\left(\mathbf{I}_N\right)vec'\left(\mathbf{I}_N\right)] \end{aligned}$$

by virtue of (C2), (C20) and (C21).

Finally, it is clear from (C3) that $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ will be a function of ς_t but not of $\mathbf{u}_t$, which immediately implies that $E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}]=\mathbf{0}$, and that

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left\{vec\left[\delta(\varsigma_t,\boldsymbol{\eta})\varsigma_t\cdot\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N\right]\mathbf{e}_{rt}'(\boldsymbol{\phi})\right\} \\ &= vec(\mathbf{I}_N)E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta})(\varsigma_t/N)-1\right]\mathbf{e}_{rt}'(\boldsymbol{\phi})\right\}. \end{aligned}$$

To obtain the expected value of the Hessian, it is also convenient to write $\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}_0)$ in (C11) as

$$
\begin{aligned}
&-4\mathbf{Z}_{st}(\boldsymbol{\theta}_0)[\mathbf{I}_N \otimes \{\delta[\varsigma_t(\boldsymbol{\theta}_0), \eta_0]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\}]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)\\
&+[\mathbf{e}_{lt}'(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_p]\frac{\partial vec}{\partial \boldsymbol{\theta}'}\left[\frac{\partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\\
&+\frac{1}{2}\{\mathbf{e}_{st}'(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)] \otimes \mathbf{I}_p\}\frac{\partial vec}{\partial \boldsymbol{\theta}'}\left\{\frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}\right\}\\
&-2\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)[\mathbf{e}_{lt}'(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0) - 2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)[\mathbf{e}_{lt}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \otimes \mathbf{I}_N]\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&-\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0) - 2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}_0) - \frac{2\partial\delta\left[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0\right]}{\partial\varsigma}\{\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&+\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)vec'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0) + \mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&+\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]vec'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)\}.
\end{aligned}
$$

Clearly, the first four lines have zero conditional expectation, and the same is true of the sixth line by virtue of (C1). As for the remaining terms, we can write them as

$$
\begin{aligned}
&-\delta(\varsigma_t, \boldsymbol{\eta}_0)\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0) - 2\partial\delta(\varsigma_t, \boldsymbol{\eta}_0)/\partial\varsigma \cdot \mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}_t'\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)\\
&-2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}_0) - 2\partial\delta(\varsigma_t, \boldsymbol{\eta}_0)/\partial\varsigma \cdot \varsigma_t^2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec(\mathbf{u}_t\mathbf{u}_t')vec'(\mathbf{u}_t\mathbf{u}_t')\mathbf{Z}_{st}'(\boldsymbol{\theta}_0),
\end{aligned}
$$

whose conditional expectation will be

$$
\begin{aligned}
&-\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}_0)E[\delta(\varsigma_t; \boldsymbol{\eta}_0) + 2(\varsigma_t/N) \cdot \partial\delta(\varsigma_t, \boldsymbol{\eta}_0)/\partial\varsigma|\boldsymbol{\eta}_0] - 2\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}_0)\\
&-\mathbf{Z}_{st}(\boldsymbol{\theta}_0)\frac{2E[\varsigma_t^2 \cdot \partial\delta(\varsigma_t, \boldsymbol{\eta}_0)/\partial\varsigma|\boldsymbol{\eta}_0]}{N(N+2)}[(\mathbf{I}_{N^2} \otimes \mathbf{K}_{NN}) + vec(\mathbf{I}_N)vec'(\mathbf{I}_N)]\mathbf{Z}_{st}'(\boldsymbol{\theta}_0).
\end{aligned}
$$

As for $\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, it follows from (C5) and (C12) that we can write it as

$$
\begin{aligned}
&\{\mathbf{Z}_{lt}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) + \mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\right]\} \cdot \partial\delta\left[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0\right]/\partial\boldsymbol{\eta}'\\
= \quad &[\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{u}_t\sqrt{\varsigma_t} + \mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{u}_t\mathbf{u}_t')\varsigma_t] \cdot \partial\delta(\varsigma_t, \boldsymbol{\eta})/\partial\boldsymbol{\eta}',
\end{aligned}
$$

whose conditional expected value will be $\mathbf{Z}_{st}(\boldsymbol{\theta}_0)vec(\mathbf{I}_N)E[(\varsigma_t/N) \cdot \partial\delta(\varsigma_t, \boldsymbol{\eta}_0)/\partial\boldsymbol{\eta}'|\boldsymbol{\eta}]$. $\square$

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

As for $\mathcal{I}(\boldsymbol{\phi}_0)$, while it is relatively straightforward to obtain closed-form expressions in conditionally homoskedastic, dynamic linear models such as multivariate regressions or VARs (see e.g. Amengual and Sentana (2010)), it is virtually impossible to do so in dynamic conditionally heteroskedastic models, as one has to resort to numerical or Monte Carlo integration methods to compute the required expected values (see e.g. Engle and Gonzalez-Rivera (1991) and Gonzalez-Rivera and Drost (1999)). Nevertheless, see Fiorentini and Sentana (2015, 2018) for closed-form expressions in the context of tests for univariate or multivariate conditional homoskedasticity, respectively.

C.4 Gaussian pseudo maximum likelihood estimators

An important special case of restricted ML estimator arises when $\bar{\boldsymbol{\eta}} = \mathbf{0}$, in which case $\tilde{\boldsymbol{\theta}}_T(\mathbf{0})$ coincides with the Gaussian PML estimator $\tilde{\boldsymbol{\theta}}_T$. Unlike what happens with other values of $\bar{\boldsymbol{\eta}}$, $\tilde{\boldsymbol{\theta}}_T$ 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 true conditional distribution of $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ is neither Gaussian nor spherical, 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 also a vector martingale difference sequence when evaluated at $\boldsymbol{\theta}_0$, a property that inherits from

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

Importantly, this property is preserved even when the standardised innovations, $\boldsymbol{\varepsilon}_t^*$, are not stochastically independent of I_{t-1}.

The asymptotic distribution of the PML estimator of $\boldsymbol{\theta}$ is stated in the following result, which specialises Proposition 1 in Bollerslev and Wooldridge (1992) to models with *i.i.d.* innovations with shape parameters $\boldsymbol{\varrho}$:

Proposition C2 *Assume that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied.*

1. *If* $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ *is i.i.d.* $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ *with* $tr[\mathcal{K}(\boldsymbol{\varrho})]<\infty$, *where* $\boldsymbol{\phi} = (\boldsymbol{\theta}',\boldsymbol{\varrho}')'$, *then* $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N\left[\mathbf{0}, \mathcal{C}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0,\mathbf{0};\boldsymbol{\phi}_0)\right]$ *with*

$$\begin{aligned}
\mathcal{C}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= \mathcal{A}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})\mathcal{B}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})\mathcal{A}^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}), \\
\mathcal{A}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{A}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\
\mathcal{A}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta};\mathbf{0})|\, I_{t-1};\boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta}), \\
\mathcal{B}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{B}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\
\mathcal{B}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\phi}) &= V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta};\mathbf{0})|\, I_{t-1};\boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\boldsymbol{\varrho})\mathbf{Z}'_{dt}(\boldsymbol{\theta}),
\end{aligned}$$

   and

$$\mathcal{K}(\boldsymbol{\varrho}) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\, I_{t-1};\boldsymbol{\phi}] = \begin{bmatrix} \mathbf{I}_N & \boldsymbol{\Phi}(\boldsymbol{\varrho}) \\ \boldsymbol{\Phi}'(\boldsymbol{\varrho}) & \boldsymbol{\Upsilon}(\boldsymbol{\varrho}) \end{bmatrix}, \tag{C22}$$

   where

$$\begin{aligned}
\boldsymbol{\Phi}(\boldsymbol{\varrho}) &= E[\boldsymbol{\varepsilon}_t^* vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\,\boldsymbol{\phi}] \\
\boldsymbol{\Upsilon}(\boldsymbol{\varrho}) &= E[vec(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}_N)vec'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}_N)|\,\boldsymbol{\phi}]
\end{aligned}$$

   depend on the multivariate third and fourth order cumulants of $\boldsymbol{\varepsilon}_t^*$, *so that* $\boldsymbol{\Phi}(\mathbf{0}) = \mathbf{0}$ *and* $\boldsymbol{\Upsilon}(\mathbf{0}) = (\mathbf{I}_{N^2} + \mathbf{K}_{NN})$ *if we use* $\boldsymbol{\varrho} = \mathbf{0}$ *to denote normality.*

2. *If* $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}_0$ *is i.i.d.* $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta}_0)$ *with* $\kappa_0 < \infty$, *then (C22) reduces to*

$$\mathcal{K}(\kappa) = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & (\kappa+1)(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{bmatrix}, \tag{C23}$$

   which only depends on the true distribution through the population coefficient of multivariate excess kurtosis

$$\kappa = E(\varsigma_t^2|\boldsymbol{\eta})/[N(N+2)] - 1. \tag{C24}$$

Proof. The proof of the first part is based on a straightforward application of Proposition 1 in Bollerslev and Wooldridge (1992) to the *i.i.d.* case. Since $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\mathbf{0}) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, and $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ is a vector martingale difference sequence, then to obtain $\mathcal{B}_t(\boldsymbol{\phi}_0)$ we only need to compute $V[\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})|I_{t-1};\boldsymbol{\phi}_0]$, which justifies (C22). Further, we will have that

$$\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta}_0,\mathbf{0})\\ \mathbf{e}_{st}(\boldsymbol{\theta}_0,\mathbf{0})\end{array}\right]=\left(\begin{array}{c}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)-\mathbf{I}_N\right]\end{array}\right)=\left[\begin{array}{c}\sqrt{\varsigma_t}\mathbf{u}_t\\ vec(\varsigma_t\mathbf{u}_t\mathbf{u}_t'-\mathbf{I}_N)\end{array}\right]$$

for any spherical distribution, with ς_t and $\mathbf{u}_t$ both mutually and serially independent. Then (C23) follows from (C1) and (C2). As for $\mathcal{A}_t(\boldsymbol{\phi}_0)$, we know that its formula, which is valid regardless of the exact nature of the true conditional distribution, coincides with the expression for $\mathcal{B}_t(\boldsymbol{\phi}_0)$ under multivariate normality by the (conditional) information matrix equality. □

C.5 Spherically symmetric semiparametric estimators

As is well known, a single scoring iteration without line searches that started from $\tilde{\boldsymbol{\theta}}_T$ and some root-T consistent estimator of $\boldsymbol{\eta}$, say $\tilde{\boldsymbol{\eta}}_T$, would suffice to yield an estimator of $\boldsymbol{\phi}$ that would be asymptotically equivalent to the full-information ML estimator $\hat{\boldsymbol{\phi}}_T$, at least up to terms of order $O_p(T^{-1/2})$. Specifically,

$$\left(\begin{array}{c}\check{\boldsymbol{\theta}}_T-\tilde{\boldsymbol{\theta}}_T\\ \check{\boldsymbol{\eta}}_T-\tilde{\boldsymbol{\eta}}_T\end{array}\right)=\left[\begin{array}{cc}\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\\ \mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta\eta}}(\boldsymbol{\phi}_0)\end{array}\right]^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T)\\ \mathbf{s}_{\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T)\end{array}\right].$$

If we use the partitioned inverse formula, then it is easy to see that

$$\check{\boldsymbol{\theta}}_T-\tilde{\boldsymbol{\theta}}_T=\left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)-\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\theta\eta}}'(\boldsymbol{\phi}_0)\right]^{-1}$$
$$\times\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T)-\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T)\right]=\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T,\tilde{\boldsymbol{\eta}}_T),$$

where

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

and

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)=\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)-\mathcal{I}_{\boldsymbol{\theta\eta}}(\boldsymbol{\phi}_0)\mathcal{I}_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\tag{C25}$$

is the residual from the unconditional theoretical regression of the score corresponding to $\boldsymbol{\theta}$, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)$, on the score corresponding to $\boldsymbol{\eta}$, $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$. This residual score is sometimes called the unrestricted parametric efficient score of $\boldsymbol{\theta}$, and its covariance matrix, $\mathcal{P}(\boldsymbol{\phi}_0)=[\mathcal{I}^{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)]^{-1}$, the marginal information matrix of $\boldsymbol{\theta}$, or the unrestricted parametric efficiency bound.

