Supplemental Material for

Normality tests for latent variables

Tincho Almuzara
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<almuzara@cemfi.edu.es>

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

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

October 2018

A Proofs

Lemmata

Lemma 1 *Let* $\mathbf{z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *be an* n_z*-dimensional real Gaussian random vector. Then,*
i) Expectation of second powers:

$$E(\mathbf{z}\mathbf{z}') = \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma},$$

ii) Expectation of third powers:

$$E[\mathbf{z}(\mathbf{z} \odot \mathbf{z})'] = \boldsymbol{\mu}(\boldsymbol{\mu} \odot \boldsymbol{\mu})' + 2(\boldsymbol{\Sigma} \odot \boldsymbol{\ell}_{n_z}\boldsymbol{\mu}') + \boldsymbol{\mu} vecd'(\boldsymbol{\Sigma}),$$

iii) Expectation of fourth powers:

$$\begin{aligned} E\left[(\mathbf{z} \odot \mathbf{z})(\mathbf{z} \odot \mathbf{z})'\right] &= (\boldsymbol{\mu} \odot \boldsymbol{\mu})(\boldsymbol{\mu} \odot \boldsymbol{\mu})' + 2(\boldsymbol{\Sigma} \odot \boldsymbol{\Sigma}) + vecd(\boldsymbol{\Sigma})vecd'(\boldsymbol{\Sigma}) \\ &\quad + 4(\boldsymbol{\Sigma} \odot \boldsymbol{\mu}\boldsymbol{\mu}') + vecd(\boldsymbol{\mu}\boldsymbol{\mu}')vecd'(\boldsymbol{\Sigma}) + vecd(\boldsymbol{\Sigma})vecd'(\boldsymbol{\mu}\boldsymbol{\mu}'), \end{aligned}$$

where $\odot$ *denotes the Hadamard (or elementwise) product,* $vecd(\cdot)$ *is the operator which stacks the diagonal elements of a square matrix in vector form and* $\boldsymbol{\ell}_{n_z}$ *is a vector of* n_z *ones.*

Proof. The proof is tedious but straightforward. □

Lemma 2 *Define* $\mathbf{m}_h : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_1 \times n_2}$ *for* $n_1, n_2 \in \mathbb{Z}_{++}$ *and* $h \in \{2, 3, 4\}$ *as*

$$\begin{aligned} \mathbf{m}_2(\mathbf{w}_1, \mathbf{w}_2) &= vec(\mathbf{w}_1\mathbf{w}_2'), \\ \mathbf{m}_3(\mathbf{w}_1, \mathbf{w}_2) &= vec[\mathbf{w}_1(\mathbf{w}_2 \odot \mathbf{w}_2)'], \\ \mathbf{m}_4(\mathbf{w}_1, \mathbf{w}_2) &= vec[(\mathbf{w}_1 \odot \mathbf{w}_1)(\mathbf{w}_2 \odot \mathbf{w}_2)'], \end{aligned}$$

where $\mathbf{w}_1 \in \mathbb{R}^{n_1}$, $\mathbf{w}_2 \in \mathbb{R}^{n_2}$, *and* $vec(\cdot)$ *is the vectorization (by columns) operator. Consider the real Gaussian random vector*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{xy}' & \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{xz}' & \boldsymbol{\Sigma}_{yz}' & \boldsymbol{\Sigma}_{zz} \end{pmatrix}\right]$$

where $\mathbf{x}$ *is* n_x*-dimensional,* $\mathbf{y}$ *is* n_y*-dimensional, and* $\mathbf{z}$ *is* n_z*-dimensional. Then,*
i) Covariance with the first power:

$$\begin{aligned} cov\left[\mathbf{x}, \mathbf{m}_2(\mathbf{y}, \mathbf{z})\right] &= \mathbf{0}, \\ cov\left[\mathbf{x}, \mathbf{m}_3(\mathbf{y}, \mathbf{z})\right] &= 2\left[\boldsymbol{\ell}_{n_x} \otimes vec'(\boldsymbol{\Sigma}_{yz})\right] \odot (\boldsymbol{\Sigma}_{xz} \otimes \boldsymbol{\ell}_{n_y}') \\ &\quad + \left[vecd'(\boldsymbol{\Sigma}_{zz}) \otimes \mathbf{1}_{n_x \times n_y}\right] \odot \left(\boldsymbol{\ell}_{n_z}' \otimes \boldsymbol{\Sigma}_{xy}\right), \\ cov\left[\mathbf{x}, \mathbf{m}_4(\mathbf{y}, \mathbf{z})\right] &= \mathbf{0}, \end{aligned}$$

