Supplemental Appendices for

Specification tests for non-Gaussian structural vector autoregressions

Dante Amengual

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

<amengual@cemfi.es>

Gabriele Fiorentini

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

<gabriele.fiorentini@unifi.it>

Enrique Sentana

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

<sentana@cemfi.es>

December 2022
Revised: March 2024

A The test in practice

We recommend following these steps for computing the discrete grid test statistics in a given sample:

1. Estimate the model by non-Gaussian PMLE assuming that the shocks follow independent univariate finite Gaussian mixtures, and compute the estimated structural residuals $\varepsilon^*_{it}(\hat{\boldsymbol{\theta}}_T)$'s evaluated at the PMLEs $\hat{\boldsymbol{\theta}}_T$ using expression (4) for the unique ordering and signs of the matrix $\mathbf{C}$ obtained using the selection procedure suggested by Ilmonen and Paindaveine (2011) and adopted by Lanne et al. (2017). Importantly, the fact the structural shocks are only identified up to permutations is numerically irrelevant for the test statistic as long as one uses the same quantile grid for all of them since they only affect their labelling. Similarly, a change in the sign of one shock is also numerically irrelevant as long as one adjusts its quantiles accordingly. In fact, there is no need for such an adjustment if one uses equally spaced quantiles (say terciles, quartiles or quintiles) for all shocks.
2. For the q version of the test, partition the $[0, 1]$ interval with knots $(0, u_1, u_2, ..., u_H, 1)$, where $u_h = \frac{1}{2}(2h-1)H^{-1}$ in the equally-spaced case, and obtain the corresponding marginal quantiles of each estimated shock $\varepsilon^*_{it}(\hat{\boldsymbol{\theta}}_T)$, namely $[k_{i1}(u_1), \ldots, k_{iH}(u_H)]$, $i = 1, \ldots, N$ using MATLAB's linear interpolation method. One could then replace the u's with the marginal empirical cdf of each shock computed at the estimated quantiles to take into account the linear interpolation method, but this would generate slightly different partitions of the unit interval for different shocks.

   For the p version of the tests, define H points, $k_1 < \cdots < k_h < \cdots < k_H$, together with $k_0 = -\infty$ and $k_{H+1} = \infty$, and estimate the marginal empirical cdf for each shock as $p_{ih} = T^{-1}\sum_{t=1}^{T} I[\varepsilon^*_{it}(\hat{\boldsymbol{\theta}}_T) \le k_h]$. One could then replace k_h with its marginal empirical quantile at the estimated p_{ih} for each shock using MATLAB's linear interpolation method, but again this would generate slightly different partitions of the real line for different shocks.
3. For the q version, estimate the joint cdf at the Cartesian product of the empirical quantiles as $q_{ij} = T^{-1}\sum_{t=1}^{T} I[\varepsilon^*_{1t}(\hat{\boldsymbol{\theta}}_T) \le k_{ih}(u_h); \ldots, \varepsilon^*_{Mt}(\hat{\boldsymbol{\theta}}_T) \le k_{Mh'}(u_{h'})]$, while for the p version do the same but evaluate them at the N-ary Cartesian product of $(k_1, \ldots, k_H)'$.
4. Compute the H^N influence functions underlying the test as the difference between the joint and the product of the marginal empirical cdfs.
5. Compute the $H^N \times H^N$ matrix whose elements are given by (13).
6. Estimate the asymptotic covariance matrix of the score and the expected Hessian of the pseudo log-likelihood function replacing the true values of the parameters $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_T$ and the expected values with sample averages in the expressions that appear in Appendices C.3 and C.4, respectively, including (C33)-(C38) and (C39)-(C44), and use them in the

sandwich formula $\mathcal{A}^{-1}\mathcal{B}\mathcal{A}$, retaining the $N \times N$ blocks corresponding to the elements of $vec(\mathbf{C})$. The consistency of the estimators of $\mathcal{A}$ and $\mathcal{B}$ follows from Lemma 4.3 in Newey and McFadden (1994), while that of $\mathcal{A}^{-1}\mathcal{B}\mathcal{A}$ from their Theorem 4.1.

7. Estimate the $H^N \times N$ expected Jacobian matrix of the influence functions with respect to the elements of $vec(\mathbf{C})$ replacing the true values of the parameters $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_T$ and the expected values with sample averages in the expressions in Lemma 1, using Silverman's (1986) robust rule-of-thumb bandwidth to obtain Gaussian kernel estimates of the true density of the shocks that appear in expression (B1). Despite involving the indicator function, the consistency of this procedure follows once again from Lemma 4.3 in Newey and McFadden (1994).
8. Estimate the $H^N \times N$ asymptotic covariance matrix between the influence functions and the scores with respect to the elements of $vec(\mathbf{C})$ replacing the true values of the parameters $\boldsymbol{\theta}_0$ with $\hat{\boldsymbol{\theta}}_T$ and the expected values with sample averages in the expressions that appear in Lemma 2, including (B3)-(B9). As before, Lemma 4.3 in Newey and McFadden (1994) guarantees the consistency of the resulting estimators despite the indicator function appearing in the influence functions.
9. Combine these matrices to estimate $\mathcal{W}$ using (15), and replace this estimated matrix in (14) to obtain the discrete grid test statistic. Theorems 2.2 and 2.3 in Newey (1985) guarantee the consistent estimation of $\mathcal{W}$ and the asymptotic χ^2 distribution of (14), respectively.