In the spherically symmetric case, we can easily prove that (C25) and its covariance matrix reduce to

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)=\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathbf{e}_{rt}(\boldsymbol{\phi}_0)\right]\tag{C26}$$

and

$$\mathcal{P}(\boldsymbol{\phi}_0)=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0)\cdot\left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathrm{M}_{sr}'(\boldsymbol{\eta}_0)\right],\tag{C27}$$

respectively, where

$$\mathbf{W}_s(\phi_0) = \mathbf{Z}_d(\boldsymbol{\theta}_0)[\mathbf{0}', vec'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\phi_0][\mathbf{0}', vec'(\mathbf{I}_N)]'$$
$$= E\left\{ \frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial \boldsymbol{\theta}} vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)] \right| \phi_0 \bigg\} = E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\phi_0] = -E\left[\frac{\partial d_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right| \phi_0 \bigg], \quad \text{(C28)}$$

It is worth noting that the last summand of (C25) coincides with $\mathbf{Z}_d(\phi_0)$ times the theoretical least squares projection of $\mathbf{e}_{dt}(\phi_0)$ on (the linear span of) $\mathbf{e}_{rt}(\phi_0)$, which is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ from Proposition 3 of Fiorentini and Sentana (2007). Such an interpretation immediately suggests alternative estimators of $\boldsymbol{\theta}$ that replace a parametric assumption on the shape of the distribution of the standardised innovations ε_t^* by a more flexible alternative. Specifically, Hodgson and Vorkink (2003), Hafner and Rombouts (2007) and other authors have suggested spherically symmetric semiparametric estimators which allow for any member of the class of spherically symmetric distribution. To derive such estimators, these authors replace the linear span of $\mathbf{e}_{rt}(\phi_0)$ by the so-called spherically symmetric tangent set, which is the Hilbert space generated by all time-invariant functions of $\varsigma_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$. The next proposition, which originally appeared as Proposition 7 in Fiorentini and Sentana (2007), provides the resulting spherically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition C3 *When $\boldsymbol{\varepsilon}_t^*|I_{t-1}, \phi$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$ with $-2/(N+2) < \kappa_0 < \infty$, the spherically symmetric semiparametric efficient score is given by:*

$$\mathring{\mathbf{s}}_{\theta t}(\phi_0) = \mathbf{s}_{\theta t}(\phi_0) - \mathbf{W}_s(\phi_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right] - \frac{2}{(N+2)\kappa_0 + 2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right]\right\}, \quad \text{(C29)}$$

while the spherically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\phi_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\phi_0) - \mathbf{W}_s(\phi_0)\mathbf{W}_s'(\phi_0) \cdot \left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]}\right\}. \quad \text{(C30)}$$

Proof. First of all, it is easy to show that for any spherical distribution

$$\begin{aligned}
\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0, \mathbf{0}) &= E\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\theta}_0, \mathbf{0}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}_0, \mathbf{0}) \end{array} \middle| \varsigma_t; \phi_0\right] = E\left\{\begin{array}{c} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right] \end{array} \middle| \varsigma_t; \phi_0\right\} \\
&= E\left[\begin{array}{c} \sqrt{\varsigma_t}\mathbf{u}_t \\ vec(\varsigma_t\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N) \end{array} \middle| \varsigma_t\right] = \left(\frac{\varsigma_t}{N} - 1\right)\left[\begin{array}{c} \mathbf{0} \\ vec(\mathbf{I}_N) \end{array}\right],
\end{aligned} \quad \text{(C31)}$$

and

$$\mathring{\mathbf{e}}_{dt}(\phi_0) = E\left[\begin{array}{c} \mathbf{e}_{lt}(\phi_0) \\ \mathbf{e}_{st}(\phi_0) \end{array} \middle| \varsigma_t; \phi_0\right]$$
$$= E\left\{\begin{array}{c} \delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\ vec\left[\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0] \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right] \end{array} \middle| \varsigma_t; \phi_0\right\}$$
$$= E\left\{\begin{array}{c} \delta(\varsigma_t, \boldsymbol{\eta}_0)\sqrt{\varsigma_t}\mathbf{u}_t \\ vec[\delta(\varsigma_t, \boldsymbol{\eta}_0)\varsigma_t\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N] \end{array} \middle| \varsigma_t\right\} = \left[\delta(\varsigma_t, \boldsymbol{\eta}_0)\frac{\varsigma_t}{N}\text{-}1\right]\left[\begin{array}{c} \mathbf{0} \\ vec(\mathbf{I}_N) \end{array}\right], \quad \text{(C32)}$$

where we have used again the fact that $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$, and ς_t and $\mathbf{u}_t$ are stochastically independent.

In addition, we can use the law of iterated expectations to show that

$$E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\phi})|\boldsymbol{\phi}]=E\{E[[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\phi})|\varsigma_t,\boldsymbol{\phi}]|\boldsymbol{\phi}\}=E[\mathbf{e}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi})|\boldsymbol{\phi}]=E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\phi})|\boldsymbol{\phi}],$$
$$E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]=E\{E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\varsigma_t,\boldsymbol{\phi}]|\boldsymbol{\phi}\}=E[\mathbf{e}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]=E[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}]$$

and

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},0)|\boldsymbol{\phi}\right] = E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right].$$

Hence, to compute these matrices we simply need three scalar moments.

In this respect, we can use (C24) to show that

$$E\left[\left(\frac{\varsigma_t}{N}-1\right)^2\middle|\boldsymbol{\eta}\right] = \frac{(N+2)\kappa+2}{N}, \tag{C33}$$

so that

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = \frac{(N+2)\kappa+2}{N}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{pmatrix} = \mathring{\mathcal{K}}(\kappa).$$

We can also use Lemma 2 in Supplemental Appendix B to show that $E(\varsigma_t^2) = N(N+2)(\kappa+1) < \infty$ implies

$$E\left[\varsigma_t^2\cdot\partial\ln h(\varsigma_t;\boldsymbol{\eta})/\partial\varsigma\middle|\,\boldsymbol{\eta}\right] = -E\left[2\varsigma_t|\boldsymbol{\eta}\right] = -2N.$$

If we then combine this result with (C20) and (C21), we will have that for any spherically symmetric distribution

$$E\left\{\left(\frac{\varsigma_t}{N}-1\right)\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]\middle|\,\boldsymbol{\eta}\right\} = \frac{2}{N}, \tag{C34}$$

so that

$$E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = \mathring{\mathcal{K}}(0),$$

which coincides with the value of $E\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right]$ under normality.

Finally, Proposition C1 immediately implies that

$$E\left\{\left[\delta(\varsigma_t,\boldsymbol{\eta}_0)\frac{\varsigma_t}{N}-1\right]^2\middle|\,\boldsymbol{\eta}\right\} = \frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta})-1. \tag{C35}$$

Therefore, it trivially follows from the expressions for $\mathring{\mathcal{K}}(0)$ and $\mathring{\mathcal{K}}(\kappa_0)$ above that

$$\begin{aligned} &E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})\middle|\,I_{t-1};\boldsymbol{\phi}\right\} \\ &= E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathring{\mathbf{e}}_{dt}'(\boldsymbol{\theta},\mathbf{0})\middle|\,I_{t-1};\boldsymbol{\phi}\right\} = \mathbf{0} \end{aligned}$$

for any spherically symmetric distribution. In addition, we also know that

$$E\left\{\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})-\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\middle|\,I_{t-1};\boldsymbol{\phi}\right\} = \mathbf{0}.$$

Thus, even though $\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}_0) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa_0)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})\right]$ is the residual from the theoretical regression of $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi})$ on a constant and $\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})$, it turns out that the second summand of (C29) belongs to the restricted tangent set, which is the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varsigma}_t(\boldsymbol{\theta}_0)$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$.

Now, if write (C29) as

$$\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_{d}(\boldsymbol{\phi})\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) + \mathbf{Z}_{d}(\boldsymbol{\phi})\mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0}),$$

then we can use the law of iterated expectations to show that the spherically symmetric semiparametric efficient score is indeed unconditionally orthogonal to the restricted tangent set.

