Supplemental Appendices for

Consistent non-Gaussian pseudo 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>

May 2019

D Auxiliary results

D.1 Some useful distribution results

A spherically symmetric random vector of dimension N, $\boldsymbol{\varepsilon}_t^\circ$, is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^\circ = e_t\mathbf{u}_t$, 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 $\boldsymbol{\varepsilon}_t^\circ$. 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 $\boldsymbol{\varepsilon}_t^\circ$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$. Specifically, if $\boldsymbol{\varepsilon}_t^\circ$ 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 $\boldsymbol{\varepsilon}_t^\circ$ 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 $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$ are given by

$$E(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}, \tag{D1}$$

$$E(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime} \otimes \boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime}) = E[vec(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^{\circ\prime})vec'(\boldsymbol{\varepsilon}_t^\circ \boldsymbol{\varepsilon}_t^\circ)] = (\kappa_0 + 1)[(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + vec\,(\mathbf{I}_N)\, vec'\,(\mathbf{I}_N)], \tag{D2}$$

where $\mathbf{K}_{mn}$ is the commutation matrix of orders m and n (see e.g. Magnus and Neudecker (1987)).

D.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 $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\,[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$ is the Jacobian and $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$.

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 $g\,[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ are differentiable, then

$$\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \partial c(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + \partial g\,[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\,/\partial\boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \tag{D3}$$

while

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})=\frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}+\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varsigma}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=[\mathbf{Z}_{lt}(\boldsymbol{\theta}),\mathbf{Z}_{st}(\boldsymbol{\theta})]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \tag{D4}$$

where

$$\begin{aligned}
\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -\mathbf{Z}_{st}(\boldsymbol{\theta})vec(\mathbf{I}_N),\\
\partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -2\{\mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]\}, && \text{(D5)}\\
\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}),\\
\mathbf{Z}_{st}(\boldsymbol{\theta}) &= \frac{1}{2}\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\cdot[\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})],\\
\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\eta}) &= \delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), && \text{(D6)}\\
\mathbf{e}_{st}(\boldsymbol{\theta},\boldsymbol{\eta}) &= vec\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\cdot\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})-\mathbf{I}_N\right\}, && \text{(D7)}
\end{aligned}$$

and

$$\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]=-2\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma \tag{D8}$$

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 E, 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:

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

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

$$\begin{aligned}
\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}+\frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{(\partial\varsigma)^2}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma}\frac{\partial^2\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, && \text{(D9)}\\
\mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) &= \partial\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\cdot\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varsigma\partial\boldsymbol{\eta}', && \text{(D10)}\\
\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) &= \partial^2 c(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'+\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}',
\end{aligned}$$

where

$$\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'=2\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{Z}_{st}'(\boldsymbol{\theta})-\frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\otimes\mathbf{I}_p\right\}\partial vec\left\{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\right\}/\partial\boldsymbol{\theta}', \tag{D11}$$

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

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 (see Kotz (1975)) and discrete scale mixture of normals, and Amengual, Fiorentini and Sentana (2013) for polynomial expansions).

D.3 Asymptotic distribution under correct specification

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 $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$,

$$\begin{array}{rcl} \mathcal{I}_t(\boldsymbol{\phi}) & = & V\left[\mathbf{s}_t(\boldsymbol{\phi})|I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\phi})\mathbf{Z}'_t(\boldsymbol{\theta}) = -E\left[\mathbf{h}_t(\boldsymbol{\phi})|I_{t-1}; \boldsymbol{\phi}\right], \\ \mathbf{Z}_t(\boldsymbol{\theta}) & = & \left(\begin{array}{cc} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{array}\right) = \left(\begin{array}{ccc} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array}\right), \end{array} \tag{D12}$$

and $\mathcal{M}(\boldsymbol{\phi}) = V\left[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right]$. In particular, Crowder (1976) requires: (i) $\boldsymbol{\phi}_0$ 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 $\boldsymbol{\phi}_0$; (iii) there is uniform convergence to the integrals involved in the computation of the mean vector and covariance matrix of $\mathbf{s}_t(\boldsymbol{\phi})$; and (iv) $-E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right] T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi}) \xrightarrow{p} \mathbf{I}_{p+q}$, where $E^{-1}\left[-T^{-1}\sum_t \mathbf{h}_t(\boldsymbol{\phi})\right]$ is positive definite on a neighbourhood of $\boldsymbol{\phi}_0$.

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 $\sqrt{T}[\tilde{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\eta}}) - \boldsymbol{\theta}_0] \to N\left[\mathbf{0}, \mathcal{I}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)\right]$, where $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)$ is the relevant block of the information matrix.

Proposition 1 in Fiorentini and Sentana (2007), which generalises Propositions 3 in Lange, Little and Taylor (1989), 1 in Fiorentini, Sentana and Calzolari (2003) and 5.2 in Hafner and Rombouts (2007), provides detailed expressions for $\mathcal{M}(\boldsymbol{\phi})$. We reproduce it here to facilitate its comparison to Proposition 2:

Proposition 8 *If $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\phi}$ is i.i.d. $s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta})$ with density* $\exp[c(\boldsymbol{\eta})+g(\varsigma_t,\boldsymbol{\eta})]$, *then*

$$\mathcal{M}(\boldsymbol{\eta})=\begin{pmatrix}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta})\\ \mathbf{0} & \mathcal{M}_{sr}'(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta})\end{pmatrix}, \tag{D13}$$

$$\mathcal{M}_{ll}(\boldsymbol{\eta})=\mathrm{M}_{ll}(\boldsymbol{\eta})\mathbf{I}_N, \tag{D14}$$

$$\mathcal{M}_{ss}(\boldsymbol{\eta})=\mathrm{M}_{ss}(\boldsymbol{\eta})\,(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+[\mathrm{M}_{ss}(\boldsymbol{\eta})-1]vec(\mathbf{I}_N)vec'(\mathbf{I}_N), \tag{D15}$$

$$\mathcal{M}_{sr}(\boldsymbol{\eta})=vec(\mathbf{I}_N)\mathrm{M}_{sr}(\boldsymbol{\eta}), \tag{D16}$$

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

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

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

Fiorentini, Sentana and Calzolari (2003) provide the relevant expressions for the multivariate standardised Student t, while the expressions for the Kotz distribution and the DSMN are given in Amengual and Sentana (2010) (The expression for $\mathrm{M}_{ss}(\kappa)$ for the Kotz distribution in Amengual and Sentana (2010) contains a typo. The correct value is $(N\kappa+2)/[(N+2)\kappa+2]$).