Given that the continuous grid test can be regarded as a regularised version of the discrete grid test computed at the finest partition of the unit interval that remains meaningful when there are T observations, its computation shares several of the elements that we have just described. Specifically:

1. Estimate the model by non-Gaussian PMLE assuming that the shocks follow independent univariate finite Gaussian mixtures, and compute the estimated structural residuals $\varepsilon^*_{it}(\hat{\boldsymbol{\theta}}_T)$'s evaluated at the PMLEs $\hat{\boldsymbol{\theta}}_T$ using expression (4) for the unique ordering and signs of the matrix $\mathbf{C}$ obtained using the selection procedure suggested by Ilmonen and Paindaveine (2011) and adopted by Lanne et al. (2017). The fact the structural shocks are only identified up to permutations and sign changes is numerically irrelevant for the continuous test statistic as it effectively depends on the homogeneous, equally-spaced "discrete" grid $u_\tau = \frac{1}{2}(2\tau - 1)T^{-1}$, $\tau = 1, \ldots, T$.
2. Compute the empirical uniform ranks using expression (17) and use them to obtain the elements of the $T \times T$ matrix $\mathcal{D}$ in (25).
3. Estimate the $T \times T$ matrix $\mathcal{C}$ by replacing the integrals in (27) by sums over the empirical cdfs of the shocks. Specifically, if we denote by $\boldsymbol{\epsilon}^*_t(\hat{\boldsymbol{\theta}}) = [\epsilon^*_{it}(\hat{\boldsymbol{\theta}}), ..., \epsilon^*_{\mathrm{Mt}}(\hat{\boldsymbol{\theta}})]$ the vector containing the empirical ranks of the t^{th} observation of each of the estimated shocks that

appear in M, we can estimate the rank one matrix $\mathcal{C}$ as

$$\widehat{\mathcal{C}} = \boldsymbol{\ell}_T\boldsymbol{\ell}_T' \cdot \sum_{\tau^1=1}^{T} \cdots \sum_{\tau^{\mathrm{M}}=1}^{T} \mathbf{j}'[\epsilon_{\tau^1}^*(\hat{\boldsymbol{\theta}}), ..., \epsilon_{\tau^{\mathrm{M}}}^*(\hat{\boldsymbol{\theta}})]\mathcal{A}^{-1}(\hat{\boldsymbol{\theta}})\mathcal{B}(\hat{\boldsymbol{\theta}})\mathcal{A}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{j}[\epsilon_{\tau^1}^*(\hat{\boldsymbol{\theta}}), ..., \epsilon_{\tau^{\mathrm{M}}}^*(\hat{\boldsymbol{\theta}})],$$

where $\boldsymbol{\ell}_T$ is a vector of T ones, while $\mathcal{A}(\hat{\boldsymbol{\theta}})$, $\mathcal{B}(\hat{\boldsymbol{\theta}})$ and $\mathbf{j}[\epsilon_{\tau^1}^*(\hat{\boldsymbol{\theta}}), ..., \epsilon_{\tau^{\mathrm{M}}}^*(\hat{\boldsymbol{\theta}})]$ are the consistent estimators that we mention in points 6 and 7 of the description of the discrete grid test, with the latter evaluated at $u_\tau = \frac{1}{2}(2\tau - 1)T^{-1-1}$, $\tau = 1, \dots, T$. Given that sums over increasingly finer grids converge to the relevant integral, $\widehat{\mathcal{C}}$ will be consistent.

4. Finally, we consistently estimate $\mathcal{E}$ by adding up the consistent estimators of $\mathcal{C}$ and $\mathcal{D}$, which we then replace in expression (26) for a given choice of the regularization parameter α. Interestingly, the fact that $\widehat{\mathcal{C}}$ is proportional to $\boldsymbol{\ell}_T\boldsymbol{\ell}_T'$ implies that the expression (26) is numerically unaffected if we replace the two $\mathcal{E}s$ that appear at the extremes of this quadratic form with $\mathcal{D}s$.

B Lemmata

Lemma 1 *If model (2) satisfies Assumption 1, then the non-zero elements of the expected Jacobian matrix of the linearised $m_t(\mathbf{u})$ evaluated at $\boldsymbol{\theta}_0$ and the estimated values of u_h^i in (10) are given by*

$$\mathrm{J}_{p_{\mathbf{h}}c_{ii'}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\varphi}_0) = -\sum_{i\in M}\sum_{i'\in M, i'\neq i}\left(\prod_{i''\in M, i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f[\varkappa(u_i)], \quad \textit{for} \quad i \neq i', \tag{B1}$$

where $\eta_{h_{i'}} = E_0[\varepsilon_{it}^* 1_{(-\infty,\varkappa(u_i))}(\varepsilon_{it}^*)]$ *for* $i \in M$.

Proof. From (12), we have that

$$\begin{aligned}\frac{\partial m_t(\mathbf{u})}{\partial\theta} &= E\left[\frac{\partial}{\partial\theta}\left[\prod_{i\in M} 1_{(-\infty,\varkappa(u_i))}(\varepsilon_{it}^*) - \prod_{i\in M} u_i\right]\right] \\ &\quad -\frac{\partial}{\partial\theta}\left\{\sum_{i\in M}\left[1_{(-\infty,\varkappa(u_i))}(\varepsilon_{it}^*) - u_i\right]\prod_{i'\in M, i'\neq i} u_{i'}\right\}\Bigg] \\ &= -\sum_{i\in M}\left[\prod_{i'\in M, i'\neq i} 1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon_{i't}^*)\right]\left[1_{(-\infty,\varkappa(u_i))}(\varepsilon_{it}^*) - u_i\right]\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon_{it}^*)}{\partial\varepsilon_{it}^*}\frac{\partial\varepsilon_{it}^*}{\partial\theta}.\end{aligned}$$

Moreover, it is worth noticing that

$$\begin{aligned}\frac{\partial\varepsilon_{it}^*(\boldsymbol{\theta})}{\partial\boldsymbol{\tau}'} &= -\mathbf{c}^{i\cdot}, \\ \frac{\partial\varepsilon_{it}^*(\boldsymbol{\theta})}{\partial\mathbf{a}_j'} &= -(\mathbf{y}_{t-j}' \otimes \mathbf{c}^{i\cdot}) \quad \text{for } j = 1, ..., p, \text{ and} \\ \frac{\partial\varepsilon_{it}^*(\boldsymbol{\theta})}{\partial\mathbf{c}'} &= -[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{c}^{i\cdot}].\end{aligned} \tag{B2}$$

