Supplemental Appendices for

Discrete Mixtures of Normals Pseudo Maximum Likelihood Estimators of Structural Vector Autoregressions

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>

October 2021

B PML estimators with cross-sectionally independent shocks

In this appendix, we derive analytical expressions for the conditional variance of the score and the expected value of the Hessian of SVAR models with cross-sectionally independent non-Gaussian shocks when the distributions assumed for estimation purposes may well be misspecified. In addition, we consider some useful reparametrisations.

B.1 Log-likelihood, its score and Hessian

Given the linear mapping between structural shocks and reduced form innovations, the contribution to the conditional log-likelihood function from observation t ($t = 1, \ldots, T$) will be given by

$$l_t(\mathbf{y}_t; \boldsymbol{\varphi}) = -\ln|\mathbf{C}| + l[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1] + \ldots + l[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N], \tag{B1}$$

where $\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}) = \mathbf{C}^{-1}(\mathbf{y}_t - \boldsymbol{\tau} - \mathbf{A}_1\mathbf{y}_{t-1} - \ldots - \mathbf{A}_p\mathbf{y}_{t-p})$ and $l(\varepsilon^*_i; \boldsymbol{\varrho}_i) = \ln f(\varepsilon^*_i; \boldsymbol{\varrho}_i)$ is the log of the univariate density function of ε^*_i, which we assume twice continuously differentiable with respect to both its arguments, although this is stronger than necessary, as the Laplace example illustrates.

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. Given that the mean vector and covariance matrix of (1) conditional on I_{t-1} are

$$\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\tau} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p}, && \text{(B2a)}\\
\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{C}\mathbf{C}', && \text{(B2b)}
\end{aligned}$$

respectively, we can use the expressions in Supplemental Appendix D.1 of Fiorentini and Sentana (2021b) with $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}) = \mathbf{C}$ to show that

$$\frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\frac{\partial vec'(\mathbf{C})}{\partial\boldsymbol{\theta}} vec(\mathbf{C}^{-1\prime}) = -\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} vec(\mathbf{C}^{-1\prime}) = -\mathbf{Z}'_{st}(\boldsymbol{\theta}) vec(\mathbf{I}_N) \tag{B3}$$

and

$$\begin{aligned}
\frac{\partial\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= -\mathbf{C}^{-1}\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}) \otimes \mathbf{C}^{-1}]\frac{\partial vec(\mathbf{C})}{\partial\boldsymbol{\theta}'} \\
&= -\{\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}'_{st}(\boldsymbol{\theta})\},
\end{aligned} \tag{B4}$$

where

$$\mathbf{Z}_{lt}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2 \times N} \end{pmatrix} \mathbf{C}^{-1\prime}, \tag{B5}$$

$$\mathbf{Z}_{st}(\boldsymbol{\theta}) = \frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} [\mathbf{I}_N \otimes \boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = \begin{pmatrix} \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{N^2 \times N^2} \\ \vdots \\ \mathbf{0}_{N^2 \times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime}), \tag{B6}$$

which confirms that the conditional mean and variance parameters are variation free. In addition,

$$\begin{aligned} \mathbf{s}_t(\boldsymbol{\phi}) &= \begin{bmatrix} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathbf{e}_{dt}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_t(\boldsymbol{\theta}) \mathbf{e}_t(\boldsymbol{\phi}), \end{aligned} \tag{B7}$$

where

$$\mathbf{e}_{lt}(\boldsymbol{\phi}) = -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} = - \begin{bmatrix} \partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]/\partial \varepsilon_1^* \\ \partial \ln f_2[\varepsilon_{2t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_2]/\partial \varepsilon_2^* \\ \vdots \\ \partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]/\partial \varepsilon_N^* \end{bmatrix}, \tag{B8}$$

$$\begin{aligned} \mathbf{e}_{st}(\boldsymbol{\phi}) &= -vec\left\{ \mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \right\} \\ &= -vec \left\{ \begin{array}{ccc} 1 + \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*} \varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*} \varepsilon_{Nt}^*(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*} \varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & 1 + \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*} \varepsilon_{Nt}^*(\boldsymbol{\theta}) \end{array} \right\} \end{aligned} \tag{B9}$$

and

$$\mathbf{e}_{rt}(\boldsymbol{\phi}) = \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}} = \left\{ \begin{array}{c} \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1} \\ \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N} \end{array} \right\} = \begin{bmatrix} \mathbf{e}_{r_1 t}(\boldsymbol{\phi}) \\ \mathbf{e}_{r_2 t}(\boldsymbol{\phi}) \\ \vdots \\ \mathbf{e}_{r_N t}(\boldsymbol{\phi}) \end{bmatrix} \tag{B10}$$

by virtue of the cross-sectional independence of the shocks, so that the derivatives involved correspond to the assumed univariate densities.

These expressions simplify when the assumed distribution of the shocks is symmetric. Some popular examples are Student t, DSMN and Laplace. In all three cases, we can write the scores as

$$-\frac{\partial \ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]}{\partial \varepsilon_i^*} = \delta[\varepsilon_{it}^{*2}(\boldsymbol{\theta}); \boldsymbol{\varrho}_i] \varepsilon_{it}^*(\boldsymbol{\theta}),$$

where $\delta[\varepsilon_{it}^{*2}(\boldsymbol{\theta}); \boldsymbol{\varrho}_i]$, a scalar function of the square of $\varepsilon_{it}^*(\boldsymbol{\theta})$, is defined in (A3).

Specifically, the log-density of a univariate Student t random variable with 0 mean, unit variance and degrees of freedom $\nu_i = \eta_i^{-1}$ is given by

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

with

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

so that

$$\delta[\varepsilon_{it}^{*2}(\boldsymbol{\theta});\boldsymbol{\varrho}_i] = \frac{\eta_i+1}{1-2\eta_i+\eta_i\varepsilon_{it}^{*2}(\boldsymbol{\theta})},$$

which converges to 1 as the Student t approaches the normal distribution.

In contrast, it becomes

$$l[\varepsilon_{it}^*(\boldsymbol{\theta});\lambda_i,\kappa_i] = c(\lambda_i,\kappa_i) + \log\left[\lambda_i\exp\left(-\frac{\varepsilon_{it}^{*2}(\boldsymbol{\theta})}{\kappa_i^*}\right) + (1-\lambda_i)\kappa_i^{-1/2}\exp\left(-\frac{\varepsilon_{it}^{*2}(\boldsymbol{\theta})}{\kappa_i^*\kappa_i}\right)\right]$$

for a two-component DSMN, with

$$\begin{aligned} c(\lambda_i,\kappa_i) &= -\frac{1}{2}\log\kappa_i^* - \log\Gamma\left(-\frac{1}{2}\right), \\ \kappa_i^* &= \frac{1}{\lambda_i+(1-\lambda_i)\kappa_i} \end{aligned}$$

and

$$\delta[\varepsilon_{it}^{*2}(\boldsymbol{\theta});\boldsymbol{\varrho}_i] = [\lambda_i + (1-\lambda_i)\kappa_i]\frac{\lambda_i\exp\left[-\frac{1}{2\kappa_i^*}\varepsilon_{it}^{*2}(\boldsymbol{\theta})\right] + (1-\lambda_i)\kappa_i^{-3/2}\exp\left[-\frac{1}{2\kappa_i^*\kappa_i}\varepsilon_{it}^{*2}(\boldsymbol{\theta})\right]}{\lambda_i\exp\left[-\frac{1}{2\kappa_i^*}\varepsilon_{it}^{*2}(\boldsymbol{\theta})\right] + (1-\lambda_i)\kappa_i^{-1/2}\exp\left[-\frac{1}{2\kappa_i^*\kappa_i}\varepsilon_{it}^{*2}(\boldsymbol{\theta})\right]}.$$

Finally, it will be

$$l[\varepsilon_{it}^*(\boldsymbol{\theta})] = -\log(2) - \sqrt{2}|\varepsilon_{it}^*(\boldsymbol{\theta})| = -\log(2) - \sqrt{2\varepsilon_{it}^{*2}(\boldsymbol{\theta})}$$

under the Laplace assumption, which introduces no additional shape parameter, so that

$$\delta[\varepsilon_{it}^{*2}(\boldsymbol{\theta});\boldsymbol{\varrho}_i] = \sqrt{\frac{2}{\varepsilon_{it}^{*2}(\boldsymbol{\theta})}} = \frac{\sqrt{2}}{|\varepsilon_{it}^*(\boldsymbol{\theta})|}.$$

Let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. Supplemental Appendix D.1 of Fiorentini and Sentana (2021b) implies that

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial\mathbf{e}_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} \\ + \left[\mathbf{e}_{lt}'(\boldsymbol{\phi})\otimes\mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi})\otimes\mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}, \end{aligned} \tag{B11}$$

where $\mathbf{Z}_{lt}(\boldsymbol{\theta})$ and $\mathbf{Z}_{st}(\boldsymbol{\theta})$ are given in (B5) and (B6), respectively. Therefore, we need to obtain $\partial vec(\mathbf{C}^{-1\prime})/\partial\boldsymbol{\theta}'$ and $\partial vec(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})/\partial\boldsymbol{\theta}'$.

Let us start with the former. Given that

$$dvec(\mathbf{C}^{-1\prime}) = -vec[\mathbf{C}^{-1\prime}d(\mathbf{C}')\mathbf{C}^{-1\prime}] = -(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1\prime})dvec(\mathbf{C}') = -(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1\prime})\mathbf{K}_{NN}dvec(\mathbf{C}),$$

where $\mathbf{K}_{NN}$ is the commutation matrix (see Magnus and Neudecker (2019)), we immediately get that

$$\frac{\partial vec(\mathbf{C}^{-1\prime})}{\partial\boldsymbol{\theta}'} = \begin{bmatrix} \mathbf{0}_{N^2\times(N+pN^2)} & -(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{bmatrix},$$

so that

$$\begin{aligned}
\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} &= \left[\mathbf{I}_N\otimes\begin{pmatrix}\mathbf{I}_N\\ \mathbf{y}_{t-1}\otimes\mathbf{I}_N\\ \vdots\\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N\\ \mathbf{0}_{N^2\times N}\end{pmatrix}\right]\frac{\partial vec(\mathbf{C}^{-1\prime})}{\partial\boldsymbol{\theta}'}\\
&= \left[\mathbf{I}_N\otimes\begin{pmatrix}\mathbf{I}_N\\ \mathbf{y}_{t-1}\otimes\mathbf{I}_N\\ \vdots\\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N\\ \mathbf{0}_{N^2\times N}\end{pmatrix}\right]\begin{bmatrix} \mathbf{0}_{N^2\times(N+pN^2)} & (\mathbf{C}^{-1}\otimes\mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{bmatrix}.
\end{aligned}$$

Similarly, given that

$$vec(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) = \{[(\mathbf{I}_N\otimes\mathbf{K}_{NN})(vec(\mathbf{I}_N)\otimes\mathbf{I}_N)]\otimes\mathbf{I}_N\}vec(\mathbf{C}^{-1\prime})$$

so that

$$\begin{aligned}
vec(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) &= ((\mathbf{I}_N\otimes\mathbf{K}_{NN})(vec(\mathbf{I}_N)\otimes\mathbf{I}_N)\otimes\mathbf{I}_N)dvec(\mathbf{C}^{-1\prime})\\
&= -\{[(\mathbf{I}_N\otimes\mathbf{K}_{NN})(vec(\mathbf{I}_N)\otimes\mathbf{I}_N)]\otimes\mathbf{I}_N\}(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1\prime})\mathbf{K}_{NN}dvec(\mathbf{C}),
\end{aligned}$$

we will have that

$$\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'} = \frac{\partial vec}{\partial\boldsymbol{\theta}'}\left[\begin{pmatrix}\mathbf{0}_{(N+pN^2)\times N^2}\\ \mathbf{I}_{N^2}\end{pmatrix}(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\right].$$

But

$$\begin{aligned}
&\left[\mathbf{I}_{N^2}\otimes\begin{pmatrix}\mathbf{0}_{(N+pN^2)\times N^2}\\ \mathbf{I}_{N^2}\end{pmatrix}\right]\frac{\partial vec(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})}{\partial\boldsymbol{\theta}'}\\
= &-\left[\mathbf{I}_{N^2}\otimes\begin{pmatrix}\mathbf{0}_{(N+pN^2)\times N^2}\\ \mathbf{I}_{N^2}\end{pmatrix}\right]\begin{bmatrix}\mathbf{0} & \{[(\mathbf{I}_N\otimes\mathbf{K}_{NN})(vec(\mathbf{I}_N)\otimes\mathbf{I}_N)]\otimes\mathbf{I}_N\}(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1\prime})\mathbf{K}_{NN}\end{bmatrix}.
\end{aligned}$$

In addition,

$$\frac{\partial\mathbf{e}_{lt}(\boldsymbol{\theta},\boldsymbol{\varrho})}{\partial\boldsymbol{\theta}'} = -\frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\frac{\partial\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = \frac{\partial^2\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \quad \text{(B12)}$$

and

$$\begin{aligned}
\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{ \mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right\} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\
&= \left\{ [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} \right] \right\} \\
&\quad \times \{ \mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \mathbf{Z}_{st}'(\boldsymbol{\theta}) \}. \qquad \text{(B13)}
\end{aligned}$$

The assumed independence across innovations implies that

$$\frac{\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{(\partial \varepsilon_1^*)^2} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{(\partial \varepsilon_N^*)^2} \end{bmatrix}, \qquad \text{(B14)}$$

which substantially simplifies the above expressions.

Moreover,

$$\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta}) \frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\varrho}'} + \mathbf{Z}_{st}(\boldsymbol{\theta}) \frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\varrho}'},$$

where

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

with

$$\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^* \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^* \partial \boldsymbol{\varrho}_N'} \end{bmatrix} \qquad \text{(B15)}$$

because of the cross-sectional independence assumption.

As for the shape parameters of the independent margins,

$$\mathbf{h}_{\boldsymbol{\varrho}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho} \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1 \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N \partial \boldsymbol{\varrho}_N'} \end{bmatrix}. \qquad \text{(B16)}$$

Finally, regarding the Jacobian term $-\ln|\mathbf{C}|$, we have that differentiating (B3) once more yields

$$-\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} dvec(\mathbf{C}^{-1\prime}) = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} (\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} dvec(\mathbf{C}),$$

so

$$\frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} \left[\begin{array}{cc} \mathbf{0}_{N^2 \times (N+pN^2)} & (\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{array} \right].$$

In the case of a restricted PMLE in which the elements of $\boldsymbol{\varrho}$ are fixed to some arbitrary parameter values $\bar{\boldsymbol{\varrho}}$, we would simply eliminate all the row and column blocks corresponding to $\boldsymbol{\varrho}$ from the expressions above.

B.2 The pseudo true values

In what follows, we maintain the assumptions that (i) $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ in (B2) are correctly specified and (ii) the true shocks $\boldsymbol{\varepsilon}_t^*$ are serially and cross-sectionally independent. Nevertheless, we continue to allow for misspecification of the marginal densities.

As usual, the pseudo true values of the parameters of a globally identified model, $\boldsymbol{\phi}_\infty$, are the unique values that maximise the expected value of the log-likelihood function over the admissible parameter space, which is a compact subset of $\mathbb{R}^{\dim(\boldsymbol{\phi})}$, where the expectation is taken with respect to the true distribution of the shocks. Under standard regularity conditions (see e.g. White (1982)), those pseudo true values will coincide with the values of the parameters that set to 0 the expected value of the pseudo-log likelihood score.

More formally, if we define $\boldsymbol{\varpi}_0$ as the true values of the shape parameters, and $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\varpi}_0)$, we would normally expect that

$$E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \mathbf{0}.$$

We have shown in Proposition 1 that the parameters $\mathbf{a}_j = vec(\mathbf{A}_j)$ $(j = 1, \ldots, p)$ and $\mathbf{j} = veco(\mathbf{J})$ are consistently estimated regardless of the true distribution. As a result, $\mathbf{a}_{j\infty} = \mathbf{a}_{j0}$ and $\mathbf{j}_\infty = \mathbf{j}_0$. In contrast, $\boldsymbol{\tau}$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$ may be inconsistently estimated, so that $\boldsymbol{\tau}_\infty \neq \boldsymbol{\tau}_0$ and $\boldsymbol{\psi}_\infty \neq \boldsymbol{\psi}_0$ in general. It is then easy to see that

$$\begin{aligned} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty) &= \boldsymbol{\Psi}_\infty^{-1}\mathbf{J}_0^{-1}(\mathbf{y}_t - \boldsymbol{\tau}_\infty - \mathbf{A}_{10}\mathbf{y}_{t-1} - \ldots - \mathbf{A}_{p0}\mathbf{y}_{t-p}) \\ &= \boldsymbol{\Psi}_\infty^{-1}[\mathbf{J}_0^{-1}(\boldsymbol{\tau}_0 - \boldsymbol{\tau}_\infty) + \boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t^*] = \boldsymbol{\Psi}_\infty^{-1}[(\boldsymbol{\upsilon}_0 - \boldsymbol{\upsilon}_\infty) + \boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t^*]. \end{aligned} \tag{B17}$$

Therefore, both $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)$ and $\mathbf{e}_t(\boldsymbol{\phi}_\infty)$ will be serially independent and not just martingale difference sequences. Moreover, given that

$$\mathbf{Z}(\boldsymbol{\theta}) = E[\mathbf{Z}_t(\boldsymbol{\theta})|\boldsymbol{\varphi}_0] = \begin{bmatrix} \mathbf{C}^{-1\prime} & \mathbf{0}_{N\times N^2} & \mathbf{0}_{N\times q} \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \vdots & \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{N^2\times N} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{q\times N} & \mathbf{0}_{q\times N^2} & \mathbf{I}_q \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_d(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} \quad \text{(B18)}$$

has full column rank,

$$E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|I_{t-1},\boldsymbol{\varphi}_0] = \mathbf{0} \quad \text{(B19)}$$

because

$$\mathbf{0} = E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = E\{E[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|I_{t-1},\boldsymbol{\varphi}_0]|\boldsymbol{\varphi}_0\} = \mathbf{Z}(\boldsymbol{\theta})E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|I_{t-1},\boldsymbol{\varphi}_0] = \mathbf{Z}(\boldsymbol{\theta})E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0].$$

Furthermore, the diagonality of $\boldsymbol{\Psi}$ means that the pseudo-shocks $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)$ will also inherit the cross-sectional independence of the true shocks $\boldsymbol{\varepsilon}_t^*$. Nevertheless, in general

$$\begin{aligned} E[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] &= \boldsymbol{\Psi}_\infty^{-1}(\boldsymbol{\upsilon}_0-\boldsymbol{\upsilon}_\infty), && \text{(B20)} \\ V[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] &= \boldsymbol{\Psi}_\infty^{-1}\boldsymbol{\Psi}_0^2\boldsymbol{\Psi}_\infty^{-1}, \text{ and} && \text{(B21)} \\ E[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] &= \boldsymbol{\Psi}_\infty^{-1}[(\boldsymbol{\upsilon}_0-\boldsymbol{\upsilon}_\infty)(\boldsymbol{\upsilon}_0-\boldsymbol{\upsilon}_\infty)'+\boldsymbol{\Psi}_0^2]\boldsymbol{\Psi}_\infty^{-1}, && \text{(B22)} \end{aligned}$$

where the diagonality of $V[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]$ confirms the cross-sectional independent nature of the shocks. Under standard regularity conditions

$$\begin{aligned} T^{-1}\sum_{t=1}^T \varepsilon_{it}^*(\hat{\boldsymbol{\theta}}) &\to E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = \frac{\upsilon_{i0}-\upsilon_{i\infty}}{\psi_{i\infty}} \text{ and} && \text{(B23)} \\ T^{-1}\sum_{t=1}^T \varepsilon_{it}^{*2}(\hat{\boldsymbol{\theta}}) &\to E[\varepsilon_{it}^{*2}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = \frac{(\upsilon_{i0}-\upsilon_{i\infty})^2+\psi_{i0}}{\psi_{i\infty}}, && \text{(B24)} \end{aligned}$$

where $\hat{\boldsymbol{\theta}}$ are the PMLEs of the conditional mean and variance parameters.

We have also shown in Proposition 1 that $\mathbf{a}$ and $\mathbf{j}$ will remain consistently estimated by the restricted PMLEs that fix the shape parameters of the assumed distributions to $\bar{\boldsymbol{\varrho}}$. To avoid confusion, we will denote by $\boldsymbol{\tau}_\infty(\bar{\boldsymbol{\varrho}})$ and $\boldsymbol{\psi}_\infty(\bar{\boldsymbol{\varrho}})$ the pseudo true values of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ in that case.

Proposition 3 shows that the unrestricted PMLEs of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ which simultaneously estimate $\boldsymbol{\varrho}$ will be consistent too when the univariate distributions used for estimation purposes are discrete mixtures of normals, in which case $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$ and $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty) = \boldsymbol{\varepsilon}_t^*$. Since the probability limits of the estimators of the shape parameters will also be affected, in what follows we will denote them by $\bar{\boldsymbol{\varrho}}_\infty$ to emphasise the distinction, so that $\bar{\boldsymbol{\phi}}_\infty = (\boldsymbol{\theta}_0', \bar{\boldsymbol{\varrho}}_\infty')'$. We could have called them $\boldsymbol{\varrho}_\infty(\boldsymbol{\theta}_0)$ to stress the fact that they would coincide with the plims of the PMLEs that estimate from (B1) the shape parameters only after fixing the mean and variance parameters to their true values, but the subsequent expressions would become too cumbersome.

Next, we will first obtain expressions for the conditional variance of the score and expected value of the Hessian for any assumed univariate distributions, but then we will simplify them to those cases, like finite normal mixtures, in which $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$. In this respect, Proposition F1 in Appendix F.1 implies that (B23) and (B24) are numerically identical to 0 and 1, respectively, in the finite normal mixture case.

B.3 The conditional variance of the score

B.3.1 General expression

The serial independence of $\mathbf{e}_t(\boldsymbol{\phi}_\infty)$ combined with (B7) immediately implies that

$$\mathcal{B}_t(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) = V[\mathbf{s}_t(\boldsymbol{\phi}_\infty)|I_{t-1}, \boldsymbol{\varphi}_0] = \mathbf{Z}_t(\boldsymbol{\theta}_\infty)\mathcal{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathbf{Z}_t'(\boldsymbol{\theta}_\infty), \tag{B25}$$

$$\mathcal{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V\left[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0\right] = V\left[\left.\begin{array}{c} \mathbf{e}_{lt}(\boldsymbol{\phi}_\infty) \\ \mathbf{e}_{st}(\boldsymbol{\phi}_\infty) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}_\infty) \end{array}\right| \boldsymbol{\varphi}_0\right]$$

$$= \left[\begin{array}{ccc} \mathcal{O}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{O}_{ls}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{O}_{lr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{O}_{ls}'(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{O}_{ss}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{O}_{sr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\ \mathcal{O}_{lr}'(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{O}_{sr}'(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \mathcal{O}_{rr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{array}\right]$$

$$= E\left[\left.\begin{array}{ccc} \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^{*\prime}} & \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} vec'\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} & -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'} \\ vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^{*\prime}} & vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} vec'\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} & -vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}'} \\ -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}^*}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^{*\prime}} & -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}^*} vec'\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} & \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}'} \end{array}\right| \boldsymbol{\varphi}_0\right].$$

Expressions (B8) and (B19), together with the fact that the pseudo shocks (B17) are cross-sectionally independent, imply that $\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ will be a diagonal matrix of order N with typical non-zero element

$$\mathrm{o}_{ll}^i(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V\left\{\left.\frac{\partial \ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_{i\infty}]}{\partial \varepsilon_i^*}\right| \boldsymbol{\varphi}_0\right\}.$$

As usual, under standard regularity conditions we can consistently estimate $\mathrm{o}_{ll}^i(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ by replacing $\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)$ with $\varepsilon_{it}^*(\hat{\boldsymbol{\theta}})$ and the population variance by its sample counterpart.