D.4 Gaussian pseudo maximum likelihood estimators

Let $\tilde{\boldsymbol{\theta}}_T=\arg\max_{\boldsymbol{\theta}}L_T(\boldsymbol{\theta},\mathbf{0})$ denote the Gaussian PML estimator of $\boldsymbol{\theta}$. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ even though the 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 $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$. 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 $\boldsymbol{\rho}$:

Proposition 9 *Assume that the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied.*

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

$N\left[\mathbf{0}, \mathcal{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathbf{0}; \varphi_0)\right]$ *with*

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

and

$$\mathcal{K}(\boldsymbol{\rho}) = V[\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})|\, I_{t-1}; \varphi] = \begin{bmatrix} \mathbf{I}_N & \boldsymbol{\Phi}(\boldsymbol{\rho}) \\ \boldsymbol{\Phi}'(\boldsymbol{\rho}) & \boldsymbol{\Upsilon}(\boldsymbol{\rho}) \end{bmatrix}, \tag{D17}$$

where

$$\boldsymbol{\Phi}(\boldsymbol{\rho}) = E[\boldsymbol{\varepsilon}^*_t vec'(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t)|\, \varphi]$$
$$\boldsymbol{\Upsilon}(\boldsymbol{\rho}) = E[vec(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t - \mathbf{I}_N)vec'(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t - \mathbf{I}_N)|\, \varphi]$$

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* $\boldsymbol{\rho} = \mathbf{0}$ *to denote normality.*

2. *If* $\boldsymbol{\varepsilon}^*_t|I_{t-1}; \boldsymbol{\phi}_0$ *is i.i.d.* $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\rho}_0)$ *with* $\kappa_0 < \infty$*, then (D17) reduces to*

$$\mathcal{K}(\kappa) = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & (\kappa+1)(\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \kappa vec(\mathbf{I}_N)vec'(\mathbf{I}_N) \end{bmatrix}, \tag{D18}$$

which only depends on the true distribution through the population coefficient of multivariate excess kurtosis κ_0*.*

D.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,

$$\begin{pmatrix} \check{\boldsymbol{\theta}}_T - \tilde{\boldsymbol{\theta}}_T \\ \check{\boldsymbol{\eta}}_T - \tilde{\boldsymbol{\eta}}_T \end{pmatrix} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) \\ \mathcal{I}'_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) & \mathcal{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0) \end{bmatrix}^{-1} \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \mathbf{s}_{\boldsymbol{\theta}t}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\eta}}_T) \\ \mathbf{s}_{\boldsymbol{\eta}t}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\eta}}_T) \end{bmatrix}.$$

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

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

where

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

and

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \tag{D19}$$

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 (D19) and its covariance matrix reduce to

$$\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\eta}t}(\boldsymbol{\phi}_0) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0) \cdot \left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathbf{e}_{rt}(\boldsymbol{\phi}_0)\right] \tag{D20}$$

and

$$\mathcal{P}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left[\mathrm{M}_{sr}(\boldsymbol{\eta}_0)\mathcal{M}_{rr}^{-1}(\boldsymbol{\eta}_0)\mathrm{M}_{sr}'(\boldsymbol{\eta}_0)\right], \tag{D21}$$

respectively, where

$$\begin{aligned} \mathbf{W}_s(\boldsymbol{\phi}_0) = \mathbf{Z}_d(\boldsymbol{\theta}_0)[\mathbf{0}', vec'(\mathbf{I}_N)]' = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0][\mathbf{0}', vec'(\mathbf{I}_N)]' \\ = E\left\{\frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}} vec[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)]\right|\boldsymbol{\phi}_0\Bigg\} = E[\mathbf{W}_{st}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0] = -E\left[\frac{\partial d_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right|\boldsymbol{\phi}_0\Bigg], \end{aligned} \tag{D22}$$

It is worth noting that the last summand of (D19) coincides with $\mathbf{Z}_d(\boldsymbol{\phi}_0)$ 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 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 $\boldsymbol{\varepsilon}_t^*$ 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 $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ 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 10 *When $\boldsymbol{\varepsilon}_t^*|I_{t-1}, \boldsymbol{\phi}$ 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:*

$$\mathring{\mathrm{s}}_{\theta t}(\boldsymbol{\phi}_0) = \mathbf{s}_{\theta t}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0), \boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right] - \frac{2}{(N+2)\kappa_0 + 2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N} - 1\right]\right\}, \tag{D23}$$

while the spherically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left\{\left[\frac{N+2}{N}\mathrm{M}_{ss}(\boldsymbol{\eta}_0) - 1\right] - \frac{4}{N\left[(N+2)\kappa_0 + 2\right]}\right\}. \tag{D24}$$

In the case of the univariate GARCH-M model (2), we estimate the model parameters using parametrisation (17), with the expressions for the score that appear in the proof of Proposition

6. On the other hand, we use the natural parametrisation of the multivariate market model in (3), so that $\boldsymbol{\theta}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')$, where $\boldsymbol{\omega} = vech(\boldsymbol{\Omega})$. Given the Jacobian matrices:

$$\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial(\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')} = \begin{pmatrix} \mathbf{I}_N & \mathbf{I}_N r_{Mt} & \mathbf{0} \end{pmatrix}, \tag{D25}$$

$$\frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial(\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{D}_N \end{pmatrix}, \tag{D26}$$

because

$$\frac{\partial vec(\boldsymbol{\Omega})}{\partial vech'(\boldsymbol{\Omega})} = \mathbf{D}_N,$$

the results in Supplemental Appendix D.2 immediately imply that

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

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

The last ingredient we need is

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

because

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

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 (2.21) in Fang, Kotz and Ng (1990) to write the density function of ς_t as

$$h(\varsigma_t; \boldsymbol{\eta}) = \frac{\pi^{N/2}}{\Gamma(N/2)} \varsigma_t^{N/2-1} \exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})], \tag{D27}$$

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 with $N = 5$, 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.