Therefore, under the independence null,

$$E\left[\frac{\partial m_t(\mathbf{u})}{\partial \theta_i}\right] = 0$$

except for the off-diagonal elements of $\mathbf{C}$, namely,

$$\begin{aligned}
& E\left\{\sum_{i\in M}\left(\sum_{i'\in M, i'\neq i} 1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon^*_{i't})\right)\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})}{\partial \varepsilon^*_{it}}\frac{\partial \varepsilon^*_{it}}{\partial \mathbf{c}'}\right\} \\
= & -E\left\{\sum_{i\in M}\left(\sum_{i'=1, i'\neq i} 1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon^*_{i't})\right)\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})}{\partial \varepsilon^*_{it}}[\boldsymbol{\varepsilon}^{*\prime}_t(\boldsymbol{\theta})\otimes \mathbf{c}^{i\cdot}]\right\} \\
= & -\sum_{i\in M}\sum_{i'\in M, i'\neq i} E\left(\prod_{i''\in M, i''\neq i'\neq i} 1_{(-\infty,\varkappa(u_{i''}))}(\varepsilon^*_{i''t})\right) \\
& \times E[1_{(-\infty,\varkappa(u_{i'}))}(\varepsilon^*_{i't})\varepsilon^*_{i't}]E\left[\frac{\partial 1_{(-\infty,\varkappa(u_i))}(\varepsilon^*_{it})}{\partial \varepsilon^*_{it}}\right](\mathbf{e}'_j\otimes \mathbf{c}^{i\cdot}) \\
= & -\sum_{i\in M}\sum_{i'\in M, i'\neq i}\left(\prod_{i''\in M, i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f[\varkappa(u_i)]
\end{aligned}$$

where the first equality uses (B2), the second one follows from the cross-sectional independence of the shocks, and the last one implicitly defines $\eta_{u_j} = E[\varepsilon^*_{jt}1_{(-\infty,\varkappa(u_j))}(\varepsilon^*_{jt})]$. □

Lemma 2 *If model (2) satisfies Assumption 1, then the non-zero elements of the covariance matrix between the linearised influence function $m_t(\mathbf{u})$ evaluated at $\boldsymbol{\theta}_0$ and the estimated values of u^i_h in (10) with the pseudo log-likelihood scores evaluated at the pseudo true values $\boldsymbol{\phi}_\infty$ is given by*

$$cov[m_t(\mathbf{u}), \mathbf{s}_{c_{ii'}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0] = E[\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0)],$$

where

$$\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}_0) & \mathbf{Z}_s(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathbf{0} \end{bmatrix},$$

where $\mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)$ is a $N^2\times 1$ vector whose entries $s = N(i-1)+i'$ for $i, i' = 1, ..., N$ are

$$\mathrm{K}_{m,s}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = -\sum_{i\in M}\sum_{i'\in M}\left(\prod_{i''\in M, i''\neq i, i''\neq i'} u_{i''}\right)\eta_{i'}E\left\{1_{(\varepsilon^*_{it}\leq\varkappa(u_i))}\cdot\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon^*_i}\Bigg|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\right\},$$

for $i\neq i'$, and zero otherwise.

Proof. We start by computing the covariance of the influence functions underlying our test with the pseudo log-likelihood scores evaluated at the pseudo true values $\boldsymbol{\phi}_\infty$, namely

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0] = \mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = E[\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0)]$$

and

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0] = \mathcal{K}_{m_\mathbf{u}}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = E[\mathcal{K}_{m_\mathbf{u}t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0)],$$

where

$$\mathcal{K}_{\cdot t}(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}_0) & \mathbf{Z}_s(\boldsymbol{\theta}_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{bmatrix} \mathcal{K}_{\cdot lt}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{K}_{\cdot st}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{K}_{\cdot rt}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \end{bmatrix}.$$

Exploiting the cross-sectional independence of the shocks, we get for the mean parameters

$$\begin{aligned} \mathrm{K}_{p_k l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\} \\ &= -E\left\{ 1_{(\varepsilon_{it}^* \leq \varkappa(u_i))} \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\}, & \text{(B3)} \\ \mathrm{K}_{p_\mathbf{k} l}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i{}^*} \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\} \\ &= -\left(\prod_{i' \in M, i' \neq i} u_{i'} \right) E\left\{ 1_{(\varepsilon_{it}^* \leq \varkappa(u_i))} \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\}, & \text{(B4)} \end{aligned}$$

and zero otherwise.

Similarly, $\mathcal{K}_{\cdot s}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)$ is a $N^2 \times 1$ vector whose entries are such that for i with $j_i > 0$,

$$\begin{aligned} \mathrm{K}_{p_k s_1}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ p(\boldsymbol{\varepsilon}_t^*; \mathbf{k}, \mathbf{u}), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\} \\ &= -E\left\{ 1_{(\varepsilon_{it}^* \geq k_{h_i})} \left[1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right] \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\}, & \text{(B5)} \\ \mathrm{K}_{p_\mathbf{k} s_1}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\} \\ &= -\left(\prod_{i' \in M, i' \neq i} \pi_{i'} \right) E\left\{ \varepsilon_{it}^* 1_{(\varepsilon_{it}^* \leq k_{h_i})} \cdot \left[1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{it}^* \right] \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\}, & \text{(B6)} \\ \mathrm{K}_{p_\mathbf{k} s_2}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= -cov\left\{ m_t(\mathbf{u}), 1 + \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \cdot \varepsilon_{i't}^* \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\} \\ &= -\left(\prod_{i'' \in M, i'' \neq i, i'' \neq i'} u_{i''} \right) \eta_{i'} E\left\{ 1_{(\varepsilon_{it}^* \leq k_{h_i})} \cdot \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*} \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\}, & \text{(B7)} \end{aligned}$$

and zero otherwise.