ii) Covariance with the second power:

$$\begin{aligned} cov\left[\mathbf{m}_2(\mathbf{x}, \mathbf{x}), \mathbf{m}_2(\mathbf{y}, \mathbf{z})\right] &= (\mathbf{1}_{n_x \times n_z} \otimes \boldsymbol{\Sigma}_{xy}) \odot \left(\boldsymbol{\Sigma}_{xz} \otimes \mathbf{1}_{n_x \times n_y}\right) \\ &\quad + (\boldsymbol{\ell}_{n_x} \otimes \boldsymbol{\Sigma}_{xz} \otimes \boldsymbol{\ell}_{n_y}') \odot \left(\boldsymbol{\ell}_{n_x}' \otimes \boldsymbol{\Sigma}_{xy} \otimes \boldsymbol{\ell}_{n_z}\right), \\ cov\left[\mathbf{m}_2(\mathbf{x}, \mathbf{x}), \mathbf{m}_3(\mathbf{y}, \mathbf{z})\right] &= \mathbf{0}, \end{aligned}$$

$$cov\left[\mathbf{m}_2(\mathbf{x},\mathbf{x}),\mathbf{m}_4(\mathbf{y},\mathbf{z})\right] = 4\left[\boldsymbol{\ell}_{n_x^2}\otimes vec'(\boldsymbol{\Sigma}_{yz})\right]\odot cov\left[\mathbf{m}_2(\mathbf{x},\mathbf{x}),\mathbf{m}_2(\mathbf{y},\mathbf{z})\right]$$
$$+2\left[\boldsymbol{\ell}_{n_x^2}\otimes\boldsymbol{\ell}'_{n_z}\otimes vecd'(\boldsymbol{\Sigma}_{yy})\right]\odot(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y})\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)$$
$$+2[\boldsymbol{\ell}_{n_x^2}\otimes vecd'(\boldsymbol{\Sigma}_{zz})\otimes\boldsymbol{\ell}'_{n_y}]\odot(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy})\odot(\boldsymbol{\ell}'_{n_y}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}),$$

iii) Covariance with the third power:

$$cov\left[\mathbf{m}_3(\mathbf{x},\mathbf{x}),\mathbf{m}_3(\mathbf{y},\mathbf{z})\right] = [vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\{\boldsymbol{\ell}_{n_x}\otimes cov\left[\mathbf{x},\mathbf{m}_3(\mathbf{y},\mathbf{z})\right]\}$$
$$+2\left(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy}\right)\odot\left[(\boldsymbol{\Sigma}_{xz}\odot\boldsymbol{\Sigma}_{xz})\otimes\mathbf{1}_{n_x\times n_y}\right]$$
$$+2[vec(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot[\boldsymbol{\ell}_{n_x^2}\otimes vecd'(\boldsymbol{\Sigma}_{zz})\otimes\boldsymbol{\ell}'_{n_y}]\odot(\boldsymbol{\ell}'_{n_y}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x})$$
$$+4[vec(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\left[\boldsymbol{\ell}_{n_x^2}\otimes vec'(\boldsymbol{\Sigma}_{yz})\right]\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)$$
$$+4(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y})\odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right),$$
$$cov\left[\mathbf{m}_3(\mathbf{x},\mathbf{x}),\mathbf{m}_4(\mathbf{y},\mathbf{z})\right] = \mathbf{0},$$

iv) Covariance with the fourth power:

$$cov\left[\mathbf{m}_4(\mathbf{x},\mathbf{x}),\mathbf{m}_4(\mathbf{y},\mathbf{z})\right] = 4cov\left[\mathbf{m}_2(\mathbf{x},\mathbf{x}),\mathbf{m}_2(\mathbf{y},\mathbf{z})\right]\odot cov\left[\mathbf{m}_2(\mathbf{x},\mathbf{x}),\mathbf{m}_2(\mathbf{y},\mathbf{z})\right]$$
$$+4[vec(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot cov\left[\mathbf{m}_2(\mathbf{x},\mathbf{x}),\mathbf{m}_4(\mathbf{y},\mathbf{z})\right]$$
$$+2[\boldsymbol{\ell}_{n_x}\otimes vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\left[\boldsymbol{\ell}_{n_x^2}\otimes\boldsymbol{\ell}_{n_z}\otimes vecd'(\boldsymbol{\Sigma}_{yy})\right]\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)$$
$$+2[\boldsymbol{\ell}_{n_x}\otimes vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\left[\boldsymbol{\ell}_{n_x^2}\otimes vecd'(\boldsymbol{\Sigma}_{zz}\otimes\boldsymbol{\ell}_{n_y})\right]\odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)$$
$$+2[vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\left[\boldsymbol{\ell}_{n_x^2}\otimes\boldsymbol{\ell}_{n_z}\otimes vecd'(\boldsymbol{\Sigma}_{yy})\right]\odot(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y})\odot(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y})$$
$$+2[vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\left[\boldsymbol{\ell}_{n_x^2}\otimes vecd'(\boldsymbol{\Sigma}_{zz}\otimes\boldsymbol{\ell}_{n_y})\right]\odot(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy})\odot(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy})$$
$$+8[\boldsymbol{\ell}_{n_x}\otimes vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\left[\boldsymbol{\ell}_{n_x^2}\otimes vec'(\boldsymbol{\Sigma}_{yz})\right]\odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\odot\left(\boldsymbol{\Sigma}_{xz}\otimes\mathbf{1}_{n_x\times n_y}\right)$$
$$+8[vecd(\boldsymbol{\Sigma}_{xx})\otimes\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\ell}'_{n_yn_z}]\odot\left[\boldsymbol{\ell}_{n_x^2}\otimes vec'(\boldsymbol{\Sigma}_{yz})\right]\odot(\mathbf{1}_{n_x\times n_z}\otimes\boldsymbol{\Sigma}_{xy})\odot(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y})$$
$$+8\left(\boldsymbol{\Sigma}_{xy}\otimes\mathbf{1}_{n_x\times n_z}\right)\odot\left(\mathbf{1}_{n_x\times n_y}\otimes\boldsymbol{\Sigma}_{xz}\right)\odot\left(\boldsymbol{\ell}'_{n_z}\otimes\boldsymbol{\Sigma}_{xy}\otimes\boldsymbol{\ell}_{n_x}\right)\odot(\boldsymbol{\ell}_{n_x}\otimes\boldsymbol{\Sigma}_{xz}\otimes\boldsymbol{\ell}'_{n_y}),$$

where $\otimes$ *denotes Kronecker product and* $\mathbf{1}_{n_1\times n_2}$ *denotes a matrix of ones of dimension* $n_1\times n_2$*.*

Proof. Again, the proof is tedious but straightforward. □

Lemma 3 *Consider the model (1)-(2) where* $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_t^{\mathrm{GH}\prime},\boldsymbol{\varepsilon}_t^{\mathrm{N}\prime})'$*, with* $\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\sim GH_R(\eta,\psi,\boldsymbol{\beta})$ *and* $\boldsymbol{\varepsilon}_t^{\mathrm{N}}\sim N(\mathbf{0};\mathbf{I}_{K-R})$*. Let* $\varsigma_t^{\mathrm{GH}} = \boldsymbol{\varepsilon}_t^{\mathrm{GH}\prime}\boldsymbol{\varepsilon}_t^{\mathrm{GH}}$ *and*

$$s_{\mathrm{k}t} = c_0 + c_1\varsigma_t^{\mathrm{GH}} + c_2(\varsigma_t^{\mathrm{GH}})^2,$$
$$\mathbf{s}_{\mathrm{s}t} = \boldsymbol{\varepsilon}_t^{\mathrm{GH}}(c_3 + \varsigma_t^{\mathrm{GH}}),$$
$$s_{\mathrm{GH}t} = s_{\mathrm{k}t} + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s}t},$$

where $c_0 = R(R+2)/4$*,* $c_1 = -(R+2)/2$*,* $c_2 = 1/4$*, and* $c_3 = -(R+2)$*. Then,*
i) For any $\boldsymbol{\beta}\in\mathbb{R}^R$ *and* $\psi > 0$*,*

$$\lim_{\eta\to 0^+}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\mathbf{E}_T|\boldsymbol{\phi})}{\partial\eta} = -\lim_{\eta\to 0^-}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\mathbf{E}_T|\boldsymbol{\phi})}{\partial\eta} = \frac{1}{T}\sum_{t=1}^{T}s_{\mathrm{GH}t},\ \text{and}$$
$$\lim_{\eta\to 0^\pm}\frac{1}{T}\frac{\partial\ln f(\mathbf{Y}_T,\mathbf{E}_T|\boldsymbol{\phi})}{\partial\psi} = 0.$$

ii) For any $\boldsymbol{\beta} \in \mathbb{R}^R$ *and* $\eta \in \mathbb{R}$,

$$\lim_{\psi \to 0^+} \frac{1}{T} \frac{\partial \ln f(\mathbf{Y}_T, \mathbf{E}_T | \boldsymbol{\phi})}{\partial \eta} = 0, \; \textit{and} \; \lim_{\psi \to 0^+} \frac{2}{T} \frac{\partial \ln f(\mathbf{Y}_T, \mathbf{E}_T | \boldsymbol{\phi})}{\partial \psi} = \frac{1}{T} \sum_{t=1}^{T} s_{\mathrm{GH}t}.$$