Finally, the expression for the semiparametric efficiency bound will be

$$\begin{aligned}
E[\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\mathring{\mathbf{s}}'_{\boldsymbol{\theta}t}(\boldsymbol{\phi})|\boldsymbol{\phi}] &= E\left[\begin{array}{c} \left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_{d}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\right\} \\ \times\left\{\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi})\right\} \end{array}\right| \boldsymbol{\phi}\right] \\
&= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right] \\
&\quad -E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\right\} \\
&\quad -E\left\{\mathbf{Z}_{d}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}'_{dt}(\boldsymbol{\phi})\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\right\} \\
&\quad +E\left\{\mathbf{Z}_{d}(\boldsymbol{\phi})\left[\mathring{\mathbf{e}}_{dt}(\boldsymbol{\phi}) - \mathring{\mathcal{K}}(0)\,\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathbf{e}}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\left[\mathring{\mathbf{e}}'_{dt}(\boldsymbol{\phi}) - \mathring{\mathbf{e}}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathring{\mathcal{K}}^{+}(\kappa)\,\mathring{\mathcal{K}}(0)\right]\mathbf{Z}'_{d}(\boldsymbol{\phi})|\boldsymbol{\phi}\right\} \\
&= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}'_s(\boldsymbol{\phi}_0)\cdot\left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}) - 1\right] - \frac{4}{N\left[(N+2)\kappa+2\right]}\right\}
\end{aligned}$$

by virtue of the law of iterated expectations. □

In the case of the univariate GARCH-M model (19), we estimate the model parameters using reparametrisation 1 in section 4. Specifically,

$$\mathbf{Z}_{lt}(\boldsymbol{\vartheta}) = \frac{\partial\mu_t(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}}{\vartheta_i^{1/2}\sigma_t^{\circ}(\boldsymbol{\vartheta}_c)} = \frac{1}{\vartheta_i^{1/2}\sigma_t^{\circ}(\boldsymbol{\vartheta}_c)}\left[\begin{array}{c} \sigma_t^{\circ}(\boldsymbol{\vartheta}_c)\frac{\partial\delta}{\partial\boldsymbol{\vartheta}_c} + \frac{\delta}{2\sigma_t^{\circ}(\boldsymbol{\vartheta}_c)}\frac{\partial\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}{\partial\boldsymbol{\vartheta}_c} \\ 0 \end{array}\right] = \frac{1}{\vartheta_i^{1/2}}\left[\begin{array}{c} \frac{\partial\delta}{\partial\boldsymbol{\vartheta}_c} + \delta W_{st}(\boldsymbol{\vartheta}_c) \\ 0 \end{array}\right],$$

$$\mathbf{Z}_{st}(\boldsymbol{\vartheta}) = \frac{\partial\sigma_t^2(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}}{2\vartheta_{ic}\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)} = \frac{1}{2\vartheta_i\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}\left[\begin{array}{c} \vartheta_i\frac{\partial\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}{\partial\boldsymbol{\vartheta}_c} \\ \sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c) \end{array}\right] = \left[\begin{array}{c} W_{st}(\boldsymbol{\vartheta}_c) \\ \frac{1}{2}\vartheta_i^{-1} \end{array}\right],$$

$$W_{st}(\boldsymbol{\vartheta}_c) = \frac{1}{2\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}\frac{\partial\sigma_t^{\circ 2}(\boldsymbol{\vartheta}_c)}{\partial\boldsymbol{\vartheta}_c}$$

and

$$\varsigma_t(\boldsymbol{\vartheta}) = \varepsilon_t^{*2}(\boldsymbol{\vartheta}) = \vartheta_i^{-1}\sigma_t^{\circ -2}(\boldsymbol{\vartheta}_c)x_t^2.$$

On the other hand, we use the natural parametrisation of the multivariate market model in (20), so that $\boldsymbol{\theta}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')$, where $\boldsymbol{\omega} = vech(\boldsymbol{\Omega})$. Given the Jacobian matrices:

$$\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial(\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')} = \left(\begin{array}{ccc} \mathbf{I}_N & \mathbf{I}_N r_{Mt} & \mathbf{0} \end{array}\right), \tag{C36}$$

$$\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial(\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')} = \left(\begin{array}{ccc} \mathbf{0} & \mathbf{0} & \mathbf{D}_N \end{array}\right), \tag{C37}$$

because $\partial vec(\mathbf{\Omega})/\partial vech'(\mathbf{\Omega})$ is the duplication matrix of order N (see Magnus and Neudecker, 1988), a direct application of (C4) immediately implies that

$$\begin{aligned}
\mathbf{s}_{\mathbf{a}t}(\boldsymbol{\theta}) &= \mathbf{\Omega}^{-1}\delta_t\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \\
\mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta}) &= \mathbf{\Omega}^{-1}r_{mt}\delta_t\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \\
\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta}) &= \frac{1}{2}\mathbf{D}_N'(\mathbf{\Omega}^{-1}\otimes\mathbf{\Omega}^{-1})vec[\delta_t\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})-\mathbf{\Omega}],
\end{aligned}$$

where $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{r}_t - \mathbf{a} - \mathbf{b}r_{mt}$.

The last ingredient we need is

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = [\mathbf{0}, \mathbf{0}, \frac{1}{2}vec'(\mathbf{\Omega}^{-1})\mathbf{D}_N]'$$

because

$$\mathbf{D}_N'(\mathbf{\Omega}^{-\frac{1}{2}\prime}\otimes\mathbf{\Omega}^{-\frac{1}{2}\prime})vec(\mathbf{I}_N) = \mathbf{D}_N'vec(\mathbf{\Omega}^{-1}).$$

In practice, $\mathbf{e}_{dt}(\boldsymbol{\phi})$ has to be replaced by a semiparametric estimate obtained from the joint density of $\boldsymbol{\varepsilon}_t^*$. However, the spherical symmetry assumption allows us to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h(\varsigma_t;\boldsymbol{\eta})$, avoiding in this way the curse of dimensionality. Specifically, if we use expression (C19), then we can estimate $\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]$ non-parametrically by exploiting that

$$-\frac{2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma} = -\frac{2\partial\ln h[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma} + \frac{N-2}{2}\frac{1}{\varsigma_t(\boldsymbol{\theta})}.$$

We can compute $h[\varsigma_t(\boldsymbol{\theta});\boldsymbol{\eta}]$ either directly by using a kernel for positive random variables (see Chen (2000)), or indirectly by using a faster standard Gaussian kernel after exploiting the Box-Cox-type transformation $v = \varsigma^k$ (see Hodgson, Linton and Vorkink (2002)). In the second case, the usual change of variable formula yields

$$p(v;\boldsymbol{\eta}) = \frac{\pi^{N/2}}{k\Gamma(N/2)}v^{-1+N/2k}\exp[c(\boldsymbol{\eta}) + g(v^{1/k};\boldsymbol{\eta})],$$

whence

$$g(v^{1/k};\boldsymbol{\eta}) = \ln p(v;\boldsymbol{\eta}) + \left(1-\frac{N}{2k}\right)\ln v - \frac{N}{2}\ln 2\pi + \ln k - \ln\Gamma(N/2) - c(\boldsymbol{\eta})$$

and

$$\frac{\partial g(v^{1/k};\boldsymbol{\eta})}{\partial v^{1/k}} = k\frac{\partial\ln f(v;\boldsymbol{\eta})}{\partial v}v^{1-1/k} + \frac{k-N/2}{v^{1/k}}.$$

We use the second procedure in our Monte Carlo simulations because the distribution of $\varsigma_t(\boldsymbol{\theta})$ becomes more normal-like as N increases, which reduces the advantages of using kernels for positive variables. Specifically, we use a cubic root transformation to improve the approximation, with a common bandwidth parameter for both the density and its first derivative. Given that a proper cross-validation procedure is extremely costly to implement in a Monte Carlo exercise, we have done some experimentation to choose the optimal bandwidth by scaling up and down the automatic choices given in Silverman (1986).

In the univariate case, there is a conceptually simpler alternative that does not require working with $\varsigma_t = \varepsilon_t^{*2}$. In particular, we can exploit the fact that the density of ε_t^* is the same as the density of $-\varepsilon_t^*$ by assigning to $\pm\varepsilon_t^*$ the equally weighted average of the non-parametric density estimates at ε_t^* and $-\varepsilon_t^*$. Likewise, we can compute the equally weighted average of the absolute value of its derivatives and assign its $\pm$ value to ε_t^* and $-\varepsilon_t^*$, respectively.

D The general case of non-spherical distributions

D.1 Likelihood, score and Hessian for non-spherical distributions

In this section, we assume that, conditional on I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed, or $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0 \sim i.i.d.\ D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho}_0)$ for short, where $\boldsymbol{\varrho}$ are some q additional parameters that determine the shape of the distribution. Importantly, this distribution could substantially depart from a multivariate normal both in terms of skewness and kurtosis. Let $f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ denote the assumed conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and those shape parameters $\boldsymbol{\varrho}$, which we assume is well defined. Let also $\boldsymbol{\phi} = (\boldsymbol{\theta}',\boldsymbol{\varrho})'$ denote the $p+q$ parameters of interest, which once again 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 $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, where $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}),\boldsymbol{\varrho}\right]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})$.

The most common choices of square root matrices are the Cholesky decomposition, which leads to a lower triangular matrix for a given ordering of $\mathbf{y}_t$, or the spectral decomposition, which yields a symmetric matrix. The choice of square root matrix is non-trivial because $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ affects the value of the log-likelihood function and its score in multivariate non-spherical contexts. In what follows, we rely mostly on the Cholesky decomposition because it is much faster to compute than the spectral one, especially when $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is time-varying. Nevertheless, we also discuss some modifications required for the spectral decomposition later on.

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 $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\varrho}$, respectively. Assuming that $\boldsymbol{\mu}_t(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ and $\ln f(\boldsymbol{\varepsilon}^*,\boldsymbol{\varrho})$ are differentiable, it trivially follows that

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\boldsymbol{\varrho}) = \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}.$$

But since

$$\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = -\frac{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}vec[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N)$$

and

$$\begin{aligned}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= -\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})]\frac{\partial vec[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} \\ &= -\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\},\end{aligned} \tag{D1}$$

where

$$\left.\begin{array}{l}\mathbf{Z}_{lt}(\boldsymbol{\theta}) = \partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) \\ \mathbf{Z}_{st}(\boldsymbol{\theta}) = \partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})]\end{array}\right\}, \tag{D2}$$

it follows that

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= [\mathbf{Z}_{lt}(\boldsymbol{\theta}),\mathbf{Z}_{st}(\boldsymbol{\theta})]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) &= \partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho} = \mathbf{e}_{rt}(\boldsymbol{\phi}),\end{aligned} \tag{D3}$$

with

$$\mathbf{e}_{dt}(\boldsymbol{\phi}) = \left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right] = \left[\begin{array}{c}-\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*, \\ -vec\left\{\mathbf{I}_N + \partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\end{array}\right]. \tag{D4}$$

Similarly, let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. Assuming twice differentiability of the different functions involved, expression (D1) implies that

$$\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'} = -\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = \frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \tag{D5}$$

because

$$d\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho}) = -d\{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\}. \tag{D6}$$

In turn,

$$\begin{aligned}d\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho}) &= -dvec\left[\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right] \\ &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]d\left\{\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right\} - \left\{\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right\}d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\end{aligned} \tag{D7}$$

implies that

$$\begin{aligned}\frac{\partial\mathbf{e}_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} &= \frac{\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'} = -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - \left\{\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right\}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \\ &= \left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N\otimes\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\right]\right\}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}.\end{aligned} \tag{D8}$$

Finally, (D6) and (D7) trivially imply that

$$\begin{aligned}\frac{\partial^2\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\varrho}'} &= -\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}, \\ \frac{\partial^2\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\varrho}'} &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}.\end{aligned}$$

Using these results, we can easily obtained the required expressions for

$$\begin{aligned}\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} \\ &\quad + \left[\mathbf{e}_{lt}'(\boldsymbol{\phi})\otimes\mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi})\otimes\mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}, & \text{(D9)} \\ \mathbf{h}_{\boldsymbol{\theta\varrho}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\partial\mathbf{e}_{lt}(\boldsymbol{\phi})/\partial\boldsymbol{\varrho}' + \mathbf{Z}_{st}(\boldsymbol{\theta})\partial\mathbf{e}_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\varrho}', & \text{(D10)} \\ \mathbf{h}_{\boldsymbol{\varrho\varrho}t}(\boldsymbol{\phi}) &= \partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'.\end{aligned}$$

In this regard, note that since (D6) and (D7) also imply that

$$\begin{aligned}\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}' &= -\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}', &\text{(D11)}\\ \partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}' &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}', &\text{(D12)}\end{aligned}$$

respectively, it is clear that

$$\begin{aligned}\mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\varrho}'}+\mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\varrho}'} &= -\{\mathbf{Z}_{lt}(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N]\}\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}\\ &= \frac{\partial\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho})}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'}\end{aligned}$$

so both ways of computing $\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi})$ indeed coincide.