E The general case of non-spherical pseudo likelihoods

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

Let $f(\boldsymbol{\varepsilon}^*;\boldsymbol{\varrho})$ denote the assumed conditional density of $\boldsymbol{\varepsilon}_t^*$ given I_{t-1} and some shape parameters $\boldsymbol{\varrho}$. 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 $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \boldsymbol{\varrho}]$, $d_t(\boldsymbol{\theta}) = \ln|\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})|$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, 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

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}.$$

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

$$\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = -\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})]\frac{\partial vec[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}$$
$$= -\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}, \tag{E28}$$

where

$$\left.\begin{array}{l} \mathbf{Z}_{lt}(\boldsymbol{\theta}) = \partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) \\ \mathbf{Z}_{st}(\boldsymbol{\theta}) = \partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]/\partial \boldsymbol{\theta} \cdot [\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] \end{array}\right\}, \tag{E29}$$

it follows that

$$\mathbf{s}_{\boldsymbol{\theta} t}(\boldsymbol{\phi}) = [\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})]\left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{array}\right] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \tag{E30}$$
$$\mathbf{s}_{\boldsymbol{\varrho} t}(\boldsymbol{\phi}) = \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varrho} = \mathbf{e}_{rt}(\boldsymbol{\phi}),$$

with

$$\mathbf{e}_{dt}(\boldsymbol{\phi}) = \left[\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \end{array}\right] = \left[\begin{array}{c} -\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^*, \\ -vec\{\mathbf{I}_N + \partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^* \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\} \end{array}\right]. \tag{E31}$$

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 (E28) implies that

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \tag{E32}$$

because

$$d\mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = -d\{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varepsilon}^*\}. \tag{E33}$$

In turn,

$$d\mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = -dvec\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \cdot \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right]$$
$$= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]d\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} d\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \tag{E34}$$

implies that

$$\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} = \frac{\partial \mathbf{e}_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}}\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\}\frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$
$$\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right]\right\}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}. \tag{E35}$$

Finally, (E33) and (E34) trivially imply that

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

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

$$\mathbf{h}_{\boldsymbol{\theta\theta} t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'}$$
$$+ \left[\mathbf{e}_{lt}'(\boldsymbol{\phi}) \otimes \mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi}) \otimes \mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}, \tag{E36}$$
$$\mathbf{h}_{\boldsymbol{\theta\varrho} t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta})\partial \mathbf{e}_{lt}(\boldsymbol{\phi})/\partial \boldsymbol{\varrho}' + \mathbf{Z}_{st}(\boldsymbol{\theta})\partial \mathbf{e}_{st}(\boldsymbol{\phi})/\partial \boldsymbol{\varrho}', \tag{E37}$$
$$\mathbf{h}_{\boldsymbol{\varrho\varrho} t}(\boldsymbol{\phi}) = \partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \boldsymbol{\varrho}\partial \boldsymbol{\varrho}'.$$

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

$$dvech(\boldsymbol{\Sigma}_t) = [\mathbf{L}_N(\boldsymbol{\Sigma}_t^{1/2}\otimes\mathbf{I}_N)\mathbf{L}_N' + \mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{1/2})\mathbf{K}_{NN}\mathbf{L}_N']dvech(\boldsymbol{\Sigma}_t^{1/2}), \tag{E38}$$

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

$$\frac{\partial vec'[\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})] = \frac{1}{2}\frac{\partial vech'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\mathbf{F}_t\mathbf{L}_N(\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2}) \tag{E39}$$

is as follows:

1. From the expression for $\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}$ we can readily obtain $\partial vech'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}$ 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

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

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 $\partial vech' [\boldsymbol{\Sigma}_t(\boldsymbol{\theta})] / \partial \boldsymbol{\theta} \cdot \mathbf{F}_t$.

E.2 Asymptotic distribution

E.2.1 Under correct specification

Proposition 11 *If $\boldsymbol{\varepsilon}_t^*|;\boldsymbol{\phi}$ is i.i.d. $D(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho})$ with density $f(\boldsymbol{\varepsilon}^*, \boldsymbol{\varrho})$, then*

$$\begin{array}{rcl} \mathcal{I}_t(\boldsymbol{\phi}) & = & \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\varrho})\mathbf{Z}_t'(\boldsymbol{\theta}), \\ \mathbf{Z}_t(\boldsymbol{\theta}) & = & \left(\begin{array}{cc} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{array} \right) = \left(\begin{array}{ccc} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array} \right), \end{array}$$

and

$$\mathcal{M}(\boldsymbol{\varrho}) = \left[\begin{array}{cc} \mathcal{M}_{dd}(\boldsymbol{\varrho}) & \mathcal{M}_{dr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{dr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{array} \right] = \left[\begin{array}{ccc} \mathcal{M}_{ll}(\boldsymbol{\varrho}) & \mathcal{M}_{ls}(\boldsymbol{\varrho}) & \mathcal{M}_{lr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{ls}'(\boldsymbol{\varrho}) & \mathcal{M}_{ss}(\boldsymbol{\varrho}) & \mathcal{M}_{sr}(\boldsymbol{\varrho}) \\ \mathcal{M}_{lr}'(\boldsymbol{\varrho}) & \mathcal{M}_{sr}'(\boldsymbol{\varrho}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{array} \right],$$

with

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

and

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

E.2.2 Under misspecification

Proposition 12 *If (14) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varphi}_0$ is i.i.d. $(\mathbf{0}, \mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\psi}$ and the true shape parameters $\boldsymbol{\rho}$, but the distribution assumed for estimation purposes does not necessarily nest the true density, then the pseudo-true value of the feasible parametric ML estimator of $\boldsymbol{\phi} = (\boldsymbol{\psi}_c', \boldsymbol{\psi}_{im}', \boldsymbol{\psi}_{ic}', \boldsymbol{\varrho})'$, $\boldsymbol{\phi}_\infty$, is such that $\boldsymbol{\psi}_{c\infty}$ is equal to the true value $\boldsymbol{\psi}_{c0}$.*