Proof. See Mencía and Sentana (2012). □

Lemma 4 *Consider the model (1)-(2) where* $\{\boldsymbol{\varepsilon}_t\}_{t=-\infty}^{\infty}$ *is white noise with identity covariance matrix. Further, assume that all the eigenvalues of* $\mathbf{F}$ *are inside the unit circle. If we observe the double-infinite sequence* $\mathbf{Y}_\infty = \{\mathbf{y}_t\}_{t=-\infty}^{\infty}$, *then the linear projection*

$$\begin{pmatrix} \hat{\boldsymbol{\xi}}_{t-1|\infty} \\ \hat{\boldsymbol{\varepsilon}}_{t|\infty} \end{pmatrix} = \mathcal{P}\left[\left(\begin{array}{c} \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\varepsilon}_t \end{array} \right) \middle| \mathbf{Y}_\infty \right] = \left[\begin{array}{c} \boldsymbol{\Psi}(L) \\ \boldsymbol{\Upsilon}(L) \end{array} \right] \mathbf{y}_t,$$

where $\boldsymbol{\Psi}$ *and* $\boldsymbol{\Upsilon}$ *are absolutely summable two-sided filters in the lag operator* L, *will be given by*

$$\begin{bmatrix} \boldsymbol{\Psi}(z) \\ \boldsymbol{\Upsilon}(z) \end{bmatrix} = \begin{bmatrix} z\mathbf{F}^{-1}(z)\mathbf{M} \\ \mathbf{I}_K \end{bmatrix} \mathbf{D}'(z^{-1}) \left[\mathbf{D}(z)\mathbf{D}'(z^{-1}) \right]^{-1},$$

where

$$\mathbf{F}^{-1}(L) = (\mathbf{I}_M - \mathbf{F}L)^{-1} = \sum_{j=0}^{\infty} \mathbf{F}^j L^j \; \textit{and} \; \mathbf{D}(L) = \mathbf{H}\mathbf{F}^{-1}(L)\mathbf{M} = \sum_{j=0}^{\infty} \mathbf{D}_j L^j$$

with $\mathbf{D}_j = \mathbf{H}\mathbf{F}^j\mathbf{M}$ *for all* j.

Proof. Given that $\mathbf{y}_t = \mathbf{D}(L)\boldsymbol{\varepsilon}_t$, the joint autocovariance generating function for $(\mathbf{y}'_t, \boldsymbol{\varepsilon}'_t)'$ is easily seen to be

$$\mathbf{G}(z) = \begin{bmatrix} \mathbf{G}_{yy}(z) & \mathbf{G}_{y\varepsilon}(z) \\ \mathbf{G}_{\varepsilon y}(z) & \mathbf{G}_{\varepsilon\varepsilon}(z) \end{bmatrix} = \begin{bmatrix} \mathbf{D}(z)\mathbf{D}'(z^{-1}) & \mathbf{D}(z) \\ \mathbf{D}'(z^{-1}) & \mathbf{I}_K \end{bmatrix}$$

for any $z \in \mathbb{C}$. The Wiener-Kolmogorov filter for $\boldsymbol{\varepsilon}_t$ is given by

$$\hat{\boldsymbol{\varepsilon}}_{t|\infty} = \mathbf{G}_{\varepsilon y}(L)\mathbf{G}_{yy}^{-1}(L)\mathbf{y}_t = \mathbf{D}'(L^{-1}) \left[\mathbf{D}(L)\mathbf{D}'(L^{-1}) \right]^{-1} \mathbf{y}_t$$

It is then easily checked that for every t, $\hat{\boldsymbol{\varepsilon}}_{t|\infty}$ is well-defined as a mean-square limit under the assumptions of the Lemma. Moreover, because

$$\boldsymbol{\xi}_{t-1} = L\mathbf{F}^{-1}(L)\mathbf{M}\boldsymbol{\varepsilon}_t,$$

the filter for $\boldsymbol{\xi}_{t-1}$ follows from the filter for $\boldsymbol{\varepsilon}_t$, so it is also well-defined. □