Importantly, while $\mathbf{Z}_{lt}(\boldsymbol{\theta})$, $\mathbf{Z}_{st}(\boldsymbol{\theta})$, $\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$ and $\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}'$ depend on the dynamic model specification, the first and second derivatives of $\ln f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ depend on the specific distribution assumed for estimation purposes.

For the standard (i.e. lower triangular) Cholesky decomposition of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, we will have that

$$dvec(\boldsymbol{\Sigma}_t)=[(\boldsymbol{\Sigma}_t^{1/2}\otimes\mathbf{I}_N)+(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}]dvec(\boldsymbol{\Sigma}_t^{1/2}).$$

Unfortunately, this transformation is singular, which means that we must find an analogous transformation between the corresponding $dvech's$. In this sense, we can write the previous expression as

$$dvech(\boldsymbol{\Sigma}_t)=[\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes\mathbf{I}_N)\mathbf{L}_N'+\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N']dvech(\boldsymbol{\Sigma}_t^{1/2}), \tag{D13}$$

where $\mathbf{L}_N$ is the elimination matrix (see Magnus, 1988). We can then use the results in chapter 5 of Magnus (1988) to show that the above mapping will be lower triangular of full rank as long as $\boldsymbol{\Sigma}_t^{1/2}$ has full rank, which means that we can readily obtain the Jacobian matrix of $vech(\boldsymbol{\Sigma}_t^{1/2})$ from the Jacobian matrix of $vech(\boldsymbol{\Sigma}_t)$.

In the case of the symmetric square root matrix, the analogous transformation would be

$$dvech(\boldsymbol{\Sigma}_t)=[\mathbf{D}_N^+(\boldsymbol{\Sigma}_t^{1/2}\otimes\mathbf{I}_N)\mathbf{D}_N+\mathbf{D}_N^+(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{D}_N]dvech(\boldsymbol{\Sigma}_t^{1/2}),$$

where $\mathbf{D}_N^+=(\mathbf{D}_N'\mathbf{D}_N)^{-1}\mathbf{D}_N'$ is the Moore-Penrose inverse of the duplication matrix (see Magnus and Neudecker, 1988).

From a numerical point of view, the calculation of both $\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes\mathbf{I}_N)\mathbf{L}_N'$ and $\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$ is straightforward. Specifically, given that $\mathbf{L}_N vec(\mathbf{A})=vech(\mathbf{A})$ for any square matrix $\mathbf{A}$, the effect of premultiplying by the $\frac{1}{2}N(N+1)\times N^2$ matrix $\mathbf{L}_N$ is to eliminate rows N+1, 2N+1 and 2N+2, 3N+1, 3N+2 and 3N+3, etc. Similarly, given that $\mathbf{L}_N\mathbf{K}_{NN}vec(\mathbf{A})=vech(\mathbf{A}')$, the effect of postmultiplying by $\mathbf{K}_{NN}\mathbf{L}_N'$ is to delete all columns but those in positions 1, N+1, 2N+1,. . . ,N+2, 2N+2,. . . , N+3, 2N+3,. . . , N^2.

Let $\mathbf{F}_t$ denote the transpose of the inverse of $\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes\mathbf{I}_N)\mathbf{L}_N'+\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$, which will be upper triangular. The fastest way to compute

$$\frac{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})]=\frac{1}{2}\frac{\partial vech'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}\mathbf{F}_t\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2}) \tag{D14}$$

is as follows:

1. From the expression for $\partial vec'[\mathbf{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$ we can readily obtain $\partial vech'[\mathbf{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$ by simply avoiding the computation of the duplicated columns

2. Then we postmultiply the resulting matrix by $\mathbf{F}_t$

3. Next, we construct the matrix

$$\mathbf{L}_N(\mathbf{I}_N \otimes \mathbf{\Sigma}_t^{1/2}) = \mathbf{L}_N \begin{pmatrix} \mathbf{\Sigma}_t^{-1/2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_t^{-1/2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma}_t^{-1/2} \end{pmatrix}$$

   by eliminating the first row from the second block, the first two rows from the third block, ..., and all the rows but the last one from the last block

4. Finally, we premultiply the resulting matrix by $\partial vech'[\mathbf{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\cdot\mathbf{F}_t$.

D.2 Asymptotic distribution

Propositions 10.1, 13, C2.1 and D3 already deal explicitly with the general case, so there is no need to generalise them. In turn, Propositions 6, 7, 8, 9 and their proofs continue to be valid if we change $\boldsymbol{\eta}$ by $\boldsymbol{\varrho}$. The same happens to Proposition 5, provided we erase the row and columns corresponding to $\mathring{\boldsymbol{\theta}}_T$ and its influence function $\mathring{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$. On the other hand, Propositions 10.2, 11, 12, C2.2 and C3 are specific to the spherically symmetric case. Therefore, the only proposition that really requires a proper generalisation is Proposition C1.

Proposition D1 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ with density $f(\boldsymbol{\varepsilon}^*,\boldsymbol{\varrho})$, then*

$$\begin{aligned}\mathcal{I}_t(\boldsymbol{\phi}) &= \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\varrho})\mathbf{Z}_t'(\boldsymbol{\theta}),\\ \mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix},\end{aligned}$$

and

$$\mathcal{M}(\boldsymbol{\varrho}) = \begin{bmatrix} \mathcal{M}_{dd}(\boldsymbol{\varrho}) & \mathcal{M}_{dr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{dr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix} = \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{lr}'(\boldsymbol{\varrho}) & \mathcal{M}_{sr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix},$$

with

$$\begin{aligned}
\mathcal{M}_{ll}(\boldsymbol{\varrho}) &= V[\mathbf{e}_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{ls}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{st}(\boldsymbol{\phi})'|\boldsymbol{\phi}] = E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t^{\prime *}\otimes\mathbf{I}_N)\right|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{ss}(\boldsymbol{\varrho}) &= V[\mathbf{e}_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\cdot\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot(\boldsymbol{\varepsilon}_t^{*\prime}\otimes\mathbf{I}_N)|\boldsymbol{\varrho}\right] - \mathbf{K}_{NN},\\
\mathcal{M}_{lr}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],\\
\mathcal{M}_{sr}(\boldsymbol{\varrho}) &= E[\mathbf{e}_{st}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[(\boldsymbol{\varepsilon}_t^*\otimes\mathbf{I}_N)\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\varrho}\right],
\end{aligned}$$

and

$$\mathcal{M}_{rr}(\boldsymbol{\varrho}) = V[\mathbf{e}_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\right].$$

Proof. Since the distribution of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} is assumed to be *i.i.d.*, then it is easy to see from (D3) that $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}'_{dt}(\boldsymbol{\phi}), \mathbf{e}'_{rt}(\boldsymbol{\phi})]'$ will inherit the martingale difference property of the score $\mathbf{s}_t(\boldsymbol{\phi}_0)$. As a result, the conditional information matrix will be given by

$$\begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{lr}(\boldsymbol{\varrho}) & \mathcal{M}'_{sr}(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix} \begin{bmatrix} \mathbf{Z}'_{lt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ll}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}'_{ls}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{ls}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}_{ss}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) \\ \mathcal{M}'_{lr}(\boldsymbol{\varrho})\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathcal{M}'_{sr}(\boldsymbol{\varrho})\mathbf{Z}'_{st}(\boldsymbol{\theta}) \end{bmatrix.}$$
$$\begin{matrix} \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{M}_{lr}(\boldsymbol{\varrho}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{matrix} \Bigg],$$

where
$$\begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{lr}(\boldsymbol{\varrho}) & \mathcal{M}'_{sr}(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix} = V \left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \\ \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \\ \mathbf{e}_{rt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \end{array} \right| \boldsymbol{\theta}, \boldsymbol{\varrho} \Bigg],$$

which confirms the variance of the score part of the proposition.

As for the expected value of the Hessian expressions, it is easy to see that

$$E[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi})|z_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{lt}(\boldsymbol{\theta}) E\left[\left. \frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right] + \mathbf{Z}_{st}(\boldsymbol{\theta}) E\left[\left. \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right]$$

because

$$E\left[\mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})|z_t, I_{t-1}; \boldsymbol{\phi}\right] = -E\left[\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* | z_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{0} \tag{D15}$$

and

$$E\left[\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho})|z_t, I_{t-1}; \boldsymbol{\phi}\right] = -E\left[vec\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\}|z_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{0}. \tag{D16}$$

Expression (D5) then leads to

$$E\left[\left. \frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right] = E\left[\left. \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right]$$
$$= E\left[\left. \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \right| \boldsymbol{\phi} \right] \mathbf{Z}'_{lt}(\boldsymbol{\theta}) + E\left[\left. \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \right| \boldsymbol{\phi} \right] \mathbf{Z}'_{st}(\boldsymbol{\theta}).$$

Likewise, equation (D8) leads to

$$E\left[\left. \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} \right| z_t, I_{t-1}; \boldsymbol{\phi} \right] = E\left[\left\{ [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right] \right\} \right.$$
$$\left. \times \{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\} \right| z_t, I_{t-1}; \boldsymbol{\phi} \Big] = E\left[\left. [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \right| \boldsymbol{\phi} \right] \mathbf{Z}'_{lt}(\boldsymbol{\theta})$$
$$+ E\left[\left. [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} [\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \right| z_t, I_{t-1}; \boldsymbol{\phi} \right] \mathbf{Z}'_{st}(\boldsymbol{\theta}) - \mathbf{K}_{NN}\mathbf{Z}'_{st}(\boldsymbol{\theta})$$

because of (D15) and (D16), which in turn implies

$$
\begin{aligned}
& E\left\{\left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\Bigg\} \\
= \; & \mathbf{K}_{NN} E\left\{\mathbf{K}_{NN}\left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\Bigg\} \\
= \; & \mathbf{K}_{NN} E\left\{\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \otimes \mathbf{I}_N\right][\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right| z_t, I_{t-1};\boldsymbol{\phi}\Bigg\} \\
= \; & \mathbf{K}_{NN} E\left\{\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N\right]\right| z_t, I_{t-1};\boldsymbol{\phi}\Bigg\} = -\mathbf{K}_{NN}
\end{aligned}
$$

in view of Theorem 3.1 in Magnus (1988).