For the same reasons,

$$\mathcal{O}_{ls}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = \left[\begin{array}{ccccc} \mathcal{O}_{ls}^1(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \ldots & \mathcal{O}_{ls}^i(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) & \ldots & \mathcal{O}_{ls}^N(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \end{array}\right],$$

where $\mathcal{O}^i_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N whose non-zero elements are

$$\text{o}^j_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (j\neq i) \text{ and}$$
$$\text{o}^i_{ls}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=cov\left\{\frac{\partial\ln f[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon^*_i},\frac{\partial\ln f[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon^*_i}\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\right\}\ (j=i).$$

As for $\mathcal{O}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$, the same argument implies that it will be given by the sum of the commutation matrix $\mathbf{K}_{NN}$ and

$$\begin{bmatrix} \boldsymbol{\Upsilon}_{11}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Upsilon}_{1i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Upsilon}_{1N}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \vdots & \ddots & \vdots & & \vdots \\ \boldsymbol{\Upsilon}_{i1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Upsilon}_{ii}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Upsilon}_{iN}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \vdots & & \vdots & \ddots & \vdots \\ \boldsymbol{\Upsilon}_{N1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Upsilon}_{Ni}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Upsilon}_{NN}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{bmatrix},$$

where $\boldsymbol{\Upsilon}_{ij}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=\boldsymbol{\Upsilon}_{ji}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ $(j\neq i)$ is a diagonal matrix of size N whose non-zero elements are

$$\begin{aligned} \text{o}^k_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]E[\varepsilon^*_{jt}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (k &\neq i,j), \\ \text{o}^i_{ls}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)E[\varepsilon^*_{jt}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (k &= i) \text{ and} \\ \text{o}^j_{ls}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)E[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (k &= j), \end{aligned}$$

while $\boldsymbol{\Upsilon}_{ii}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is another diagonal matrix of the same size whose non-zero elements are

$$\text{o}^j_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon^{*2}_{it}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (j\neq i) \text{ and}$$
$$\text{o}^i_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=V\left\{\frac{\partial\ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon^*_i}\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\right\}-1\ (j=i).$$

It is worth noting that the off-diagonal elements of $\mathbf{K}_{NN}$ reflect the fact that

$$E\left[\left(\frac{\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_2)}{\partial\varepsilon^*_i}\varepsilon^*_{jt}\right)\left(\frac{\partial\ln f_k(\varepsilon^*_{kt};\boldsymbol{\varrho}_2)}{\partial\varepsilon^*_k}\varepsilon^*_{lt}\right)\right]=1$$

when $i=k$ and $j=l$ despite the fact that $i\neq j$.

In turn, $\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is an $N\times q$ block diagonal matrix with typical diagonal block of size $1\times q_i$

$$\text{o}^i_{lr}(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0)=-cov\left[\frac{\partial\ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon^*_i},\frac{\partial\ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\boldsymbol{\varrho}_i}\right|\boldsymbol{\varphi}_0\right],$$

while

$$\mathcal{O}'_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=\begin{bmatrix}\mathcal{O}^{1\prime}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{O}^{i\prime}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{O}^{N\prime}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\end{bmatrix},$$

where $\mathcal{O}^i_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is another block diagonal matrix of order $N\times q$ whose non-zero blocks of size $1 \times q_j$ will be

$$\mathrm{o}^{j}_{lr}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) E[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0], \;\; i \neq j \text{ and}$$

$$\mathrm{o}^{i}_{sr}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0) = -cov\left[\frac{\partial \ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_{i\infty}]}{\partial \varepsilon^*_i}\varepsilon^*_{it}(\boldsymbol{\theta}_\infty), \frac{\partial \ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_{i\infty}]}{\partial \boldsymbol{\varrho}_i}\right| \boldsymbol{\varphi}_0\right], \;\; i = j.$$

Finally, $\mathcal{O}_{rr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)$ is a $q \times q$ block diagonal matrix with typical diagonal block of size $q_i \times q_i$

$$\mathrm{o}^{i}_{rr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) = V\left\{\frac{\partial \ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_{i\infty}]}{\partial \boldsymbol{\varrho}_i}\right| \boldsymbol{\varphi}_0\right\}.$$

Let us now combine these expressions with the special structure of $\mathbf{Z}_{lt}$ and $\mathbf{Z}_{st}$ to obtain the conditional covariance matrix of the score for model (1) in more detail.

If we expand (B25), we end up with

$$\left[\begin{array}{cc} \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{O}'_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ +\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{O}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}'_{st}(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{O}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}'_{st}(\boldsymbol{\theta}) & +\mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{O}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathcal{O}'_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}'_{lt}(\boldsymbol{\theta}) + \mathcal{O}'_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}'_{st}(\boldsymbol{\theta}) & \mathcal{O}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{array}\right].$$

Thus, diagonal block of the covariance matrix of the score corresponding to $\boldsymbol{\theta}$ will be

$$\begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1}\otimes\mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{pmatrix} \mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1}\begin{pmatrix} \mathbf{I}_N & \mathbf{y}'_{t-1}\otimes\mathbf{I}_N & \dots & \mathbf{y}'_{t-p}\otimes\mathbf{I}_N & \mathbf{0}_{N\times N^2} \end{pmatrix}$$

$$+\begin{pmatrix} \mathbf{0}_{N\times N^2} \\ \mathbf{0}_{N^2\times N^2} \\ \vdots \\ \mathbf{0}_{N^2\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix}(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}'_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1}\begin{pmatrix} \mathbf{I}_N & \mathbf{y}'_{t-1}\otimes\mathbf{I}_N & \dots & \mathbf{y}'_{t-p}\otimes\mathbf{I}_N & \mathbf{0}_{N\times N^2} \end{pmatrix}$$

$$+\begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1}\otimes\mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{pmatrix} \mathbf{C}^{-1\prime}\mathcal{O}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1})\begin{pmatrix} \mathbf{0}_{N^2\times N} & \mathbf{0}_{N^2\times N^2} & \dots & \mathbf{0}_{N^2\times N^2} & \mathbf{I}_{N^2} \end{pmatrix}$$

$$+\begin{pmatrix} \mathbf{0}_{N\times N^2} \\ \mathbf{0}_{N^2\times N^2} \\ \vdots \\ \mathbf{0}_{N^2\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix}(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1})\begin{pmatrix} \mathbf{0}_{N^2\times N} & \mathbf{0}_{N^2\times N^2} & \dots & \mathbf{0}_{N^2\times N^2} & \mathbf{I}_{N^2} \end{pmatrix}$$

$$
= \left[\begin{array}{ccc}
\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & \mathbf{y}'_{t-1}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & \cdots \\
\mathbf{y}_{t-1}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & \mathbf{y}_{t-1}\mathbf{y}'_{t-1}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & \cdots \\
\vdots & \vdots & \ddots \\
\mathbf{y}_{t-p}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & \mathbf{y}_{t-p}\mathbf{y}'_{t-1}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & \cdots \\
(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}'_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}'_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1}(\mathbf{y}'_{t-1}\otimes\mathbf{I}_N) & \cdots
\end{array}\right.
$$

$$
\left.\begin{array}{cc}
\mathbf{y}'_{t-p}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & \mathbf{C}^{-1\prime}\mathcal{O}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1}) \\
\mathbf{y}_{t-1}\mathbf{y}'_{t-p}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{O}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1}) \\
\vdots & \vdots \\
\mathbf{y}_{t-p}\mathbf{y}'_{t-p}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1} & (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{O}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1}) \\
(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}'_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1}(\mathbf{y}'_{t-p}\otimes\mathbf{I}_N) & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1})
\end{array}\right]. \quad \text{(B26)}
$$

As a result, the block of the unconditional covariance matrix of the score corresponding to the conditional mean parameters $\boldsymbol{\tau}$ and $\mathbf{a}$ will be

$$
E\left[\begin{pmatrix} 1 \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p} \end{pmatrix}\begin{pmatrix} 1 & \mathbf{y}'_{t-1} & \cdots & \mathbf{y}'_{t-p} \end{pmatrix}\right]\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1}
$$

$$
= \begin{bmatrix} 1 & \boldsymbol{\mu}' & \cdots & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & \boldsymbol{\Gamma}(0)+\boldsymbol{\mu}\boldsymbol{\mu}' & \cdots & \boldsymbol{\Gamma}(p-1)+\boldsymbol{\mu}\boldsymbol{\mu}' \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu} & \boldsymbol{\Gamma}'(p-1)+\boldsymbol{\mu}\boldsymbol{\mu}' & \cdots & \boldsymbol{\Gamma}(0)+\boldsymbol{\mu}\boldsymbol{\mu}' \end{bmatrix}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{C}^{-1}, \quad \text{(B27)}
$$

where $\boldsymbol{\Gamma}(j)$ is the j^{th} autocovariance matrix of $\mathbf{y}_t$.

In turn, the off-diagonal $\boldsymbol{\theta\varrho}$ block of the conditional covariance matrix of the score will be

$$
\begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1}\otimes\mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{pmatrix}\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) + \begin{pmatrix} \mathbf{0}_{N\times N^2} \\ \mathbf{0}_{N^2\times N^2} \\ \vdots \\ \mathbf{0}_{N^2\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix}(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)
$$

$$
= \begin{bmatrix} \mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \vdots \\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{bmatrix} = \begin{bmatrix} \mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathbf{y}_{t-1}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{bmatrix}, \quad \text{(B28)}
$$

whose unconditional expectation is trivially

$$
\begin{bmatrix} \mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \boldsymbol{\mu}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \vdots \\ \boldsymbol{\mu}\otimes\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{bmatrix}.
$$

Unfortunately, there seems to be no obvious simplification to the matrices

$$\begin{aligned}
&\mathbf{C}^{-1\prime}\mathcal{O}_{ll}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)\mathbf{C}^{-1},\\
&\mathbf{C}^{-1\prime}\mathcal{O}_{ls}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1}),\\
&(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{ss}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)(\mathbf{I}_N\otimes\mathbf{C}^{-1}),\\
&\mathbf{C}^{-1\prime}\mathcal{O}_{lr}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0),\ \text{and}\\
&(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathcal{O}_{sr}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)
\end{aligned}$$

unless $\mathbf{C}$ is diagonal (see, e.g., the discussion in the proof of Proposition 14 in Fiorentini and Sentana (2021b)). In principle, we could effectively make $\mathbf{C}$ equal to the identity matrix by premultiplying $\mathbf{y}_t$ in (1) by $\mathbf{C}^{-1}$, which would preserve the vector autoregressive structure with the similar autoregressive matrices $\mathbf{C}^{-1}\mathbf{A}_j\mathbf{C}$ for $j=1,\ldots,p$. Moreover, we could also effectively set the drifts to $\mathbf{0}$ by subtracting the unconditional mean $\boldsymbol{\mu}$ from the observations. However, we shall not pursue any of these avenues.

B.3.2 Special case: $\boldsymbol{\theta}$ consistently estimated

Many of the elements of $\mathcal{O}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)$ simplify considerably when $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ are consistently estimated, in which case $v_{i\infty}-v_{i0}=0$ and $\psi_{i0}/\psi_{i\infty}=1$ for all i. Specifically, Amengual et al (2021b) show that $\mathcal{O}_{ls}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)=\mathrm{O}_{ls}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)\mathbf{E}_N'$, where $\mathrm{O}_{ls}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N with typical element

$$\mathrm{o}_{ls}^{i}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)=cov\left\{\frac{\partial\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{\partial\varepsilon_i^*},\frac{\partial\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{\partial\varepsilon_i^*}\varepsilon_{it}^*(\boldsymbol{\theta}_0)\right|\boldsymbol{\varphi}_0\right\}.$$

Similarly, they show that $\boldsymbol{\Upsilon}$ will be a block diagonal matrix of order N^2 in which each of the diagonal blocks $\boldsymbol{\Upsilon}_{ii}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N whose non-zero elements are

$$\begin{aligned}
\mathrm{o}_{ll}^{j}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)&=V\left\{\frac{\partial\ln f_j[\varepsilon_{jt}^*(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{j\infty}]}{\partial\varepsilon_j^*}\right|\boldsymbol{\varphi}_0\right\}\ (j\neq i)\ \text{and}\\
\mathrm{o}_{ss}^{i}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)&=V\left\{\frac{\partial\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{\partial\varepsilon_i^*}\varepsilon_{it}^*(\boldsymbol{\theta}_0)\right|\boldsymbol{\varphi}_0\right\}-1\ (j=i).
\end{aligned}$$

Finally, they prove that $\mathcal{O}_{sr}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)=\mathbf{E}_N\mathrm{O}_{sr}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)$, where $\mathrm{O}_{sr}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)$ is another block diagonal matrix of order $N\times q$ with typical block of size $1\times q_i$

$$\mathrm{o}_{sr}(\bar{\boldsymbol{\phi}}_{\infty};\boldsymbol{\varphi}_0)=-cov\left[\frac{\partial\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{\partial\varepsilon_i^*}\varepsilon_{it}^*(\boldsymbol{\theta}_{\infty}),\frac{\partial\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{\partial\boldsymbol{\varrho}_i}\right|\boldsymbol{\varphi}_0\right].$$

B.4 The conditional expected value of the Hessian

B.4.1 General expression

Given that $\mathbf{Z}_{dt}(\boldsymbol{\theta}_{\infty})\in I_{t-1}$, the first thing to note is that (B19) sets to 0 the conditional expectation of the last two terms of (B11). Moreover, the serial independence of $\mathbf{e}_t(\boldsymbol{\phi}_{\infty})$ in

(B17), together with (B12) and (B13) implies that

$$
\begin{aligned}
-E[\mathbf{h}_t(\boldsymbol{\phi}_\infty)|I_{t-1},\boldsymbol{\varphi}_0] &= \left[\begin{array}{ccc} \mathbf{Z}_{lt}(\boldsymbol{\theta}_\infty) & \mathbf{Z}_{st}(\boldsymbol{\theta}_\infty) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array}\right] \\
\times E &\left[\begin{array}{ccc}
-\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}} & -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N] & \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}^{\prime}} \\
-[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}} - \left[\mathbf{I}_N\otimes\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\right] & -\left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N\otimes\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\right]\right\}[\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N] & [\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}^{\prime}} \\
\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varepsilon}^{*\prime}} & \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varepsilon}^{*\prime}}[\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N] & -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}^{\prime}}
\end{array}\right|\left.\boldsymbol{\varphi}_0\right] \\
&\times \left[\begin{array}{cc} \mathbf{Z}_{lt}^\prime(\boldsymbol{\theta}_\infty) & \mathbf{0} \\ \mathbf{Z}_{st}^\prime(\boldsymbol{\theta}_\infty) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{array}\right].
\end{aligned}
$$

But

$$
-\left[\mathbf{I}_N\otimes\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\right] = [\mathbf{I}_N\otimes\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)],
$$

whose expected value is clearly 0 in view of (B19). In turn, if we now focus on $\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)$, (B19) also implies that the expected value of

$$
-\left[\mathbf{I}_N\otimes\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N]
$$

will be $\mathbf{K}_{NN}$ because

$$
\begin{aligned}
-\mathbf{K}_{NN}\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N\right] &= -\mathbf{K}_{NN}\left[\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\otimes\mathbf{I}_N\right][\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N] \\
&= -\left[\mathbf{I}_N\otimes\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\right][\boldsymbol{\varepsilon}_t^{\prime *}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N].
\end{aligned}
$$

Consequently, we can write

$$
\mathcal{A}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = -E[\mathbf{h}_{\boldsymbol{\phi}\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|I_{t-1},\boldsymbol{\varphi}_0] = \mathbf{Z}_t(\boldsymbol{\theta}_\infty)\mathcal{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_t^\prime(\boldsymbol{\theta}_\infty), \tag{B29}
$$

where

$$
\mathcal{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = \left[\begin{array}{ccc} \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathcal{H}_{ls}^\prime(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathcal{H}_{lr}^\prime(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{H}_{sr}^\prime(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{H}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{array}\right]
$$

with

$$
\begin{aligned}
\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= -E\left\{\left.\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\right|\boldsymbol{\varphi}_0\right\},\\
\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= -E\left\{\left.\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N]\right|\boldsymbol{\varphi}_0\right\},\\
\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= -E\left\{\left.[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N)\frac{\partial^2 f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}[\boldsymbol{\varepsilon}_t^{\prime*}(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N]\right|\boldsymbol{\varphi}_0\right\}+\mathbf{K}_{NN}\\
\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E\left\{\left.\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}^{\prime}}\right|\boldsymbol{\varphi}_0\right\},\\
\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E\left\{\left.[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty)\otimes\mathbf{I}_N]\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varrho}^{\prime}}\right|\boldsymbol{\varphi}_0\right\},\ \text{and}\\
\mathcal{H}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= -E\left\{\left.\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}^{\prime}}\right|\boldsymbol{\varphi}_0\right\}
\end{aligned}
$$

Expression (B14) implies that $\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ will be a diagonal matrix of order N with typical non-zero element.

$$
\mathrm{H}_{ll}^i(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-E\left\{\left.\frac{\partial^2 \ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{(\partial\varepsilon_i^*)^2}\right|\boldsymbol{\varphi}_0\right\}.
$$

Once again, under standard regularity conditions we can consistently estimate $\mathrm{H}_{ll}^i(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ by replacing $\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)$ with $\varepsilon_{it}^*(\hat{\boldsymbol{\theta}})$ and the expected value by its sample counterpart.

For the same reason,

$$
\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=[\ \mathcal{H}_{ls}^1(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\quad\dots\quad\mathcal{H}_{ls}^i(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\quad\dots\quad\mathcal{H}_{ls}^N(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\ ],
$$

where $\mathcal{H}_{ls}^i(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N whose non-zero elements are

$$
\begin{gathered}
\mathrm{H}_{ll}^j(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (j\neq i)\ \text{and}\\
\mathrm{H}_{ls}^i(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-E\left[\left.\frac{\partial^2 \ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{(\partial\varepsilon_i^*)^2}\cdot\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\right]\ (j=i).
\end{gathered}
$$

As for the first summand of $\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$, the cross-sectional independence implies that it will be given by

$$
\begin{bmatrix}
\boldsymbol{\Gamma}_{11}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Gamma}_{1i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Gamma}_{1N}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
\vdots & \ddots & \vdots & & \vdots\\
\boldsymbol{\Gamma}_{i1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Gamma}_{ii}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Gamma}_{iN}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
\vdots & & \vdots & \ddots & \vdots\\
\boldsymbol{\Gamma}_{N1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Gamma}_{Ni}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \boldsymbol{\Gamma}_{NN}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)
\end{bmatrix},
$$

where $\boldsymbol{\Gamma}_{ij}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=\boldsymbol{\Gamma}_{ji}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ $(j\neq i)$ is a diagonal matrix of size N whose non-zero elements are

$$\mathrm{H}_{ll}^{k}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]E[\varepsilon_{jt}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (k\neq i,j),$$
$$\mathrm{H}_{ls}^{i}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)E[\varepsilon_{jt}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (k=i) \text{ and}$$
$$\mathrm{H}_{ls}^{j}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (k=j),$$

while $\boldsymbol{\Gamma}_{ii}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is another diagonal matrix of the same size whose non-zero elements are

$$\mathrm{H}_{ll}^{j}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon_{it}^{*2}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (j\neq i) \text{ and}$$
$$\mathrm{H}_{ss}^{i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=-E\left\{\frac{\partial^2\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{(\partial\varepsilon_i)^2}\varepsilon_{it}^{*2}(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\Bigg\}\ (j=i).$$

In turn, $\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is an $N\times q$ block diagonal matrix with typical diagonal block of size $1\times q_i$

$$\mathrm{H}_{lr}^{i}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=E\left[\frac{\partial^2\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon_i^*\partial\boldsymbol{\varrho}_i'}\right|\boldsymbol{\varphi}_0\Bigg],$$

while

$$\mathcal{H}_{sr}'(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=\left[\begin{array}{ccccc}\mathcal{H}_{sr}^{1\prime}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \ldots & \mathcal{H}_{sr}^{i\prime}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \ldots & \mathcal{H}_{sr}^{N\prime}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\end{array}\right],$$

where $\mathcal{H}_{sr}^{i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is another block diagonal matrix of order $N\times q$ whose non-zero blocks of size $1\times q_j$ will be

$$\mathrm{H}_{lr}^{j}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0],\ i\neq j \text{ and}$$
$$\mathrm{H}_{sr}^{i}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=E\left\{\frac{\partial^2\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon_i^*\partial\boldsymbol{\varrho}_i'}\varepsilon_i^*(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\Bigg\},\ i\neq j.$$

Finally, $\mathcal{H}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is a $q\times q$ block diagonal matrix with typical block of size $q_i\times q_i$

$$\mathrm{H}_{rr}^{i}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-E\left[\frac{\partial^2\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\boldsymbol{\varrho}_i\partial\boldsymbol{\varrho}_i'}\right|\boldsymbol{\varphi}_0\Bigg].$$

As a result, if we expand (B29), we end up with

$$\left[\begin{array}{cc}\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{H}_{ls}'(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta}) & \mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\ +\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta})+\mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}) & +\mathbf{Z}_{st}(\boldsymbol{\theta})\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\ \mathcal{H}_{lr}'(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_{lt}'(\boldsymbol{\theta})+\mathcal{H}_{sr}'(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\mathbf{Z}_{st}'(\boldsymbol{\theta}) & \mathcal{O}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\end{array}\right],$$

which in turn adopt expressions entirely analogous to the ones we have obtained before for the conditional covariance matrix of the score in Appendix B.3.

B.4.2 Special case: $\boldsymbol{\theta}$ consistently estimated

Again, many of the elements of $\mathcal{H}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ will also simplify considerably if $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ are consistently estimated. Specifically, Amengual et al (2021b) show that $\mathcal{H}_{ls}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0)=\mathrm{H}_{ls}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0)\mathbf{E}_N'$, where $\mathrm{H}_{ls}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N with typical element

$$\mathrm{H}_{ls}^{i}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0)=-E\left[\frac{\partial^2\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{(\partial\varepsilon_i^*)^2}\cdot\varepsilon_{it}^*(\boldsymbol{\theta}_0)\right|\boldsymbol{\varphi}_0\Bigg].$$

Similarly, they show that $\mathbf{\Gamma}$ will be a block diagonal matrix of order N^2 in which each of the diagonal blocks $\mathbf{\Gamma}_{ii}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N whose non-zero elements are

$$\mathrm{H}^j_{ll}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0) = -E\left\{ \left. \frac{\partial^2 \ln f_j[\varepsilon^*_{it}(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{(\partial\varepsilon_j)^2} \right| \boldsymbol{\varphi}_0 \right\} \;(j\neq i) \text{ and}$$
$$\mathrm{H}^i_{ss}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0) = -E\left\{ \left. \frac{\partial^2 \ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{(\partial\varepsilon_i)^2}\varepsilon^{*2}_{it}(\boldsymbol{\theta}_0) \right| \boldsymbol{\varphi}_0 \right\} \;(j = i).$$

Finally, they prove that $\mathcal{H}_{sr}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0) = \mathbf{E}_N \mathrm{H}_{sr}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0)$, where $\mathrm{H}_{sr}(\boldsymbol{\theta}_0,\bar{\boldsymbol{\varrho}}_\infty;\boldsymbol{\varphi}_0)$ another block diagonal matrix of order $N\times q$ with typical block of size $1\times q_i$

$$\mathrm{H}_{sr}(\bar{\boldsymbol{\phi}}_\infty;\boldsymbol{\varphi}_0) = E\left\{ \left. \frac{\partial^2 \ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_0);\bar{\boldsymbol{\varrho}}_{i\infty}]}{\partial\varepsilon^*_i \partial\boldsymbol{\varrho}'_i}\varepsilon^*_i(\boldsymbol{\theta}_0) \right| \boldsymbol{\varphi}_0 \right\}.$$

B.5 Asymptotic distribution

B.5.1 Robust standard errors for the PMLEs

For simplicity, we assume henceforth that there are no unit roots in the autoregressive polynomial, so that the SVAR model (1) generates a covariance stationary process in which $rank(\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p) = N$. If the autoregressive polynomial $(\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)$ had some unit roots, then $\mathbf{y}_t$ would be a (co-) integrated process, and the estimators of the conditional mean parameters would have non-standard asymptotic distributions, as some (linear combinations) of them would converge at the faster rate T. In contrast, the distribution of the ML estimators of the conditional variance parameters would remain standard (see, e.g., Phillips and Durlauf (1986)).

We also assume that the regularity conditions A1-A6 in White (1982) are satisfied. These conditions are only slightly stronger than those in Crowder (1976), which guarantee that MLEs will be consistent and asymptotically normally distributed under correct specification. 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 $\mathbb{R}^{\dim(\phi)}$; (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 $\boldsymbol{\phi}_0$.