Lemma 5 *Consider the model (1)-(2). The score of the asymmetric GH with respect to the parameter* τ *when* $\tau = 0$ *for fixed values of the skewness parameters* $\boldsymbol{\beta}$ *is given by*

$$\begin{aligned} \bar{s}_{\mathrm{GH}T}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= \frac{1}{T} \sum_{t=1}^{T} \left[s_{\mathrm{k}t|T}(\boldsymbol{\theta}) + \boldsymbol{\beta}' \mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta}) \right], \\ s_{\mathrm{k}t|T}(\boldsymbol{\theta}) &= \mathbf{b}'_{\mathrm{k}t|T}(\boldsymbol{\theta}) \mathbf{m}_{\mathrm{k}t|T}(\boldsymbol{\theta}), \\ \mathbf{s}_{\mathrm{s}t|T}(\boldsymbol{\theta}) &= \mathbf{b}'_{\mathrm{s}t|T}(\boldsymbol{\theta}) \mathbf{m}_{\mathrm{s}t|T}(\boldsymbol{\theta}), \end{aligned}$$

where

$$\mathbf{m}_{\mathrm{k}t|T}(\boldsymbol{\theta}) = \begin{pmatrix} 1 \\ \mathbf{m}_{2t|T}(\boldsymbol{\theta}) \\ \mathbf{m}_{4t|T}(\boldsymbol{\theta}) \end{pmatrix}, \quad \mathbf{b}_{\mathrm{k}t|T}(\boldsymbol{\theta}) = \begin{pmatrix} b_{0t|T}(\boldsymbol{\theta}) \\ \mathbf{b}_{2t|T}(\boldsymbol{\theta}) \\ \mathbf{b}_{4t|T}(\boldsymbol{\theta}) \end{pmatrix},$$

$$\mathbf{m}_{\mathrm{s}t|T}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{m}_{1t|T}(\boldsymbol{\theta}) \\ \mathbf{m}_{3t|T}(\boldsymbol{\theta}) \end{pmatrix}, \quad \mathbf{b}_{\mathrm{s}t|T}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{b}_{1t|T}(\boldsymbol{\theta}) \\ \mathbf{b}_{3t|T}(\boldsymbol{\theta}) \end{pmatrix},$$

$$\begin{aligned}
b_{0t|T}(\boldsymbol{\theta}) &= c_0 + \{c_1 + c_2 tr[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\} tr[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})] + 2c_2 tr\{[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]^2\}, \\
\mathbf{b}_{1t|T}(\boldsymbol{\theta}) &= [c_3 + tr(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\mathbf{S}'_{RK} + 2\mathbf{S}'_{RK}\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}), \\
\mathbf{b}_{2t|T}(\boldsymbol{\theta}) &= \{c_1 + 2c_2 tr[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\} \left(\mathbf{S}'_{RK} \otimes \mathbf{S}'_{RK}\right) vec(\mathbf{I}_R) + 4c_2 \left(\mathbf{S}'_{RK} \otimes \mathbf{S}'_{RK}\right) vec[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})], \\
\mathbf{b}_{3t|T}(\boldsymbol{\theta}) &= \mathbf{S}'_{RK}\boldsymbol{\ell}_R \otimes \mathbf{S}'_{RK}, \\
\mathbf{b}_{4t|T}(\boldsymbol{\theta}) &= c_2 \left(\mathbf{S}'_{RK} \otimes \mathbf{S}'_{RK}\right) \boldsymbol{\ell}_{R^2},
\end{aligned}$$

with $c_0 = R(R+2)/4$, $c_1 = -(R+2)/2$, $c_2 = 1/4$, $c_3 = -(R+2)$ *and* $\boldsymbol{\ell}_H$ *a vector of* H *ones.*

Proof. From Lemma 3, we can obtain the expression for the score with respect to τ for a fixed value of the skewness parameter vector $\boldsymbol{\beta}$, $s_{\mathrm{GH}t} = s_{\mathrm{k}t} + \boldsymbol{\beta}'\mathbf{s}_{\mathrm{s}t}$, which corresponds to the M-step of the EM algorithm. Next, we can apply the E-step to each of the components separately.