Finally, $\mathcal{K}_{kr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = \mathrm{K}_{kr}' vecd(\mathbf{I}_n)$, where K_{kr} another block diagonal matrix of order $N \times q$ with typical block of size $1 \times q_i$,

$$\begin{aligned} \mathrm{K}_{p_k r}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= cov\left\{ m_t(\mathbf{u}), \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\} \\ &= E\left\{ 1_{(\varepsilon_{it}^* \leq \varkappa(u_i))} \cdot \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i'} \middle| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0 \right\} & \text{(B8)} \end{aligned}$$

$$\begin{aligned}
\mathrm{K}_{p_{\mathbf{k}}r}(\boldsymbol{\varrho}_{\infty},\boldsymbol{\upsilon}_0) &= cov\left\{ m_t(\mathbf{u}), \left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}'_i}\right| \boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0 \right\} \\
&= \left(\prod_{i'\in M, i'\neq i} u_{i'}\right) E\left\{ 1_{(\varepsilon^*_{it}\leq k_{h_i})} \cdot \left.\frac{\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}'_i}\right| \boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0 \right\} \qquad \text{(B9)}
\end{aligned}$$

and zero otherwise, again because of the cross-sectional independence of the shocks and the fact that $E[\partial \ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{\infty})/\partial\varepsilon^*_i|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = 0$.

Next, to obtain the covariance of the influence function evaluated at $\boldsymbol{\theta}_0$ and the estimated values of u^i_h in (10) with the pseudo log-likelihood scores evaluated at the true values $\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0$, we can make use of (12) to write

$$\begin{aligned}
cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] &= cov\{p(\boldsymbol{\varepsilon}^*_t;\mathbf{k},\mathbf{u}), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\} \qquad \text{(B10)} \\
&\quad - \sum_{i\in M}\left(\prod_{i'\in M, i'\neq i} u_{i'}\right) cov\left\{p_{k_i}(\varepsilon^*_{it}), \mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0\right\}.
\end{aligned}$$

Then, substituting (B3) and (B4) into (B10), we get

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\tau}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = \mathbf{0}$$

and

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\mathbf{a}_j t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = \mathbf{0}, \text{ for } j = 1,...,p.$$

Similarly, substituting (B5) and (B6) into (B10), we get

$$cov[m_t(\mathbf{u}), \mathbf{s}_{c_{ii}t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = 0, \text{ for } i = 1,...,N;$$

and substituting (B8) and (B9) into (B10), we get

$$cov[m_t(\mathbf{u}), \mathbf{s}_{\boldsymbol{\varrho}_i t}(\boldsymbol{\phi}_{\infty})|\boldsymbol{\theta}_0,\boldsymbol{\upsilon}_0] = \mathbf{0}, \text{ for } i = 1,...,N.$$

Finally, substituting (B5) and (B7) into (B10), we get the result stated in the statement. □

Lemma 3 *If model (2) satisfies Assumption 1, then the adjustment of the covariance operator that accounts for the estimation of $\boldsymbol{\theta}$ is given by (27).*

Proof. From 1, the expected Jacobian with respect to $\boldsymbol{\theta}$ of the influence functions linearised with respect to the $\varkappa$'s can be written as

$$E\left[\frac{\partial n_t(\mathbf{u}_M)}{\partial \boldsymbol{\theta}'}\right] = -\sum_{i\in M}\sum_{i'\in M, i'\neq i}\left(\prod_{i''\in M, i''\neq i'\neq i} u_{i''}\right)\eta_{u_{i'}} f[\varkappa(u_i)](\mathbf{e}'_{i'}\otimes \mathbf{c}^{i\cdot}),$$

where

$$n_t(\mathbf{u}_M) = \left[\prod_{i\in M} 1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - \prod_{i\in M} u_i\right] - \sum_{i\in M}\left[1_{(-\infty,\varkappa_i(u_i))}(\varepsilon^*_{it}) - u_i\right]\prod_{i'\in M, i'\neq i} u_{i'}.$$

We are after

$$\int_{[0,1]^M} \left\{ n_t(\mathbf{u}_M) - E\left[\frac{\partial n_t(\mathbf{u}_M)}{\partial \boldsymbol{\theta}'}\right] \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right\} \left\{ n_s(\mathbf{u}_M) - E\left[\frac{\partial n_s(\mathbf{u}_M)}{\partial \boldsymbol{\theta}'}\right] \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right\} d\mathbf{u}_M.$$

Let us consider each of the four terms separately. The first one, namely

$$\int_{[0,1]^M} n_t(\mathbf{u}_M) n_s(\mathbf{u}_M) d\mathbf{u}_M,$$

is given in (13). Next, we have the cross-terms, which are of the form

$$-\int_{[0,1]^M} E\left[\frac{\partial n_s(\mathbf{u}_M)}{\partial \boldsymbol{\theta}'}\right] \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) n_t(\mathbf{u}_M) d\mathbf{u}_M.$$

If we then use the fact that

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \sqrt{T}\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\bar{\mathbf{s}}_{\boldsymbol{\theta}} + o_p(1) = \mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\frac{\sqrt{T}}{T}\sum_{t=1}^{T} \mathbf{s}_{\boldsymbol{\theta}t} + o_p(1),$$

we can see that

$$-\frac{1}{\sqrt{T}} \int_{[0,1]^M} E\left[\frac{\partial n_s(\mathbf{u}_M)}{\partial \boldsymbol{\theta}'}\right] \left(\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\sum_{\tau=1}^{T} \mathbf{s}_{\boldsymbol{\theta}\tau}\right) n_t(\mathbf{u}_M) d\mathbf{u}_M = o_p(1)$$

because of the scaling factor $1/\sqrt{T}$ and the fact that the ε's entering into $\mathbf{s}_{\boldsymbol{\theta}\tau}(\boldsymbol{\phi})$ are asymptotically independent of the ones that appear in $n_t(\mathbf{u}_M)$ and $E\left[\partial n_s(\mathbf{u}_M)/\partial \boldsymbol{\theta}'\right]$. Therefore, the covariance of the linearised influence function with the pseudo log-likelihood scores evaluated at the pseudo true values $\boldsymbol{\phi}_\infty$ is asymptotically negligible.