As a result, the information matrix equality implies that

$$
\begin{aligned}
\mathcal{M}_{ll}(\boldsymbol{\varrho}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\right|\boldsymbol{\phi}\Big\} \\
\mathcal{M}_{ls}(\boldsymbol{\varrho}) &= E\left\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right|\boldsymbol{\phi}\Big\} \\
\mathcal{M}_{ss}(\boldsymbol{\varrho}) &= E\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}\cdot[\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\right|\boldsymbol{\phi}\Big\} - \mathbf{K}_{NN}
\end{aligned}
$$

Similarly, equation (D10) implies that

$$
E[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi})|z_t, I_{t-1};\boldsymbol{\phi}] = E[\mathbf{Z}_{lt}(\boldsymbol{\theta})\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}' + \mathbf{Z}_{st}(\boldsymbol{\theta})\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t, I_{t-1};\boldsymbol{\phi}].
$$

But then the information matrix equality together with equations (D11) and (D12) imply that

$$
\begin{aligned}
E[\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t, I_{t-1};\boldsymbol{\phi}] &= -E\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\} = \mathcal{M}_{lr}(\boldsymbol{\varrho}), \\
E[\partial\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\varrho})/\partial\boldsymbol{\varrho}'|z_t, I_{t-1};\boldsymbol{\phi}] &= -E\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\} = \mathcal{M}_{sr}(\boldsymbol{\varrho}).
\end{aligned}
$$

Finally, the information matrix equality also implies that

$$
\mathcal{M}_{rr}(\boldsymbol{\varrho}) = -E\{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'|\boldsymbol{\phi}\},
$$

as required. □

D.3 Cross-sectionally independent disturbances

Let us now specialise the results in the previous two subsections for the case in which the disturbances are cross-sectionally independent. Specifically, we assume that the conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and the shape parameters $\boldsymbol{\varrho}$ can be factorised as

$$
\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}),\boldsymbol{\varrho}] = \sum_{i=1}^N \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}),\boldsymbol{\varrho}_i],
$$

where $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = [\varepsilon_{1t}^*(\boldsymbol{\theta}),\ldots,\varepsilon_{Nt}^*(\boldsymbol{\theta})]'$ and $\boldsymbol{\varrho} = (\boldsymbol{\varrho}_1,\ldots,\boldsymbol{\varrho}_N)$, with $\dim(\boldsymbol{\varrho}_i) = q_i$ and $\sum_{i=1}^N q_i = q$.

The main simplification in the expressions for the scores result from the fact that

$$
\mathbf{e}_{lt}(\boldsymbol{\phi}) = \left\{\begin{array}{c} -\frac{\partial f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial\varepsilon_1^*} \\ \vdots \\ -\frac{\partial f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial\varepsilon_N^*} \end{array}\right\},
$$

$$\mathbf{e}_{st}(\boldsymbol{\phi})=-vec\left\{\begin{array}{ccc} 1+\frac{\partial \ln f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & 1+\frac{\partial \ln f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \end{array}\right\}$$

and

$$\mathbf{e}_{rt}(\boldsymbol{\phi})=\left\{\begin{array}{c} \frac{\partial \ln f[\varepsilon_{1t}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1} \\ \vdots \\ \frac{\partial \ln f[\varepsilon_{Nt}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N} \end{array}\right\},$$

so that the derivatives involved correspond to the underlying univariate densities.

When any of the N distributions is symmetric, then these expressions simplify further as

$$-\frac{\partial f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}=\delta(\varepsilon_{it}^{*2};\boldsymbol{\varrho}_i)\varepsilon_{it}^*.$$

Additional simplifications in the expressions for the Hessian arise because $\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}$, $\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}'$ and $\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'$ are (block) diagonal matrices with representative elements $\partial^2 \ln f[\varepsilon_{it}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_i]/\partial\varepsilon_i^*\partial\varepsilon_i^*$, $\partial^2 \ln f[\varepsilon_{it}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_i]/\partial\varepsilon_i^*\partial\boldsymbol{\varrho}_i'$ and $\partial^2 \ln f[\varepsilon_{it}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_i]/\partial\boldsymbol{\varrho}_i\partial\boldsymbol{\varrho}_i'$, respectively.

As for the information matrix, Proposition D1 simplifies to

Proposition D2 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\varrho})$ with density $f(\boldsymbol{\varepsilon}^*,\boldsymbol{\varrho})=\prod_{i=1}^N f(\varepsilon_{it}^*,\boldsymbol{\varrho}_i)$, then the information matrix will be given by a special case of Proposition D1 in which $\mathcal{M}_{ll}$ will be a diagonal matrix of order N with typical element*

$$\mathrm{M}_{ll}(\boldsymbol{\varrho}_i)=V\left[\left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{ls}=\mathrm{M}_{ls}\mathbf{E}_N'$, where M_{ls} also a diagonal matrix of order N with typical element

$$\mathrm{M}_{ls}(\boldsymbol{\varrho}_i)=cov\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*},\left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^*\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{ss}$ is the sum of the commutation matrix $\mathbf{K}_{NN}$ and a block diagonal matrix $\boldsymbol{\Upsilon}$ of order N^2 in which each of the N diagonal blocks is a diagonal matrix of size N with the following structure:

$$\boldsymbol{\Upsilon}_i=\left[\begin{array}{ccccccc} \mathrm{M}_{ll}(\boldsymbol{\varrho}_1) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{M}_{ll}(\boldsymbol{\varrho}_{i-1}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{M}_{ss}(\boldsymbol{\varrho}_i)-1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{M}_{ll}(\boldsymbol{\varrho}_{i+1}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathrm{M}_{ll}(\boldsymbol{\varrho}_N) \end{array}\right],$$

where

$$\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)=V\left[\left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^*\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{lr}$ is an $N\times q$ block diagonal matrix with typical diagonal block of size $1\times q_i$

$$\mathrm{M}_{lr}(\boldsymbol{\varrho}_i)=-cov\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*},\left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right|\boldsymbol{\varrho}\right],$$

$\mathcal{M}_{sr} = \mathbf{E}_N \mathrm{M}_{sr}$, where$\mathrm{M}_{sr}$ another block diagonal matrix of order $N \times q$ with typical block of size $1 \times q_i$

$$\mathrm{M}_{sr}(\boldsymbol{\varrho}_i) = cov\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^*, \left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right| \boldsymbol{\varrho}\right],$$

and $\mathcal{M}_{rr}$ is an $q \times q$ block diagonal matrix with typical block of size $q_i \times q_i$

$$\mathrm{M}_{rr}(\boldsymbol{\varrho}_i) = V\left[\left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right| \boldsymbol{\varrho}_i\right].$$

Proof. The expression for $\mathcal{M}_{ll}$ follows trivially from the fact that

$$cov\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}, \left.\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_j)}{\partial \varepsilon_j^*}\right| \boldsymbol{\varrho}\right] = 0$$

for $i \neq j$ because of the cross-sectional independence of the shocks.

The same property also implies that $\mathcal{M}_{ls} = \mathrm{M}_{ls}\mathbf{E}_N'$ because for $i \neq j \neq k$

$$\begin{aligned}
E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_j)}{\partial \varepsilon_j^*}\varepsilon_{it}^*\right| \boldsymbol{\varrho}\right] &= 0 \text{ since } E\left[\left.\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_j^*}\right| \boldsymbol{\varrho}\right] = 0, \\
E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{jt}^*\right| \boldsymbol{\varrho}\right] &= 0 \text{ since } E\left(\varepsilon_{jt}^*\right| \boldsymbol{\varrho}\right) = 0, \\
E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\left(\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_j)}{\partial \varepsilon_j^*}\varepsilon_{jt}^* + 1\right)\right| \boldsymbol{\varrho}\right] &= 0 \text{ since } E\left[\left.\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\right| \boldsymbol{\varrho}\right] = 0
\end{aligned}$$

and

$$E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_j)}{\partial \varepsilon_j^*}\varepsilon_{kt}^*\right| \boldsymbol{\varrho}\right] = 0 \text{ since } E\left(\varepsilon_{kt}^*\right| \boldsymbol{\varrho}) = 0.$$

The expression for $\mathcal{M}_{ss}$ is slightly more involved. First, most but not all the off-diagonal terms will be 0. Specifically, when $i \neq j$

$$E\left[\left(\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^* + 1\right)\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_j)}{\partial \varepsilon_j^*}\varepsilon_{it}^*\right| \boldsymbol{\varrho}\right] = 0 \text{ since } E\left[\left.\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_j)}{\partial \varepsilon_j^*}\right| \boldsymbol{\varrho}\right] = 0,$$

$$E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^*\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{jt}^*\right| \boldsymbol{\varrho}\right] = 0 \text{ since } E\left(\varepsilon_{jt}^*\right| \boldsymbol{\varrho}\right) = 0$$

and

$$E\left[\left(\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^* + 1\right)\left(\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_j^*}\varepsilon_{jt}^* + 1\right)\right| \boldsymbol{\varrho}\right] = 0 \text{ since } E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^* + 1\right| \boldsymbol{\varrho}\right] = 0$$

However,

$$E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{jt}^*\frac{\partial \ln f(\varepsilon_{jt}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_j^*}\varepsilon_{it}^*\right| \boldsymbol{\varrho}\right] = 1 \text{ since } E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^* + 1\right| \boldsymbol{\varrho}\right] = 0.$$

In contrast, the diagonal terms, which can only take two forms, are different from 0. Specifically, they will be either

$$E\left[\left(\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^* + 1\right)^2\right| \boldsymbol{\varrho}\right] = \mathrm{M}_{ss}(\boldsymbol{\varrho}_i) \text{ since } E\left[\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{it}^* + 1\right| \boldsymbol{\varrho}\right] = 0$$

or

$$E\left[\left(\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\varepsilon_{jt}^*\right)^2\Bigg|\boldsymbol{\varrho}\right]=E\left[\left(\frac{\partial \ln f(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)}{\partial \varepsilon_i^*}\right)^2\Bigg|\boldsymbol{\varrho}\right]=\mathrm{M}_{ll}(\boldsymbol{\varrho}_i) \text{ since } E\left(\varepsilon_{jt}^{*2}\big|\,\boldsymbol{\varrho}\right)=1.$$

As a result, we can write $\mathcal{M}_{ss}=\mathbf{K}_{NN}+\boldsymbol{\Upsilon}$.