As for $s_{\mathrm{k}t}$, we have that $\boldsymbol{\varepsilon}_t|\,\mathbf{Y}_T, \boldsymbol{\theta} \sim N[\boldsymbol{\varepsilon}_{t|T}(\boldsymbol{\theta}), \boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})]$ under the null of normality, so that

$$s_{\mathrm{k}t|T}(\boldsymbol{\theta}) = c_0 + c_1 E\left[\varsigma^{\mathrm{GH}}_t\middle|\,\mathbf{Y}_T, \boldsymbol{\theta}\right] + c_2 E\left[(\varsigma^{\mathrm{GH}}_t)^2\middle|\,\mathbf{Y}_T, \boldsymbol{\theta}\right]$$

involves the computation of $E\left[\varsigma_t|\,\mathbf{Y}_T, \boldsymbol{\theta}\right]$ and $E\left[\varsigma^2_t\middle|\,\mathbf{Y}_T, \boldsymbol{\theta}\right]$. To compute the first expectation, we can write

$$\begin{aligned}
E\left[\varsigma^{\mathrm{GH}}_t\middle|\,\mathbf{Y}_T\right] &= E\left[\boldsymbol{\varepsilon}^{\mathrm{GH}\prime}_t\boldsymbol{\varepsilon}^{\mathrm{GH}}_t\middle|\,\mathbf{Y}_T, \boldsymbol{\theta}\right] \\
&= tr\left\{E\left[\boldsymbol{\varepsilon}^{\mathrm{GH}}_t\boldsymbol{\varepsilon}^{\mathrm{GH}\prime}_t\,|\mathbf{Y}_T, \boldsymbol{\theta}\right]\right\} \\
&= tr[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})] + vec(\mathbf{I}_R)'vec[\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})'],
\end{aligned}$$

where the first equality follows from the fact that $tr(A'B) = tr(BA')$, and the second one from Lemma 1.i. As for the second expectation,

$$\begin{aligned}
E\left[(\varsigma^{\mathrm{GH}}_t)^2\,|\mathbf{Y}_T, \boldsymbol{\theta}\right] &= E\{\left[\boldsymbol{\varepsilon}^{\mathrm{GH}}_t \odot \boldsymbol{\varepsilon}^{\mathrm{GH}}_t\right]' \mathbf{1}_{R\times R}\left[\boldsymbol{\varepsilon}^{\mathrm{GH}}_t \odot \boldsymbol{\varepsilon}^{\mathrm{GH}}_t\right]|\mathbf{Y}_T, \boldsymbol{\theta}\} \\
&= tr[\mathbf{1}_{R\times R}E\{\left[\boldsymbol{\varepsilon}^{\mathrm{GH}}_t \odot \boldsymbol{\varepsilon}^{\mathrm{GH}}_t\right]\left[\boldsymbol{\varepsilon}^{\mathrm{GH}}_t \odot \boldsymbol{\varepsilon}^{\mathrm{GH}}_t\right]'|\mathbf{Y}_T, \boldsymbol{\theta}\}] \\
&= 2\boldsymbol{\ell}'_{R^2}vec[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})] \\
&+ \boldsymbol{\ell}'_{R^2}vec\{vecd[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]vecd[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]'\} \\
&+ 4\boldsymbol{\ell}'_{R^2}vec[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})'] \\
&+ \boldsymbol{\ell}'_{R^2}vec\{vecd[(\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]vecd'[\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})']\} \\
&+ \boldsymbol{\ell}'_{R^2}vec\{vecd[\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})']vecd'[\boldsymbol{\Omega}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]\} \\
&+ \boldsymbol{\ell}'_{R^2}vec\{[\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})][\boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}^{\mathrm{GH}}_{t|T}(\boldsymbol{\theta})]'\},
\end{aligned}$$

where the first equality is a rewriting of $(\varsigma_t^{\mathrm{GH}})^2$, the second one follows from the aforementioned property of the trace, and the third one from Lemma 1.iii. Finally, to obtain the expression for $s_{\mathrm{k},t|T}(\boldsymbol{\theta})$, we have made use of the following identities:

$$\begin{aligned}
\boldsymbol{\ell}'_{R^2}vec[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\odot\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})] &= vec'[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]vec[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]\\
&= tr[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})] = tr\{[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]^2\}\\
\boldsymbol{\ell}'_{R^2}vec\{vecd[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]vecd'[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]\} &= tr^2[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]\\
\boldsymbol{\ell}'_{R^2}vec\{\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\odot[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})']\} &= vec'[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]vec[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})']\\
\boldsymbol{\ell}'_{R^2}vec\{vecd[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]vecd'[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})']\} &= tr[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]vec'(\mathbf{I}_R)vec[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})'],
\end{aligned}$$

together with

$$\begin{aligned}
vec[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})'] &= (\mathbf{S}_{RK}\otimes\mathbf{S}_{RK})\,\mathbf{m}_{2,t|T}(\boldsymbol{\theta})\\
vec\{[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\odot\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})][\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\odot\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]'\} &= (\mathbf{S}_{RK}\otimes\mathbf{S}_{RK})\,\mathbf{m}_{4,t|T}(\boldsymbol{\theta}).
\end{aligned}$$