Finally, regarding the last term, we obtain (27), as desired. □

C ML 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, but all the parameters that characterise the conditional mean and covariance functions are consistently estimated, as in the case of finite normal mixtures. Fiorentini and Sentana (2023) consider the general case.

C.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 will be given by

$$l_t(\mathbf{y}_t; \boldsymbol{\varphi}) = -\ln|\mathbf{C}| + l[\varepsilon^*_{1t}(\boldsymbol{\theta}); \boldsymbol{\varrho}_1] + \ldots + l[\varepsilon^*_{Nt}(\boldsymbol{\theta}); \boldsymbol{\varrho}_N], \quad \text{(C11)}$$

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 $l(\varepsilon_{it}^*; \boldsymbol{\varrho}_i) = \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_i)$ is the log of the univariate density function of ε_{it}^*, 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 (2) conditional on I_{t-1} are

$$\begin{aligned} \boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\tau} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p}, && \text{(C12a)} \\ \boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \mathbf{C}\mathbf{C}', && \text{(C12b)} \end{aligned}$$

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

$$\frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\frac{\partial vec'(\mathbf{C})}{\partial\boldsymbol{\theta}} vec(\mathbf{C}^{-1\prime}) = -\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{N^2} \end{pmatrix} vec(\mathbf{C}^{-1\prime}) = -\mathbf{Z}_{st}'(\boldsymbol{\theta}) vec(\mathbf{I}_N) \tag{C13}$$

and

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= -\mathbf{C}^{-1}\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{C}^{-1}]\frac{\partial vec(\mathbf{C})}{\partial\boldsymbol{\theta}'} \\ &= -\{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}, \end{aligned} \tag{C14}$$

where

$$\mathbf{Z}_{lt}(\boldsymbol{\theta}) = \frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1}\otimes\mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p}\otimes\mathbf{I}_N \\ \mathbf{0}_{N^2\times N} \end{pmatrix}\mathbf{C}^{-1\prime}, \tag{C15}$$

$$\mathbf{Z}_{st}(\boldsymbol{\theta}) = \frac{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}[\mathbf{I}_N\otimes\boldsymbol{\Sigma}_t^{-1/2\prime}(\boldsymbol{\theta})] = \begin{pmatrix} \mathbf{0}_{N\times N^2} \\ \mathbf{0}_{N^2\times N^2} \\ \vdots \\ \mathbf{0}_{N^2\times N^2} \\ \mathbf{I}_{N^2} \end{pmatrix}(\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}), \tag{C16}$$

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

$$\begin{aligned} \mathbf{s}_t(\boldsymbol{\phi}) &= \begin{bmatrix} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) \\ \mathbf{s}_{\boldsymbol{\varrho}t}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix}\begin{bmatrix} \mathbf{e}_{lt}(\boldsymbol{\phi}) \\ \mathbf{e}_{st}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix}\begin{bmatrix} \mathbf{e}_{dt}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{bmatrix} = \mathbf{Z}_t(\boldsymbol{\theta})\mathbf{e}_t(\boldsymbol{\phi}), \end{aligned} \tag{C17}$$

where

$$\mathbf{e}_{lt}(\boldsymbol{\phi}) = -\frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*} = -\begin{bmatrix} \partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]/\partial \varepsilon_1^* \\ \partial \ln f_2[\varepsilon_{2t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_2]/\partial \varepsilon_2^* \\ \vdots \\ \partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]/\partial \varepsilon_N^* \end{bmatrix}, \tag{C18}$$

$$\begin{aligned} \mathbf{e}_{st}(\boldsymbol{\phi}) &= -vec\left\{\mathbf{I}_N + \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\} \\ &= -vec\left\{\begin{array}{ccc} 1 + \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{1t}^*(\boldsymbol{\theta}) & \dots & 1 + \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^*}\varepsilon_{Nt}^*(\boldsymbol{\theta}) \end{array}\right\} \end{aligned} \tag{C19}$$

and

$$\mathbf{e}_{rt}(\boldsymbol{\phi}) = \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho}} = \left\{\begin{array}{c} \frac{\partial \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1} \\ \vdots \\ \frac{\partial \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N} \end{array}\right\} = \begin{bmatrix} \mathbf{e}_{r_1 t}(\boldsymbol{\phi}) \\ \mathbf{e}_{r_2 t}(\boldsymbol{\phi}) \\ \vdots \\ \mathbf{e}_{r_N t}(\boldsymbol{\phi}) \end{bmatrix} \tag{C20}$$

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

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 (2021) implies that

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} \\ + \left[\mathbf{e}_{lt}'(\boldsymbol{\phi}) \otimes \mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{lt}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} + \left[\mathbf{e}_{st}'(\boldsymbol{\phi}) \otimes \mathbf{I}_{N+(p+1)N^2}\right]\frac{\partial vec[\mathbf{Z}_{st}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'}, \end{aligned} \tag{C21}$$

where $\mathbf{Z}_{lt}(\boldsymbol{\theta})$ and $\mathbf{Z}_{st}(\boldsymbol{\theta})$ are given in (C15) and (C16), 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

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

$$= \left[\mathbf{I}_N \otimes \begin{pmatrix} \mathbf{I}_N \\ \mathbf{y}_{t-1} \otimes \mathbf{I}_N \\ \vdots \\ \mathbf{y}_{t-p} \otimes \mathbf{I}_N \\ \mathbf{0}_{N^2 \times N} \end{pmatrix}\right] \begin{bmatrix} \mathbf{0}_{N^2 \times (N+pN^2)} & (\mathbf{C}^{-1} \otimes \mathbf{C}^{-1\prime})\mathbf{K}_{NN} \end{bmatrix}.$$