The cross-sectional independence of the shocks also implies the block diagonal structure of $\mathcal{M}_{lr}$ and $\mathcal{M}_{rr}$, as well as the fact that $\mathcal{M}_{sr}=\mathbf{E}_N\mathrm{M}_{sr}$. As expected, the same expressions are obtained by taking the expected value of the (minus) Hessian. □

When one of the univariate distributions is symmetric, then $\mathrm{M}_{ls}(\boldsymbol{\varrho}_i)=\mathrm{M}_{lr}(\boldsymbol{\varrho}_i)=0$. One popular example will be the univariate standardised Student t distribution with $\nu=\eta^{-1}$ degrees of freedom, which is such that

$$\ln f[\varepsilon_{it}^*(\boldsymbol{\theta});\eta_i]=c(\eta_i)-\left(\frac{\eta_i+1}{2\eta_i}\right)\log\left[1+\frac{\eta_i}{1-2\eta_i}\varepsilon_{it}^{*2}(\boldsymbol{\theta})\right],$$

with

$$c(\eta_i)=\log\left(\frac{\eta_i+1}{2\eta_i}\right)-\log\left[\Gamma\left(\frac{1}{2\eta_i}\right)\right]-\frac{1}{2}\log\left(\frac{1-2\eta_i}{\eta_i}\right)-\frac{1}{2}\log\pi.$$

Here,

$$\delta(\varepsilon_t^{*2};\eta)=\frac{\eta+1}{1-2\eta+\eta\varepsilon_t^{*2}}$$

and

$$\begin{aligned}\frac{\partial \ln f(\varepsilon_{it}^*;\eta)}{\partial\eta} &= \frac{1}{2\eta(1-2\eta)}-\frac{1}{2\eta^2}\left[\psi\left(\frac{\eta+1}{2\eta}\right)-\psi\left(\frac{1}{2\eta}\right)\right]\\ &\quad-\frac{\eta+1}{1-2\eta+\eta\varepsilon_{it}^{*2}}\frac{\varepsilon_{it}^{*2}}{2\eta(1-2\eta)}+\frac{1}{2\eta^2}\ln\left(1+\frac{\eta}{1-2\eta}\varepsilon_{it}^{*2}\right).\end{aligned}$$

In addition

$$\begin{aligned}\mathrm{M}_{ll}(\boldsymbol{\varrho}_i) &= \frac{\nu_i(\nu_i+1)}{(\nu_i-2)(\nu_i+3)},\\ \mathrm{M}_{ss}(\boldsymbol{\varrho}_i) &= \frac{2\nu_i}{\nu_i+3},\\ \mathrm{M}_{sr}(\boldsymbol{\varrho}_i) &= -\frac{6\nu_i^2}{(\nu_i-2)(\nu_i+1)(\nu_i+3)}\end{aligned}$$

and

$$\mathrm{M}_{rr}(\boldsymbol{\varrho}_i)=\frac{\nu_i^4}{4}\left[\psi'\left(\frac{\nu_i}{2}\right)-\psi'\left(\frac{\nu_i+1}{2}\right)\right]-\frac{\nu_i^4(\nu_i-3)(\nu_i+4)}{2(\nu_i-2)^2(\nu_i+1)(\nu_i+3)},$$

where $\psi'(x)=\partial^2\ln\Gamma(x)/\partial x^2$ is the so-called tri-gamma function (Abramowitz and Stegun 1964), which reduce to 1, 1, 0 and 3/2 respectively, under normality (see Fiorentini, Sentana and Calzolari (2003)). As a result, when all shocks are in fact Gaussian, $\mathcal{M}_{ss}=\mathbf{K}_{NN}+\mathbf{I}_{N^2}$, which confirms that not all elements of $\mathbf{C}$ can be identified with a Gaussian log-likelihood function because rank$(\mathbf{K}_{NN}+\mathbf{I}_{N^2})=N(N+1)/2$ (see section 4 in Magnus and Sentana (2020) for a general expression for the eigenvalues of $(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})$.

D.4 Semiparametric estimators

In Supplemental Appendix C.5 we interpreted the last summand of (C25) as $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ times the theoretical least squares projection of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on (the linear span of) $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$, which is conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$ from Proposition 3 in Fiorentini and Sentana (2007). Such an interpretation allowed Gonzalez-Rivera and Drost (1999) to replace a parametric assumption on the shape of the distribution of the standardised innovations $\boldsymbol{\varepsilon}_t^*$ by a fully non-parametric alternative. Specifically, in a univariate context they replaced the linear span of $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ by the so-called unrestricted tangent set, which is the Hilbert space generated by all the time-invariant functions of $\boldsymbol{\varepsilon}_t^*$ with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})$. The next proposition, which originally appeared as Proposition 6 in Fiorentini and Sentana (2007), describes the resulting semiparametric efficient score and the corresponding efficiency bound for multivariate conditionally heteroskedastic models whose conditionally mean is not identically zero:

Proposition D3 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta},\boldsymbol{\rho}$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\rho})$ with density function $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ denotes the possibly infinite dimensional vector of shape parameters and $\boldsymbol{\rho}=\mathbf{0}$ normality, and both its Fisher information matrix for location and scale,*

$$\mathcal{M}_{dd}(\boldsymbol{\theta},\boldsymbol{\rho}) = V\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})|I_{t-1};\boldsymbol{\theta},\boldsymbol{\rho}\right]$$
$$= V\left\{\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\rho})\\ \mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\rho})\end{array}\right]\middle|\boldsymbol{\theta},\boldsymbol{\rho}\right\} = V\left\{\left[\begin{array}{c}-\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*\\ -vec\left\{\mathbf{I}_N+\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*\cdot\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\end{array}\right]\middle|\boldsymbol{\theta},\boldsymbol{\rho}\right\}$$

and the matrix of third and fourth order central moments $\mathcal{K}(\boldsymbol{\rho})$ in (C22) are bounded, then the semiparametric efficient score will be given by:

$$\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\rho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\rho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right], \tag{D17}$$

while the semiparametric efficiency bound is

$$\ddot{\mathcal{S}}(\boldsymbol{\phi}) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\rho})\left[\mathcal{M}_{dd}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\rho})\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta},\boldsymbol{\rho}), \tag{D18}$$

where $+$ denotes Moore-Penrose inverses and $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta},\boldsymbol{\rho}) = E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{M}_{dd}(\boldsymbol{\theta},\boldsymbol{\rho})\mathbf{Z}_{dt}'(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\rho}\right]$.

Proof. It trivially follows from expressions (B3) and (C22) in appendices B and C, respectively, that

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}'(\boldsymbol{\theta},\mathbf{0})\,|I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} = \mathbf{0}$$

for any distribution. In addition, we also know that

$$E\left\{\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]|I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}\right\} = \mathbf{0}.$$

Hence, the second summand of (D17), which can be interpreted as $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ times the residual from the theoretical regression of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on a constant and $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, belongs to the unrestricted tangent set, which is the Hilbert space spanned by all the time-invariant functions of $\boldsymbol{\varepsilon}_t^*$ with zero conditional means and bounded second moments that are conditionally orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$.

Now, if we write (D17) as

$$[\mathbf{Z}_{dt}(\boldsymbol{\theta}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})]\,\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) + \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\mathcal{K}\,(0)\,\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}),$$

then we can use the law of iterated expectations to show that the semiparametric efficient score (D17) evaluated at the true parameter values will be unconditionally orthogonal to the unrestricted tangent set because so is $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$, and $E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right] = \mathbf{0}$.

Finally, the expression for the semiparametric efficiency bound will be

$$\begin{aligned}
E&\left[\left.\begin{array}{c}\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\}\\ \times\{\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})'\mathbf{Z}'_{dt}(\boldsymbol{\theta}) - \left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{e}'_{dt}(\boldsymbol{\theta},0)\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})\}\end{array}\right|\boldsymbol{\theta},\boldsymbol{\varrho}\right]\\
&= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\varrho}\right]\\
&-E\left\{\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{e}'_{dt}(\boldsymbol{\theta},\mathbf{0})\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}\\
&-E\left\{\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right]\mathbf{e}_{dt}(\boldsymbol{\phi})'\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}\\
&+E\left\{\mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathbf{e}_{dt}(\boldsymbol{\theta},0)\right]\left[\mathbf{e}'_{dt}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{e}'_{dt}(\boldsymbol{\theta},0)\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})|\boldsymbol{\theta},\boldsymbol{\varrho}\right\}\\
&= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\boldsymbol{\varrho}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\varrho})\left[\mathcal{M}_{dd}\,(\boldsymbol{\varrho}) - \mathcal{K}\,(0)\,\mathcal{K}^+\,(\boldsymbol{\varrho})\,\mathcal{K}\,(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta},\boldsymbol{\varrho})
\end{aligned}$$

by virtue of (C22), (B3) and the law of iterated expectations. □

In the case of the univariate GARCH-M model (19), we estimate the model parameters using reparametrisation 2 in section 4. Specifically, expressions (D2) and (D4) become

$$\mathbf{Z}_{lt}(\boldsymbol{\varphi}) = \frac{\partial\mu_t(\boldsymbol{\varphi})/\partial\boldsymbol{\varphi}}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)} = \frac{1}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}\left[\begin{array}{c}\frac{1}{2}\varphi_{im}\sigma_t^{\diamond-1}(\boldsymbol{\varphi}_c)\partial\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)/\partial\boldsymbol{\varphi}_c\\ \sigma_t^{\diamond}(\boldsymbol{\varphi}_c)\\ 0\end{array}\right] = \left[\begin{array}{c}\varphi_{im}\varphi_{ic}^{-1/2}\mathbf{W}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi}_c)\\ \varphi_{ic}^{-1/2}\\ 0\end{array}\right],$$

$$\mathbf{Z}_{st}(\boldsymbol{\varphi}) = \frac{\partial\sigma_t^2(\boldsymbol{\varphi})/\partial\boldsymbol{\varphi}}{2\varphi_{ic}\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)} = \frac{1}{2\varphi_{ic}\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)}\left[\begin{array}{c}\varphi_{ic}\partial\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)/\partial\boldsymbol{\varphi}_c\\ 0\\ \sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)\end{array}\right] = \left[\begin{array}{c}\mathbf{W}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi}_c)\\ 0\\ \frac{1}{2}\varphi_{ic}^{-1}\end{array}\right]$$

and

$$\begin{aligned}
e_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho}) &= -\frac{\partial\ln f[\epsilon_t(\boldsymbol{\varphi});\boldsymbol{\rho}]}{\partial\varepsilon},\\
e_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho}) &= -\left\{1 + \epsilon_t(\boldsymbol{\varphi})\frac{\partial\ln f[\epsilon_t(\boldsymbol{\varphi});\boldsymbol{\rho}]}{\partial\varepsilon}\right\},
\end{aligned}$$

respectively, where

$$\epsilon_t(\boldsymbol{\varphi}) = \frac{\epsilon_t^{\diamond}(\boldsymbol{\varphi}_c) - \varphi_{im}}{\varphi_{ic}^{1/2}} = \frac{x_t}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)} - \frac{\varphi_{im}}{\varphi_{ic}^{1/2}} = \frac{x_t - \varphi_{im}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}{\varphi_{ic}^{1/2}\sigma_t^{\diamond}(\boldsymbol{\varphi}_c)}$$

and

$$\mathbf{W}_{\boldsymbol{\varphi}_c t}(\boldsymbol{\varphi}_c) = \frac{1}{2\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)}\frac{\partial\sigma_t^{\diamond 2}(\boldsymbol{\varphi}_c)}{\partial\boldsymbol{\varphi}_c}.$$