We can use the law of iterated expectations to compute

$$\mathcal{A}(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0) = E[-\mathbf{h}_{\phi\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0,\boldsymbol{\varphi}_0] = E\left[\mathcal{A}_t(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0)\right]$$

and

$$V[\mathbf{s}_{\phi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] = \mathcal{B}(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0) = E\left[\mathcal{B}_t(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi})\right].$$

In this context, the asymptotic distribution of the PMLEs of $\boldsymbol{\phi}$ under the regularity conditions

A1-A6 in White (1982) will be given by

$$\sqrt{T}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_{\infty}) \to N[\mathbf{0}, \mathcal{A}^{-1}(\boldsymbol{\phi}_{\infty}, \boldsymbol{\varphi}_0)\mathcal{B}(\boldsymbol{\phi}_{\infty}, \boldsymbol{\varphi}_0)\mathcal{A}^{-1}(\boldsymbol{\phi}_{\infty}, \boldsymbol{\varphi}_0)].$$

As we explained before, analogous expressions apply *mutatis mutandi* to a restricted PML estimator of $\boldsymbol{\theta}$ that fixes $\boldsymbol{\varrho}$ some a priori chosen value to $\bar{\boldsymbol{\varrho}}$. In that case, we would simply need to replace $\boldsymbol{\theta}_{\infty}$ by $\boldsymbol{\theta}_{\infty}(\bar{\boldsymbol{\varrho}})$ and eliminate the rows and columns corresponding to the shape parameters $\boldsymbol{\varrho}$ from the $\mathcal{A}$ and $\mathcal{B}$ matrices.

B.5.2 The information matrix under correct distributional specification

If the distribution of the shocks were correctly specified, then $\bar{\boldsymbol{\varrho}}_{\infty} = \boldsymbol{\varpi}_0$ and the information matrix equality holds, so that $\mathcal{H}(\boldsymbol{\varphi}_0; \boldsymbol{\varphi}_0) = \mathcal{O}(\boldsymbol{\varphi}_0; \boldsymbol{\varphi}_0) = \mathcal{M}(\boldsymbol{\varrho}_0)$ and $\mathcal{A}_t(\boldsymbol{\phi}_{\mathbf{0}}, \boldsymbol{\varphi}_0) = \mathcal{B}_t(\boldsymbol{\phi}_{\mathbf{0}}, \boldsymbol{\varphi}_0) = \mathcal{I}_t(\boldsymbol{\phi}_0)$ (see Proposition D.2 in Fiorentini and Sentana (2021b)).

In this context, we can see more clearly the structure of $\mathcal{M}(\boldsymbol{\varrho})$ by appropriately re-arranging the elements of $\mathbf{e}_{st}(\boldsymbol{\phi})$. Expression (A5) allows us to conveniently re-write

$$\mathbf{e}_{st}(\boldsymbol{\phi}) = \begin{pmatrix} \mathbf{E}_N & \boldsymbol{\Delta}_N \end{pmatrix} \begin{bmatrix} \mathbf{E}'_N \mathbf{e}_{st}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}'_N \mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix},$$

where

$$\mathbf{E}'_N \mathbf{e}_{dt}(\boldsymbol{\phi}) = -vecd\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\right\} = -\left\{\begin{array}{c} 1 + \frac{\partial \ln f_1[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon^*_1}\varepsilon^*_{1t}(\boldsymbol{\theta}) \\ \vdots \\ 1 + \frac{\partial \ln f_N[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon^*_N}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \end{array}\right\}$$

and

$$\boldsymbol{\Delta}'_N \mathbf{e}_{dt}(\boldsymbol{\phi}) = -veco\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\right\} = -\left\{\begin{array}{c} \frac{\partial \ln f_1[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon^*_1}\varepsilon^*_{1t}(\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial \ln f_N[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon^*_N}\varepsilon^*_{1t}(\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial \ln f_1[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon^*_1}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial \ln f_N[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon^*_N}\varepsilon^*_{Nt}(\boldsymbol{\theta}) \end{array}\right\}.$$

This vector is such that

$$\begin{aligned} \begin{pmatrix} \mathbf{E}_N & \boldsymbol{\Delta}_N \end{pmatrix} \begin{bmatrix} \mathbf{E}'_N \mathbf{e}_{st}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}'_N \mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} &= -vec\left[dg\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\right\}\right] \\ -vec\left[\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\right\}\right. &- \left. dg\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\right\}\right], \end{aligned}$$

where dg is the operator which transforms a square matrix into a diagonal one by setting its off-diagonal elements to zero.

Then, we can use Propositions 6 and 7 in Magnus and Sentana (2020) to prove that

$$V\left[\begin{array}{c}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]=\left(\begin{array}{c}\mathbf{E}_N'\\ \boldsymbol{\Delta}_N'\end{array}\right)V[\mathbf{e}_{st}(\boldsymbol{\phi})]\left(\begin{array}{cc}\mathbf{E}_N & \boldsymbol{\Delta}_N\end{array}\right)=\left(\begin{array}{c}\mathbf{E}_N'\\ \boldsymbol{\Delta}_N'\end{array}\right)(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\left(\begin{array}{cc}\mathbf{E}_N & \boldsymbol{\Delta}_N\end{array}\right)$$
$$=\left[\begin{array}{cc}\mathbf{E}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\mathbf{E}_N & \mathbf{E}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N\\ \boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\mathbf{E}_N & \boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N\end{array}\right]=\left[\begin{array}{cc}\mathrm{M}_{ss} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N\end{array}\right],$$

where $\mathrm{M}_{ss}=(\mathbf{I}_N+\mathbf{E}_N'\boldsymbol{\Upsilon}\mathbf{E}_N)$ is a diagonal matrix of order N with typical element $\mathrm{M}_{ss}(\boldsymbol{\varrho}_i)$.

Hence,

$$V\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\\ \mathbf{e}_{rt}(\boldsymbol{\phi})\end{array}\right]=\left(\begin{array}{ccc}\mathbf{I}_N & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_N' & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Delta}_N' & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{array}\right)V\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi}_0)\\ \mathbf{e}_{st}(\boldsymbol{\phi}_0)\\ \mathbf{e}_{rt}(\boldsymbol{\phi}_0)\end{array}\right]\left(\begin{array}{cccc}\mathbf{I}_N & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_N & \boldsymbol{\Delta}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{array}\right)$$
$$=\left[\begin{array}{cccc}\mathcal{M}_{ll} & \mathrm{M}_{ls} & \mathbf{0} & \mathcal{M}_{lr}\\ \mathrm{M}_{ls} & \mathrm{M}_{ss} & \mathbf{0} & \mathrm{M}_{sr}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}_N'(\mathbf{K}_{NN}+\boldsymbol{\Upsilon})\boldsymbol{\Delta}_N & \mathbf{0}\\ \mathcal{M}_{lr}' & \mathrm{M}_{sr}' & \mathbf{0} & \mathcal{M}_{rr}\end{array}\right]. \tag{B30}$$

Therefore, $\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})$ is orthogonal to all the other elements of the score.

Importantly, when any of the N distributions is symmetric, it is easy to see that bot $\mathrm{M}_{ls}(\boldsymbol{\varrho}_i)$ and $\mathrm{M}_{lr}(\boldsymbol{\varrho}_i)$ will be equal to 0, so (B30) simplifies even further.

B.6 Reparametrisations

A convenient property of the expressions for $\mathbf{s}_t(\boldsymbol{\phi}_\infty)$, $\mathcal{A}_t(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0)$ and $\mathcal{B}_t(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0)$ above is that reparametrisations only have an effect on the $\mathbf{Z}_{lt}$ and $\mathbf{Z}_{st}$ matrices, which only involve first derivatives of the conditional mean vector and covariance matrix functions.

B.6.1 Unconditional mean

The first reparametrisation that we will consider involves rewriting the drift $\boldsymbol{\tau}$ as $(\mathbf{I}_N-\mathbf{A}_1-\ldots-\mathbf{A}_p)\boldsymbol{\mu}$, which we can always do under our maintained assumption of covariance stationarity. The Jacobian from one vector of parameters to the other is

$$\frac{\partial\left(\begin{array}{c}\boldsymbol{\tau}\\ \mathbf{a}\end{array}\right)}{\partial(\boldsymbol{\mu}',\mathbf{a}')}=\left(\begin{array}{cccc}\mathbf{I}_N-\mathbf{A}_1-\ldots-\mathbf{A}_p & -\boldsymbol{\mu}'\otimes\mathbf{I}_N & \ldots & -\boldsymbol{\mu}'\otimes\mathbf{I}_N\\ \mathbf{0} & \mathbf{I}_{N^2} & \ldots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_{N^2}\end{array}\right).$$

Consequently, $\mathbf{Z}_{lt}(\boldsymbol{\theta})$ for $(\boldsymbol{\mu}',\mathbf{a}',\mathbf{c}')$ becomes

$$\left[\begin{array}{c}(\mathbf{I}_N-\mathbf{A}_1-\ldots-\mathbf{A}_p)\mathbf{C}^{-1\prime}\\ (\mathbf{y}_{t-1}-\boldsymbol{\mu})\otimes\mathbf{C}^{-1\prime}\\ \vdots\\ (\mathbf{y}_{t-p}-\boldsymbol{\mu})\otimes\mathbf{C}^{-1\prime}\\ \mathbf{0}_{N^2\times N}\end{array}\right].$$

Given that $E(\mathbf{y}_t) = \boldsymbol{\mu}\ \forall t$, and that the rest of the elements of $\mathbf{Z}_{lt}$ and $\mathbf{Z}_{st}$ are constant over time, it is clear from (B26) and (B28) that both $\mathcal{A}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0)$ and $\mathcal{B}(\boldsymbol{\phi}_\infty, \boldsymbol{\varphi}_0)$ will become block-diagonal between the elements of $\mathbf{a}$ and the rest. In addition, it is also easy to see from the same expressions that

$$\begin{aligned}
\mathcal{A}_{\boldsymbol{\mu\mu}} &= (\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)\mathbf{C}^{-1\prime}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathbf{C}^{-1}(\mathbf{I}_N - \mathbf{A}_1 - \ldots - \mathbf{A}_p)', \\
\mathcal{A}_{\mathbf{aa}} &= \begin{bmatrix} \boldsymbol{\Gamma}(0) & \ldots & \boldsymbol{\Gamma}(p-1) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}'(p-1) & \ldots & \boldsymbol{\Gamma}(0) \end{bmatrix} \otimes \mathbf{C}^{-1\prime}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\mathbf{C}^{-1},
\end{aligned}$$

where $\boldsymbol{\Gamma}(j)$ is the j^{th} autocovariance matrix of $\mathbf{y}_t$, with analogous expressions for $\mathcal{B}_{\boldsymbol{\mu\mu}}$ and $\mathcal{B}_{\mathbf{aa}}$. Consequently, the asymptotic variances of the restricted and unrestricted ML estimators of $\mathbf{a}$ will be given by

$$\begin{bmatrix} \boldsymbol{\Gamma}(0) & \ldots & \boldsymbol{\Gamma}(p-1) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}'(p-1) & \ldots & \boldsymbol{\Gamma}(0) \end{bmatrix}^{-1} \otimes \mathbf{C}\mathcal{H}_{ll}^{-1}(\phi_\infty; \varphi_0)\mathcal{O}_{ll}(\phi_\infty; \varphi_0)\mathcal{H}_{ll}^{-1}(\phi_\infty; \varphi_0)\mathbf{C}'.$$

But since the PML estimators of $\mathbf{a}$ are the same regardless of whether we estimate the model in terms of $\boldsymbol{\tau}$ or $\boldsymbol{\mu}$, the asymptotic variance of $\mathbf{a}$ above is valid under covariance stationarity.

B.6.2 Standard deviations of shocks

Another reparametrisation that is very relevant in this paper is the one that expresses $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$. In this case, the parameters of interest become

$$\begin{aligned}
\mathbf{j} &= veco(\mathbf{J} - \mathbf{I}_N) = (c_{21}/c_{11}, ..., c_{N1}/c_{11}, ..., c_{1N}/c_{NN}, ..., c_{N-1N}/c_{NN})', \\
\boldsymbol{\psi} &= vecd(\boldsymbol{\Psi}) = (c_{11}, ..., c_{NN})'.
\end{aligned}$$

The product rule for differentials $d\mathbf{C} = (d\mathbf{J})\boldsymbol{\Psi} + \mathbf{J}(d\boldsymbol{\Psi})$ immediately implies that

$$dvec(\mathbf{C}) = (\boldsymbol{\Psi} \otimes \mathbf{I}_N)\boldsymbol{\Delta}_N dveco(\mathbf{J}) + (\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N dveco(\boldsymbol{\Psi}).$$

Therefore, the Jacobian will be

$$\frac{\partial vec(\mathbf{C})}{\partial(\mathbf{j}', \boldsymbol{\psi}')} = \left[\ (\boldsymbol{\Psi} \otimes \mathbf{I}_N)\boldsymbol{\Delta}_N \quad (\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N\ \right] = \left[\ \boldsymbol{\Delta}_N(\boldsymbol{\Psi} \otimes \mathbf{I}_{N-1}) \quad (\mathbf{I}_N \otimes \mathbf{J})\mathbf{E}_N\ \right], \tag{B31}$$

where we have used that $\boldsymbol{\Upsilon}\boldsymbol{\Delta}_N = \boldsymbol{\Delta}_N(\boldsymbol{\Delta}_N'\boldsymbol{\Upsilon}\boldsymbol{\Delta}_N)$ for any diagonal matrix $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Delta}_N'(\boldsymbol{\Psi} \otimes \mathbf{I}_N)\boldsymbol{\Delta}_N = (\boldsymbol{\Psi} \otimes \mathbf{I}_{N-1})$ (see Proposition 6 in Magnus and Sentana (2020)).

As a result, the chain rule for derivatives leads to

$$\begin{aligned}
\begin{bmatrix} \mathbf{s}_{\mathbf{j}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\phi}) \end{bmatrix} &= \begin{pmatrix} \partial\mathbf{c}'/\partial\mathbf{j} \\ \partial\mathbf{c}'/\partial\boldsymbol{\psi} \end{pmatrix} \mathbf{s}_{\mathbf{c}}(\boldsymbol{\phi}) \\
= -\begin{bmatrix} (\boldsymbol{\Psi}\otimes\mathbf{I}_{N-1})\boldsymbol{\Delta}_N' \\ \mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}') \end{bmatrix} &(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\mathbf{I}_N\otimes\boldsymbol{\Psi}^{-1})vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \\
= -\begin{bmatrix} (\boldsymbol{\Psi}\otimes\mathbf{I}_{N-1})\boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\mathbf{I}_N\otimes\boldsymbol{\Psi}^{-1}) \\ \boldsymbol{\Psi}^{-1}\mathbf{E}_N' \end{bmatrix} &vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \\
= -\begin{bmatrix} \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1}) \\ \boldsymbol{\Psi}^{-1}\mathbf{E}_N' \end{bmatrix} &vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \\
= \begin{bmatrix} -veco\left\{\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Psi} + \mathbf{J}^{-1\prime}\right\} \\ -\boldsymbol{\Psi}^{-1}vecd\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) + \mathbf{I}_N\right\} \end{bmatrix}&, \qquad \text{(B32)}
\end{aligned}$$

where we have used the fact that $\boldsymbol{\Delta}_N' vec(\mathbf{A}) = veco(\mathbf{A})$ and $\mathbf{E}_N' vec(\mathbf{A}) = vecd(\mathbf{A})$ for any square matrix $\mathbf{A}$ of order N.

In particular,

$$s_{\psi_i t}(\boldsymbol{\phi}) = -\frac{1}{\psi_i}\left\{\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta});\boldsymbol{\varrho}_i]}{\partial\varepsilon_i^*}\varepsilon_{it}^*(\boldsymbol{\theta}) + 1\right\} \quad (i = 1,\ldots,N). \qquad \text{(B33)}$$

Once again, we can use (B31) to transform $\mathcal{A}(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0)$ and $\mathcal{B}(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0)$ appropriately.

B.6.3 Unconditional means of the shocks

A third possible parametrisation associated to the previous one would replace $\boldsymbol{\tau}$ by $\boldsymbol{\upsilon}$, where $\boldsymbol{\tau} = \mathbf{J}\boldsymbol{\upsilon}$. This is slightly more involved than before because it combines mean and variance parameters. The steps, though, are otherwise standard. In particular, given that $d\boldsymbol{\tau} = (d\mathbf{J})\boldsymbol{\upsilon} + \mathbf{J}d\boldsymbol{\upsilon}$, we will have that after vectorising

$$d\boldsymbol{\tau} = (\boldsymbol{\upsilon}'\otimes\mathbf{I}_N)\boldsymbol{\Delta}_N dveco(\mathbf{J}) + \mathbf{J}d\boldsymbol{\upsilon},$$

whence

$$\frac{\partial\boldsymbol{\tau}}{\partial(\mathbf{j}',\boldsymbol{\upsilon}')} = [\ (\boldsymbol{\upsilon}'\otimes\mathbf{I}_N)\boldsymbol{\Delta}_N \quad \mathbf{J}\ ]. \qquad \text{(B34)}$$

Hence, if we combine this expression with the Jacobian in (B31) we will have

$$\left[\begin{array}{c} \mathbf{s}_{\mathbf{j}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\boldsymbol{\upsilon}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\phi}) \end{array}\right] = \left(\begin{array}{cc} \partial\boldsymbol{\tau}'/\partial\mathbf{j} & \partial\mathbf{c}'/\partial\mathbf{j} \\ \partial\boldsymbol{\tau}'/\partial\boldsymbol{\upsilon} & \mathbf{0} \\ \mathbf{0} & \partial\mathbf{c}'/\partial\boldsymbol{\psi} \end{array}\right) \left[\begin{array}{c} \mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\mathbf{c}}(\boldsymbol{\phi}) \end{array}\right]$$

$$= \left[\begin{array}{cc} \boldsymbol{\Delta}_N'(\boldsymbol{\upsilon}\otimes\mathbf{I}_N) & \boldsymbol{\Delta}_N'(\boldsymbol{\Psi}\otimes\mathbf{I}_N) \\ \mathbf{J}' & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{J}') \end{array}\right] \left[\begin{array}{c} \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*} \\ (\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\mathbf{I}_N\otimes\boldsymbol{\Psi}^{-1})vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \end{array}\right]$$

$$= -\left[\begin{array}{cc} \boldsymbol{\Delta}_N'(\boldsymbol{\upsilon}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1} & \boldsymbol{\Delta}_N'(\mathbf{I}_N\otimes\mathbf{J}^{-1\prime})(\boldsymbol{\Psi}\otimes\boldsymbol{\Psi}^{-1}) \\ \boldsymbol{\Psi}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{-1}\mathbf{E}_N' \end{array}\right] \left[\begin{array}{c} \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*} \\ vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \end{array}\right]$$

$$= -\left[\begin{array}{c} \boldsymbol{\Delta}_N'(\boldsymbol{\upsilon}\otimes\mathbf{I}_N)\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*} + veco\left\{\mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Psi} + \mathbf{J}^{-1\prime}\right\} \\ \boldsymbol{\Psi}^{-1}\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*} \\ \boldsymbol{\Psi}^{-1}vecd\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) + \mathbf{I}_N\right\} \end{array}\right]$$

This expression makes immediately clear that $\boldsymbol{\upsilon}$ and $\boldsymbol{\psi}$ mop up any potential biases in the MLEs based on a misspecified distribution, as we saw in Proposition 1.

B.6.4 Non-standardised shocks

So far, we have maintained the identifying assumption that the distribution for the shocks used for estimation purposes is such that they all have 0 mean and unit variance by construction. However, this restriction might be cumbersome to impose with certain distributions. For example, while in Appendix H.1 we explain how to recursively standardise a discrete mixture of normals with K components, the expressions become tedious for $K > 3$.

For that reason, it may sometimes be preferable to work with non-standardised distributions. For identification purposes, though, one would need to eliminate $\boldsymbol{\upsilon}$ (or $\boldsymbol{\tau}$) and $\boldsymbol{\psi}$ from the set of parameters to estimate, leaving only $\mathbf{a}$ and $\mathbf{j}$ plus an extended set of shape parameters $\boldsymbol{\rho} = (\boldsymbol{\rho}_1', \ldots, \boldsymbol{\rho}_N')'$ that implicitly capture the mean and variance of the structural shocks. In the case of a finite Gaussian mixture with K_i components, those extended set parameters would be

$$\boldsymbol{\rho}_i = (\boldsymbol{\mu}_i', \boldsymbol{\sigma}_i', \boldsymbol{\lambda}_i')' = (\mu_{i1}, \ldots, \mu_{iK_i}; \sigma_{i1}^2, \ldots, \sigma_{iK_i}^2; \lambda_{i1}, \ldots, \lambda_{iK_i-1})'$$

if we impose the adding up constraint on the mixing probabilities by making $\lambda_{iK_i} = 1 - \lambda_{i1} - \ldots - \lambda_{iK_i-1}$.

In this context, it is easy to see that the expressions for the covariance matrix of the score and the expected value of the Hessian that we derived in this appendix continue to hold if we define the estimated pseudo-standardised shocks as $\boldsymbol{\varepsilon}_t^*(\mathbf{0}, \mathbf{a}, \mathbf{j}, \ell_N) = \mathbf{J}^{-1}(\mathbf{y}_t - \mathbf{A}_1\mathbf{y}_{t-1} - \ldots - \mathbf{A}_p\mathbf{y}_{t-p})$, so that in this respect it is as if $\boldsymbol{\tau} = \mathbf{0}$ and $\boldsymbol{\psi} = \ell_N$. Nevertheless, it is important to remember that the consistency of the PMLEs of $\mathbf{a}$ and $\mathbf{j}$ hinges on the shape parameters effectively being estimated to mop up the biases in the estimation of the mean and variance of the shocks.

Having estimated the parameters in this way, it would be straightforward to go back to the standardised parametrisation by means of the delta method if we express the mean and variance of each shock as a function of its extended set of shape parameters $\boldsymbol{\rho}_i$. For example, in the case of finite mixtures of normals,

$$\upsilon_i = \lambda_{i1}\mu_{i1} + \ldots + \lambda_{iK_i-1}\mu_{iK_i-1} + (1 - \lambda_{i1} - \ldots - \lambda_{iK_i-1})\mu_{iK_i}$$

so

$$\frac{\partial \upsilon_i}{\partial \boldsymbol{\mu}_i'} = \begin{pmatrix} \lambda_{i1} & \ldots & \lambda_{iK_i-1} & 1 - \lambda_{i1} - \ldots - \lambda_{iK_i-1} \end{pmatrix}$$

while

$$\frac{\partial \upsilon_i}{\partial \boldsymbol{\lambda}_i'} = \begin{pmatrix} \mu_{i1} - \mu_{iK_i} & \cdots & \mu_{iK_i-1} - \mu_{iK_i} \end{pmatrix}.$$

Similarly, given that

$$\psi_i = \sqrt{\lambda_{i1}(\mu_{i1}^2 + \sigma_{i1}^2) + \ldots + \lambda_{i,K_i-1}(\mu_{i,K_i-1}^2 + \sigma_{iK_i-1}^2) + (1 - \lambda_{i1} - \ldots - \lambda_{iK_i-1})(\mu_{i,K_i}^2 + \sigma_{i,K_i}^2) - \upsilon_{ii}^2},$$

we will have that $\partial\psi_i/\partial\rho_i' = .5\psi_i \cdot \partial\psi_i/\partial\rho_i'$, where

$$\begin{aligned}
\frac{d\psi_i^2}{d\boldsymbol{\mu}_i'} &= 2\begin{bmatrix} \lambda_{i1}\mu_{i1} & \ldots & \lambda_{iK_i-1}\mu_{iK_i-1} & (1 - \lambda_{i1} - \ldots - \lambda_{iK_i-1})\mu_{iK_i} \end{bmatrix} - 2\upsilon_i \frac{\partial \upsilon_i}{\partial \boldsymbol{\mu}_i'} \\
&= 2\begin{bmatrix} \lambda_{i1}(\mu_{i1} - \upsilon_i) & \ldots & \lambda_{i,K_i-1}(\mu_{iK_i-1} - \upsilon_i) & (1 - \lambda_{i1} - \ldots - \lambda_{iK_i-1})(\mu_{i,K_i} - \upsilon_i) \end{bmatrix},
\end{aligned}$$

$$\frac{\partial \psi_i^2}{\partial \boldsymbol{\sigma}_i'} = \begin{pmatrix} \lambda_{i1} & \ldots & \lambda_{i,K_i-1} & 1 - \lambda_{i1} - \ldots - \lambda_{iK_i-1} \end{pmatrix},$$

and

$$\frac{\partial \psi_i^2}{\partial \boldsymbol{\lambda}_i'} = \begin{bmatrix} (\mu_{i1}^2 + \sigma_{i1}^2) - (\mu_{iK_i}^2 + \sigma_{iK_i}^2) & \ldots & (\mu_{iK_i-1}^2 + \sigma_{iK_i-1}^2) - (\mu_{iK_i}^2 + \sigma_{iK_i}^2) \end{bmatrix} - 2\upsilon_i \frac{\partial \upsilon_i}{\partial \boldsymbol{\lambda}_i'}$$

$$\begin{bmatrix} (\mu_{i1} + \mu_{iK_i} - 2\upsilon_i)(\mu_{i1} - \mu_{iK_i}) + (\sigma_{i1}^2 - \sigma_{iK_i}^2) & \ldots & (\mu_{iK_i-1} + \mu_{iK_i} - 2\upsilon_i)(\mu_{iK_i-1} - \mu_{iK_i}) + (\sigma_{iK_i-1}^2 - \sigma_{iK_i}^2) \end{bmatrix}.$$

C The FS consistent estimators

In Fiorentini and Sentana (2007) we proposed consistent estimators for the mean and variance parameters of conditionally heteroskedastic dynamic regression models that are inconsistently estimated by a distributionally misspecified log-likelihood function. An important advantage of these estimators is that they can be written in closed-form as simple means, variances and covariances of residuals, which make them very easy to code. In turn, in Fiorentini and Sentana (2019) we studied in detail their statistical properties. The purpose of this appendix is to derive analogous estimators in the context of the structural vector autoregression in (1) when its shocks are cross-sectionally independent.