Proof. We can directly work in terms of the $\boldsymbol{\psi}$ parameters thanks to our assumptions on the mapping $\mathbf{r}_g(.)$. Let us initially keep $\boldsymbol{\varrho}$ fixed to some admissible value. The parametric score vector for the remaining parameters will then be given by (E30), with $\mathbf{Z}_{\psi_{ic}lt}(\psi) = \mathbf{0}$ and $\mathbf{Z}_{\psi_{im}st}(\psi) = \mathbf{0}$.

Since we are systematically working with lower triangular square root decompositions, we can write

$$\begin{aligned} \mathbf{Z}_{\psi_c st}(\psi) &= \partial vech'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\psi_c)]/\partial \psi_c \cdot \mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2} \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\psi_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}], \\ \mathbf{Z}_{\psi_{ic}s}(\psi) &= \partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})/\partial \psi_{ic} \cdot \mathbf{L}_N[\mathbf{I}_N \otimes \boldsymbol{\Psi}_{ic}^{-1/2\prime}]. \end{aligned}$$

Given that $\boldsymbol{\Psi}_{ic}^{1/2\prime}$ is upper triangular, $\boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond -1/2}(\psi_c)$ is lower triangular and $\mathbf{I}_N$ is diagonal, Theorem 5.7.i in Magnus (1988) implies that

$$\begin{aligned} [\boldsymbol{\Psi}_{ic}^{1/2\prime} \otimes \boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond -1/2}(\psi_c)]\mathbf{L}_N' &= \mathbf{L}_N'\mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2\prime} \otimes \boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Sigma}_t^{\diamond -1/2}(\psi_c)]\mathbf{L}_N', \\ (\mathbf{I}_N \otimes \boldsymbol{\Psi}_{ic}^{-1/2})\mathbf{L}_N' &= \mathbf{L}_N'\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Psi}_{ic}^{-1/2})\mathbf{L}_N', \end{aligned}$$

whence

$$\begin{aligned} \mathbf{Z}_{\psi_c st}(\psi) &= \frac{\partial vech'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\psi_c)]}{\partial \psi_c}\mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2} \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\psi_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}]\mathbf{L}_N'\mathbf{L}_N, \\ \mathbf{Z}_{\psi_{ic}s}(\psi) &= \frac{\partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})}{\partial \psi_{ic}}\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Psi}_{ic}^{-1/2\prime})\mathbf{L}_N'\mathbf{L}_N. \end{aligned}$$

As a result,

$$\begin{aligned} \mathbf{s}_{\psi_{ic}t}(\psi, \varrho) &= -\frac{\partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})}{\partial \psi_{ic}}\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Psi}_{ic}^{-1/2\prime})\mathbf{L}_N' vech\left\{\mathbf{I}_N + \frac{\partial \ln f[\varepsilon_t^*(\psi); \varrho]}{\partial \varepsilon^*}\varepsilon_t^{*\prime}(\psi)\right\} \\ \mathbf{s}_{\psi_{im}t}(\psi, \varrho) &= -\boldsymbol{\Psi}_{ic}^{-1/2\prime}\frac{\partial \ln f[\varepsilon_t^*(\psi); \varrho]}{\partial \varepsilon^*} \end{aligned}$$

and

$$\mathbf{s}_{\psi_c t}(\psi, \varrho) = \left\{\frac{\partial \boldsymbol{\mu}_t^{\diamond\prime}(\psi_c)}{\partial \psi_c} + \frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\psi_c)]}{\partial \psi_c}(\psi_{im} \otimes \mathbf{I}_N)\right\}\boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\psi_c)\mathbf{s}_{\psi_{im}t}(\psi, \varrho)$$
$$-\frac{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond 1/2}(\psi_c)]}{\partial \psi_c} \cdot \mathbf{L}_N[\boldsymbol{\Psi}_{ic}^{1/2} \otimes \boldsymbol{\Sigma}_t^{\diamond -1/2\prime}(\psi_c)\boldsymbol{\Psi}_{ic}^{-1/2\prime}]\mathbf{L}_N' vech\left\{\mathbf{I}_N + \frac{\partial \ln f[\varepsilon_t^*(\psi); \varrho]}{\partial \varepsilon^*}\varepsilon_t^{*\prime}(\psi)\right\}$$

since $vech(\mathbf{A}) = \mathbf{L}_N vec(\mathbf{A})$ for any $N \times N$ square matrix $\mathbf{A}$ regardless of its structure.