Then, a direct application of (D3) yields

$$\mathbf{s}_{\boldsymbol{\varphi}t}(\boldsymbol{\phi}) = \left[\begin{array}{cc}\mathbf{Z}_{lt}(\boldsymbol{\varphi}) & \mathbf{Z}_{st}(\boldsymbol{\varphi})\end{array}\right]\left[\begin{array}{c}e_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ e_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{array}\right] = \left[\begin{array}{c}\mathbf{W}_t(\boldsymbol{\varphi}_c)\mathbf{r}'(\boldsymbol{\varphi}_i)\\ \boldsymbol{\Delta}(\varphi_{ic})\end{array}\right]\left[\begin{array}{c}e_{lt}(\boldsymbol{\varphi},\boldsymbol{\varrho})\\ e_{st}(\boldsymbol{\varphi},\boldsymbol{\varrho})\end{array}\right],$$

where

$$\mathbf{r}(\boldsymbol{\varphi}_i) = \begin{pmatrix} \varphi_{im}\varphi_{ic}^{-1/2} & 1 \end{pmatrix}'$$

and

$$\boldsymbol{\Delta}(\varphi_{ic}) = \begin{pmatrix} \varphi_{ic}^{-1/2} & 0 \\ 0 & \frac{1}{2}\varphi_{ic}^{-1} \end{pmatrix}.$$

On the other hand, we use again the natural parametrisation of the multivariate market model in (20). As a result, the Jacobian matrix (C36) in Supplemental Appendix C remains relevant, so that

$$\begin{aligned} \mathbf{s}_{\mathbf{a}t}(\boldsymbol{\theta}) &= -\boldsymbol{\Omega}^{-1/2}\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*, \\ \mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta}) &= -\boldsymbol{\Omega}^{-1/2} r_{mt}\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^*, \end{aligned}$$

where $\boldsymbol{\Omega}^{1/2}$ is a matrix square root of $\boldsymbol{\Omega}$.

If we choose the Cholesky decomposition, we can use expression (D14) to obtain

$$\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{D}_N'\mathbf{F}\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-\frac{1}{2}})vec\left\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\rho}]/\partial\boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\},$$

where $\mathbf{F}$ denotes the transpose of the inverse of $\mathbf{L}_N(\boldsymbol{\Omega}^{1/2} \otimes \mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{1/2})\mathbf{K}_{NN}\mathbf{L}_N'$.

Finally, it is worth noting that it is possible to avoid the use of explicit Moore-Penrose generalised inverses in the computation of the correction by exploiting the fact that

$$\mathcal{K}(\boldsymbol{\rho}) = \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_N \end{pmatrix} \begin{bmatrix} \mathbf{I}_N & E[\boldsymbol{\varepsilon}_t^* vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \\ E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})\boldsymbol{\varepsilon}_t^{\prime *}|\,\boldsymbol{\theta},\boldsymbol{\varrho}] & E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}) - \mathbf{I}_N|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \end{bmatrix} \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_N' \end{pmatrix}$$

and

$$\mathcal{K}(\mathbf{0}) = \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N^2} + \mathbf{K}_{NN} \end{pmatrix}$$

imply that

$$\begin{aligned} \mathcal{K}(0)\,\mathcal{K}^+(\boldsymbol{\rho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}) &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{D}^{+\prime} \end{pmatrix} \\ &\times \begin{bmatrix} \mathbf{I}_N & E[\boldsymbol{\varepsilon}_t^* vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \\ E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})\boldsymbol{\varepsilon}_t^{\prime *}|\,\boldsymbol{\theta},\boldsymbol{\varrho}] & E[vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime})vech'(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime}) - \mathbf{I}_N|\,\boldsymbol{\theta},\boldsymbol{\varrho}] \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\varepsilon}_t^* \\ vech(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\prime} - \mathbf{I}) \end{bmatrix}. \end{aligned}$$

Nevertheless, $f(\boldsymbol{\varepsilon}_t^*;\boldsymbol{\rho})$ has to be replaced by a nonparametric estimator, which increasingly suffers from the curse of dimensionality as the cross-sectional dimension N increases. In line with the usual practice, we employ a standard multivariate Gaussian kernel. Once again, we have done some experimentation to choose optimal bandwidths by scaling up and down the automatic choices given in Silverman (1986) because a proper cross-validation procedure is extremely costly to implement in a Monte Carlo exercise when $N = 3$.

E Other results

E.1 Standardised two component mixtures of multivariate normals

Consider the following mixture of two multivariate normals

$$\varepsilon_t \sim \begin{cases} N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) & \text{with probability} \quad \lambda, \\ N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) & \text{with probability} \quad 1-\lambda. \end{cases}$$

Let d_t denote a Bernoulli variable which takes the value 1 with probability λ and 0 with probability $1-\lambda$. As is well known, the unconditional mean vector and covariance matrix of the observed variables are:

$$\begin{aligned} E(\varepsilon_t) &= E[E(\varepsilon_t|d_t)] = \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2, \\ V(\varepsilon_t) &= V[E(\varepsilon_t|d_t)] + E[V(\varepsilon_t|d_t)] = \lambda(1-\lambda)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)' + \lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2. \end{aligned}$$

Therefore, this random vector will be standardised if and only if

$$\begin{aligned} \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2 &= \mathbf{0}, \\ \lambda(1-\lambda)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)' + \lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 &= \mathbf{I}. \end{aligned}$$

Let us initially assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$ but that the mixture is not degenerate, so that $\lambda \neq 0, 1$. Let $\boldsymbol{\Sigma}_{1L}\boldsymbol{\Sigma}_{1L}'$ and $\boldsymbol{\Sigma}_{2L}\boldsymbol{\Sigma}_{2L}'$ denote the Cholesky decompositions of the covariance matrices of the two components. Then, we can write

$$\lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_{1L}[\lambda\mathbf{I}_N + (1-\lambda)\boldsymbol{\Sigma}_{1L}^{-1}\boldsymbol{\Sigma}_{2L}\boldsymbol{\Sigma}_{2L}'\boldsymbol{\Sigma}_{1L}^{-1\prime}]\boldsymbol{\Sigma}_{1L}' = \boldsymbol{\Sigma}_{1L}(\lambda\mathbf{I}_N + \mathbf{K}_L\mathbf{K}_L')\boldsymbol{\Sigma}_{1L}',$$

where $\mathbf{K}_L = \sqrt{1-\lambda}\boldsymbol{\Sigma}_{1L}^{-1}\boldsymbol{\Sigma}_{2L}$ remains a lower triangular matrix. Given that $\mathbf{I}_N = \mathbf{e}_1\mathbf{e}_1 + \ldots + \mathbf{e}_N\mathbf{e}_N$, where $\mathbf{e}_i$ is the i^{th} vector of the canonical basis, the Cholesky decomposition of $\lambda\mathbf{I}_N + \mathbf{K}_L\mathbf{K}_L'$, say $\mathbf{J}_L\mathbf{J}_L'$, can be computed by means of N rank-one updates that sequentially add $\sqrt{\lambda}\mathbf{e}_i\sqrt{\lambda}\mathbf{e}_i'$ for $i = 1, \ldots, N$. The special form of those vectors can be efficiently combined with the usual rank-one update algorithms to speed up this process (see e.g. Sentana (1999) and the references therein). In any case, the elements of $\mathbf{J}_L$ will be functions of λ and the $N(N+1)/2$ elements in $\mathbf{K}_L$. If we then choose $\boldsymbol{\Sigma}_{1L} = \mathbf{J}_L^{-1}$, we will guarantee that $\lambda\boldsymbol{\Sigma}_1 + (1-\lambda)\boldsymbol{\Sigma}_2 = \mathbf{I}_N$. Therefore, we can achieve a standardised two-component mixture of two multivariate normals with 0 means by drawing with probability λ one random variable from a distribution with covariance matrix $\mathbf{J}_L^{-1\prime}\mathbf{J}_L^{-1}$, and with probability $1-\lambda$ from another distribution with covariance matrix $(1-\lambda)^{-1}\mathbf{K}_L\mathbf{K}_L'$.

Let us now turn to the case in which the means of the components are no longer 0. The zero unconditional mean condition is equivalent to $\boldsymbol{\mu}_1 = (1-\lambda)\boldsymbol{\delta}$ and $\boldsymbol{\mu}_2 = -\lambda\boldsymbol{\delta}$, so that $\boldsymbol{\delta}$ measures the difference between the two means. Thus, the unconditional covariance matrix will be $\lambda(1-\lambda)\boldsymbol{\delta}\boldsymbol{\delta}' + \mathbf{I}_N$ after imposing the restrictions on $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in the previous paragraph. Once again, the Cholesky decomposition of this matrix is very easy to obtain because it can be regarded as a positive rank-one update of the identity matrix, whose decomposition is trivial.

Thus, we can parametrise a standardised mixture of two multivariate normals, which usually involves $2N$ mean parameters, $2N(N+1)/2$ covariance parameters and one mixing parameter, in terms of the N mean difference parameters in $\boldsymbol{\delta}$, the $N(N+1)/2$ relative variance parameters in $\mathbf{K}_L$ and the mixing parameter λ, the remaining N mean parameters and $N(N+1)/2$ covariance ones freed up to target any unconditional mean vector and covariance matrix.

Mencía and Sentana (2009) explain how to standardise Bernoulli location-scale mixtures of normals, which are a special case of the two component mixtures we have just discussed in which $\boldsymbol{\Sigma}_2 = \varkappa\boldsymbol{\Sigma}_1$. Straightforward algebra confirms that the standardisation procedure described above simplifies to the one they provide in their Proposition 1.

E.2 Non-causal ARMA models

Consider the following AR(2) process:

$$(1-\alpha_1 L)(1-\alpha_2 L)x_t = \mu + \xi_t, \tag{E1}$$

where ξ_t is a possibly non-Gaussian *i.i.d.* sequence, $\alpha_1, \alpha_2 \in \mathbb{R}$, $|\alpha_1| < 1$, $|\alpha_2| > 1$ but $\alpha_2 \neq \alpha_1^{-1}$. Higher order process with possibly complex roots can be handled analogously, but the algebra gets messier. Brockwell and Davis (1987) showed that x_t can be written as the following doubly infinite MA process

$$x_t = \frac{-\alpha_2^{-1}\mu}{(1-\alpha_1)(1-\alpha_2^{-1})} - (\ldots + \alpha_2^{-2}L^{-3} + \alpha_2^{-1}L^{-2} + L^{-1} + \alpha_1 + \alpha_1^2 L + \alpha_1^3 L^2 + \alpha_1^4 L^3 + \ldots)\frac{\xi_t}{\alpha_2 - \alpha_1},$$

which they called mixed causal/non-causal because x_t effectively depends on past, present and future values of the underlying innovations. Nevertheless, by looking at the spectral density of x_t they also showed that this process has the following purely causal AR(2) representation:

$$(1-\alpha_1 L)(1-\alpha_2^{-1} L)x_t = \nu + u_t, \tag{E2}$$

where u_t is a white noise but not necessarily serially independent sequence, with variance $\sigma_u^2 = \alpha_2^{-2}\sigma_\xi^2$ and $\nu = -\alpha_2^{-1}\mu$. Thus, the situation is entirely analogous to the well known multiple invertible and non-invertible representations of MA processes.