C.1 The estimators

There are in fact two types of consistent FS estimators. The first one assumes that the true distribution of the shocks is symmetric, and exploits the fact that the Gaussian pseudo score for

$\boldsymbol{\psi}$ is

$$\mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\theta};\mathbf{0}) = \boldsymbol{\Psi}^{-1} vecd[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N] \tag{C35}$$

to propose

$$\tilde{\psi}_i^2 = \psi_{it}(\hat{\boldsymbol{\tau}}, \hat{\mathbf{a}}, \hat{\jmath}),\ i = 1, \ldots, N,$$

as in (3), but using the consistent PMLEs of $\boldsymbol{\tau}$, $\mathbf{a}$ and $\mathbf{j}$ to compute $\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\tau}}, \hat{\mathbf{a}}, \hat{\jmath}, \ell_N) = \widehat{\mathbf{J}}^{-1}(\mathbf{y}_t - \hat{\boldsymbol{\tau}} - \hat{\mathbf{A}}_1\mathbf{y}_{t-1} - \ldots - \hat{\mathbf{A}}_p\mathbf{y}_{t-p})$.

Let $\tilde{\boldsymbol{\Psi}} = diag(\tilde{\psi}_1, \ldots, \tilde{\psi}_N)$ denote the symmetric FS estimators obtained in this manner. Then the FS consistent estimator of $\mathbf{C}$ is obtained as

$$\tilde{\mathbf{C}} = \hat{\mathbf{J}}\tilde{\boldsymbol{\Psi}}. \tag{C36}$$

In turn, the asymmetric FS estimators exploit the fact that the sample mean of the Gaussian pseudo scores for $\boldsymbol{\tau}$, which are given by

$$\mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\theta};\mathbf{0}) = \mathbf{J}^{-1\prime}\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}), \tag{C37}$$

will be 0 if and only if the sample average of $\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}, \mathbf{a})$ is 0 too. This yields

$$\widetilde{\tau}_i = \tau_i(\hat{\mathbf{a}}),\ i = 1, \ldots, N,$$

as in (2), but using the consistent PMLEs of $\mathbf{a}$ to compute $\boldsymbol{\varepsilon}_t(\mathbf{0}, \hat{\mathbf{a}}) = (\mathbf{y}_t - \hat{\mathbf{A}}_1\mathbf{y}_{t-1} - \ldots - \hat{\mathbf{A}}_p\mathbf{y}_{t-p})$. Naturally, if we parametrise the model in terms of $\boldsymbol{\upsilon}$ instead, then

$$\tilde{\upsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}^*(\mathbf{0}, \hat{\mathbf{a}}, \hat{\jmath}, \ell_N),$$

while we would use $(\mathbf{I} - \hat{\mathbf{A}}_1 - \ldots - \hat{\mathbf{A}}_p)^{-1}\tilde{\boldsymbol{\tau}} = (\mathbf{I} - \hat{\mathbf{A}}_1 - \ldots - \hat{\mathbf{A}}_p)^{-1}\mathbf{J}\tilde{\boldsymbol{\upsilon}}$ if we focused on $\boldsymbol{\mu}$.

As for the standard deviations, in the asymmetric case the estimator of $\boldsymbol{\psi}$ would be

$$\tilde{\psi}_i^2 = \psi_{it}(\tilde{\boldsymbol{\tau}}, \hat{\mathbf{a}}, \hat{\jmath}),\ i = 1, \ldots, N,$$

with $\boldsymbol{\varepsilon}_t^*(\tilde{\boldsymbol{\tau}}, \hat{\mathbf{a}}, \hat{\jmath}, \ell_N) = \hat{\mathbf{J}}^{-1}(\mathbf{y}_t - \tilde{\boldsymbol{\tau}} - \hat{\mathbf{A}}_1\mathbf{y}_{t-1} - \ldots - \hat{\mathbf{A}}_p\mathbf{y}_{t-p})$, while the estimator of $\mathbf{c}$ would again be given by (C36).

In both cases, the fact that the first step estimators of $\mathbf{j}$ are based on a non-Gaussian log-likelihood function allows the use of the Gaussian scores for $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ despite the fact that the Gaussian log-likelihood function is unable to point identify $\mathbf{j}$.

C.2 The asymptotic covariance matrix

To obtain the asymptotic covariance matrices, we can follow Proposition 13 of the supplemental appendix of Fiorentini and Sentana (2019) and treat the sequential estimators as exactly identified GMM estimators based on the extended set of moment conditions:

$$\begin{bmatrix} \mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\tau},\mathbf{a},\mathbf{j},\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{s}_{\mathbf{a}t}(\boldsymbol{\tau},\mathbf{a},\mathbf{j},\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{s}_{\mathbf{j}t}(\boldsymbol{\tau},\mathbf{a},\mathbf{j},\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\tau},\mathbf{a},\mathbf{j},\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\tau},\mathbf{a},\mathbf{j},\boldsymbol{\psi},\boldsymbol{\varrho}) \\ \mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0}) \\ \mathbf{s}_{\bar{\boldsymbol{\psi}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0}) \end{bmatrix}, \tag{C38}$$

where we have introduced $\bar{\boldsymbol{\tau}}$ and $\bar{\boldsymbol{\psi}}$ to denote the parameters that are consistently estimated in the second step, which differ from their namesakes in the first stage, whose pseudo true values are $\boldsymbol{\tau}_\infty$ and $\boldsymbol{\psi}_\infty$. In this respect, it is also convenient to distinguish between the shocks estimated with the first step estimators, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\tau},\mathbf{a},\mathbf{j},\boldsymbol{\psi})$, and the FS standardised shocks

$$\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) = \bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\left(\mathbf{y}_t - \bar{\boldsymbol{\tau}} - \sum_{j=1}^{p}\mathbf{A}_j\mathbf{y}_{t-j}\right), \tag{C39}$$

Appendices B.3 and B.4 give us the covariance matrix and expected Jacobian of the first five components of the influence functions in (C38), which we combined to obtain the asymptotic covariance matrix of the first step estimators. To obtain the joint distribution of the first- and second-step estimators, we need all the remaining elements. In this respect, it is worth mentioning that quite a few other blocks of the Jacobian will be identically 0 because $\mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$ and $\mathbf{s}_{\bar{\boldsymbol{\psi}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$ do not depend on $\boldsymbol{\varrho}$ or the inconsistent first-step estimators of $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$.

As for the remaining blocks of the expected Jacobian, (C37) leads to

$$\begin{gathered} ds_{\boldsymbol{\tau}}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0}) = d(\mathbf{J}^{-1\prime})\cdot\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) + \mathbf{J}^{-1\prime}\cdot d(\bar{\boldsymbol{\Psi}}^{-1})\cdot\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) \\ +\mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}d\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) = -\mathbf{J}^{-1\prime}\cdot d\mathbf{J}'\cdot\mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) \\ -\mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}\cdot d\bar{\boldsymbol{\Psi}}^{-1}\cdot\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) + \mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}d\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) \\ = -[\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1\prime}\otimes\mathbf{J}^{-1\prime}]dvec(\mathbf{J}') - [\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}^{-1}\otimes\mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}]dvec(\bar{\boldsymbol{\Psi}}) \\ +\mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}d\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) = -[\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1\prime}\otimes\mathbf{J}^{-1\prime}]\mathbf{K}_{NN}dvec(\mathbf{J}) \\ -[\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}^{-1}\otimes\mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}]\mathbf{E}_N dvecd(\bar{\boldsymbol{\Psi}}) + \mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}d\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}) \\ = -[\mathbf{J}^{-1\prime}\otimes\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1\prime}]\boldsymbol{\Delta}_N dveco(\mathbf{J}) \\ -[\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}^{-1}\otimes\mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}]\mathbf{E}_N dvecd(\bar{\boldsymbol{\Psi}}) + \mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}d\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}). \end{gathered}$$

Given that (C39) implies that

$$d\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})=d(\bar{\boldsymbol{\Psi}}^{-1})\cdot\mathbf{J}^{-1}\left(\mathbf{y}_t-\bar{\boldsymbol{\tau}}-\sum_{j=1}^{p}\mathbf{A}_j\mathbf{y}_{t-j}\right)+\bar{\boldsymbol{\Psi}}^{-1}\cdot d(\mathbf{J}^{-1})\cdot\left(\mathbf{y}_t-\bar{\boldsymbol{\tau}}-\sum_{j=1}^{p}\mathbf{A}_j\mathbf{y}_{t-j}\right)$$

$$-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\cdot d\bar{\boldsymbol{\tau}}-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\sum_{j=1}^{p}(d\mathbf{A}_j)\cdot\mathbf{y}_{t-j}=-\bar{\boldsymbol{\Psi}}^{-1}\cdot d\bar{\boldsymbol{\Psi}}\cdot\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\left(\mathbf{y}_t-\bar{\boldsymbol{\tau}}-\sum_{j=1}^{p}\mathbf{A}_j\mathbf{y}_{t-j}\right)$$

$$-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\cdot d\mathbf{J}\cdot\mathbf{J}^{-1}\left(\mathbf{y}_t-\bar{\boldsymbol{\tau}}-\sum_{j=1}^{p}\mathbf{A}_j\mathbf{y}_{t-j}\right)-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\cdot d\bar{\boldsymbol{\tau}}$$

$$-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\sum_{j=1}^{p}(d\mathbf{A}_j)\cdot\mathbf{y}_{t-j}=-\bar{\boldsymbol{\Psi}}^{-1}\cdot d\bar{\boldsymbol{\Psi}}\cdot\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\cdot d\mathbf{J}\cdot\bar{\boldsymbol{\Psi}}\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})$$

$$-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\cdot d\bar{\boldsymbol{\tau}}-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\sum_{j=1}^{p}(d\mathbf{A}_j)\cdot\mathbf{y}_{t-j}=-[\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\otimes\bar{\boldsymbol{\Psi}}^{-1}]dvec(\bar{\boldsymbol{\Psi}})$$

$$-[\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}\otimes\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}]dvec(\mathbf{J})-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}d\bar{\boldsymbol{\tau}}-\sum_{j=1}^{p}(\mathbf{y}_{t-j}'\otimes\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1})dvec(\mathbf{A}_j)$$

$$=-\ [\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\otimes\bar{\boldsymbol{\Psi}}^{-1}]\mathbf{E}_N dvecd(\bar{\boldsymbol{\Psi}})$$

$$-[\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}\otimes\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}]\boldsymbol{\Delta}_N dveco(\mathbf{J})-\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}d\bar{\boldsymbol{\tau}}-\sum_{j=1}^{p}(\mathbf{y}_{t-j}'\otimes\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1})dvec(\mathbf{A}_j),$$

if we take into account that $E[\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0)|\boldsymbol{\varphi}_0]=\mathbf{0}$, then we are only left with the following non-zero blocks for the expected Jacobian of $\mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$:

$$\begin{aligned}E\left[\left.\frac{\partial s_{\boldsymbol{\tau}}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})}{\partial\mathbf{a}_j'}\right|\boldsymbol{\varphi}_0\right] &= -\mathbf{J}_0^{-1\prime}\bar{\boldsymbol{\Psi}}_0^{-1}(\mathbf{y}_{t-j}'\otimes\bar{\boldsymbol{\Psi}}_0^{-1}\mathbf{J}_0^{-1}).\\ E\left[\left.\frac{\partial s_{\boldsymbol{\tau}}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})}{\partial\bar{\boldsymbol{\tau}}'}\right|\boldsymbol{\varphi}_0\right] &= -\mathbf{J}_0^{-1\prime}\bar{\boldsymbol{\Psi}}_0^{-2}\mathbf{J}_0^{-1}.\end{aligned}$$

Similarly, we have from (C35) that

$$ds_{\bar{\psi}_i}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})=-\bar{\psi}_i^2[\varepsilon_{it}^2(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})-1]d\bar{\psi}_i+\frac{2\varepsilon_{it}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})}{\bar{\psi}_i}d\varepsilon_{it}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}}).$$

As a result, the expected value of the Jacobian of $s_{\bar{\psi}_i}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$ with respect to $\bar{\boldsymbol{\tau}}$ and $\mathbf{a}$ evaluated at the true parameter values will be 0. In addition,

$$\frac{ds_{\bar{\psi}_i}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})}{d\bar{\psi}_i}=-\bar{\psi}_i^2[\varepsilon_{it}^2(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})-1]-\frac{2\varepsilon_{it}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})^2}{\bar{\psi}_i^2}$$

so that

$$E\left[\left.\frac{\partial\mathbf{s}_{\bar{\boldsymbol{\psi}}t}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})}{\partial\bar{\boldsymbol{\psi}}'}\right|\boldsymbol{\varphi}_0\right]=-2\bar{\boldsymbol{\Psi}}^{-2}.$$

Finally,

$$\begin{aligned}\frac{ds_{\bar{\boldsymbol{\psi}}}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})}{d\mathbf{j}'} &= 2diag[\bar{\boldsymbol{\Psi}}_0^{-1}\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})]\frac{d\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})}{d\mathbf{j}'}\\ &= -2diag[\bar{\boldsymbol{\Psi}}_0^{-1}\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})][\boldsymbol{\varepsilon}_t'(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}})\bar{\boldsymbol{\Psi}}\otimes\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}]\boldsymbol{\Delta}_N\end{aligned}$$

so that

$$E\left[\frac{ds_{\bar{\boldsymbol{\psi}}}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})}{d\mathbf{j}'}\right] = -2diag_N\left(\bar{\boldsymbol{\Psi}}^{-1}\mathbf{J}^{-1}\right),$$

where $diag_N$ is the N block-diagonalisation operator in Yang (2000), which is such that

$$diag_N\left(\mathbf{C}^{-1}\right) = \sum_{i=1}^{N}\mathbf{e}_i\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{H}_i$$

when the blocks are of dimension $1\times N$, with $\mathbf{e}_i$ being the i^{th} canonical vector and $\mathbf{H}_1 = [\mathbf{I}_N, \mathbf{0}_{N\times(N^2-N)}]$, $\mathbf{H}_2 = [\mathbf{0}_{N\times N}, \mathbf{I}_N, \mathbf{0}_{N\times(N^2-2N)}]$, …, $\mathbf{H}_N = [\mathbf{0}_{N\times(N^2-N)}, \mathbf{I}_N]$.

Some of the expressions for the expected Jacobian would be different if in the first-step one estimates the model in terms of $\mathbf{c}$ rather than $\mathbf{j}$ and $\boldsymbol{\psi}$. Nevertheless, the relevant terms can be easily obtained from the ones above by using the chain rule for first-derivatives

$$\frac{ds_{\bar{\boldsymbol{\psi}}}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})}{d\mathbf{c}'} = \frac{ds_{\bar{\boldsymbol{\psi}}}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})}{d(\mathbf{j}',\boldsymbol{\psi}')}\frac{d(\mathbf{j}',\boldsymbol{\psi}')'}{d\mathbf{c}'},$$

with the last term computed from the inverse of (B31).

Let us now turn to the covariance matrix of the influence functions. The covariance between $\mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$ and $\mathbf{s}_{\bar{\boldsymbol{\psi}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$ is straightforward in view of (C35) and (C37) and the cross-sectional independence of the shocks. Specifically,

$$\begin{aligned}V[\mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\theta}_0;\mathbf{0})|\boldsymbol{\varphi}_0] &= \mathbf{J}_0^{-1\prime}\bar{\boldsymbol{\Psi}}_0^{-2}\mathbf{J}_0^{-1},\\ cov[\mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\theta}_0;\mathbf{0}),\mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\theta}_0;\mathbf{0})] &= \mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}_0^{-2}\mathrm{K}_{ls}\end{aligned}$$

and

$$V[\mathbf{s}_{\boldsymbol{\psi}t}(\boldsymbol{\theta}_0;\mathbf{0})|\boldsymbol{\varphi}_0] = \bar{\boldsymbol{\Psi}}_0^{-2}\mathrm{K}_{ss},$$

where K_{ls} and K_{ss} are the diagonal matrices of order N with typical element $\varphi(\boldsymbol{\varrho}_i)$ and $\kappa(\boldsymbol{\varrho}_i)-1$ defined in the proof of Proposition 5.

To compute the blocks of the covariance matrix of $\mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$ and $\mathbf{s}_{\bar{\boldsymbol{\psi}}t}(\bar{\boldsymbol{\tau}},\mathbf{a},\mathbf{j},\bar{\boldsymbol{\psi}},\mathbf{0})$ with the non-Gaussian scores with respect to all the other parameters, we must take into account that the latter scores will be evaluated at inconsistent parameter estimators for $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$. Specifically,

$$\begin{aligned}&E\left\{\mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})\left[\begin{array}{cc}\mathbf{s}_{\boldsymbol{\tau}t}'(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty) & \mathbf{s}_{\mathbf{a}t}'(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty)\end{array}\right]\Big|\boldsymbol{\varphi}_0\right\}\\ =\ &\mathbf{J}_0^{-1\prime}\bar{\boldsymbol{\Psi}}_0^{-1}\mathrm{D}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\bar{\boldsymbol{\Psi}}_0^{-1}\mathbf{J}_0^{-1}\left(\begin{array}{cccc}\mathbf{I}_N & \boldsymbol{\mu}_0'\otimes\mathbf{I}_N & \dots & \boldsymbol{\mu}_0'\otimes\mathbf{I}_N\end{array}\right),\end{aligned}$$

where

$$\mathrm{D}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = -E\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^{*\prime}}\right|\boldsymbol{\varphi}_0\right]$$

is a diagonal matrix with typical element

$$\mathrm{D}_{ll}^i(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = -E\left\{\varepsilon_{it}^*(\boldsymbol{\theta}_0)\frac{\partial \ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial \varepsilon_i^*}\right|\boldsymbol{\varphi}_0\right\} \neq -1$$

thanks to the cross-sectional independence of both the true and pseudo-standardised shocks (B17).

Similarly,

$$\begin{aligned}
& E\left\{\mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})\left[\begin{array}{cc}\mathbf{s}'_{\mathbf{j}t}(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty) & \mathbf{s}'_{\boldsymbol{\psi}t}(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty)\end{array}\right]\right|\boldsymbol{\varphi}_0\right\} \\
= \; & \mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}\mathcal{D}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\left[\begin{array}{cc}\mathbf{E}_N\boldsymbol{\Psi}_\infty^{-1} & (\boldsymbol{\Psi}_\infty\otimes\boldsymbol{\Psi}_\infty^{-1})(\mathbf{I}_N\otimes\mathbf{J}_0^{-1})\boldsymbol{\Delta}_N\end{array}\right],
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{D}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= -E\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)vec'\left\{\mathbf{I}_N+\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_\infty)\right\}\right|\boldsymbol{\varphi}_0\right] \\
&= [\begin{array}{ccccc}\mathcal{D}_{ls}^1(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \ldots & \mathcal{D}_{ls}^i(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \ldots & \mathcal{D}_{ls}^N(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\end{array}],
\end{aligned}$$

and $\mathcal{D}_{ls}^i(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N whose non-zero elements are

$$\begin{gathered}
\mathrm{D}_{ll}^j(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] \; (j\neq i) \text{ and} \\
\mathrm{D}_{ls}^i(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0) = -E\left\{\varepsilon_{it}^*(\boldsymbol{\theta}_0)\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial \varepsilon_i^*}\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\right\} \; (j=i).
\end{gathered}$$

If in the first-step one estimates the model in terms of $\mathbf{c}$ rather than $\mathbf{j}$ and $\boldsymbol{\psi}$, then the relevant covariances would change, but once again they could be obtained from the previous expression by using the chain rule for derivatives.

In addition,

$$E\left\{\mathbf{s}_{\bar{\boldsymbol{\tau}}t}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})\mathbf{s}'_{\boldsymbol{\varrho}t}(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty)'\right|\boldsymbol{\varphi}_0\right\} = \mathbf{J}^{-1\prime}\bar{\boldsymbol{\Psi}}^{-1}\mathcal{D}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0),$$

where

$$\mathcal{D}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) = E\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\frac{\partial \ln f_i[\varepsilon_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial \boldsymbol{\varrho}'}\right|\boldsymbol{\varphi}_0\right],$$

an $N\times q$ block diagonal matrix with typical diagonal block of size $1\times q_i$

$$\mathrm{D}_{lr}^i(\boldsymbol{\phi}_\infty,\boldsymbol{\varphi}_0) = E\left[\varepsilon_{it}^*(\boldsymbol{\theta}_0)\frac{\partial \ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial \boldsymbol{\varrho}_i'}\right|\boldsymbol{\varphi}_0\right].$$

In turn,

$$\begin{aligned}
& E\left\{\mathbf{s}_{\bar{\boldsymbol{\psi}}t}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})\left[\begin{array}{cc}\mathbf{s}'_{\boldsymbol{\tau}t}(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty) & \mathbf{s}'_{\mathbf{a}t}(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty)\end{array}\right]\right|\boldsymbol{\varphi}_0\right\} \\
= \; & \bar{\boldsymbol{\Psi}}_0^{-1}\mathbf{E}_N'\mathcal{D}_{sl}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\bar{\boldsymbol{\Psi}}_0^{-1}\mathbf{J}_0^{-1}\left(\begin{array}{cccc}\mathbf{I}_N & \boldsymbol{\mu}_0'\otimes\mathbf{I}_N & \ldots & \boldsymbol{\mu}_0'\otimes\mathbf{I}_N\end{array}\right) \\
= \; & \bar{\boldsymbol{\Psi}}_0^{-1}\mathrm{D}_{sl}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\bar{\boldsymbol{\Psi}}_0^{-1}\mathbf{J}_0^{-1}\left(\begin{array}{cccc}\mathbf{I}_N & \boldsymbol{\mu}_0'\otimes\mathbf{I}_N & \ldots & \boldsymbol{\mu}_0'\otimes\mathbf{I}_N\end{array}\right)
\end{aligned}$$

in view of (A5), where

$$\mathcal{D}_{sl}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=-E\left[vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)-\mathbf{I}_N\right]\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^{*\prime}}\right|\boldsymbol{\varphi}_0\Bigg]=\mathbf{E}_N\mathrm{D}_{sl}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0),$$

and $\mathrm{D}_{sl}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is a diagonal matrix of order N whose diagonal elements are

$$\mathrm{D}_{sl}^{i}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)=-E\left\{\varepsilon_{it}^{*2}(\boldsymbol{\theta}_0)\frac{\partial\ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon_i^*}\right|\boldsymbol{\varphi}_0\Bigg\}.$$

Moreover,

$$\begin{aligned}&E\left\{\mathbf{s}_{\bar{\psi}t}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})\left[\begin{array}{cc}\mathbf{s}_{\mathbf{j}t}'(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty) & \mathbf{s}_{\boldsymbol{\psi}t}'(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty)\end{array}\right]\right|\boldsymbol{\varphi}_0\Bigg\}\\ =\;&\bar{\boldsymbol{\Psi}}^{-1}\mathbf{E}_N'\mathcal{D}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\left[\begin{array}{cc}\mathbf{E}_N\boldsymbol{\Psi}_\infty^{-1} & (\boldsymbol{\Psi}_\infty\otimes\boldsymbol{\Psi}_\infty^{-1})(\mathbf{I}_N\otimes\mathbf{J}_0^{-1})\boldsymbol{\Delta}_N\end{array}\right],\end{aligned}$$

where

$$\begin{aligned}\mathcal{D}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= -E\left[vec\left[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)-\mathbf{I}_N\right]vec'\left\{\mathbf{I}_N+\frac{\partial\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_\infty)\right\}\right|\boldsymbol{\varphi}_0\Bigg]\\ &= \left[\begin{array}{ccccc}\mathcal{D}_{ss}^{11}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{D}_{ss}^{1i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{D}_{ss}^{1N}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\ \vdots & \ddots & \vdots & & \vdots\\ \mathcal{D}_{ss}^{i1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{D}_{ss}^{ii}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{D}_{ss}^{iN}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\ \vdots & & \vdots & \ddots & \vdots\\ \mathcal{D}_{ss}^{N1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{D}_{ss}^{Ni}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \dots & \mathcal{D}_{ss}^{NN}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\end{array}\right].\end{aligned}$$

Although this matrix contains many 0's, its pattern is a bit more complex than in Appendix B.4. For example, the elements of this matrix that correspond to the off-diagonal 1's in $\mathbf{K}_{NN}$ are no longer 1. Specifically, they will be

$$-E\left\{\varepsilon_{it}^*(\boldsymbol{\theta}_0)\varepsilon_{jt}^*(\boldsymbol{\theta}_0)\frac{\partial\ln f[\varepsilon_{jt}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon_j^*}\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\Bigg\}=\mathrm{D}_{ll}^{j}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon_{it}^*(\boldsymbol{\theta}_0)\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]$$

because

$$E[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]=E\{\boldsymbol{\varepsilon}_t^*[(\boldsymbol{v}_0'-\boldsymbol{v}_\infty')+\boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\Psi}_0]\boldsymbol{\Psi}_\infty^{-1}|\boldsymbol{\varphi}_0\}=\boldsymbol{\Psi}_0\boldsymbol{\Psi}_\infty^{-1}\neq\mathbf{I}_N.$$

But all the remaining elements of $\mathcal{D}_{ss}^{ij}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ $(j\neq i)$ are 0 except its i^{th} diagonal element, which will be given by

$$\mathrm{D}_{sl}^{i}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\varphi}_0)E[\varepsilon_{jt}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0],$$

while $\mathcal{D}_{ss}^{ii}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is another diagonal matrix of the same size whose non-zero elements are

$$\begin{gathered}\mathrm{D}_{ll}^{j}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)E[\varepsilon_{it}^*(\boldsymbol{\theta}_0)\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0]\ (j\neq i)\text{ and}\\ \mathrm{D}_{ss}^{i}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=-E\left\{\varepsilon_{it}^{*2}(\boldsymbol{\theta}_0)\frac{\partial\ln f_i[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\varepsilon_i^*}\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)\right|\boldsymbol{\varphi}_0\Bigg\}-1\ (j=i).\end{gathered}$$

Nevertheless, since the scores for $\bar{\boldsymbol{\psi}}$ only involve $vecd[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)-\mathbf{I}_N]$, the calculation of some of the above elements is unnecessary.