Let $\psi_{im\infty}(\varrho)$ and $\psi_{ic\infty}(\varrho)$ denote the solution to the implicit system of $N + N(N+1)/2$ equations (A13), which we assume is such that $\boldsymbol{\Psi}_{ic\infty}(\varrho)$ is p.d. Given the expression for $\varepsilon_t^*(\psi)$ in (A14), we can immediately see that $\varepsilon_t^*(\psi_{c0}, \psi_{im}, \psi_{ic})$ will be $i.i.d.[\boldsymbol{\Psi}_{ic}^{-1/2}(\psi_{im0} - \psi_{im}), \boldsymbol{\Psi}_{ic}^{-1/2}\boldsymbol{\Psi}_{ic0}\boldsymbol{\Psi}_{ic}^{-1/2\prime}]$ conditional on I_{t-1}. This, together with the full rank of $\boldsymbol{\Psi}_{ic}^{-1/2\prime}$ implies that

$$E\left[\left.\frac{\partial \ln f[\varepsilon_t^*[\psi_{c0}, \psi_{im\infty}(\varrho), \psi_{ic\infty}(\varrho)]; \varrho]}{\partial \varepsilon^*}\right| I_{t-1}; \varphi_0\right] = \mathbf{0}.$$

In addition, we know from Theorem 5.6 in Magnus (1988) that the matrix

$$\mathbf{L}_N(\mathbf{I}_N \otimes \boldsymbol{\Psi}_{ic}^{-1/2\prime})\mathbf{L}_N'$$

will be upper triangular of full rank. Similarly, given that we have defined $\boldsymbol{\psi}_{ic} = vech(\boldsymbol{\Psi}_{ic})$, the matrix $\partial vech'(\boldsymbol{\Psi}_{ic}^{1/2})/\partial\boldsymbol{\psi}_{ic}$ would also be of full rank in view of the discussion that follows expression (E38).

As a result, we will also have that

$$vech\left\{E\left[\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})]; \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})]\right| I_{t-1}; \boldsymbol{\varphi}_0\right]\right\} = \mathbf{0}.$$

Consequently,

$$E\{\mathbf{s}_{\boldsymbol{\psi}t}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}]|I_{t-1}; \boldsymbol{\varphi}_0\} = \mathbf{0}, \tag{E40}$$

which confirms that $\boldsymbol{\psi}_{c0}$, $\boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho})$ and $\boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho})$ will be the pseudo-true values corresponding to a restricted PML estimator that keeps $\boldsymbol{\varrho}$ fixed.

If we define $\boldsymbol{\varrho}_\infty$ as the solution to the q equations

$$E\{\mathbf{s}_{\boldsymbol{\varrho}t}[\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}), \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho}), \boldsymbol{\varrho}]|\boldsymbol{\varphi}_0\} = \mathbf{0},$$

which we assume lies in the interior of the admissible parameter space, then it is clear that $\boldsymbol{\psi}_{c0}$, $\boldsymbol{\psi}_{im\infty} = \boldsymbol{\psi}_{im\infty}(\boldsymbol{\varrho}_\infty)$, $\boldsymbol{\psi}_{ic\infty} = \boldsymbol{\psi}_{ic\infty}(\boldsymbol{\varrho}_\infty)$ and $\boldsymbol{\varrho}_\infty$ will be the pseudo-true values of the parameters corresponding to an unrestricted PMLE that also estimates $\boldsymbol{\varrho}$. $\square$

If we further assume that the true conditional mean of $\mathbf{y}_t$ is $\mathbf{0}$, and this restriction is imposed in estimation, then $\boldsymbol{\psi}_{im}$ becomes unnecessary, thereby generalising the second part of Theorem 1 in Newey and Steigerwald (1997).

The next result, which extends propositions 2 and 4, contains the ingredients necessary to compute the joint asymptotic covariance matrix of the consistent estimators $\boldsymbol{\psi}_{im}(\hat{\boldsymbol{\psi}}_{cT})$ and $\boldsymbol{\psi}_{ic}(\hat{\boldsymbol{\psi}}_{cT})$ defined in (21) and (22), respectively, and $\hat{\boldsymbol{\phi}}_T$:

Proposition 13 *If (14) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1}; \boldsymbol{\varphi}_0$ is i.i.d. $(\mathbf{0}, \mathbf{I}_N)$, where $\boldsymbol{\varphi}$ includes $\boldsymbol{\psi}$ and the true shape parameters, but the distribution assumed for estimation purposes does not necessarily nest the true density, then:*

1.

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathbf{0} \\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\boldsymbol{\phi}} & \mathcal{A}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} \end{pmatrix} = \begin{pmatrix} \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{A}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{im}} & \mathcal{A}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_{ic}\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{A}'_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_{ic}\boldsymbol{\psi}_{ic}} & \mathcal{A}_{\boldsymbol{\psi}_{im}\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}'_{\boldsymbol{\psi}_c\boldsymbol{\varrho}} & \mathcal{A}'_{\boldsymbol{\psi}_{im}\boldsymbol{\varrho}} & \mathcal{A}'_{\boldsymbol{\psi}_{ic}\boldsymbol{\varrho}} & \mathcal{A}_{\boldsymbol{\varrho}\boldsymbol{\varrho}} & \mathbf{0} & \mathbf{0} \\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_{im}\boldsymbol{\psi}_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{A}_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathbf{0} \\ \mathcal{A}_{\bar{\boldsymbol{\psi}}_{ic}\boldsymbol{\psi}_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{A}_{\bar{\boldsymbol{\psi}}_{ic}\bar{\boldsymbol{\psi}}_{ic}} \end{pmatrix},$$

$$\mathcal{B} = \begin{pmatrix} \mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathcal{B}_{\boldsymbol{\phi}\bar{\boldsymbol{\psi}}_i} \\ \mathcal{B}'_{\boldsymbol{\phi}\bar{\boldsymbol{\psi}}_i} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_i\bar{\boldsymbol{\psi}}_i} \end{pmatrix} = \begin{pmatrix} \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_c} & \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_c\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{im}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\boldsymbol{\psi}_{ic}} & \mathcal{B}'_{\boldsymbol{\psi}_{im}\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\boldsymbol{\psi}_{ic}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\boldsymbol{\eta}} & \mathcal{B}'_{\boldsymbol{\psi}_{im}\boldsymbol{\eta}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\eta}\boldsymbol{\eta}} & \mathcal{B}_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\psi}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{ic}} \\ \mathcal{B}'_{\boldsymbol{\psi}_c\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{im}} & \mathcal{B}'_{\boldsymbol{\psi}_{ic}\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}'_{\boldsymbol{\eta}\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}'_{\bar{\boldsymbol{\psi}}_{im}\bar{\boldsymbol{\psi}}_{ic}} & \mathcal{B}_{\bar{\boldsymbol{\psi}}_{ic}\bar{\boldsymbol{\psi}}_{ic}} \end{pmatrix},$$

with detailed expressions for all the elements in the proof.