Breidt et al (1991) showed that a non-Gaussian log-likelihood function based on the assumption that the distribution of ξ_t is *i.i.d.* with 0 mean and finite variance σ_ξ^2 will be able to consistently estimate the values of the two autoregressive roots that appear in (E1) as well as the true drift and variance of the innovations. In contrast, a Gaussian log-likelihood function, which effectively exploits the information in the spectral density of x_t, can only consistently estimate the parameters in (E2).

At first sight, it might appear that one cannot apply the procedures we have developed in the paper to assess the adequacy of the non-Gaussian distribution chosen for the purposes of estimating the "structural" parameters because the Gaussian pseudo log-likelihood cannot consistently estimate them. However, under correct specification, the non-Gaussian log-likelihood function will also estimate α_1, α_2^{-1}, $-\alpha_2^{-1}\mu$ and $\alpha_2^{-2}\sigma_\xi^2$ consistently. Therefore, one can easily develop a DWH specification test to check the validity of the distributional assumption for ξ_t by comparing the non-Gaussian coefficient estimators of those "reduced form" parameters with the Gaussian ones. The score versions of those tests that we discussed in section 2.1 are also straightforward. As we have argued in section 3.7, power gains may be obtained by focusing on ν and σ_u^2.

E.3 Additional Monte Carlo results

In this section, we look at the sampling distribution of the estimators we used in section 4 to compute the DWH tests of the univariate GARCH-M model and the multivariate market model.

Univariate GARCH-M Table 1S displays the Monte Carlo medians and interquartile ranges of the estimators. The results broadly confirm the theoretical predictions in terms of bias and relative efficiency. It is worth noticing that the bias of the restricted (unrestricted) Student t maximum likelihood estimators of the scale parameter is negative (positive) when the log-likelihood is misspecified, which suggests that our tests will have good power for pairwise comparisons involving this parameter, at least for the distributions considered in the exercise. In turn, the location parameter estimators are biased only when the true distribution is asymmetric.

Multivariate market model Table 2S displays the Monte Carlo medians and interquartile ranges of the estimators for several representative parameters in addition to the global scale parameter $\vartheta_i = |\mathbf{\Omega}|^{1/N}$. Specifically, we exploit the exchangeability of our design to pool the results of all the elements of the vectors of intercepts $\mathbf{a}$ and slopes $\mathbf{b}$, and the "vectors" of residual covariance parameters $vecd(\mathbf{\Omega}^\circ)$, $vecl(\mathbf{\Omega}^\circ)$, $vecd(\mathbf{\Omega})$ and $vecl(\mathbf{\Omega})$. Once again, the results are in line with the theoretical predictions. Moreover, the biases of the restricted and unrestricted Student t maximum likelihood estimators of the global scale parameter have opposite signs, as in the univariate case. Finally, the location parameters are only biased in the asymmetric distribution simulations. Therefore, we expect tests that involve the intercepts to increase power in that case, but to result in a waste of degrees of freedom otherwise.

Additional references

Abramowitz, M., and Stegun, I.A. (1964): *Handbook of mathematical functions*, AMS 55, National Bureau of Standards.

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

Breidt, J., Davis, R.A., Lh, K.S. and Rosenblatt, M. (1991): "Maximum likelihood estimation for noncausal autoregressive processes", *Journal of Multivariate Analysis* 36, 175-198.

Brockwell, P.J. and Davis, R.A. (1987): *Time series: theory and methods* (2nd ed.), Springer.

Chen, S.X. (2000): "Probability density function estimation using gamma kernels", *Annals of the Institute of Statistical Mathematics* 52, 471-480.

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

Fiorentini, G. and Sentana, E. (2015): "Tests for serial dependence in static, non-Gaussian factor models", in S.J. Koopman and N. Shephard (eds.) *Unobserved Components and Time Series Econometrics*, pp. 118-189, Oxford University Press.

Magnus, J.R. and Neudecker, H. (2019): *Matrix differential calculus with applications in Statistics and Econometrics*, 3rd ed., Wiley.

Mencía, J. and Sentana, E. (2009): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics* 153, 105-121.

Sentana, E. (1999): "Econometric applications of positive rank-one modifications of the symmetric factorization of a positive semi-definite matrix", *Spanish Economic Review* 1, 79-90.

Silverman B.W. (1986): *Density estimation*, Chapman and Hall.

TABLE 1S: Univariate GARCH-M: Parameter estimators.

Parameter		β	γ	δ, φ_{im}	$\vartheta_i, \varphi_{ic}$	$\eta = 1/\nu$
True value		0.85	0.1	0.05	1.0	
	RML	0.8467	0.0960	0.0506	1.0404	0.0833
		(0.0375)	(0.0348)	(0.0314)	(0.4132)	
	UML	0.8467	0.0959	0.0507	1.0397	0.0815
Student t_{12}		(0.0376)	(0.0350)	(0.0315)	(0.4125)	(0.0276)
	PML	0.8464	0.0956	0.0508	1.0420	
		(0.0392)	(0.0363)	(0.0324)	(0.4331)	
	RML	0.8467	0.0956	0.0505	1.0137	0.0833
		(0.0383)	(0.0344)	(0.0315)	(0.3986)	
	UML	0.8468	0.0959	0.0504	1.0392	0.1232
Student t_8		(0.0381)	(0.0343)	(0.0314)	(0.4077)	(0.0276)
	PML	0.8460	0.0955	0.0504	1.0439	
		(0.0423)	(0.0384)	(0.0333)	(0.4539)	
	RML	0.8461	0.0955	0.0506	0.8706	0.0833
		(0.0437)	(0.0383)	(0.0278)	(0.3817)	
	UML	0.8470	0.0967	0.0502	1.3990	0.3604
GC(0,3.2)		(0.0371)	(0.0338)	(0.0254)	(0.5748)	(0.0264)
	PML	0.8460	0.0956	0.0506	1.0425	
		(0.0429)	(0.0377)	(0.0327)	(0.4476)	
	RML	0.8460	0.0956	0.1117	0.8601	0.0833
		(0.0436)	(0.0386)	(0.0358)	(0.3848)	
	UML	0.8475	0.0970	0.1723	1.5853	0.3865
GC(-.9,3.2)		(0.0356)	(0.0321)	(0.0380)	(0.6728)	(0.0265)
	PML	0.8459	0.0956	0.0511	1.0453	
		(0.0431)	(0.0381)	(0.0326)	(0.4626)	

Monte Carlo medians and (interquartile ranges) of RML (Student t-based maximum likelihood with 12 degrees of freedom), UML (unrestricted Student t-based maximum likelihood), and PML (Gaussian pseudo maximum likelihood) estimators. GC (Gram-Charlier expansion). Sample length=2,000. Replications=20,000.

TABLE 2S: Multivariate market model: Parameter estimators.

Parameter		**a**	**b**	ϑ_i	ω°_{ii}	ω°_{ij}	ω_{ii}	ω_{ij}	$\eta = 1/\nu$
True value		0.112	1	2.8917	1.0845	0.3253	3.136	0.9408	
	RML	0.1124	0.9989	2.8702	1.0872	0.3262	3.1215	0.9355	0.0833
		(0.1040)	(0.1173)	(0.1696)	(0.0808)	(0.0705)	(0.2955)	(0.2115)	
Student t_{12}	UML	0.1123	0.9989	2.8674	1.0872	0.3262	3.1176	0.9347	0.0810
		(0.1041)	(0.1174)	(0.1815)	(0.0808)	(0.0706)	(0.3043)	(0.2117)	(0.0280)
	PML	0.1124	0.9998	2.8646	1.0873	0.3262	3.1147	0.9341	
		(0.1066)	(0.1213)	(0.1807)	(0.0849)	(0.0738)	(0.3125)	(0.2200)	
	RML	0.1127	0.9989	2.7652	1.0874	0.3259	3.0077	0.9008	0.0833
		(0.1015)	(0.1148)	(0.1763)	(0.0832)	(0.0723)	(0.2980)	(0.2078)	
Student t_8	UML	0.1126	0.9989	2.8683	1.0875	0.3259	3.1211	0.9352	0.1233
		(0.1013)	(0.1144)	(0.2088)	(0.0831)	(0.0718)	(0.3301)	(0.2171)	(0.0304)
	PML	0.1126	0.9988	2.8618	1.0877	0.3259	3.1129	0.9318	
		(0.1075)	(0.1219)	(0.2085)	(0.0927)	(0.0803)	(0.3649)	(0.2391)	
	RML	0.1123	0.9995	2.0600	1.0882	0.3264	2.2402	0.6705	0.0833
		(0.0803)	(0.0912)	(0.1989)	(0.0975)	(0.0853)	(0.2945)	(0.1886)	
DSMN(0.2,0.1)	UML	0.1125	0.9997	3.5341	1.0878	0.3262	3.8521	1.1545	0.3474
		(0.0775)	(0.0874)	(0.8393)	(0.0877)	(0.0765)	(0.9692)	(0.3848)	(0.0372)
	PML	0.1119	1.0000	2.8483	1.0907	0.3266	3.1071	0.9280	
		(0.1071)	(0.1202)	(0.3197)	(0.1241)	(0.1077)	(0.4966)	(0.3275)	
	RML	-0.0003	1.0004	2.0275	1.0829	0.3140	2.1962	0.6351	0.0833
		(0.0830)	(0.0891)	(0.1984)	(0.0991)	(0.0868)	(0.2980)	(0.1900)	
DLSMN(0.2,0.1,0.5)	UML	-0.0576	1.0006	3.7270	1.0753	0.2986	4.0177	1.1141	0.3616
		(0.0831)	(0.0854)	(0.9916)	(0.0880)	(0.0763)	(1.1239)	(0.4204)	(0.0373)
	PML	0.1119	1.0010	2.8485	1.0908	0.3271	3.1067	0.9295	
		(0.1065)	(0.1204)	(0.3152)	(0.1252)	(0.1097)	(0.4948)	(0.3306)	

Monte Carlo medians and (interquartile ranges) of RML (Student t-based maximum likelihood with 12 degrees of freedom), UML (unrestricted Student t-based maximum likelihood), and PML (Gaussian pseudo maximum likelihood) estimators. DSMN (discrete scale mixture of two normals), DLSMN (discrete location-scale mixture of two normals). Sample length=500. Replications=20,000.