Finally,

$$E\left\{\mathbf{s}_{\bar{\psi}t}(\bar{\boldsymbol{\tau}}_0,\mathbf{a}_0,\mathbf{j}_0,\bar{\boldsymbol{\psi}}_0,\mathbf{0})\mathbf{s}'_{\boldsymbol{\varrho}t}(\boldsymbol{\tau}_\infty,\mathbf{a}_0,\mathbf{j}_0,\boldsymbol{\psi}_\infty,\boldsymbol{\varrho}_\infty)'\right|\boldsymbol{\varphi}_0\right\}=\bar{\boldsymbol{\Psi}}^{-1}\mathbf{E}'_N\mathcal{D}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=\bar{\boldsymbol{\Psi}}^{-1}\mathrm{D}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0),$$

where

$$\mathcal{D}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=E\left[vec[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta}_0)-\mathbf{I}_N]\frac{\partial\ln f[\boldsymbol{\varepsilon}^*_t(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_\infty]}{\partial\boldsymbol{\varrho}'}\right|\boldsymbol{\varphi}_0\right]=\mathbf{E}_N\mathrm{D}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0),$$

and $\mathrm{D}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ is another block diagonal matrix of order $N\times q$ with typical block of size $1\times q_i$

$$\mathrm{D}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)=E\left\{\varepsilon^{2*}_{it}(\boldsymbol{\theta}_0)\frac{\partial\ln f_i[\varepsilon^*_{it}(\boldsymbol{\theta}_\infty);\boldsymbol{\varrho}_{i\infty}]}{\partial\boldsymbol{\varrho}_i}\right|\boldsymbol{\varphi}_0\right\}.$$

Once again, sample versions of the previous expressions will consistently estimate their population counterparts under standard regularity conditions as long as we replace $\boldsymbol{\theta}_\infty$ by the first-step estimator and $\boldsymbol{\theta}_0$ by the FS one in evaluating $\varepsilon^*_{it}(\boldsymbol{\theta}_\infty)$ and $\varepsilon^*_{it}(\boldsymbol{\theta}_0)$, respectively.

We can use the delta method to obtain the covariance matrix of the FS estimator of $\mathbf{c}$ in (C36). Specifically, if we define

$$\bar{\mathbf{C}}=\mathbf{J}\bar{\boldsymbol{\Psi}}=\mathbf{C}\boldsymbol{\Psi}^{-1}\bar{\boldsymbol{\Psi}},$$

it immediately follows that

$$d\bar{\mathbf{C}}=d\mathbf{C}\cdot\boldsymbol{\Psi}^{-1}\bar{\boldsymbol{\Psi}}+\mathbf{C}\cdot d(\boldsymbol{\Psi}^{-1})\cdot\bar{\boldsymbol{\Psi}}+\mathbf{C}\boldsymbol{\Psi}^{-1}\cdot d\bar{\boldsymbol{\Psi}}=d\mathbf{C}\cdot\boldsymbol{\Psi}^{-1}\bar{\boldsymbol{\Psi}}-\mathbf{C}\boldsymbol{\Psi}^{-1}\cdot d\boldsymbol{\Psi}\cdot\boldsymbol{\Psi}^{-1}\bar{\boldsymbol{\Psi}}+\mathbf{C}\boldsymbol{\Psi}^{-1}\cdot d\bar{\boldsymbol{\Psi}},$$

which after vectorising yields

$$d\bar{\mathbf{c}}=(\bar{\boldsymbol{\Psi}}\boldsymbol{\Psi}^{-1}\otimes\mathbf{I}_N)d\mathbf{c}-(\bar{\boldsymbol{\Psi}}\boldsymbol{\Psi}^{-1}\otimes\mathbf{C}\boldsymbol{\Psi}^{-1})\mathbf{E}_N d\boldsymbol{\psi}+(\mathbf{I}_N\otimes\mathbf{C}\boldsymbol{\Psi}^{-1})\mathbf{E}_N d\bar{\boldsymbol{\psi}},$$

and hence

$$\begin{aligned}\frac{\partial\bar{\mathbf{c}}}{\partial\mathbf{c}'}&=(\bar{\boldsymbol{\Psi}}\boldsymbol{\Psi}^{-1}\otimes\mathbf{I}_N),\\ \frac{\partial\bar{\mathbf{c}}}{\partial\boldsymbol{\psi}'}&=-(\bar{\boldsymbol{\Psi}}\boldsymbol{\Psi}^{-1}\otimes\mathbf{C}\boldsymbol{\Psi}^{-1})\mathbf{E}_N,\\ \frac{\partial\bar{\mathbf{c}}}{\partial\bar{\boldsymbol{\psi}}'}&=(\mathbf{I}_N\otimes\mathbf{C}\boldsymbol{\Psi}^{-1})\mathbf{E}_N.\end{aligned}$$

But since $\boldsymbol{\psi}=vecd(\mathbf{C})$,

$$\frac{d\bar{\mathbf{c}}}{d\mathbf{c}'}=(\bar{\boldsymbol{\Psi}}\boldsymbol{\Psi}^{-1}\otimes\mathbf{I}_N)-(\bar{\boldsymbol{\Psi}}\boldsymbol{\Psi}^{-1}\otimes\mathbf{C}\boldsymbol{\Psi}^{-1})\mathbf{E}_N\mathbf{E}'_N.$$

D Semiparametric estimators with cross-sectionally independent shocks

In this appendix, we initially provide a new characterisation of the unrestricted ML parametric estimators under correct specification, which we then exploit to derive SP estimators that impose the cross-sectional independence of shocks, first in general and then under symmetry.

D.1 The parametric efficient score

Let

$$\mathcal{I}(\boldsymbol{\phi}) = E[\mathcal{I}_t(\boldsymbol{\phi})] = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta\varrho}}(\boldsymbol{\phi}) \\ \mathcal{I}'_{\boldsymbol{\theta\varrho}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\varrho\varrho}}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} E[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{M}_{dd}(\boldsymbol{\varrho})\mathbf{Z}'_{dt}(\boldsymbol{\theta})] & \mathbf{Z}_d(\boldsymbol{\phi})\mathcal{M}_{dr}(\boldsymbol{\varrho}) \\ \mathcal{M}'_{dr}(\boldsymbol{\varrho})\mathbf{Z}'_d(\boldsymbol{\phi}) & \mathcal{M}_{rr}(\boldsymbol{\varrho}) \end{bmatrix},$$

with $\mathbf{Z}_d(\boldsymbol{\phi})$ defined in (B18), denote the unconditional information matrix. 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 $\boldsymbol{\varrho}$, $\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}_0)$, namely

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\varrho}t}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) &= \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathcal{I}_{\boldsymbol{\theta\varrho}}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}_{\boldsymbol{\varrho\varrho}}(\boldsymbol{\phi}_0)\mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) \\ &= \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\mathcal{M}_{dr}(\boldsymbol{\varrho})\mathcal{M}^{-1}_{rr}(\boldsymbol{\varrho})\mathbf{e}_{rt}(\boldsymbol{\phi}), \end{aligned} \tag{D40}$$

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.

We can interpret the second summand of (D40) as $\mathbf{Z}_d(\boldsymbol{\phi})$ 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). In this respect, it is important to note that the different $\mathbf{e}_{r_i t}(\boldsymbol{\phi})$ in (B10) are orthogonal to each other, so the projection of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ on the linear span of $\mathbf{e}_{rt}(\boldsymbol{\phi}_0)$ coincides with the sum of the projections of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ onto the linear spans of $\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)/\partial \boldsymbol{\varrho}_i$ $(i = 1, \ldots, N)$. In fact, the special structure of $\mathcal{M}_{lr}$, $\mathcal{M}_{sr}$ and $\mathcal{M}_{rr}$ discussed in Appendices B.3.2 and B.5.2 implies that we simply need to project each

$$\mathbf{e}_{d_i t}(\boldsymbol{\phi}_0) = -\begin{bmatrix} \partial \ln f_i(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i0})/\partial \varepsilon^*_i \\ [1 + \varepsilon^*_{it}\partial \ln f_i(\varepsilon^*_{it}; \boldsymbol{\varrho}_{i0})/\partial \varepsilon^*_i] \end{bmatrix}$$

onto the linear span of $\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)/\partial \boldsymbol{\varrho}_i$. Thus, we will end up with the following projections

$$\begin{aligned} -\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i} \quad &: \quad \mathrm{M}_{lr}(\boldsymbol{\varrho}_i)\mathrm{M}^{-1}_{rr}(\boldsymbol{\varrho}_i)\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}, \\ -(1 + \frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}) \quad &: \quad \mathrm{M}_{sr}(\boldsymbol{\varrho}_i)\mathrm{M}^{-1}_{rr}(\boldsymbol{\varrho}_i)\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}, \end{aligned}$$

where

$$\begin{aligned} \mathrm{M}_{lr}(\boldsymbol{\varrho}_i) &= cov\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}, \frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right| \boldsymbol{\varrho}_i\bigg], \\ \mathrm{M}_{sr}(\boldsymbol{\varrho}_i) &= cov\left[\frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}, \frac{\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)}{\partial \boldsymbol{\varrho}_i}\right| \boldsymbol{\varrho}_i\bigg], \end{aligned}$$

and

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

In contrast, the projection of the off-diagonal terms $\varepsilon^*_{jt}\partial \ln f(\varepsilon^*_{it}; \boldsymbol{\varrho}_i)/\partial \varepsilon^*_i$ onto the linear span of $\partial \ln f(\varepsilon^*_{kt}; \boldsymbol{\varrho}_k)/\partial \boldsymbol{\varrho}_k$ is 0 for all possible combinations of i, j and k with $i \neq j$.

We can compactly re-write these expressions in matrix form by working with the rearranged version of $\mathbf{e}_{st}(\boldsymbol{\phi})$. Specifically, we can use (B30) to write he parametric efficient score as

$$
\begin{aligned}
\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\varrho}t}(\boldsymbol{\phi}) &= \mathbf{V}_{dt}(\boldsymbol{\phi}) \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} - \mathbf{V}_d(\boldsymbol{\theta}) \begin{pmatrix} \mathcal{M}_{lr} \\ \mathrm{M}_{sr} \end{pmatrix} \mathcal{M}_{rr}^{-1}\mathbf{e}_{rt}(\boldsymbol{\phi}) \\
&= \begin{bmatrix} \mathbf{C}^{-1\prime}[\mathbf{e}_{lt}(\boldsymbol{\phi}) - \mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathbf{e}_{rt}(\boldsymbol{\phi})] \\ (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathbf{e}_{lt}(\boldsymbol{\phi}) - (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathbf{e}_{rt}(\boldsymbol{\phi}) \\ \vdots \\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathbf{e}_{lt}(\boldsymbol{\phi}) - (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathbf{e}_{rt}(\boldsymbol{\phi}) \\ (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N[\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) - \mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathbf{e}_{rt}(\boldsymbol{\phi})] + (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\boldsymbol{\Delta}_N\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix},
\end{aligned}
$$

where $\mathcal{M}_{lr}$, M_{sr} and $\mathcal{M}_{rr}$ are block-diagonal matrices with typical blocks $\mathrm{M}_{lr}(\boldsymbol{\varrho}_i)$, $\mathrm{M}_{sr}(\boldsymbol{\varrho}_i)$ and $\mathrm{M}_{rr}(\boldsymbol{\varrho}_i)$, respectively,

$$
\mathbf{V}_{dt}(\boldsymbol{\phi}) = \mathbf{Z}_{dt}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_N & \boldsymbol{\Delta}_N \end{pmatrix} = \begin{bmatrix} \mathbf{C}^{-1\prime} & \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N(N-1)} \\ (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N} & \mathbf{0}_{N^2\times N(N-1)} \\ \vdots & \vdots & \vdots \\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N} & \mathbf{0}_{N^2\times N(N-1)} \\ \mathbf{0}_{N^2\times N} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\boldsymbol{\Delta}_N \end{bmatrix}
$$

and

$$
\mathbf{V}_d(\boldsymbol{\theta}) = \mathbf{Z}_d(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_N \end{pmatrix} = \begin{bmatrix} \mathbf{C}^{-1\prime} & \mathbf{0}_{N\times N} \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N} \\ \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N} \\ \mathbf{0}_{N^2\times N} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N \end{bmatrix}.
$$

The covariance matrix of this score then becomes

$$
\begin{aligned}
V\{\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\varrho}t}(\boldsymbol{\phi})\} &= E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi})\mathbf{e}_{dt}'(\boldsymbol{\phi})\mathbf{Z}_{dt}(\boldsymbol{\theta})|\boldsymbol{\phi}\right] \\
&-E\left\{\mathbf{V}_{dt}(\boldsymbol{\theta}) \begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi}) \end{bmatrix} \mathbf{e}_{rt}'(\boldsymbol{\phi}) \begin{pmatrix} \mathcal{M}_{lr}' & \mathrm{M}_{sr}' \end{pmatrix} \mathbf{V}_d'(\boldsymbol{\phi}) \middle| \boldsymbol{\phi}\right\} \\
&-E\left\{\mathbf{V}_d(\boldsymbol{\phi}) \begin{pmatrix} \mathcal{M}_{lr} \\ \mathrm{M}_{sr} \end{pmatrix} \mathcal{M}_{rr}^{-1}\mathbf{e}_{rt}(\boldsymbol{\phi}) \begin{bmatrix} \mathbf{e}_{lt}'(\boldsymbol{\phi}) & \mathbf{e}_{st}'(\boldsymbol{\phi})\mathbf{E}_N & \mathbf{e}_{st}'(\boldsymbol{\phi})\boldsymbol{\Delta}_N \end{bmatrix} \mathbf{V}_{dt}'(\boldsymbol{\theta}) \middle| \boldsymbol{\phi}\right\} \\
&+E\left\{\mathbf{V}_d(\boldsymbol{\phi}) \begin{pmatrix} \mathcal{M}_{lr} \\ \mathrm{M}_{sr} \end{pmatrix} \mathcal{M}_{rr}^{-1}\mathbf{e}_{rt}(\boldsymbol{\phi})\mathbf{e}_{rt}'(\boldsymbol{\phi})\mathcal{M}_{rr}^{-1} \begin{pmatrix} \mathcal{M}_{lr}' & \mathrm{M}_{sr}' \end{pmatrix} \mathbf{V}_d'(\boldsymbol{\phi}) \middle| \boldsymbol{\phi}\right\} \\
&= \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \mathbf{V}_d(\boldsymbol{\phi}) \left[\begin{pmatrix} \mathcal{M}_{lr} \\ \mathrm{M}_{sr} \end{pmatrix} \mathcal{M}_{rr}^{-1} \begin{pmatrix} \mathcal{M}_{lr}' & \mathrm{M}_{sr}' \end{pmatrix}\right] \mathbf{V}_d'(\boldsymbol{\phi}) = \mathcal{P}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}). \qquad \text{(D41)}
\end{aligned}
$$

If we then re-write this expression as

$$
\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathcal{P}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})=\left[\begin{array}{cc}
\mathbf{C}^{-1\prime} & \mathbf{0}_{N\times N}\\
(\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N}\\
\vdots & \vdots\\
(\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N}\\
\mathbf{0}_{N^2\times N} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N
\end{array}\right]\left(\begin{array}{cc}
\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}' & \mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'\\
\mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}' & \mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'
\end{array}\right)
$$

$$
\times\left[\begin{array}{cccc}
\mathbf{C}^{-1} & \mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & \dots & \mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & \mathbf{0}\\
\mathbf{0}_{N\times N} & \mathbf{0}_{N\times N^2} & \dots & \mathbf{0}_{N\times N^2} & \mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{C}^{-1})
\end{array}\right]
$$

$$
=\left[\begin{array}{cccc}
\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1} & \mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & \dots & \mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & \mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{C}^{-1})\\
(\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1} & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{C}^{-1})\\
\vdots & & \ddots & & \vdots\\
(\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1} & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{C}^{-1})\\
(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N\mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N\mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & \dots & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N\mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}'\mathbf{C}^{-1}(\boldsymbol{\mu}'\otimes\mathbf{I}_N) & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N\mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'\mathbf{E}_N'(\mathbf{I}_N\otimes\mathbf{C}^{-1})
\end{array}\right],
$$

we can easily see that there will be some functions of the original parameters $\boldsymbol{\theta}$ which can be estimated equally efficiently by the restricted and unrestricted parametric estimators because this matrix has less than full rank. Specifically, given that $\mathbf{V}_d(\boldsymbol{\theta})$ has column rank $2N$, and that

$$
\left(\begin{array}{cc}
\mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}' & \mathcal{M}_{lr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'\\
\mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathcal{M}_{lr}' & \mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}'
\end{array}\right),
$$

whose four blocks are diagonal matrices of order N, has rank q at most, the rank of the difference between $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ and $\mathcal{P}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ will be $\min(2N,q)$. Therefore, a relevant question is which transformations of parameters this rank deficiency suggests. Proposition 5 implies that the unrestricted and restricted MLEs of $\mathbf{a}$ and $\mathbf{j}$ will be equally efficient both when $q\geq 2N$ and when $q<2N$. In this last case, which arises when some of the shocks follow for instance skew-normal distributions, whose standardised densities depend on a single parameter, or when they follow a distribution like the Laplace that contains no shape parameters, there will be other functions of $\boldsymbol{\theta}$ for which unrestricted and restricted MLEs are equally efficient, the extreme example being $q=0$.

The analysis of the rank of the difference between $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ and $\mathcal{P}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ changes when the distributions of the shocks are symmetric, such as a Student t. The reason is because $\mathcal{M}_{lr}=\mathbf{0}$

in that case, which means that

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \mathcal{P}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$$

$$= \begin{bmatrix} \mathbf{0}_{N\times N} \\ \mathbf{0}_{N^2\times N} \\ \vdots \\ \mathbf{0}_{N^2\times N} \\ (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime})\mathbf{E}_N \end{bmatrix} \mathrm{M}_{sr}\mathcal{M}_{rr}^{-1}\mathrm{M}_{sr}' \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N^2} & \dots & \mathbf{0}_{N\times N^2} & \mathbf{E}_N'(\mathbf{I}_N \otimes \mathbf{C}^{-1}) \end{bmatrix},$$

whose only non-zero block, which appears in the last position, is

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

a matrix of rank $\min(N, q)$. Once again, this rank deficiency implies there will be some additional combinations of the original parameters $\boldsymbol{\theta}$ which can be estimated equally efficiently by the restricted and unrestricted parametric estimators. Proposition 6 shows that the additional parameters will be $\boldsymbol{\tau}$ both when $q \geq N$ and when $q < N$, as explained in Proposition 14 in Fiorentini and Sentana (2021b).

Finally, a closer inspection of the information matrix $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ indicates that in this symmetric case the unrestricted estimators of the mean parameters $\boldsymbol{\tau}$ and $\mathbf{a}$ are asymptotically orthogonal to the corresponding estimators for $\mathbf{c}$ and $\boldsymbol{\varrho}$ precisely because $\mathcal{M}_{lr} = \mathbf{0}$.

D.2 The cross-sectionally independent SP efficient score

The interpretation of the difference between $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $\mathbf{s}_{\boldsymbol{\theta}|\boldsymbol{\varrho}t}(\boldsymbol{\theta}, \boldsymbol{\varrho})$ in terms of projections that we derived above allows us to replace the parametric assumption on the shape of each of the distributions of the standardised innovations in $\boldsymbol{\varepsilon}_t^*$ by a non-parametric alternative. Specifically, we can replace the linear span of $\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_i)/\partial \boldsymbol{\varrho}_i$ by the so-called marginal unrestricted tangent set, which is the Hilbert space generated by all time-invariant functions of ε_{it}^* with bounded second moments that have zero conditional means and are conditionally orthogonal to $\mathbf{e}_{d_it}'(\boldsymbol{\theta}_0, \mathbf{0}) = (\varepsilon_{it}^*, \varepsilon_{it}^{*2} - 1)$. Moreover, since the different shocks are stochastically independent of each other, we can continue to compute the multivariate projection as the sum of the univariate projections, as explained by Chen and Bickel (2006). The main difference with their approach is that they normalised the scale of the shocks so that $med[|\varepsilon_{it}^*|] = 1$ while we use $E(\varepsilon_{it}^{*2}) = 1$, so their tangent set differs from ours in that respect (see Appendix D.2.1 below for further details).

In practice, we simply need to theoretically regress $\mathbf{e}_{d_it}(\boldsymbol{\phi})$ onto $\mathbf{e}_{d_it}(\boldsymbol{\theta}, \mathbf{0})$ and retain the projection error because the projection of $\mathbf{e}_{d_it}(\boldsymbol{\phi}_0)$ onto $\mathbf{e}_{d_kt}(\boldsymbol{\theta}_0, \mathbf{0})$ for $k \neq i$ will be 0 under independence. In this respect, please note that we do need to regress the off-diagonal elements of $\mathbf{e}_{dt}(\boldsymbol{\phi}_0)$ onto any $\mathbf{e}_{d_kt}(\boldsymbol{\theta}_0, \mathbf{0})$ because the projection of $\varepsilon_{jt}^* \partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_i)/\partial \varepsilon_i^*$ onto the linear span of $\mathbf{e}_{d_kt}(\boldsymbol{\theta}_0, \mathbf{0})$ is also 0 for all possible combinations of i, j and k with $i \neq j$ thanks to independence.