Similarly, in order to compute

$$\mathbf{s}_{\mathrm{st}|T}(\boldsymbol{\theta}) = c_3E\left[\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\,|\mathbf{Y}_T,\boldsymbol{\theta}\right] + E\left[\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\varsigma_t^{\mathrm{GH}}\,|\mathbf{Y}_T,\boldsymbol{\theta}\right],$$

we need the expectation of the first component, which is trivially $E\left[\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\,|\mathbf{Y}_T,\boldsymbol{\theta}\right] = \boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})$. We also need

$$\begin{aligned}
E\left[\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\varsigma_t\,|\mathbf{Y}_T,\boldsymbol{\theta}\right] &= E[\boldsymbol{\varepsilon}_t^{\mathrm{GH}}(\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\odot\boldsymbol{\varepsilon}_t^{\mathrm{GH}})'\,|\mathbf{Y}_T,\boldsymbol{\theta}]\boldsymbol{\ell}_R\\
&= 2\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta}) + tr[\boldsymbol{\Omega}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\\
&+ \boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\odot\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]'\boldsymbol{\ell}_R,
\end{aligned}$$

where we have used the fact that $\varsigma_t^{\mathrm{GH}} = \left[\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\odot\boldsymbol{\varepsilon}_t^{\mathrm{GH}}\right]'\boldsymbol{\ell}_R$ in the first equality, and applied Lemma 1.ii. in the last one. Finally, we obtain the desired result by exploiting the fact that

$$vec\{\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})\odot\boldsymbol{\varepsilon}_{t|T}^{\mathrm{GH}}(\boldsymbol{\theta})]'\} = (\mathbf{S}_{RK}\otimes\mathbf{S}_{RK})\,\mathbf{m}_{3,t|T}(\boldsymbol{\theta}),$$

after re-arranging terms. □

Lemma 6 *Let*

$$\boldsymbol{\kappa}_i(\boldsymbol{\theta}) = \sum_{j=-\infty}^{\infty} cov[\mathbf{m}_{it}(\boldsymbol{\theta}),\mathbf{m}_{it-j}(\boldsymbol{\theta})],$$

denote the autocovariance generating function of $\mathbf{m}_{it}(\boldsymbol{\theta})$ *evaluated at one. Then,*

i) The asymptotic variance of $\bar{s}_{\mathrm{k}T}(\boldsymbol{\theta})$ *is given by*

$$\mathcal{C}_{\mathrm{k}}(\boldsymbol{\theta}) = \mathbf{b}'_4(\boldsymbol{\theta})\boldsymbol{\kappa}_4(\boldsymbol{\theta})\mathbf{b}'_4(\boldsymbol{\theta}) - \mathbf{b}'_2(\boldsymbol{\theta})\boldsymbol{\kappa}_2(\boldsymbol{\theta})\mathbf{b}'_2(\boldsymbol{\theta}).$$

ii) The asymptotic variance of $\bar{s}_{\mathrm{s}T}(\boldsymbol{\theta})$ *is given by*

$$\mathcal{C}_{\mathrm{s}|\infty}(\boldsymbol{\theta}) = \mathbf{b}'_3(\boldsymbol{\theta})\boldsymbol{\kappa}_3(\boldsymbol{\theta})\mathbf{b}'_3(\boldsymbol{\theta}) - \mathbf{b}'_1(\boldsymbol{\theta})\boldsymbol{\kappa}_1(\boldsymbol{\theta})\mathbf{b}'_1(\boldsymbol{\theta}),$$

iii) $\sqrt{T}\bar{s}_{\mathrm{k}T}(\boldsymbol{\theta})$ *and* $\sqrt{T}\bar{\mathbf{s}}_{\mathrm{s}T}(\boldsymbol{\theta})$ *are asymptotically independent.*