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,

$$\frac{\partial \mathbf{e}_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho})}{\partial \boldsymbol{\theta}'} = -\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\} \quad \text{(C22)}$$

and

$$\begin{aligned} \frac{\partial \mathbf{e}_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} &= -[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \left\{\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right\} \frac{\partial \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ &= \left\{[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N] \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} + \left[\mathbf{I}_N \otimes \frac{\partial \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^*}\right]\right\} \\ &\quad \times \{\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + [\boldsymbol{\varepsilon}_t'^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta})\}. \end{aligned} \quad \text{(C23)}$$

The assumed independence across innovations implies that

$$\frac{\ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varepsilon}^{*\prime}} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{(\partial \varepsilon_1^*)^2} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{(\partial \varepsilon_N^*)^2} \end{bmatrix}, \quad \text{(C24)}$$

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

$$\frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \varepsilon_1^* \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \varepsilon_N^* \partial \boldsymbol{\varrho}_N'} \end{bmatrix} \tag{C25}$$

because of the cross-sectional independence assumption.

As for the shape parameters of the independent margins,

$$\mathbf{h}_{\boldsymbol{\varrho}\boldsymbol{\varrho}t}(\boldsymbol{\phi}) = \frac{\partial^2 \ln f[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]}{\partial \boldsymbol{\varrho} \partial \boldsymbol{\varrho}'} = \begin{bmatrix} \frac{\partial^2 \ln f_1[\varepsilon_{1t}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_1]}{\partial \boldsymbol{\varrho}_1 \partial \boldsymbol{\varrho}_1'} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & 0 & \frac{\partial^2 \ln f_N[\varepsilon_{Nt}^*(\boldsymbol{\theta}); \boldsymbol{\varrho}_N]}{\partial \boldsymbol{\varrho}_N \partial \boldsymbol{\varrho}_N'} \end{bmatrix}. \tag{C26}$$

Finally, regarding the Jacobian term $-\ln|\mathbf{C}|$, we have that differentiating (C13) 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

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

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

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

Let us now consider the alternative parametrisation $\mathbf{C} = \mathbf{J\Psi}$ studied in Fiorentini and Sentana (2021, 2023), so that the parameters of interest become $\boldsymbol{\tau}$, $\mathbf{a}_j = vec(\mathbf{A}_j)$ $(j = 1, \ldots, p)$, $\mathbf{j} = veco(\mathbf{J})$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$, where $veco(.)$ stacks by columns all the elements of the zero-diagonal matrix $\mathbf{J} - \mathbf{I}_N$ except those that appear in its diagonal, and $vecd(.)$ places the elements in the main diagonal of $\boldsymbol{\Psi}$ in a column vector (see Magnus and Sentana (2020) for some useful properties of these operators). Given that a pseudo log-likelihood function based on finite Gaussian mixtures for the shocks will lead to consistent estimators for all these parameters regardless of the true distribution, $\mathbf{e}_t(\boldsymbol{\phi}_\infty)$ will be serially independent and not just martingale difference sequences. Moreover, given that

$$\mathbf{Z}(\boldsymbol{\theta}) = E[\mathbf{Z}_t(\boldsymbol{\theta})|\boldsymbol{\varphi}_0] = \begin{bmatrix} \mathbf{C}^{-1\prime} & \mathbf{0}_{N\times N^2} & \mathbf{0}_{N\times q} \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \vdots & \vdots & \vdots \\ (\boldsymbol{\mu}\otimes\mathbf{I}_N)\mathbf{C}^{-1\prime} & \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{N^2\times N} & (\mathbf{I}_N\otimes\mathbf{C}^{-1\prime}) & \mathbf{0}_{N^2\times q} \\ \mathbf{0}_{q\times N} & \mathbf{0}_{q\times N^2} & \mathbf{I}_q \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_d(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix} \tag{C27}$$

has full column rank,

$$E[\mathbf{e}_t(\boldsymbol{\phi}_\infty)|I_{t-1}, \boldsymbol{\varphi}_0] = \mathbf{0} \tag{C28}$$

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 $\boldsymbol{\varepsilon}_t^*$. In addition, given that the estimators of $\boldsymbol{\theta}$ that we consider are consistent, we will have that under standard regularity conditions

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^*(\hat{\boldsymbol{\theta}}) \rightarrow E[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = 0 \text{ and} \tag{C29}$$

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^{*2}(\hat{\boldsymbol{\theta}}) \rightarrow E[\varepsilon_{it}^{*^2}(\boldsymbol{\theta}_\infty)|\boldsymbol{\varphi}_0] = 1, \tag{C30}$$

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

C.2 Asymptotic distribution

For simplicity, we assume henceforth that there are no unit roots in the autoregressive polynomial, so that the SVAR model (2) 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, although like in his Theorems 3.1 and 3.2, we drop Assumption A3(b) when talking about the negative definiteness of the expected Hessian or the asymptotic normality of the PML estimators because they are both local rather than global results. 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

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

and

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

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.

If we write $\mathbf{C} = \mathbf{J}\boldsymbol{\Psi}$, then the chain rule for first derivatives implies that the gradient with respect to the parameters in $\mathbf{C}$ will be a linear combination of those corresponding to $\mathbf{j} = veco(\mathbf{J} - \mathbf{I}_N)$ and $\boldsymbol{\psi} = vecd(\boldsymbol{\Psi})$.