2. *If in addition (16) holds, then both $\mathcal{A}$ and $\mathcal{B}$ become block diagonal between $\boldsymbol{\psi}_c$ and $(\boldsymbol{\psi}_{im}, \boldsymbol{\psi}_{ic}, \boldsymbol{\varrho}, \bar{\boldsymbol{\psi}}_{im}, \bar{\boldsymbol{\psi}}_{ic})$.*

Proof. To obtain the asymptotic distribution of the unrestricted pseudo ML estimators $\hat{\boldsymbol{\psi}}_T$ and $\hat{\boldsymbol{\varrho}}_T$, we need the asymptotic covariance matrix of the average scores as well as the expected value of the average Hessian matrix evaluated at the pseudo true values $\boldsymbol{\phi}'_\infty = (\boldsymbol{\psi}'_{c0}, \boldsymbol{\psi}'_{im\infty}, \boldsymbol{\psi}'_{ic\infty}, \boldsymbol{\varrho}'_\infty)$. Given that $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}_\infty)$ only depends on $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_{c0}, \boldsymbol{\psi}_{im\infty}, \boldsymbol{\psi}_{ix\infty})$, which is *i.i.d.* over time, it follows that

$$E[\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] = \mathbf{0}, \tag{E41}$$

which in conjunction with (9) proves the martingale difference nature of the spherical score evaluated at the pseudo-true values. As a result, we only need the contemporaneous covariance matrix of the component of the score corresponding to the t^{th} observation, which in turn depends on the contemporaneous covariance matrix of $\mathbf{e}_{dt}(\boldsymbol{\phi}_\infty)$ and $\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)$. Given the expression for $\mathbf{e}_{dt}(\boldsymbol{\phi}_\infty)$ in (E31), it immediately follows that

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^{*\prime}}\bigg|\boldsymbol{\varphi}_0\right\} = \mathcal{M}^O_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \tag{E42}$$

$$\begin{aligned} E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\phi}_\infty)] &= E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\right. \\ &\left.\times vec'\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\psi}_\infty)\right\}\bigg|\boldsymbol{\varphi}_0\right\} = \mathcal{M}^O_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \end{aligned} \tag{E43}$$

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\phi}_\infty)] &= E\left\{vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\psi}_\infty)\right\}\right. \\ &\left.\times vec'\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\psi})\right\}\bigg|\boldsymbol{\varphi}_0\right\} = \mathcal{M}^O_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0). \end{aligned} \tag{E44}$$

Similarly,

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = E\left\{-\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}\bigg|\boldsymbol{\varphi}_0\right\} = \mathcal{M}^O_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \tag{E45}$$

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)] &= E\left\{-vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\cdot\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}\right\}\right. \\ &\left.\times[\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\psi}_\infty)\otimes\mathbf{I}_N])\big|\,\boldsymbol{\varphi}_0\right\} = \mathcal{M}^O_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{aligned} \tag{E46}$$

and

$$E[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{rt}(\boldsymbol{\phi}_\infty)] = E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}}\frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}|\boldsymbol{\varphi}_0\right\} = \mathcal{M}^O_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0). \tag{E47}$$

Hence, we will have that $\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}} = E[\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)]$, where

$$\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = V[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] = \mathbf{Z}_t(\boldsymbol{\psi}_\infty)\mathcal{M}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_t(\boldsymbol{\psi}_\infty), \tag{E48}$$

and $\mathcal{M}^O(\boldsymbol{\phi};\boldsymbol{\varphi}) = V[\mathbf{e}_t(\boldsymbol{\phi})|\boldsymbol{\varphi}]$.

Tedious algebra shows that $\mathcal{A}_{\phi\phi} = E[\mathcal{A}_t(\phi_\infty;\varphi_0)]$, where

$$\mathcal{A}_t(\phi_\infty;\varphi_0) = -E[\mathbf{h}_t(\phi_\infty)|I_{t-1};\varphi_0] = \mathbf{Z}_t(\psi_\infty)\mathcal{M}^H(\phi_\infty;\varphi_0)\mathbf{Z}_t(\psi_\infty), \tag{E49}$$

and $\mathcal{M}^H(\phi_\infty;\varphi_0)$ contains the following elements

$$\mathcal{M}^H_{ll}(\phi_\infty;\varphi_0) = E\left\{\partial^2 \ln f[\varepsilon^*_t(\psi_\infty);\varrho_\infty]/\partial\varepsilon^*\partial\varepsilon^{*\prime}\big|\,\varphi_0\right\}, \tag{E50}$$

$$\mathcal{M}^H_{ls}(\phi;\varphi) = E\left\{\partial^2 \ln f[\varepsilon^*_t(\psi);\varrho]/\partial\varepsilon^*\partial\varepsilon^{*\prime}\cdot[\varepsilon^{*\prime}_t(\psi)\otimes\mathbf{I}_N])\big|\,\varphi\right\}, \tag{E51}$$

$$\mathcal{M}^H_{ss}(\phi;\varphi) = E\left\{[\varepsilon^*_t(\psi)\otimes\mathbf{I}_N]\cdot\partial^2 \ln f[\varepsilon^*_t(\psi);\varrho]/\partial\varepsilon^*\partial\varepsilon^{*\prime}\cdot[\varepsilon^{*\prime}_t(\psi)\otimes\mathbf{I}_N]|\varphi\right\} - \mathbf{K}_{NN} \tag{E52}$$

$$\mathcal{M}^H_{lr}(\phi;\varphi) = -E\left[\partial^2 \ln f[\varepsilon^*_t(\psi);\varrho]/\partial\varepsilon^*\partial\varrho'|\varphi\right], \tag{E53}$$

$$\mathcal{M}^H_{sr}(\phi;\varphi) = -E\left[[\varepsilon^*_t(\psi)\otimes\mathbf{I}_N]\partial^2 \ln f[\varepsilon^*_t(\psi);\varrho]/\partial\varepsilon^*\partial\varrho'|\varphi\right\}, \tag{E54}$$

and

$$\mathcal{M}^H_{rr}(\phi;\varphi) = -E\left\{\partial^2 \ln f[\varepsilon^*_t(\psi);\varrho]/\partial\varrho\partial\varrho'|\varphi\right\}. \tag{E55}$$