We also need

$$V[\mathbf{e}_{d_it}(\boldsymbol{\theta}_0,\mathbf{0})] = V\left(\begin{array}{c}\varepsilon^*_{it}\\ \varepsilon^{*2}_{it}-1\end{array}\right) = \left(\begin{array}{cc}1 & \phi_i(\boldsymbol{\varrho}_i)\\ \phi_i(\boldsymbol{\varrho}_i) & \kappa_i(\boldsymbol{\varrho}_i)-1\end{array}\right) = \mathcal{K}_{ii}(\boldsymbol{\varrho}_i), \tag{D42}$$

where $\varphi(\boldsymbol{\varrho}_i) = E(\varepsilon^{*3}_{it}|\boldsymbol{\varrho}_i)$ and $\kappa_{ii}(\boldsymbol{\varrho}_i) = E(\varepsilon^{*4}_{it}|\boldsymbol{\varrho}_i)$ are the usual coefficients of skewness and kurtosis of ε^*_{it}. In the Student t case, $\varphi(\boldsymbol{\varrho}_i) = 0$ and $\kappa_{ii}(\boldsymbol{\varrho}_i) = 3(\nu_i-2)/(\nu_i-4)$, while under normality $\varphi(\mathbf{0}) = \kappa_{ii}(\mathbf{0}) = 0$.

It is then easy to see that

$$\mathcal{K}^{-1}_{ii}(\boldsymbol{\varrho}) = \frac{1}{\kappa_i(\boldsymbol{\varrho}_i)-1-\phi^2_i(\boldsymbol{\varrho}_i)}\left(\begin{array}{cc}\kappa_i(\boldsymbol{\varrho}_i)-1 & -\phi_i(\boldsymbol{\varrho}_i)\\ -\phi_i(\boldsymbol{\varrho}_i) & 1\end{array}\right)$$

is generally well defined because the Cauchy-Schwarz inequality implies that $\kappa_i(\boldsymbol{\varrho}_i) \geq 1+\phi^2_i(\boldsymbol{\varrho}_i)$ for all distributions, with equality when ε^*_{it} is a centred and standardised Bernoulli. The generalised hyperbolic family of distributions can also reach the skewness-kurtosis bound, but it does so in limiting cases in which its standardised members approach a centred and standardised Bernoulli.

We can also use the generalised information matrix equality to show that

$$\begin{aligned}
-E\left[\varepsilon^*_{it}\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_i)/\partial\varepsilon^*_i\right] &= E(\partial\varepsilon^*_{it}/\partial\varepsilon^*_i) = 1\\
-E\left[(\varepsilon^{*2}_{it}-1)\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_i)/\partial\varepsilon^*_i\right] &= E[\partial(\varepsilon^{*2}_{it}-1)/\partial\varepsilon^*_i] = 0,\\
-E\{\varepsilon^*_{it}[1+\varepsilon^*_{it}\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_i)/\partial\varepsilon^*_i]\} &= E(\partial\varepsilon^{*2}_{it}/\partial\varepsilon^*_i) = 0,\\
-E\{(\varepsilon^{*2}_{it}-1)[1+\varepsilon^*_{it}\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_i)/\partial\varepsilon^*_i]\} &= -1-E\left[\varepsilon^{*3}_{it}\partial\ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_i)/\partial\varepsilon^*_i\right]\\
&= -1+E(\partial\varepsilon^{*3}_{it}/\partial\varepsilon^*_i) = 2.
\end{aligned}$$

so that

$$E[\mathbf{e}_{d_it}(\boldsymbol{\phi}_0)\mathbf{e}'_{d_it}(\boldsymbol{\theta}_0,\mathbf{0})] = \mathcal{K}_{ii}(\mathbf{0}).$$

On this basis, we can easily find the required projections

$$\begin{aligned}
-\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial\varepsilon^*_i} \quad &: \quad \frac{[\kappa_i(\boldsymbol{\varrho}_i)-1]\varepsilon^*_{it}-\phi_i(\boldsymbol{\varrho}_i)(\varepsilon^{*2}_{it}-1)}{\kappa_i(\boldsymbol{\varrho}_i)-1-\phi^2_i(\boldsymbol{\varrho}_i)},\\
-(1+\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial\varepsilon^*_i}\varepsilon^*_{it}) \quad &: \quad \frac{-2\phi_i(\boldsymbol{\varrho}_i)\varepsilon^*_{it}+2(\varepsilon^{*2}_{it}-1)}{\kappa_i(\boldsymbol{\varrho}_i)-1-\phi^2_i(\boldsymbol{\varrho}_i)}.
\end{aligned}$$

To express these in matrix notation it is convenient to remember that under cross-sectional independence of the shocks

$$\mathcal{K}(\boldsymbol{\varrho}) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\, I_{t-1};\boldsymbol{\phi}]$$

$$= \left[\begin{array}{cc}\mathbf{I}_N & E[\boldsymbol{\varepsilon}^*_t vec'(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t-\mathbf{I}_N)|\,\boldsymbol{\phi}]\\ E[vec(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t-\mathbf{I}_N)\boldsymbol{\varepsilon}^{*\prime}_t|\,\boldsymbol{\phi}] & E[vec(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t-\mathbf{I}_N)vec'(\boldsymbol{\varepsilon}^*_t\boldsymbol{\varepsilon}^{*\prime}_t-\mathbf{I}_N)|\,\boldsymbol{\phi}]\end{array}\right] = \left[\begin{array}{cc}\mathbf{I}_N & \mathrm{K}_{ls}\mathbf{E}'_N\\ \mathbf{E}_N\mathrm{K}_{ls} & \mathbf{K}_{NN}+\boldsymbol{\Lambda}\end{array}\right],$$

where K_{ls} is the diagonal matrix of order N with typical element $\varphi(\boldsymbol{\varrho}_i)$ defined in the proof of Proposition 5 and $\boldsymbol{\Lambda}$ is a block diagonal matrix of order N^2 in which each of the N blocks is diagonal matrix of size N with the following structure:

$$\boldsymbol{\Lambda}_{ii} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa_{ii}(\boldsymbol{\varrho}_i) - 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In addition, it is also well known that he covariance between $\mathbf{e}_{dt}(\boldsymbol{\phi})$ and $\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})$ is given by

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

As usual, the off-diagonal elements of $\mathbf{K}_{NN}$ in these two covariance matrices simply reflect the fact that

$$-E\left[\frac{\partial \ln f_i(\varepsilon_{it}^*; \boldsymbol{\varrho}_2)}{\partial \varepsilon_i^*} \varepsilon_{jt}^* \varepsilon_{kt}^* \varepsilon_{lt}^*\right] = E(\varepsilon_{it}^* \varepsilon_{jt}^* \varepsilon_{kt}^* \varepsilon_{lt}^*) = 1$$

when $i = k$ and $j = l$ despite the fact that $i \neq j$.

It is then easy to see that the covariance matrix of the rearranged vector of Gaussian scores will be given by

$$\begin{aligned} V\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\theta}, \mathbf{0}) \\ \mathbf{E}_N' \mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0}) \\ \boldsymbol{\Delta}_N' \mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0}) \end{bmatrix} &= \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_N' \\ \mathbf{0} & \boldsymbol{\Delta}_N' \end{pmatrix} \begin{bmatrix} \mathbf{I}_N & \mathrm{K}_{ls} \mathbf{E}_N' \\ \mathbf{E}_N \mathrm{K}_{ls} & \mathbf{K}_{NN} + \boldsymbol{\Lambda} \end{bmatrix} \begin{pmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_N & \boldsymbol{\Delta}_N \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_N & \mathrm{K}_{ls} & \mathbf{0} \\ \mathrm{K}_{ls} & \mathrm{K}_{ss} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}_N' \mathbf{K}_{NN} \boldsymbol{\Delta}_N + \mathbf{I}_{N(N-1)} \end{pmatrix}, \end{aligned} \quad \text{(D43)}$$

where $\mathrm{K}_{ss} = \mathbf{I}_N + \mathbf{E}_N' \boldsymbol{\Lambda} \mathbf{E}_N$ is the diagonal matrix with typical element $\kappa_{ii}(\boldsymbol{\varrho}_i) - 1$ defined in the proof of Proposition 5. Notice that this matrix contains the same information as the $\mathcal{K}_{ii}(\boldsymbol{\varrho}_i)'s$, but in arranged in different order. Notice also that even though $\boldsymbol{\Delta}_N' \mathbf{K}_{NN} \boldsymbol{\Delta}_N$ is a symmetric orthogonal matrix (see Proposition 5 in Magnus and Sentana (2020)), $\boldsymbol{\Delta}_N' \mathbf{K}_{NN} \boldsymbol{\Delta}_N + \mathbf{I}_{N(N-1)} = \boldsymbol{\Delta}_N' (\mathbf{K}_{NN} + \mathbf{I}_{N^2}) \boldsymbol{\Delta}_N$ is a singular matrix because $(\mathbf{K}_{NN} + \mathbf{I}_{N^2})$ is singular of rank $N(N+1)/2$ (see Theorem 3.11 in Magnus and Neudecker (2019)). Intuitively, $\boldsymbol{\Delta}_N' \mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0})$ only contains $N(N-1)/2$ non-duplicated elements, so its covariance matrix must necessarily be singular.

Similarly,

$$E\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})\\ \boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]\left[\begin{array}{ccc}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0}) & \mathbf{e}'_{st}(\boldsymbol{\theta},\mathbf{0})\mathbf{E}_N & \mathbf{e}'_{st}(\boldsymbol{\theta},\mathbf{0})\boldsymbol{\Delta}_N\end{array}\right]$$
$$=\left(\begin{array}{cc}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{E}'_N\\ \mathbf{0} & \boldsymbol{\Delta}'_N\end{array}\right)\left(\begin{array}{cc}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{K}_{NN}+\mathbf{I}_{N^2}\end{array}\right)\left(\begin{array}{ccc}\mathbf{I}_N & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_N & \boldsymbol{\Delta}_N\end{array}\right)$$
$$=\left(\begin{array}{ccc}\mathbf{I}_N & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & 2\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}'_N\mathbf{K}_{NN}\boldsymbol{\Delta}_N+\mathbf{I}_{N^2}\end{array}\right). \tag{D44}$$

Thus, the projections of $\mathbf{e}_{lt}(\boldsymbol{\phi})$ and $\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})$ onto the linear span of $\mathbf{e}_{lt}(\boldsymbol{\phi})$ and $\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})$ can be written in matrix notation as

$$\left(\begin{array}{cc}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & 2\mathbf{I}_N\end{array}\right)\left(\begin{array}{cc}\mathbf{I}_N & \mathrm{K}_{ls}\\ \mathrm{K}_{ls} & \mathrm{K}_{ss}\end{array}\right)^{-1}\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right]$$
$$=\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right]$$

because

$$\left(\begin{array}{cc}\mathbf{I}_N & \mathrm{K}_{ls}\\ \mathrm{K}_{ls} & \mathrm{K}_{ss}\end{array}\right)^{-1}=\left[\begin{array}{cc}\mathbf{I}_N+\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathrm{K}_{ls} & -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & (\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]$$

by virtue of the partitioned inverse formula and $\mathbf{I}_N+\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathrm{K}_{ls}=\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}$ in view of the diagonality of all the matrices involved.

Consequently, the cross-sectionally independent SP efficient score will be

$$\ddot{\ddot{\mathbf{s}}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})=\mathbf{V}_{dt}(\boldsymbol{\phi})\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})\\ \boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]$$
$$-\mathbf{V}_d(\boldsymbol{\theta})\left\{\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]-\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\left(\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right)\right\}$$
$$=\left\{\begin{array}{c}\mathbf{C}^{-1\prime}[\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})-\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})]\\ [(\mathbf{y}_{t-1}-\boldsymbol{\mu})\otimes\mathbf{I}_N]\mathbf{C}^{-1\prime}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ +(\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}[\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})-\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})]\\ \vdots\\ [(\mathbf{y}_{t-p}-\boldsymbol{\mu})\otimes\mathbf{I}_N]\mathbf{C}^{-1\prime}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ +(\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}[\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})-\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})]\\ (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N[-2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ +2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})]+(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\boldsymbol{\Delta}_N\boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right\}. \tag{D45}$$

To see why, we can make use of the fact that

$$E\left[\left\{\left[\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]-\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\left(\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right)\right]\right\}\right.$$
$$\left.\times\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right]\right|I_{t-1};\boldsymbol{\phi}\right]=\mathbf{0}$$

for any distribution because the projection errors are orthogonal to the regressors by construction.

In addition, we also know that

$$E\left\{\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]-\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\left(\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right)\right| I_{t-1};\boldsymbol{\phi}\right\}=\mathbf{0}$$

because both $\mathbf{e}_{dt}(\boldsymbol{\phi})$ and $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$ are martingale difference sequences. Hence, the second summand of (D45), which can be interpreted as $\mathbf{V}_d(\boldsymbol{\phi}_0)$ times the residual from the theoretical regression of $\mathbf{e}_{lt}(\boldsymbol{\phi})$ and $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})$ (and $\mathbf{0}$) on the linear span generated by a constant, $\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})$ and $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$, belongs to the cross-sectionally independent tangent set, which is the orthogonal sum for $i=1,\ldots,N$ of the marginal unrestricted tangent sets defined above. Importantly, note that $\mathbf{e}_{rt}(\boldsymbol{\phi})$ trivially belongs to this cross-sectionally independent tangent set because it is conditionally orthogonal to all the elements of $\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})$ from Proposition 3 in Fiorentini and Sentana (2007).

The expression for the cross-sectionally independent SP efficiency bound will then become

$$\begin{array}{c}
V\{\ddddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})\}=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]\\
-E\left[\begin{array}{c}\mathbf{V}_{dt}(\boldsymbol{\theta})\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]\\
\times\left\{\begin{array}{c}\left[\begin{array}{cc}\mathbf{e}_{lt}'(\boldsymbol{\phi}) & \mathbf{e}_{st}'(\boldsymbol{\phi})\mathbf{E}_N\end{array}\right]-\left[\begin{array}{cc}\mathbf{e}_{lt}'(\boldsymbol{\theta},\mathbf{0}) & \mathbf{e}_{st}'(\boldsymbol{\theta},\mathbf{0})\mathbf{E}_N\end{array}\right]\\
\times\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\end{array}\right\}\mathbf{V}_d'(\boldsymbol{\phi})\end{array}\right|\boldsymbol{\phi}\right]\\
-E\left[\mathbf{V}_d(\boldsymbol{\phi})\left\{\begin{array}{c}\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]\\
-\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\left(\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right)\\
\times\left[\begin{array}{ccc}\mathbf{e}_{lt}'(\boldsymbol{\phi}) & \mathbf{e}_{st}'(\boldsymbol{\phi})\mathbf{E}_N & \mathbf{e}_{st}'(\boldsymbol{\phi})\boldsymbol{\Delta}_N\end{array}\right]\mathbf{V}_{dt}'(\boldsymbol{\theta})\end{array}\right\}\right|\boldsymbol{\phi}\right]\\
+E\left[\begin{array}{c}\mathbf{V}_d(\boldsymbol{\phi})\left\{\begin{array}{c}\left[\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\end{array}\right]\\
-\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\left(\begin{array}{c}\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\end{array}\right)\end{array}\right\}\\
\times\left\{\begin{array}{c}\left[\begin{array}{cc}\mathbf{e}_{lt}'(\boldsymbol{\phi}) & \mathbf{e}_{st}'(\boldsymbol{\phi})\mathbf{E}_N\end{array}\right]-\left[\begin{array}{cc}\mathbf{e}_{lt}'(\boldsymbol{\theta},\mathbf{0}) & \mathbf{e}_{st}'(\boldsymbol{\theta},\mathbf{0})\mathbf{E}_N\end{array}\right]\\
\times\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 2(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\end{array}\right\}\mathbf{V}_d'(\boldsymbol{\phi})\end{array}\right|\boldsymbol{\phi}\right]\\
=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})\\
-\mathbf{V}_d(\boldsymbol{\phi})\left\{\left(\begin{array}{cc}\mathcal{M}_{ll} & \mathrm{M}_{ls}\\ \mathrm{M}_{ls}' & \mathrm{M}_{ss}\end{array}\right)-\left[\begin{array}{cc}\mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 4(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1}\end{array}\right]\right\}\mathbf{V}_d'(\boldsymbol{\phi})\\
=\ddddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}).
\end{array}$$

where we have exploited the fact that $\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})$ is orthogonal both to $\mathbf{e}_{lt}(\boldsymbol{\phi})$ and $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})$ and to $\mathbf{e}_{lt}(\boldsymbol{\theta}, \mathbf{0})$ and $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0})$. Importantly,

$$\begin{pmatrix} \mathcal{M}_{ll} & \mathrm{M}_{ls} \\ \mathrm{M}_{ls}' & \mathrm{M}_{ss} \end{pmatrix} - \begin{bmatrix} \mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} \\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 4(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} \end{bmatrix}$$

is the residual covariance matrix in the projection of $\mathbf{e}_{lt}(\boldsymbol{\phi})$ and $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})$ onto the linear span of $\mathbf{e}_{lt}(\boldsymbol{\theta}, \mathbf{0})$ and $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta}, \mathbf{0})$ because

$$\begin{aligned} & \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathbf{I}_N & \mathrm{K}_{ls} \\ \mathrm{K}_{ls} & \mathrm{K}_{ss} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I}_N \end{pmatrix} \\ = & \begin{bmatrix} \mathrm{K}_{ss}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} \\ -2\mathrm{K}_{ls}(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} & 4(\mathrm{K}_{ss}-\mathrm{K}_{ls}^2)^{-1} \end{bmatrix}. \end{aligned}$$

D.2.1 The Chen and Bickel (2006) approach

Chen and Bickel (2006) consider a model in which the mean is correctly specified to be 0 so that the only parameters of interest are the elements of the so-called unmixing matrix $\mathbf{W} = \mathbf{C}^{-1}$, which we denote by $\mathbf{w} = vec(\mathbf{W})$.

Given that $d\mathbf{C} = -\mathbf{W}^{-1}(d\mathbf{W})\mathbf{W}^{-1}$, we will have that $dvec(\mathbf{C}) = -(\mathbf{W}^{-1\prime} \otimes \mathbf{W}^{-1})dvec(\mathbf{W})$ so that $\partial\mathbf{c}/\partial\mathbf{w}' = -(\mathbf{W}^{-1\prime} \otimes \mathbf{W}^{-1})$. But since $\mathbf{s}_{ct}(\mathbf{c}) = (\mathbf{I}_N \otimes \mathbf{C}^{-1\prime})\mathbf{e}_{st}(\boldsymbol{\phi})$, the chain rule immediately implies that

$$\begin{aligned} \mathbf{s}_{\mathbf{w}t}(\mathbf{w}) &= \frac{\partial\mathbf{c}'}{\partial\mathbf{w}}\mathbf{s}_{ct}(\mathbf{c}) = -(\mathbf{W}^{-1} \otimes \mathbf{W}^{-1\prime})(\mathbf{I}_N \otimes \mathbf{W}')\mathbf{e}_{st}(\boldsymbol{\phi}) \\ &= -(\mathbf{W}^{-1} \otimes \mathbf{I}_N)\mathbf{e}_{st}(\boldsymbol{\phi}) = vec\left[\left\{\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial\boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) + \mathbf{I}_N\right\}\mathbf{W}^{-1}\right]. \end{aligned}$$

They also normalise the shocks so that $med(|\varepsilon_i^*|) = 1$, which is equivalent to $P(|\varepsilon_i^*| \leq 1) = \frac{1}{2}$, which is in turn equivalent to

$$2\int_{-1}^{1} f_i(\varepsilon_i^*)d\varepsilon_i^* = 1,$$

which they finally re-write in terms of the following equivalent moment condition

$$E\left[2I(|\varepsilon_i^*| \leq 1) - 1\right] = E[d(\varepsilon_i^*)] = 0. \tag{D46}$$

Thus, Chen and Bickel (2006) impose three restrictions on the density of ε_i^*: a) it must integrate to 1; b) it must have 0 mean, and c) it must be such that (D46) holds.

They obtain the efficient score for $\mathbf{w}$ by projecting the entries of $\mathbf{s}_{\mathbf{w}t}(\mathbf{w})$ corresponding to the diagonal elements onto the linear span of 1, ε_i^* and $d(\varepsilon_i^*)$ and retaining the projection errors.

Let

$$V\left\{\begin{bmatrix} \varepsilon_i^* \\ d(\varepsilon_i^*) \end{bmatrix}\right\} = \begin{pmatrix} \sigma_i^2 & \upsilon_i \\ \upsilon_i & 1 \end{pmatrix}$$

because $[2I(|\varepsilon_i^*| \leq 1) - 1]^2 = 4I^2(|\varepsilon_i^*| \leq 1) + 1 - 4I(|\varepsilon_i^*| \leq 1) = 1$ regardless of the distribution of ε_i^*.

Similarly, let

$$E\left\{-\left[1+\frac{\partial \ln f_i(\varepsilon_i^*)}{\partial \varepsilon_i^*}\varepsilon_i^*\right]\left[\begin{array}{c}\varepsilon_i^* \\ d(\varepsilon_i^*)\end{array}\right]\right\}=\left(\begin{array}{c}0 \\ \omega_i-1\end{array}\right),$$

with

$$\omega_i=E\left[-\frac{\partial \ln f_i(\varepsilon_i^*)}{\partial \varepsilon_i^*}\varepsilon_i^* 2I(|\varepsilon_i^*|\leq 1)\right],$$

because

$$E\left[(\varepsilon_{it}^{*2}\partial \ln f_i(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)/\partial \varepsilon_i^*\right]=E[\partial \varepsilon_{it}^{*2}/\partial \varepsilon_i^*]=0$$

by the generalised information matrix equality.

Hence, the coefficients of the projection of $-[1+\varepsilon_{it}^{*2}\partial \ln f_i(\varepsilon_{it}^*;\boldsymbol{\varrho}_i)/\partial \varepsilon_i^*]$ onto the linear span of 1, ε_i^* and $d(\varepsilon_i^*)$ will be given by

$$\left(\begin{array}{cc}\sigma_i^2 & \upsilon_i \\ \upsilon_i & 1\end{array}\right)^{-1}\left(\begin{array}{c}0 \\ \omega_i\end{array}\right)=\frac{\omega_i-1}{\sigma_i^2-\upsilon_i^2}\left(\begin{array}{c}-\upsilon_i \\ \sigma_i^2\end{array}\right)$$

so the projection error will be

$$-\left[1+\frac{\partial \ln f_i(\varepsilon_i^*)}{\partial \varepsilon_i^*}\varepsilon_i^*\right]+\frac{(\omega_i-1)\upsilon_i}{\sigma_i^2-\upsilon_i^2}\varepsilon_i^*-\frac{(\omega_i-1)\sigma_i^2}{\sigma_i^2-\upsilon_i^2}d(\varepsilon_i^*).$$

On this basis, the efficient score finally becomes $-(\mathbf{W}^{-1}\otimes \mathbf{I}_N)$ times the vec of a square matrix of order N which has

$$-\left[1+\frac{\partial \ln f_i(\varepsilon_i^*)}{\partial \varepsilon_i^*}\varepsilon_i^*\right]-\left\{-\left[1+\frac{\partial \ln f_i(\varepsilon_i^*)}{\partial \varepsilon_i^*}\varepsilon_i^*\right]+\frac{(\omega_i-1)\upsilon_i}{\sigma_i^2-\upsilon_i^2}\varepsilon_i^*-\frac{(\omega_i-1)\sigma_i^2}{\sigma_i^2-\upsilon_i^2}d(\varepsilon_i^*)\right\}$$
$$=-\frac{(\omega_i-1)\upsilon_i}{\sigma_i^2-\upsilon_i^2}\varepsilon_i^*+\frac{(\omega_i-1)\sigma_i^2}{\sigma_i^2-\upsilon_i^2}d(\varepsilon_i^*)$$

as diagonal elements and

$$-\frac{\partial \ln f_i(\varepsilon_i^*)}{\partial \varepsilon_i^*}\varepsilon_j^*$$

as off-diagonal ones.

D.3 The unrestricted SP efficient score

The procedure described in Appendix D.2 is different from the unrestricted efficient SP procedure described in Appendix D.4 of Fiorentini and Sentana (2021b) even if one makes use of the cross-sectional independence of the shocks in computing the different expressions. The reason is that the unrestricted tangent space 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 conditional orthogonal to $\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})$. This means that they must be orthogonal not only to ε_{it}^* and $(\varepsilon_{it}^{*2}-1)$ for $i=1,\ldots,N$, but also to cross-product terms of the form $\varepsilon_{it}^*\varepsilon_{jt}^*$ with $i\neq j$. The addition of these cross-products has two effects. First, the covariance matrix of $\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})$, $\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$ and $\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$ in (D43) is an augmented version of the covariance matrix of the first two elements, which an additional singular block diagonal term because of the duplicated cross-product terms. Similarly, the covariance matrix of $\mathbf{e}_{lt}(\boldsymbol{\phi})$, $\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})$ and $\boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\phi})$ with $\mathbf{e}_{lt}(\boldsymbol{\theta},\mathbf{0})$, $\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$ and $\boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$ in (D44) is also an augmented version of the covariance matrix between the first two blocks of each vector, which exactly the same additional singular block. As a result, we must take into account the projections of $\boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\phi})$ onto the linear span of the Gaussian scores $\boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$.