Therefore, we can invoke Proposition 3 in Fiorentini and Sentana (2023), which shows the consistency of the Gaussian mixture-based Pseudo MLEs of $\mathbf{j}$ and $\boldsymbol{\psi}$, to show that

$$E\left[\left.\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \varepsilon_i^*}\right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\right] = 0$$

and

$$E\left[1 + \left.\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \varepsilon_i^*}\varepsilon_{it}^*(\boldsymbol{\theta}_\infty)\right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\right] = 0 \tag{C31}$$

for $i = 1, ..., N$. Moreover, the maintained assumption of cross-sectional independence of the shocks also implies that

$$E\left[\left.\frac{\partial \ln f[\varepsilon_{it}^*(\boldsymbol{\theta}_\infty); \boldsymbol{\varrho}_\infty]}{\partial \varepsilon_i^*}\varepsilon_{jt}^*(\boldsymbol{\theta}_\infty)\right| \boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0\right] = 0$$

As a consequence,

$$E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0] = \mathbf{0} \quad \text{and} \quad E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0] = \mathbf{0}.$$

C.3 Variance of the score

If we maintain that $\boldsymbol{\theta}_\infty = \boldsymbol{\theta}_0$ because of the aforementioned consistency, and adapt Proposition D.2 in Fiorentini and Sentana (2023) to a PMLE context, we can show that

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

where

$$\mathcal{B}_t(\boldsymbol{\phi}_\infty, \boldsymbol{\upsilon}_0) = \mathbf{Z}_t(\boldsymbol{\theta}_\infty)\mathcal{O}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)\mathbf{Z}_t'(\boldsymbol{\theta}_\infty), \tag{C32}$$

$$\mathbf{Z}_t(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_s(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix},$$

and

$$\mathcal{O}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = \begin{bmatrix} \mathcal{O}_{ll}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) & \mathcal{O}_{ls}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) & \mathcal{O}_{lr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{O}_{ls}'(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) & \mathcal{O}_{ss}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) & \mathcal{O}_{sr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \\ \mathcal{O}_{lr}'(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) & \mathcal{O}_{sr}'(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) & \mathcal{O}_{rr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) \end{bmatrix},$$

with

$$\begin{aligned}
\mathcal{O}_{ll}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= V[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0], \\
\mathcal{O}_{ls}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}_{st}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0], \\
\mathcal{O}_{ss}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= V[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0], \\
\mathcal{O}_{lr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= E[\mathbf{e}_{lt}(\boldsymbol{\phi}_\infty)\mathbf{e}_{rt}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0], \\
\mathcal{O}_{sr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= E[\mathbf{e}_{st}(\boldsymbol{\phi}_\infty)\mathbf{e}_{rt}'(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0], \text{ and} \\
\mathcal{O}_{rr}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) &= V[\mathbf{e}_{rt}(\boldsymbol{\phi}_\infty)|\boldsymbol{\theta}_0, \boldsymbol{\upsilon}_0].
\end{aligned}$$

$\mathcal{O}_{ll}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)$ will be a diagonal matrix of order N with typical element

$$\mathrm{o}_{ll}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\upsilon}_0) = V\left[\left.\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}\right| \boldsymbol{\upsilon}_0\right], \tag{C33}$$

$\mathcal{O}_{ls}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0) = \mathrm{O}_{ls}\mathbf{E}_N'$, where $\mathbf{E}_N'$ is the so-called diagonalization matrix and O_{ls} is a diagonal matrix of order N with typical element

$$\mathrm{o}_{ls}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{\upsilon}_0) = cov\left[\left.\frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}, \frac{\partial \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^*}\varepsilon_{it}^*\right| \boldsymbol{\upsilon}_0\right], \tag{C34}$$

$\mathcal{O}_{ss}(\boldsymbol{\varrho}_\infty, \boldsymbol{\upsilon}_0)$ is the sum of the commutation matrix $\mathbf{K}_{NN}$ and a block diagonal matrix $\boldsymbol{\Upsilon}$

of order N^2 in which each of the N diagonal blocks is a diagonal matrix of size N with the following structure:

$$\boldsymbol{\Upsilon}_i(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)=\begin{bmatrix} \mathrm{o}_{ll,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{o}_{ll,i-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{o}_{ss}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)-1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{o}_{ll,i+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathrm{o}_{ll,N} \end{bmatrix},$$

where $\mathrm{o}_{ll,i}=\mathrm{o}_{ll}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)$ to shorten the expressions and

$$\mathrm{o}_{ss}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)=V\left[\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial\varepsilon^*_i}\varepsilon^*_{it}\middle|\boldsymbol{\upsilon}_0\right], \tag{C35}$$

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

$$\mathrm{o}_{lr}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)=-cov\left[\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial\varepsilon^*_i},\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial\boldsymbol{\varrho}_i}\middle|\boldsymbol{\upsilon}_0\right], \tag{C36}$$

$\mathcal{O}_{sr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)=\mathbf{E}_N\mathrm{O}_{sr}$, where O_{sr} another block diagonal matrix of order $N\times q$ with typical block of size $1\times q_i$

$$\mathrm{o}_{sr}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)=-cov\left[\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial\varepsilon^*_i}\varepsilon^*_{it},\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial\boldsymbol{\varrho}_i}\middle|\boldsymbol{\upsilon}_0\right], \tag{C37}$$

and $\mathcal{O}_{rr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)$ is a $q\times q$ block diagonal matrix with typical block of size $q_i\times q_i$

$$\mathrm{o}_{rr}(\boldsymbol{\varrho}_{i\infty},\boldsymbol{\upsilon}_0)=V\left[\frac{\partial\ln f(\varepsilon^*_{it};\boldsymbol{\varrho}_{i\infty})}{\partial\boldsymbol{\varrho}_i}\middle|\boldsymbol{\upsilon}_0\right]. \tag{C38}$$