Let us now turn to our consistent estimators of ψ_{ic} and ψ_{im}. The fact that the Gaussian pseudo score for these parameters are influence functions that only depend on ψ_c and $\bar{\psi}_i$ trivially implies that

$$\frac{\partial \mathbf{s}_{\psi_i t}(\psi_c,\bar{\psi}_i;\mathbf{0})}{\partial\psi_i'} = \mathbf{0} \text{ and } \frac{\partial \mathbf{s}_{\psi_i t}(\psi_c,\bar{\psi}_i;\mathbf{0})}{\partial\varrho'} = \mathbf{0}.$$

For analogous reasons,

$$\frac{\partial \mathbf{s}_{\psi_c t}(\psi_c,\psi_i,\varrho)}{\partial\bar{\psi}_i'} = \mathbf{0},\ \frac{\partial \mathbf{s}_{\psi_i t}(\psi_c,\psi_i,\varrho)}{\partial\bar{\psi}_i'} = \mathbf{0}, \frac{\partial \mathbf{s}_{\varrho t}(\psi_c,\psi_i,\varrho)}{\partial\bar{\psi}_i'} = \mathbf{0},$$

We will also have that

$$\frac{\partial \mathbf{s}'_{\psi_i t}(\psi_c,\bar{\psi}_i;\mathbf{0})}{\partial\psi_c} = \mathbf{h}'_{\psi_c\psi_i t}(\psi,\mathbf{0})$$

and

$$\frac{\partial \mathbf{s}'_{\psi_i t}(\psi_c,\bar{\psi}_i;\mathbf{0})}{\partial\bar{\psi}_i} = \mathbf{h}'_{\psi_i\psi_i t}(\psi,\mathbf{0}).$$

But since we are evaluating these expressions at consistent estimators of ψ, we will have that $\varepsilon^*_t(\psi_0) = \varepsilon^*_t$, whence we can obtain the remaining elements of $\mathcal{A}$. In particular, given that (A14) implies that for a fixed value of ψ_c we could understand the Gaussian log-likelihood function of $\mathbf{y}_t$ as a Gaussian log-likelihood for the pseudo-standardised residuals $\varepsilon^\diamond_t(\psi_c)$ with mean ψ_{im} and covariance matrix Ψ_{ic}, it immediately follows that $\mathcal{A}_{\bar{\psi}_{im}\bar{\psi}_{ic}} = \mathbf{0}$.

Next, we need to find out the asymptotic covariance matrix of the sample averages of $\mathbf{s}_{\psi_{ic}t}(\psi_0;\mathbf{0})$ and $\mathbf{s}_{\psi_{im}t}(\psi_0;\mathbf{0})$, as well as their asymptotic covariances with the sample averages of $\mathbf{s}_{\psi t}(\phi_\infty)$ and $\mathbf{s}_{\varrho t}(\phi_\infty)$, which coincide with contemporaneous variance and covariances of these influence functions because they are martingale difference sequences. In turn, they depend on the covariance matrix of $\mathbf{e}_{dt}(\psi_0,\mathbf{0})$, which is given by (D17), as well as on the covariances of this vector with $\mathbf{e}_{dt}(\phi_\infty)$ and $\mathbf{e}_{rt}(\phi_\infty)$. Specifically, the required additional elements are

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\psi}_0,\mathbf{0})] = E\left\{\left.\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0)\right| \boldsymbol{\varphi}_0\right\} = \mathcal{M}_{l\bar{l}}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \tag{E56}$$

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\psi}_0,\mathbf{0})] &= E\left\{\left. vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_\infty)\right\}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0)\right| \boldsymbol{\varphi}_0\right\} \\ &= \mathcal{M}_{s\bar{l}}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \end{aligned} \tag{E57}$$

$$E[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{lt}(\boldsymbol{\psi}_0,\mathbf{0})] = E\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0)|\boldsymbol{\varphi}_0\right\} = \mathcal{M}_{r\bar{l}}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \tag{E58}$$

and

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\psi}_0,\mathbf{0})] = E\left\{\left.\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*} vec'\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) - \mathbf{I}_N\right]\right| \boldsymbol{\varphi}_0\right\} = \mathcal{M}_{l\bar{s}}^O(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0), \tag{E59}$$

$$\begin{aligned} E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{st}(\boldsymbol{\psi}_0,\mathbf{0})] &= E\left\{ vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_\infty)\right\}\right. \\ &\left.\times vec'\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) - \mathbf{I}_N\right] \right| \boldsymbol{\varphi}\right\} = \mathcal{M}_{s\bar{s}}^O(\boldsymbol{\phi};\boldsymbol{\varphi}), \end{aligned} \tag{E60}$$

$$E[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)\mathbf{e}'_{ct}(\boldsymbol{\psi}_0,\mathbf{0})] = E\left\{\left.\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'} vec'\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\psi}_0) - \mathbf{I}_N\right]\right| \boldsymbol{\varphi}\right\} = \mathcal{M}_{r\bar{s}}^O(\boldsymbol{\phi};\boldsymbol{\varphi}). \tag{E61}$$