Using the same notation as in the previous sections of this appendix, the unrestricted SP efficient score will be given by:

$$\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\left[\mathbf{e}_{dt}(\boldsymbol{\phi}) - \mathcal{K}(0)\,\mathcal{K}^{+}(\boldsymbol{\rho})\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})\right], \tag{D47}$$

while the unrestricted SP efficiency bound is

$$\ddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) - \mathbf{Z}_d(\boldsymbol{\phi})\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}) - \mathcal{K}(0)\,\mathcal{K}^{+}(\boldsymbol{\rho})\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\phi}), \tag{D48}$$

where $+$ denotes Moore-Penrose inverses.

The fact that the residual variance of a multivariate regression cannot increase as we increase the number of regressors immediately implies that $\dddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ is at least as large (in the positive semidefinite matrix sense) as $\ddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0)$, reflecting the fact that the relevant tangent sets become increasing larger.

D.4 The cross-sectionally independent symmetric SP efficient score

Assuming that each shock is symmetrically distributed, we can also consider another SP estimator which exploits not only the cross-sectional independence of the structural shocks, but also their symmetry. This estimator is different from the spherically symmetric SP estimator discussed in Appendix C.5 of Fiorentini and Sentana (2021b), which would in fact be inconsistent in this case since the joint distribution of the shocks is not spherically symmetric unless all shocks are Gaussian even though all their marginal distributions are symmetric.

To derive this score, we need to define the marginal spherically symmetric tangent sets, which are the Hilbert spaces generated all time-invariant functions of ε^{*2}_{it} with bounded second moments that have zero conditional means and are conditionally orthogonal to $(\varepsilon^{*2}_{it}-1)$. Once again, given that the different shocks are stochastically independent of each other, we can continue to compute the multivariate projection as the sum of the univariate projections.

To obtain the cross-sectionally independent symmetric SP efficient score, the first thing we need to note is that we only need to correct $\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\phi})$ as both $\mathbf{e}_{lt}(\boldsymbol{\phi})$ and $\boldsymbol{\Delta}'_N\mathbf{e}_{st}(\boldsymbol{\phi})$ are orthogonal to $\mathbf{E}'_N\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})$ under the assumption that the shocks are not only cross-sectionally independent but also symmetric.

Note also that in this case

$$-\left[1+\frac{\partial \ln f_i(\varepsilon^*_{it};\boldsymbol{\varrho}_i)}{\partial \varepsilon^*_i}\varepsilon^*_{it}\right] = -[1+\delta(\varepsilon^{*2}_{it};\boldsymbol{\varrho}_i)\varepsilon^{*2}_{it}],$$

with $\delta[\varepsilon_{it}^{*2}(\boldsymbol{\theta});\boldsymbol{\varrho}_i]$ defined in (A3), is a function of the shocks through ε_i^{*2} only. As a result, the projections are trivial to find. Given that we have proved before that

$$-E\{(\varepsilon_{it}^{*2}-1)[1+\varepsilon_{it}^{*}\partial\ln f_i(\varepsilon_{it}^{*};\boldsymbol{\varrho}_i)/\partial\varepsilon_i^{*}]\}=2$$

and that $V(\varepsilon_{it}^{*2})=\kappa(\boldsymbol{\varrho}_i)-1$, each projection will be given by $2[\kappa(\boldsymbol{\varrho}_i)-1]^{-1}\varepsilon_{it}^{*2}$.

Once again, we can use matrix notation to write the cross-sectionally independent symmetric SP efficient score in compact form.

Let

$$\mathbf{U}_d(\boldsymbol{\theta})=\mathbf{Z}_d(\boldsymbol{\theta})\begin{pmatrix}\mathbf{0}\\ \mathbf{E}_N\end{pmatrix}=\begin{bmatrix}\mathbf{0}_{N\times N^2}\\ \mathbf{0}_{N^2\times N^2}\\ \vdots\\ \mathbf{0}_{N^2\times N^2}\\ (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\end{bmatrix}.$$

Then

$$\begin{aligned}\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \mathbf{V}_{dt}(\boldsymbol{\phi})\begin{bmatrix}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\\ \boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\end{bmatrix}-\mathbf{U}_d(\boldsymbol{\theta})\left[\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\phi})-2\mathrm{K}_{ss}^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})\right]\\ &= \begin{bmatrix}\mathbf{C}^{-1\prime}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ (\mathbf{y}_{t-1}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ \vdots\\ (\mathbf{y}_{t-p}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime}\mathbf{e}_{lt}(\boldsymbol{\phi})\\ (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\mathbf{E}_N2\mathrm{K}_{ss}^{-1}\mathbf{E}_N'\mathbf{e}_{st}(\boldsymbol{\theta},\mathbf{0})+(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime})\boldsymbol{\Delta}_N\boldsymbol{\Delta}_N'\mathbf{e}_{st}(\boldsymbol{\phi})\end{bmatrix}.\end{aligned}\tag{D49}$$

Analogous derivations to the ones in previous sections of this appendix show that the associated efficiency bound will be

$$V[\ddot{\mathbf{s}}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})]=\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})-\mathbf{U}_d(\boldsymbol{\phi})(\mathrm{M}_{ss}-4\mathrm{K}_{ss}^{-1})\mathbf{U}_d'(\boldsymbol{\phi})=\dddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}).$$

The fact that the residual variance of a multivariate regression cannot increase as we increase the number of regressors immediately implies that $\dddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$ is at least as large (in the positive semidefinite matrix sense) as $\ddddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})$, reflecting the fact that the relevant tangent sets become increasing larger.

Therefore, we will have the following ranking of efficiency bounds

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})\geq\mathcal{P}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})\geq\dddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})\geq\ddddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi})\geq\ddot{\mathcal{S}}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}_0).$$

In fact, one can show that the corresponding estimators have the Matryoshka doll covariance structure discussed by Fiorentini and Sentana (2021b). The only difference is that here we do not consider the Gaussian PMLE because it fails to point identify the parameters of the matrix $\mathbf{J}$.

E Mittnik and Zadrozny (1993) standard errors for IRFs and FEVDs

As is well known, assuming covariance stationarity, we can re-write (1) as

$$\mathbf{y}_t = \boldsymbol{\mu} + (\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L)^{-1} \mathbf{C} \boldsymbol{\varepsilon}_t^* = (\mathbf{I}_N + \mathbf{B}_1 L + \mathbf{B}_2 L^2 + \ldots) \mathbf{C} \boldsymbol{\varepsilon}_t^*,$$

where

$$(\mathbf{I}_N + \mathbf{B}_1 L + \mathbf{B}_2 L^2 + \ldots)(\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L) = \mathbf{I}_N.$$

Hence, the impulse response function (IRF) of the structural shocks in model (1) is given by

$$\mathbf{B}_k \mathbf{C} = \mathbf{B}_k \mathbf{J} \boldsymbol{\Psi}.$$

Mittnik and Zadrozny (1993) explain how to use the delta method to compute asymptotically valid standard errors for these expressions once we know the asymptotic distribution of the estimators of the SVAR parameters.

For simplicity of exposition, let us consider the simplest possible case in which $p = 1$. Then we know that the MA coefficient matrices satisfy the recursion $\mathbf{B}_k = \mathbf{A}_1 \mathbf{B}_{k-1} = \mathbf{A}_1^k$ with initial condition $\mathbf{B}_0 = \mathbf{I}_N$. Following Magnus et al (2021), we can use the product rule for differentials to show that

$$d\mathbf{B}_k = d(\mathbf{A}_1 \mathbf{B}_{k-1}) = (d\mathbf{A}_1)\mathbf{B}_{k-1} + \mathbf{A}_1(d\mathbf{B}_{k-1})$$

and

$$dvec(\mathbf{B}_k) = (\mathbf{B}_{k-1}' \otimes \mathbf{I}_N)\mathbf{a}_1 + (\mathbf{I}_N \otimes \mathbf{A}_1) dvec(\mathbf{B}_{k-1}),$$

so that

$$\frac{\partial vec(\mathbf{B}_k)}{\partial \mathbf{a}_1'} = (\mathbf{B}_{k-1}' \otimes \mathbf{I}_N) + (\mathbf{I}_N \otimes \mathbf{A}_1) \frac{\partial vec(\mathbf{B}_{k-1})}{\partial \mathbf{a}'},$$

with initial condition $\partial vec(\mathbf{B}_1)/\partial \mathbf{a}_1' = \mathbf{I}_{N^2}$, which leads to

$$\frac{\partial vec(\mathbf{B}_k)}{\partial \mathbf{a}_1'} = \sum_{j=0}^{k} [(\mathbf{A}_1')^{k-j} \otimes \mathbf{A}_1^j].$$

For example,

$$\begin{aligned}
\frac{\partial vec(\mathbf{B}_2)}{\partial \mathbf{a}_1'} &= (\mathbf{A}_1' \otimes \mathbf{I}_N) + (\mathbf{I}_N \otimes \mathbf{A}_1), \\
\frac{\partial vec(\mathbf{B}_3)}{\partial \mathbf{a}_1'} &= (\mathbf{A}_1^2 \otimes \mathbf{I}_N)' + (\mathbf{A}_1' \otimes \mathbf{A}_1) + (\mathbf{I}_N \otimes \mathbf{A}_j^2).
\end{aligned}$$

The product rule for differentials also implies that

$$\begin{aligned}
d(\mathbf{B}_k \mathbf{C}) &= (d\mathbf{B}_k)\mathbf{C} + \mathbf{B}_k(d\mathbf{C}), \\
d(\mathbf{B}_k \mathbf{J} \boldsymbol{\Psi}) &= (d\mathbf{B}_k)\mathbf{J}\boldsymbol{\Psi} + \mathbf{B}_k(d\mathbf{J})\boldsymbol{\Psi} + \mathbf{B}_k \mathbf{J}(d\boldsymbol{\Psi}),
\end{aligned}$$

so that

$$\begin{aligned}
\frac{\partial vec(\mathbf{B}_k\mathbf{C})}{\partial \boldsymbol{\phi}'} &= (\mathbf{C}'\otimes \mathbf{I}_N)\frac{\partial vec(\mathbf{B}_k)}{\partial \boldsymbol{\phi}'} + (\mathbf{I}_N\otimes \mathbf{B}_k)\frac{\partial vec(\mathbf{C})}{\partial \boldsymbol{\phi}'}\\
\frac{\partial vec(\mathbf{B}_k\mathbf{J}\boldsymbol{\Psi})}{\partial \boldsymbol{\phi}'} &= (\boldsymbol{\Psi}\mathbf{J}'\otimes \mathbf{I}_N)\frac{\partial vec(\mathbf{B}_k)}{\partial \boldsymbol{\phi}'} + (\boldsymbol{\Psi}\otimes \mathbf{B}_k)\frac{\partial vec(\mathbf{J})}{\partial \boldsymbol{\phi}'} + (\mathbf{I}_N\otimes \mathbf{B}_k\mathbf{J})\frac{\partial vec(\boldsymbol{\Psi})}{\partial \boldsymbol{\phi}'}.
\end{aligned}$$

As expected, the delta method may lead to a singular covariance matrix if we simultaneously consider multiple values of k. Specifically, a singularity will arise when $k > 3$ because $\mathbf{B}_1$ and $\mathbf{B}_2$ already contain as many elements as $\mathbf{A}_1$ and $\mathbf{C}$.

Entirely analogous calculations apply to the forecast error variance decomposition of those shocks, which are given by

$$\mathbf{B}_k\mathbf{C}\mathbf{B}_k' = \mathbf{B}_k\mathbf{J}\boldsymbol{\Psi}^2\mathbf{J}'\mathbf{B}_k'.$$

F Discrete mixture of normals

F.1 General mixtures and their ML estimators

Let $\mathbf{s} = (s_1, \ldots, s_k, \ldots, s_K)$ denote a categorical random variable of dimension K, which is nothing other than a collection of K mutually exclusive Bernoulli random variables with $P(s_k = 1) = \lambda_k$ such that $\sum_{k=1}^K \lambda_k = 1$. If $\mathbf{z}|\mathbf{s}$ is $N(\mathbf{0}, \mathbf{I}_N)$, then

$$\mathbf{x} = \textstyle\sum_{k=1}^K s_k(\boldsymbol{\mu}_k + \boldsymbol{\Sigma}_k^{1/2}\mathbf{z}), \tag{F50}$$

is a K-component mixture of normals, whose first two unconditional moments are

$$\boldsymbol{\pi} = E(\mathbf{x}) = \textstyle\sum_{k=1}^K \lambda_k\boldsymbol{\mu}_k = E[E(\mathbf{x}|\mathbf{s})], \textit{ and} \tag{F51}$$

$$\boldsymbol{\Omega} = V(\mathbf{x}) = \textstyle\sum_{k=1}^K \lambda_k[(\boldsymbol{\mu}_k\boldsymbol{\mu}_k') + \boldsymbol{\Sigma}_k] - (\sum_{k=1}^K \lambda_k\boldsymbol{\mu}_k)(\sum_{k=1}^K \lambda_k\boldsymbol{\mu}_k') = E[V(\mathbf{x}|\mathbf{s})] + V[E(\mathbf{x}|\mathbf{s})]. \tag{F52}$$

The model parameters are $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k, \ldots, \lambda_K)$, subject to the unit simplex restrictions $\lambda_k \geq 0\ \forall k$ and $\sum_{k=1}^K \lambda_k = 1$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_k', \ldots, \boldsymbol{\mu}_K')'$ and $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1', \ldots, \boldsymbol{\sigma}_k', \ldots, \boldsymbol{\sigma}_K')'$, where $\boldsymbol{\sigma}_k = vech(\boldsymbol{\Sigma}_k)$. The representation in (F50) is very general, and may give rise to substantially deviations from multivariate normality through higher order moments. In particular, it nests random vectors consisting of N independent univariate mixtures with K_i components each, in which case $K = \prod_{i=1}^N K_i$, which play an important role in our analysis of Svars with cross-sectional independent structural shocks in section 2. Third- and fourth-order raw moments, defined as $E[vec(\mathbf{x}\mathbf{x}')\mathbf{x}']$ and $E[vec(\mathbf{x}\mathbf{x}')vec(\mathbf{x}\mathbf{x}')']$ respectively, can be readily obtained as convex combinations of the third- and fourth-order raw moments of the K underlying Gaussian components using the law of iterated expectations. Subtracting the corresponding moments for a $N(\boldsymbol{\pi}, \boldsymbol{\Omega})$ random vectors yields the third- and fourth-order cumulants.

If we observe a random sample of size T on $\mathbf{x}$, ML estimation of the model parameters by numerical methods is conceptually straightforward. Nevertheless, the log-likelihood function of a finite normal mixture has a pole for each observation. Specifically, it will go to infinity if we set $\hat{\boldsymbol{\mu}}_1 = \mathbf{x}_t$ and let $|\hat{\boldsymbol{\Sigma}}_1|$ and $\hat{\lambda}_1$ go to 0 and $1/T$, respectively. As a result, the ML estimator must be defined as the consistent root of the first order conditions (see Kiefer (1978)). In practice, one may deal with this issue by starting the numerical algorithm from many different values. In addition, there is a trivial identification issue that arises by exchanging the labels of the components, but this is also easy to fix. Boldea and Magnus (2009) provide analytical expressions for the score and Hessian matrix, and compare several numerical algorithms and asymptotic covariance matrix estimators, while Amengual et al (2021c) exploit the EM principle to compute the score and the expected value of the Hessian in an intuitive and fast way.

However, it is usually convenient to start the recursions from sensibly chosen values. In this respect, the EM algorithm discussed by Dempster et al (1977) allows us to obtain initial values as close to the MLEs as desired. In the unrestricted case, the recursions are as follows:

$$\hat{\boldsymbol{\mu}}_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}}\frac{1}{T}\sum_{t=1}^{T} w_k(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)})\mathbf{x}_t, \tag{F53a}$$

$$\hat{\boldsymbol{\Sigma}}_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}}\frac{1}{T}\sum_{t=1}^{T} w_k(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)})\mathbf{x}_t\mathbf{x}_t' - \hat{\boldsymbol{\mu}}_k^{(n}\hat{\boldsymbol{\mu}}_k^{(n\prime}, \tag{F53b}$$

$$\hat{\lambda}_k^{(n} = \frac{1}{T}\sum\nolimits_{t=1}^{T} w_k(\mathbf{x}_t;\boldsymbol{\varphi}^{(n-1)}) \tag{F53c}$$

where

$$w_k(\mathbf{x}_t;\boldsymbol{\varphi}) = P(s_{kt}=1|\mathbf{x}_t) = \frac{\lambda_k|\boldsymbol{\Sigma}_k|^{-N/2}\phi_N[\boldsymbol{\Sigma}_k^{-1/2}(\mathbf{x}_t-\boldsymbol{\mu}_k)]}{\sum_{j=1}^{K}\lambda_j|\boldsymbol{\Sigma}_j|^{-N/2}\phi_N[\boldsymbol{\Sigma}_j^{-1/2}(\mathbf{x}_t-\boldsymbol{\mu}_j)]} \tag{F54}$$

is the posterior probability that observation t comes from the k^{th} component, and $\phi_N(.)$ the spherical normal density of dimension N. These recursions had been proposed by several authors without appealing to the EM principle. For example, Hassenblad (1966) shows that they coincide with the steepest descent recursions, which confirms that they always lead to improvements in the log-likelihood function (see also Wolfe (1970) and Peters and Walker (1978)).

The EM algorithm might get stuck in at least two situations. First, if one starts up the recursions with $\hat{\boldsymbol{\mu}}_k^{(0} = \boldsymbol{\mu}^{(0}$ and $\hat{\boldsymbol{\Sigma}}_k^{(0} = \boldsymbol{\Sigma}^{(0}$ for all k, then $w_k(\mathbf{x}_t;\boldsymbol{\varphi}^{(0)}) = \lambda_k^{(0}$ and the parameter values do not get updated because priors and posteriors coincide. The second undesirable situation may arise when the mean of one component equals $\mathbf{x}_t$ for some t. In this case, the algorithm may be irresistibly attracted to one of the log-likelihood poles mentioned before.

The following proposition, which generalises the univariate result in Behboodian (1970), is instrumental for our consistency results:

Proposition F1 *The (pseudo) maximum likelihood estimators of the unconditional mean vector (F51) and covariance matrix (F52) of the discrete unrestricted mixture of K multivariate normals in (F50) are given by the sample mean vector and covariance matrix (with denominator T) of $\mathbf{x}_t$, respectively.*

Proof. It is easy to check that the EM recursions (F53a)-(F53c) imply that

$$\begin{aligned}
\hat{\boldsymbol{\pi}}^{(n} &= \textstyle\sum_{k=1}^{K} \hat{\boldsymbol{\mu}}_k^{(n} \hat{\lambda}_k^{(n} = \dfrac{1}{T}\textstyle\sum_{t=1}^{T} \mathbf{x}_t, \\
\hat{\boldsymbol{\Omega}}^{(n} &= \textstyle\sum_{k=1}^{K} (\hat{\boldsymbol{\mu}}_k^{(n} \hat{\boldsymbol{\mu}}_k^{(n\prime} + \hat{\boldsymbol{\Sigma}}_k^{(n)}) \hat{\lambda}_k^{(n} - \hat{\boldsymbol{\pi}}^{(n} \hat{\boldsymbol{\pi}}^{(n\prime} = \dfrac{1}{T}\textstyle\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t - \left(\dfrac{1}{T}\textstyle\sum_{t=1}^{T} \mathbf{x}_t\right)\left(\dfrac{1}{T}\textstyle\sum_{t=1}^{T} \mathbf{x}_t\right)',
\end{aligned}$$

for all T regardless of the values of $\boldsymbol{\varphi}^{(n-1}$. Since the ML estimators constitute the fixed point of the EM recursions, (i.e. $\hat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}^{(\infty)}$), it follows that $\hat{\boldsymbol{\pi}}^{(n}$ and $\hat{\boldsymbol{\Omega}}^{(n}$ coincide with the Gaussian PML estimators. □

As a result, if we reparametrise the model as $\mathbf{x}_t = \boldsymbol{\pi} + \boldsymbol{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*$, where $\boldsymbol{\varepsilon}_t^*$ is a standardised discrete mixture of normals, then we can maximise the log-likelihood function with respect to $\boldsymbol{\lambda}$ and the free elements of this distribution keeping $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\Omega}}$ fixed at their Gaussian pseudo ML values. Interestingly, this somewhat surprising result will continue to be true even in a complete log-likelihood situation in which we would observe not only $\mathbf{x}_t$ but also $\mathbf{s}_t$. Appendix H.2 first explains how to parametrise the distribution of $\boldsymbol{\varepsilon}_t^*$ so as to ensure that $E(\boldsymbol{\varepsilon}_t^*) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t^*) = \mathbf{I}_N$ when $K = 2$ as a function of N mean difference parameters $\boldsymbol{\delta}$, $N(N+1)/2$ relative variance parameters $\mathbf{K}$ and a single probability parameter λ, and then generalises this procedure for any K.

Given that Proposition F1 is a numerical result that holds for any sample size T and does not depend in any way on the true distribution of the data, the discrete mixture of normals PMLEs of $\boldsymbol{\pi}$ and $\boldsymbol{\Omega}$ will continue to be consistent for $E(\mathbf{x})$ and $V(\mathbf{x})$ under distributional misspecification.

F.2 Scale mixtures and their ML estimators

Given that they are rather popular in empirical research, for completeness we also analyse scale mixtures of normals, which as we will see below, inherit the consistency properties of general mixtures under distributional misspecifications that preserve ellipticity.

The random vector $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\sqrt{\varsigma}\mathbf{u}$, where $\mathbf{u}$ is uniform on the unit sphere surface in $\mathbb{R}^N$, is distributed as a discrete scale mixture of normals (DSMN) if

$$\varsigma = \textstyle\sum_{k=1}^{K} s_k \kappa_k^{1/2} \varsigma_i^o, \tag{F55}$$

where $\varsigma^o|\mathbf{s}$ is $\boldsymbol{\chi}_N^2$. This is a special case of (F50) in which $\boldsymbol{\mu}_k = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_k = \kappa_k \boldsymbol{\Sigma}\ \forall k$. Therefore, its unconditional mean is $\boldsymbol{\mu}$ while its unconditional variance will be

$$\begin{aligned}
\boldsymbol{\Omega} &= V(\mathbf{x}) = \varpi\boldsymbol{\Sigma} = E[V(\mathbf{x}|\mathbf{s})], \\
\varpi &= E(\varsigma/N) = \textstyle\sum_{k=1}^{K} \lambda_k \kappa_k.
\end{aligned} \tag{F56}$$

As a result, we can easily standardise $\mathbf{x}$ by assuming that $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}_N$ and defining the relative variance parameters

$$\kappa_k^* = \kappa_k/\varpi,\ k = 1, \ldots, K.$$

DSMNs with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_N$ are a particular case of spherically symmetric random vectors. Therefore, all their odd central moments will be 0, while their fourth-order moments, which exceed those of the multivariate normal, depend on a single parameter known as the multivariate coefficient of excess kurtosis, which is given by $E(\varsigma^2)/N(N+1) - 1$.

DSMNs approach the multivariate normal when $\kappa_k^* \to 1$ for all k, or when any $\lambda_k \to 1$. Near the limit, though, the distributions can be radically different. For instance, given that we can choose $\kappa_2/\kappa_1 \in (0, 1]$ when $K = 2$ without loss of generality, when $\lambda \to 0^+$ there are very few observations with very large variance ("outliers case"), while when $\lambda \to 1^-$ the opposite happens, very few observations with very small variance ("inliers case") (see Amengual and Sentana (2011) for further details).