C.4 Expected Hessian

We can also show that

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

where

$$\mathcal{A}_t(\boldsymbol{\phi}_\infty,\boldsymbol{\upsilon}_0)=\mathbf{Z}_t(\boldsymbol{\theta}_0)\mathcal{H}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)\mathbf{Z}'_t(\boldsymbol{\theta}_0),$$

$$\mathcal{H}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0)=\begin{bmatrix} \mathcal{H}_{ll}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{H}_{ls}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{H}_{lr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) \\ \mathcal{H}'_{ls}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{H}_{ss}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{H}_{sr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) \\ \mathcal{H}'_{lr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{H}'_{sr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) & \mathcal{H}_{rr}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) \end{bmatrix},$$

$$\begin{aligned}
\mathcal{H}_{ll}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) &= -E\left[\frac{\partial^2\ln f(\boldsymbol{\varepsilon}^*_t;\boldsymbol{\varrho}_\infty)}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}\middle|\boldsymbol{\upsilon}_0\right] \\
\mathcal{H}_{ls}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) &= -E\left[\frac{\partial^2\ln f(\boldsymbol{\varepsilon}^*_t;\boldsymbol{\varrho}_\infty)}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}(\boldsymbol{\varepsilon}^{*\prime}_t\otimes\mathbf{I}_N)\middle|\boldsymbol{\upsilon}_0\right] \\
\mathcal{H}_{ss}(\boldsymbol{\varrho}_\infty,\boldsymbol{\upsilon}_0) &= -E\left[\left\{[\boldsymbol{\varepsilon}^*_t\otimes\mathbf{I}_N]\frac{\partial^2\ln f(\boldsymbol{\varepsilon}^*_t;\boldsymbol{\varrho}_\infty)}{\partial\boldsymbol{\varepsilon}^*\partial\boldsymbol{\varepsilon}^{*\prime}}+\left[\mathbf{I}_N\otimes\frac{\partial\ln f(\boldsymbol{\varepsilon}^*_t;\boldsymbol{\varrho}_\infty)}{\partial\boldsymbol{\varepsilon}^*}\right]\right\}[\boldsymbol{\varepsilon}'^*_t\otimes\mathbf{I}_N]\middle|\boldsymbol{\upsilon}_0\right]
\end{aligned}$$

$$\mathcal{H}_{lr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = E\left[\left.\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho}_\infty)}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'}\right| \boldsymbol{v}_0\right]$$

$$\mathcal{H}_{sr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = E\left[\left.[\boldsymbol{\varepsilon}_t^* \otimes \mathbf{I}_N]\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho}_\infty)}{\partial \boldsymbol{\varepsilon}^* \partial \boldsymbol{\varrho}'}\right| \boldsymbol{v}_0\right]$$

$\mathcal{H}_{ll}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0)$ will be a diagonal matrix of order N with typical element

$$\mathrm{H}_{ll}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = -E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{(\partial \varepsilon_i^*)^2}\right| \boldsymbol{v}_0\right], \tag{C39}$$

$\mathcal{H}_{ls}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = \mathrm{H}_{ls}\mathbf{E}'_N$, H_{ls} is a diagonal matrix of order N with typical element

$$\mathrm{H}_{ls}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = -E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{(\partial \varepsilon_i^*)^2} \cdot \varepsilon_{it}^*\right| \boldsymbol{v}_0\right], \tag{C40}$$

Given (C31),

$$-E\left[\left.\left\{\left[\mathbf{I}_N \otimes \frac{\partial \ln f(\boldsymbol{\varepsilon}_t^*; \boldsymbol{\varrho}_\infty)}{\partial \boldsymbol{\varepsilon}^*}\right]\right\}[\boldsymbol{\varepsilon}_t'^* \otimes \mathbf{I}_N]\right| \boldsymbol{v}_0\right] = \mathbf{K}_{NN},$$

so $\mathcal{H}_{ss}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0)$ will be the sum of the commutation matrix $\mathbf{K}_{NN}$ and a block diagonal matrix $\boldsymbol{\Gamma}$ of order N^2 in which each of the N diagonal blocks is a diagonal matrix of size N with the following structure:

$$\boldsymbol{\Gamma}_i(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = \begin{bmatrix} \mathrm{H}_{ll,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{H}_{ll,i-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{H}_{ss}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{H}_{ll,i+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathrm{H}_{ll,N} \end{bmatrix},$$

where $\mathrm{H}_{ll,i} = \mathrm{H}_{ll}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0)$ to shorten the expressions and

$$\mathrm{H}_{ss}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = -E\left\{\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_\infty)}{(\partial \varepsilon_i^2)}(\varepsilon_{it}^*)^2\right| \boldsymbol{v}_0\right\}. \tag{C41}$$

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

$$\mathrm{H}_{lr}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*; _{i\infty})}{\partial \varepsilon_i^* \partial \boldsymbol{\varrho}_i'}\right| \boldsymbol{v}_0\right], \tag{C42}$$

$\mathcal{H}_{sr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0) = \mathbf{E}_N \mathrm{H}_{sr}$, where H_{sr} another block diagonal matrix of order $N \times q$ with typical block of size $1 \times q_i$

$$\mathrm{H}_{sr}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \varepsilon_i^* \partial \boldsymbol{\varrho}_i'}\varepsilon_i^*\right| \boldsymbol{v}_0\right], \tag{C43}$$

and $\mathcal{H}_{rr}(\boldsymbol{\varrho}_\infty, \boldsymbol{v}_0)$ is a $q \times q$ block diagonal matrix with typical block of size $q_i \times q_i$

$$\mathrm{H}_{rr}(\boldsymbol{\varrho}_{i\infty}, \boldsymbol{v}_0) = -E\left[\left.\frac{\partial^2 \ln f(\varepsilon_{it}^*; \boldsymbol{\varrho}_{i\infty})}{\partial \boldsymbol{\varrho}_i \partial \boldsymbol{\varrho}_i'}\right| \boldsymbol{v}_0\right]. \tag{C44}$$

Additional references

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

Fiorentini, G., and Sentana, E. (2021): "Specification tests for non-Gaussian maximum likelihood estimators", *Quantitative Economics* 12, 683-742.

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

Magnus, J.R. and Sentana, E. (2020): "Zero-diagonality as a linear structure", *Economics Letters* 196, 109513.

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

Phillips, P.C.B. and Durlauf, S.N. (1986): "Multiple time series regression with integrated processes", *Review of Economic Studies* 53, 473–495.

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

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