Finally, we can tediously show that the conditions for block-diagonality of the expected value of the Hessian and the covariance matrix of the score are that $E[\mathbf{Z}_{\psi_c lt}(\boldsymbol{\psi}_\infty)|\boldsymbol{\varphi}_0]$ and $E[\mathbf{Z}_{\psi_c st}(\boldsymbol{\psi}_\infty)|\boldsymbol{\varphi}_0]$ are both 0. But given that

$$\begin{aligned} \mathbf{Z}_{\psi_c lt}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic}) &= \left[\partial\boldsymbol{\mu}_t^{\diamond\prime}(\boldsymbol{\psi}_{c0})/\partial\boldsymbol{\psi}_c\cdot\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_{c0})\right]\boldsymbol{\Psi}_{ic}^{-1/2\prime} \\ &\quad+\left\{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond1/2}(\boldsymbol{\psi}_{c0})]/\partial\boldsymbol{\psi}_c\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_{c0})]\right\}(\boldsymbol{\psi}_{im}\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime}), \\ \mathbf{Z}_{\psi_c st}(\boldsymbol{\psi}_{c0},\boldsymbol{\psi}_{im},\boldsymbol{\psi}_{ic}) &= \left\{\partial vec'[\boldsymbol{\Sigma}_t^{\diamond1/2}(\boldsymbol{\psi}_{c0})]/\partial\boldsymbol{\psi}_c\cdot[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{\diamond-1/2\prime}(\boldsymbol{\psi}_{c0})]\right\}(\boldsymbol{\Psi}_{ic}^{1/2}\otimes\boldsymbol{\Psi}_{ic}^{-1/2\prime}), \end{aligned}$$

those conditions will be satisfied if (16) holds in view of the full rank of $\boldsymbol{\Psi}_{ic}$. □

E.3 Semiparametric estimators

In Supplemental Appendix D.5 we interpreted the last summand of (D19) 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 14 *If $\varepsilon_t^*|I_{t-1};\boldsymbol{\theta},\boldsymbol{\rho}$ is i.i.d. $D(\mathbf{0},\mathbf{I}_N,\boldsymbol{\rho})$ with density function $f(\varepsilon_t^*;\boldsymbol{\rho})$, 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,*

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

and the matrix of third and fourth order central moments $\mathcal{K}(\boldsymbol{\rho})$ in (D17) are bounded, then the semiparametric efficient score will be given by:

$$\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\rho})\left[\mathbf{e}_{dt}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathcal{K}(0)\mathcal{K}^+(\boldsymbol{\rho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right], \tag{E62}$$

while the semiparametric efficiency bound is

$$\ddot{\mathcal{S}}(\boldsymbol{\phi}) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathbf{Z}_d(\boldsymbol{\theta},\boldsymbol{\rho})\left[\mathcal{M}_{dd}(\boldsymbol{\theta},\boldsymbol{\rho}) - \mathcal{K}(0)\mathcal{K}^+(\boldsymbol{\rho})\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta},\boldsymbol{\rho}), \tag{E63}$$

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

In the case of the univariate GARCH-M model (2), we estimate the model parameters using parametrisation (17), with the expressions for the score that appear in the proof of Proposition 6. On the other hand, we use again the natural parametrisation of the multivariate market model in (3). As a result, the Jacobian matrix (D25) remains relevant, so that

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

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

If we choose the Cholesky decomposition, we can use expression (E39) in Supplemental Appendix E.1 to obtain

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

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

$$\mathcal{K}(\boldsymbol{\rho}) = \begin{pmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{D}_N\end{pmatrix}\left[\begin{array}{cc}\mathbf{I}_N & E[\varepsilon_t^* vech'(\varepsilon_t^*\varepsilon_t^{*\prime})|\boldsymbol{\varphi}]\\ E[vech(\varepsilon_t^*\varepsilon_t^{*\prime})\varepsilon_t^{\prime *}|\boldsymbol{\varphi}] & E[vech(\varepsilon_t^*\varepsilon_t^{*\prime})vech'(\varepsilon_t^*\varepsilon_t^{*\prime})-\mathbf{I}_N|\boldsymbol{\varphi}]\end{array}\right]\begin{pmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{D}_N'\end{pmatrix}$$

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

$$\begin{aligned}\mathcal{K}(0)\mathcal{K}^+(\boldsymbol{\rho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0}) &= \begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & 2\mathbf{D}^{+\prime}\end{pmatrix}\\ &\times\left[\begin{array}{cc}\mathbf{I}_N & E[\varepsilon_t^* vech'(\varepsilon_t^*\varepsilon_t^{*\prime})|\boldsymbol{\varphi}]\\ E[vech(\varepsilon_t^*\varepsilon_t^{*\prime})\varepsilon_t^{\prime *}|\boldsymbol{\varphi}] & E[vech(\varepsilon_t^*\varepsilon_t^{*\prime})vech'(\varepsilon_t^*\varepsilon_t^{*\prime})-\mathbf{I}_N|\boldsymbol{\varphi}]\end{array}\right]^{-1}\left[\begin{array}{c}\varepsilon_t^*\\ vech(\varepsilon_t^*\varepsilon_t^{*\prime}-\mathbf{I})\end{array}\right].\end{aligned}$$

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 $N = 5$.

Additional references

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

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

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

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

Gonzalez-Rivera, G. and Drost, F.C. (1999): "Efficiency comparisons of maximum-likelihood-based estimators in GARCH models", *Journal of Econometrics* 93, 93-111.

Hodgson, D.J. and Vorkink, K.P. (2003): "Efficient estimation of conditional asset pricing models", *Journal of Business and Economic Statistics* 21, 269-283.

Kotz, S. (1975): "Multivariate distributions at a cross-road", in G. P. Patil, S. Kotz and J.K. Ord (eds.) *Statistical distributions in scientific work*, vol. I, 247-270, Reidel.

Lange, K.L., Little, R.J.A., and Taylor, J.M.G. (1989): "Robust statistical modeling using the t distribution", *Journal of the American Statistical Association* 84, 881-896.

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