It is also possible to apply the EM algorithm to DSMNs but the recursions are different. Specifically, they become:

$$\hat{\boldsymbol{\mu}}^{(n} = \frac{\sum_{t=1}^{T} w_k(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)})(\kappa_k^{(n})^{-1}\mathbf{x}_t}{\sum_{t=1}^{T}\sum_{j=1}^{K} w_j(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)})(\kappa_j^{(n})^{-1}}, \tag{F57a}$$

$$\hat{\boldsymbol{\Sigma}}^{(n} = \frac{\sum_{t=1}^{T} w_k(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)})(\kappa_k^{(n})^{-1}(\mathbf{x}_t - \hat{\boldsymbol{\mu}}^{(n)})(\mathbf{x}_t - \hat{\boldsymbol{\mu}}^{(n)})'}{\sum_{t=1}^{T}\sum_{j=1}^{K} w_j(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)})(\kappa_j^{(n})^{-1}}, \tag{F57b}$$

$$\kappa_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}}\frac{1}{TN}\sum_{t=1}^{T} w_k(\mathbf{x}_t; \boldsymbol{\varphi}^{(n-1)})(\mathbf{x}_t - \hat{\boldsymbol{\mu}}^{(n)})'(\hat{\boldsymbol{\Sigma}}^{(n)})^{-1}(\mathbf{x}_t - \hat{\boldsymbol{\mu}}^{(n)}) \tag{F57c}$$

with $\hat{\lambda}_k^{(n}$ and $w_k(\mathbf{x}_t; \boldsymbol{\varphi})$ still given by (F53c) and (F54), respectively. Some overall scale normalisation is obviously required. For example, we could fix one κ_1 to 1, work with the relative variance parameters κ_k^* subject to the restriction $\sum_{k=1}^{K} \lambda_k \kappa_j^* = 1$ or fix $|\boldsymbol{\Omega}| = 1$, as explained in appendix B of Fiorentini and Sentana (2019). In the first case, the recursions (F57a)-(F57c) continue to be valid after excluding the relevant element. Given the invariance properties of ML estimators, we recommend the first normalisation, which can be changed after convergence has been achieved.

But if we keep $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ fixed, then the recursions for the $\lambda's$ and $\kappa's$ simplify considerably. To understand why, it is convenient to work with the log-likelihood function of ς, which is a discrete mixture of K gamma random variables with common shape parameter $N/2$ and scale parameters $2\kappa_k$, so that their means are $N\kappa_k$.

Let

$$h_\varsigma(\varsigma_t; \boldsymbol{\eta}) = \frac{\varsigma_t^{N/2-1}}{2^{N/2}\Gamma(N/2)} \sum_{k=1}^{K} \lambda_k \kappa_k^{-N/2} \exp(-.5\kappa_k^{-1}\varsigma)$$

denote the marginal density of ς, where $\boldsymbol{\eta}$ contains the free elements of $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$. In this context, the EM recursions are given by

$$\kappa_k^{(n} = \frac{1}{\hat{\lambda}_k^{(n}} \frac{1}{NT} \sum_{t=1}^{T} w_k(\boldsymbol{\varsigma}_t; \boldsymbol{\eta}^{(n-1)}) \boldsymbol{\varsigma}_t, \tag{F58a}$$

$$\hat{\lambda}_k^{(n} = \frac{1}{T} \sum\nolimits_{t=1}^{T} w_k(\boldsymbol{\varsigma}_t; \boldsymbol{\eta}^{(n-1)}) \tag{F58b}$$

where

$$w_k(\boldsymbol{\varsigma}_t; \boldsymbol{\varphi}) = P(s_{kt} = 1|\varsigma_t) = \frac{\lambda_k \kappa_k^{-N/2} \exp(-.5\kappa_k^{-1}\varsigma)}{\sum_{j=1}^{K} \lambda_j \kappa_j^{-N/2} \exp(-.5\kappa_j^{-1}\varsigma)} \tag{F59}$$

is the posterior probability that observation t comes from the k^{th} component. Not surprisingly, (F58a) and (F59) coincide with (F57c) and (F54), respectively, when

$$\varsigma_t(\hat{\boldsymbol{\mu}}^{(n}, \hat{\boldsymbol{\sigma}}^{(n)}) = (\mathbf{x}_t - \hat{\boldsymbol{\mu}}^{(n)})'(\hat{\boldsymbol{\Sigma}}^{(n)})^{-1}(\mathbf{x}_t - \hat{\boldsymbol{\mu}}^{(n)}).$$

The following proposition, which is the counterpart to Proposition F1, is also instrumental for our consistency results in the spherically symmetric case:

Proposition F2 *The (pseudo) maximum likelihood estimators of the unconditional mean (F56) of the discrete unrestricted mixture of K gammas with common shape parameter $N/2$ and scale parameters $2\kappa_k$ in (F55) is given by the sample mean of ς_t.*

Proof. It is easy to check that the EM recursions (F58a)-(F58b) imply that

$$\hat{\varpi}^{(n} = \sum\nolimits_{k=1}^{K} \hat{\kappa}_k^{(n} \hat{\lambda}_k^{(n} = \frac{1}{T} \sum\nolimits_{t=1}^{T} \varsigma_t,$$

for all T regardless of the values of $\boldsymbol{\eta}^{(n-1}$. Since the ML estimators constitute the fixed point of the EM recursions, (i.e. $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}^{(\infty)}$), it follows that $\hat{\varpi}$ coincides with the sample mean of ς_t. □

Given that Proposition F2 is a numerical result that holds for any sample size T and does not depend in any way on the true distribution of the data, the discrete scale mixture of normals PMLE of ϖ will continue to be consistent for $E(\varsigma/N)$ under distributional misspecification for any spherically symmetric distribution. Once again, this somewhat surprising result will continue to be true even in a complete log-likelihood situation in which we would observe not only ς_t but also $\mathbf{s}_t$.

G Multivariate dynamic regression models with time-varying variances and covariances

G.1 Model specification

In a multivariate dynamic regression model with time-varying variances and covariances, the vector of N observed variables, $\mathbf{y}_t$, is typically assumed to be generated as:

$$\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*,$$

where $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}(I_{t-1};\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(I_{t-1};\boldsymbol{\theta})$, $\boldsymbol{\mu}()$ and $vech[\boldsymbol{\Sigma}()]$ are $N \times 1$ and $N(N+1)/2 \times 1$ vector functions describing the conditional mean vector and covariance matrix known up to the $p \times 1$ vector of parameters $\boldsymbol{\theta}$, I_{t-1} denotes the information set available at $t-1$, which contains past values of $\mathbf{y}_t$ and possibly some contemporaneous conditioning variables, and $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is some particular "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2\prime}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. To focus on the effect of distributional misspecification, we maintain the assumption that the conditional mean and variance are correctly specified, in the sense that there is a true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, such that $E(\mathbf{y}_t|I_{t-1}) = \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)$ and $V(\mathbf{y}_t|I_{t-1}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$. We also maintain the high level regularity conditions in Bollerslev and Wooldridge (1992) because we want to leave unspecified the conditional mean vector and covariance matrix in order to achieve full generality. Primitive conditions for specific multivariate models can be found for example in Ling and McAleer (2003).

To complete the model, a researcher needs to specify the conditional distribution of $\boldsymbol{\varepsilon}_t^*$. For the sake of generality, we initially consider a situation in which she makes the assumption that, conditional on I_{t-1}, the distribution of $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed with mean vector equal to 0 and covariance matrix equal to the identity. Nevertheless, we can obtain stronger results below by assuming that that this vector follows some particular member of the spherical family with a well defined density, or $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\theta},\boldsymbol{\eta} \sim i.i.d.\ s(\mathbf{0},\mathbf{I}_N,\boldsymbol{\eta})$ for short, where $\boldsymbol{\eta}$ denotes q additional shape parameters which effectively characterise the distribution of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. As is well known, spherical symmetry reduces to ordinary symmetry in the univariate case ($N = 1$).

In the general case, we follow Fiorentini and Sentana (2019) in assuming that it is possible to rewrite the model in this form:

Reparametrisation 1 *A homeomorphic transformation $\mathbf{r}_g(.) = [\mathbf{r}'_{gc}(.), \mathbf{r}'_{gim}(.), \mathbf{r}'_{gic}(.)]'$ of the mean-variance parameters $\boldsymbol{\theta}$ into an alternative set $\boldsymbol{\phi} = (\boldsymbol{\phi}'_c, \boldsymbol{\phi}'_{im}, \boldsymbol{\phi}'_{ic},)'$, where $\boldsymbol{\phi}_{im}$ is $N \times 1$, $\boldsymbol{\phi}_{ic} = vech(\boldsymbol{\Phi}_{ic})$, $\boldsymbol{\Phi}_{ic}$ is an unrestricted positive definite symmetric matrix of order N and $\mathbf{r}_g(\boldsymbol{\theta})$ is twice continuously differentiable in a neighbourhood of $\boldsymbol{\theta}_0$ with* rank $\left[\partial \mathbf{r}'_g(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\right] = p$, *such that*

$$\left.\begin{array}{c} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\phi}_c) + \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\phi}_c)\boldsymbol{\phi}_{im} \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{\diamond 1/2}(\boldsymbol{\phi}_c)\boldsymbol{\Phi}_{ic}\boldsymbol{\Sigma}_t^{\diamond 1/2\prime}(\boldsymbol{\phi}_c) \end{array}\right\} \quad \forall t. \tag{G60}$$

This parametrisations simply requires the pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\phi}_c) = \boldsymbol{\Sigma}_t^{\diamond -1/2}(\boldsymbol{\phi}_c)[\mathbf{y}_t - \boldsymbol{\mu}_t^{\diamond}(\boldsymbol{\phi}_c)] \tag{G61}$$

to be $i.i.d.$ with mean vector $\boldsymbol{\phi}_{im}$ and covariance matrix $\boldsymbol{\Phi}_{ic}$.

In the spherically case, in contrast, we are able to consider the existence of a less restricted reparametrisation.

Reparametrisation 2 *A homeomorphic transformation $\mathbf{r}_s(.) = [\mathbf{r}'_{sc}(.), r'_{si}(.)]'$ of the mean-variance parameters $\boldsymbol{\theta}$ into an alternative set $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_c, \vartheta'_i)'$, where ϑ_i is a positive scalar, and $\mathbf{r}_s(\boldsymbol{\theta})$ is twice continuously differentiable with* rank $[\partial \mathbf{r}'_s(\boldsymbol{\theta})/\partial\boldsymbol{\theta}] = p$ *in a neighbourhood of $\boldsymbol{\theta}_0$, such that*

$$\left.\begin{array}{c} \boldsymbol{\mu}_t(\boldsymbol{\theta}) = \boldsymbol{\mu}_t(\boldsymbol{\vartheta}_c), \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \vartheta_i\boldsymbol{\Sigma}_t^{\circ}(\boldsymbol{\vartheta}_c) \end{array}\right\} \quad \forall t. \tag{G62}$$

Expression (G62) simply requires that one can construct pseudo-standardised residuals

$$\boldsymbol{\varepsilon}_t^{\circ}(\boldsymbol{\vartheta}_c)=\boldsymbol{\Sigma}_t^{\circ-1/2}(\boldsymbol{\vartheta}_c)[\mathbf{y}_t-\boldsymbol{\mu}_t^{\circ}(\boldsymbol{\vartheta}_c)] \tag{G63}$$

which are $i.i.d.$ $s(\mathbf{0},\vartheta_i\mathbf{I}_N,\boldsymbol{\eta})$, where ϑ_i is a global scale parameter, a condition satisfied by most static and dynamic models.

G.2 Consistency of discrete mixtures of normals ML estimators

Proposition 1 in Fiorentini and Sentana (2019) states that if (G62) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varpi}_0$, is $i.i.d.$ $s(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\varpi}$ includes $\boldsymbol{\vartheta}$ and the true shape parameters, but the spherical distribution assumed for estimation purposes does not necessarily nest the true density, then the pseudo true value of the joint ML estimator of $\boldsymbol{\varphi}=(\boldsymbol{\vartheta}_c',\vartheta_i,\boldsymbol{\eta})'$, $\boldsymbol{\varphi}_\infty$, is such that $\boldsymbol{\vartheta}_{c\infty}$ is equal to the true value $\boldsymbol{\vartheta}_{c0}$. In this context, in Fiorentini and Sentana (2007) we proposed to estimate ϑ_i by $\vartheta_{iT}(\hat{\boldsymbol{\vartheta}}_{cT})$, where

$$\vartheta_{iT}(\boldsymbol{\vartheta}_c)=\frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\varsigma_t^{\circ}(\boldsymbol{\vartheta}_c). \tag{G64}$$

The rationale for this estimator comes from the fact that under normality the score for ϑ_i simplifies to:

$$\mathbf{s}_{\vartheta_i t}(\boldsymbol{\vartheta},\mathbf{0})=\frac{1}{2\vartheta_i}\left[\varsigma_t(\boldsymbol{\vartheta})-N\right], \tag{G65}$$

whose expected value when evaluated at $\boldsymbol{\vartheta}_0$ is 0 because the expected value of $\varsigma_t^{\circ}(\boldsymbol{\vartheta}_{c0})=\boldsymbol{\varepsilon}_t^{\circ\prime}(\boldsymbol{\vartheta}_c)\boldsymbol{\varepsilon}_t^{\circ}(\boldsymbol{\vartheta}_c)$ in (G63) is precisely $N\vartheta_{i0}$.

However, it turns out that Proposition F2 above implies that (G64) numerically coincides the MLE of ϑ_i when the assumed spherical distribution is a discrete scale mixture of normals, so it is irrelevant whether we replace it or not. As a result, the ML estimators based on a discrete scale mixture of normals are consistent for all the parameters when the true distribution is spherical. In addition, (G64) also gives us the elliptically symmetric, SP estimator of ϑ_i when $\boldsymbol{\vartheta}_c$ is replaced by its iterated, elliptically symmetric, SP estimator.

In turn, Proposition 2 in Fiorentini and Sentana (2019) states that if (G60) holds, and $\boldsymbol{\varepsilon}_t^*|I_{t-1};\boldsymbol{\varpi}_0$ is $i.i.d.$ $(\mathbf{0},\mathbf{I}_N)$, where $\boldsymbol{\varpi}$ includes $\boldsymbol{\phi}$ and the true shape parameters, but the distribution assumed for estimation purposes does not necessarily nest the true density, then the pseudo true value of the joint ML estimator of $\boldsymbol{\varphi}=(\boldsymbol{\phi}_c',\boldsymbol{\phi}_i',\boldsymbol{\varrho})'$, $\boldsymbol{\varphi}_\infty$, is such that $\boldsymbol{\phi}_{c\infty}$ is equal to the true value $\boldsymbol{\phi}_{c0}$. In this context, in Fiorentini and Sentana (2007) we proposed to estimate $\boldsymbol{\phi}_{im}$ and $\boldsymbol{\phi}_{ic}$ as $\boldsymbol{\phi}_{imT}(\hat{\boldsymbol{\phi}}_{cT})$ and $\boldsymbol{\phi}_{icT}(\hat{\boldsymbol{\phi}}_{cT})$, respectively, where

$$\boldsymbol{\phi}_{imT}(\boldsymbol{\phi}_c) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\phi}_c), \tag{G66}$$

$$\boldsymbol{\phi}_{icT}(\boldsymbol{\phi}_c) = vech\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\phi}_c)-\boldsymbol{\phi}_{imT}(\boldsymbol{\phi}_c)\right]\left[\boldsymbol{\varepsilon}_t^{\diamond}(\boldsymbol{\phi}_c)-\boldsymbol{\phi}_{imT}(\boldsymbol{\phi}_c)\right]'\right\}. \tag{G67}$$

Once again, the rationale for these estimators comes from the fact that under normality the scores for $\boldsymbol{\phi}_{im}$ and $\boldsymbol{\phi}_{ic}$ simplify to:

$$\begin{aligned}
\mathbf{s}_{\psi_{im}t}(\boldsymbol{\phi},\mathbf{0}) &= \frac{1}{2}\boldsymbol{\Phi}_{ic}^{-1/2\prime}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi}), \\
\mathbf{s}_{\psi_{ic}t}(\boldsymbol{\phi},\mathbf{0}) &= \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Phi}_{ic}^{-1/2\prime}\otimes\boldsymbol{\Phi}_{ic}^{-1/2\prime})vec\left\{\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\phi})-\mathbf{I}_N\right\},
\end{aligned}$$

where $\mathbf{D}_N$ is the duplication matrix (see Magnus and Neudecker (2019)), whose expected values at $\boldsymbol{\phi}_0$ are 0 because the expected value of

$$\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi}_{c0},\boldsymbol{\phi}_i)=\boldsymbol{\Phi}_{ic}^{-1/2}(\boldsymbol{\phi}_{im0}-\boldsymbol{\phi}_{im})+\boldsymbol{\Phi}_{ic}^{-1/2}\boldsymbol{\Phi}_{ic0}^{1/2}\boldsymbol{\varepsilon}_t^*$$

is 0 and the expected value of $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\phi}_{c0},\boldsymbol{\phi}_i)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\phi}_{c0},\boldsymbol{\phi}_i)$ is $\mathbf{I}_N$ when $\boldsymbol{\phi}_i=\boldsymbol{\phi}_{i0}$.

However, it turns out that Proposition F1 above implies that (G66) and (G67) numerically coincide with the MLEs of $\boldsymbol{\phi}_{im}$ and $\boldsymbol{\phi}_{ic}$ when the assumed distribution is an unrestricted discrete mixture of normals, so once again, it is irrelevant whether we replace them or not. As a result, the ML estimators based on an unrestricted discrete mixture of normals are consistent for all the parameters regardless of the true distribution. Moreover, (G66) and (G67) also give us the SP estimators of $\boldsymbol{\phi}_{im}$ and $\boldsymbol{\phi}_{ic}$ when $\boldsymbol{\phi}_c$ is replaced by its iterated SP estimator. In this respect, the results in this appendix provide an alternative justification for the model-specific consistency results in Lee and Lee (2009) and Ha and Lee (2011).

H Standardised random variables

H.1 Univariate discrete location scale mixtures of normals

Let s_t denote an *i.i.d.* Bernoulli variate with $P(s_t=1)=\lambda$. If $z_t|s_t$ is *i.i.d.* $N(0,1)$, then

$$\varepsilon_t^*=\frac{1}{\sqrt{1+\lambda(1-\lambda)\delta^2}}\left[\delta(s_t-\lambda)+\frac{s_t+(1-s_t)\sqrt{\kappa}}{\sqrt{\lambda+(1-\lambda)\kappa}}z_t\right],$$

where $\delta\in\mathbb{R}$ and $\kappa>0$, is a two component mixture of normals whose first two unconditional moments are 0 and 1, respectively. The intuition is as follows. First, note that $\delta(s_t-\lambda)$ is a shifted and scaled Bernoulli random variable with 0 mean and variance $\lambda(1-\lambda)\delta^2$. But since

$$\frac{s_t+(1-s_t)\sqrt{\kappa}}{\sqrt{\lambda+(1-\lambda)\kappa}}z_t$$

is a discrete scale mixture of normals with 0 unconditional mean and unit unconditional variance that is orthogonal to $\delta(s_t-\lambda)$, the sum of the two random variables will have variance $1+\lambda(1-\lambda)\delta^2$, which explains the scaling factor.

An equivalent way to define and simulate the same standardised random variable is as follows

$$\varepsilon_t^*=\begin{cases} N[\mu_1^*(\boldsymbol{\varrho}),\sigma_1^{*2}(\boldsymbol{\varrho})] \text{ with probability } \lambda \\ N[\mu_2^*(\boldsymbol{\varrho}),\sigma_2^{*2}(\boldsymbol{\varrho})] \text{ with probability } 1-\lambda \end{cases} \tag{H1}$$

where $\boldsymbol{\varrho} = (\delta, \kappa, \lambda)'$ and

$$\begin{aligned}
\mu_1^*(\boldsymbol{\varrho}) &= \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \\
\mu_2^*(\boldsymbol{\varrho}) &= -\frac{\delta\lambda}{\sqrt{1+\lambda(1-\lambda)\delta^2}} = -\frac{\lambda}{1-\lambda}\mu_1^*(\boldsymbol{\varrho}), \\
\sigma_1^{*2}(\boldsymbol{\varrho}) &= \frac{1}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\kappa]}, \\
\sigma_2^{*2}(\boldsymbol{\varrho}) &= \frac{\kappa}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\kappa]} = \kappa\sigma_1^{*2}(\boldsymbol{\varrho}).
\end{aligned}$$

Therefore, we can immediately interpret κ as the ratio of the two variances. Similarly, since

$$\delta = \frac{\mu_1^*(\boldsymbol{\varrho}) - \mu_2^*(\boldsymbol{\varrho})}{\sqrt{\lambda\sigma_1^{*2}(\boldsymbol{\varrho}) + (1-\lambda)\sigma_1^{*2}(\boldsymbol{\varrho})}},$$

we can also interpret δ as the parameter that regulates the distance between the means of the two underlying components relative to the mean of the two conditional variances.

Finally, note that we can also use the above expressions to generate a two component mixture of normals with mean π and variance ω^2 as

$$y_t = \begin{cases} N(\mu_1, \sigma_1^2) \text{ with probability } \lambda \\ N(\mu_2, \sigma_2^2) \text{ with probability } 1-\lambda \end{cases}$$

with

$$\mu_1 = \pi + \omega\mu_1^*(\boldsymbol{\varrho}), \quad \mu_2 = \pi + \omega\mu_2^*(\boldsymbol{\varrho}), \quad \sigma_1^2 = \omega\sigma_1^{*2}(\boldsymbol{\varrho}), \quad \sigma_2^2 = \omega\sigma_2^{*2}(\boldsymbol{\varrho}).$$

Interestingly, the expressions for κ and δ above continue to be valid if we replace $\mu_1^*(\boldsymbol{\varrho})$, $\mu_2^*(\boldsymbol{\varrho})$, $\sigma_1^{*2}(\boldsymbol{\varrho})$ and $\sigma_2^{*2}(\boldsymbol{\varrho})$ by μ_1, μ_2, σ_1^2 and σ_2^2.

We can trivially extend this procedure to define and simulate standardised mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (H1) by a $(K-1)$-component normal mixture with mean and variance given by $\mu_1^*(\boldsymbol{\varrho})$ and $\sigma_1^{*2}(\boldsymbol{\varrho})$, respectively, then the resulting random variable will be a K-component Gaussian mixture with zero mean and unit variance.

H.2 Standardised multivariate discrete location scale mixtures of normals

Consider the following mixture of two multivariate normals

$$\boldsymbol{\varepsilon}_t \sim \begin{cases} N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) & \text{with probability} \quad \lambda, \\ N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) & \text{with probability} \quad 1-\lambda. \end{cases} \tag{H2}$$

Let s_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:

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

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 $\mathbf{G}_L\mathbf{G}'_L$, 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 $\mathbf{G}_L$ will be functions of λ and the $N(N+1)/2$ elements in $\mathbf{K}_L$. If we then choose $\boldsymbol{\Sigma}_{1L} = \mathbf{G}_L^{-1}$, 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 $\mathbf{G}_L^{-1\prime}\mathbf{G}_L^{-1}$, 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

$\mathbf{\Sigma}_2 = \kappa\mathbf{\Sigma}_1$. Straightforward algebra confirms that the standardisation procedure described above simplifies to the one they provide in their Proposition 1.

As in the univariate case, we can trivially extend this procedure to define and simulate standardised mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (H2) by a $(K-1)$-component normal mixture with mean and variance given by $\boldsymbol{\mu}_1^*(\boldsymbol{\varrho})$ and $\mathbf{\Sigma}_1^*(\boldsymbol{\varrho})$, respectively, then the resulting random variable will be a K-component Gaussian mixture with zero mean and unit variance.

Additional references

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

Amengual, D., Fiorentini, F. and Sentana (2021c): "Information matrix tests for Gaussian mixtures", mimeo, CEMFI.

Boldea, O. and Magnus, J.R. (2009): "Maximum likelihood estimation of the multivariate normal mixture model", *Journal of the American Statistical Association* 104, 1539-1549.

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

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

Dempster, A., Laird, N., and Rubin, D. (1977): "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society B* 39, 1-38.

Ha, J. and Lee, T. (2011): "NM–QELE for Arma-Garch models with non-Gaussian innovations", *Statistics and Probability Letters* 81, 694–703.

Hasselblad, V. (1966): "Estimation of parameters for a mixture of normal distributions", *Technometrics* 8, 431-444.

Kiefer, N.M. (1978): "Discrete parameter variation: efficient estimation of a switching regression model", *Econometrica* 46, 427-434.

Lee, T. and Lee, S. (2009): "Normal mixture quasi-maximum likelihood estimator for Garch models", *Scandinavian Journal of Statistics* 36, 157–170.

Ling, S. and McAleer, M. (2003): "Asymptotic theory for a vector Arma-Garch model", *Econometric Theory* 19, 280-310.

Peters, B.C. and Walker, H.F. (1978): "An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions", *SIAM Journal of Applied Mathematics* 35, 362-378.

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.

White, H. (1982): "Maximum likelihood estimation of misspecified models", *Econometrica* 50, 1–25.

Wolfe, J.H. (1970): "Pattern clustering by multivariate mixture analysis", *Multivariate Behavioral Research* 5, 329-350.

Yang M. (2000): "Some properties of vector autoregressive processes with Markov-switching coefficients", *Econometric Theory* 16, 23-43.