Proof. Following the same steps as in Lemma 5, but conditioning on $\mathbf{Y}_\infty$ instead of $\mathbf{Y}_T$, we can obtain $\mathbf{s}_{kt|\infty}(\boldsymbol{\theta}) = E[\mathbf{s}_{kt}(\boldsymbol{\theta})|\mathbf{Y}_\infty, \boldsymbol{\theta}]$ and $\mathbf{s}_{st|\infty}(\boldsymbol{\theta}) = E[\mathbf{s}_{st}(\boldsymbol{\theta})|\mathbf{Y}_\infty, \boldsymbol{\theta}]$. Specifically, we can write

$$\begin{bmatrix} s_{kt|\infty}(\boldsymbol{\theta}) - b_0(\boldsymbol{\theta}) \\ \mathbf{s}_{st|\infty}(\boldsymbol{\theta}) \end{bmatrix} = \mathbf{B}'(\boldsymbol{\theta})\mathbf{m}_t(\boldsymbol{\theta}) \text{ where } \mathbf{B}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0} & \mathbf{b}_1(\boldsymbol{\theta}) \\ \mathbf{b}_2(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_3(\boldsymbol{\theta}) \\ \mathbf{b}_4(\boldsymbol{\theta}) & \mathbf{0} \end{bmatrix},$$

and $\mathbf{m}_t(\boldsymbol{\theta}) = [\mathbf{m}_{1t}(\boldsymbol{\theta}), \mathbf{m}_{2t|}(\boldsymbol{\theta}), \mathbf{m}_{3t}(\boldsymbol{\theta}), \mathbf{m}_{4t}(\boldsymbol{\theta})]'$, where

$$\begin{aligned}
b_0(\boldsymbol{\theta}) &= c_0 + \{c_1 + tr[\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})]c_2\}tr[\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})] + 2c_2 tr\{[\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})]^2\},\\
\mathbf{b}_1(\boldsymbol{\theta}) &= \{c_3 + tr[\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})]\}\mathbf{S}'_{RK} + 2\mathbf{S}'_{RK}\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta}),\\
\mathbf{b}_2(\boldsymbol{\theta}) &= \{c_1 + 2tr[\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})]c_2\}(\mathbf{S}'_{RK} \otimes \mathbf{S}'_{RK})vec(\mathbf{I}_R) + 4c_2(\mathbf{S}'_{RK} \otimes \mathbf{S}'_{RK})vec[\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta})],\\
\mathbf{b}_3(\boldsymbol{\theta}) &= \mathbf{S}'_{RK}\boldsymbol{\ell}_R \otimes \mathbf{S}'_{RK},\\
\mathbf{b}_4(\boldsymbol{\theta}) &= c_2\left[\mathbf{S}'_{RK} \otimes \mathbf{S}'_{RK}\right]\boldsymbol{\ell}_{R^2},
\end{aligned}$$

with $\boldsymbol{\Omega}_\infty^{\text{GH}}(\boldsymbol{\theta}) = \mathbf{S}_{RK}\boldsymbol{\Omega}_\infty(\boldsymbol{\theta})\mathbf{S}'_{RK}$ and

$$\begin{aligned}
\mathbf{m}_{1t}(\boldsymbol{\theta}) &= \boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta}),\\
\mathbf{m}_{2t}(\boldsymbol{\theta}) &= vec[\boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta})'],\\
\mathbf{m}_{3t}(\boldsymbol{\theta}) &= vec\{\boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta})]'\},\\
\mathbf{m}_{4t}(\boldsymbol{\theta}) &= vec\{[\boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta})][\boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta}) \odot \boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta})]'\}.
\end{aligned}$$

Next, we can use Lemma 4 to obtain $\boldsymbol{\Gamma}_j = E[\boldsymbol{\varepsilon}_{t|\infty}^{\text{GH}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_{t-j|\infty}^{\text{GH}}(\boldsymbol{\theta})']$, which corresponds to the j^{th} order autocovariance matrix of the Wiener-Kolmogorov filter for $\boldsymbol{\varepsilon}_t$ based on $\mathbf{Y}_\infty$ for any integer j. Further, we can apply Lemma 2 to obtain:

i) Covariance matrices with the first power:

$$cov[\mathbf{m}_{1t}(\boldsymbol{\theta}), \mathbf{m}_{2t-j}(\boldsymbol{\theta})] = \mathbf{0}, \tag{A1}$$

$$\begin{aligned}
cov[\mathbf{m}_{1t}(\boldsymbol{\theta}), \mathbf{m}_{3t-j}(\boldsymbol{\theta})] &= 2\left[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)\right] \odot \left(\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right)\\
&\quad + \left[vecd'(\boldsymbol{\Gamma}_0) \otimes \mathbf{1}_{K\times K}\right] \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j\right),
\end{aligned} \tag{A2}$$

$$cov[\mathbf{m}_{1t}(\boldsymbol{\theta}), \mathbf{m}_{4t-j}(\boldsymbol{\theta})] = \mathbf{0}, \tag{A3}$$

ii) Covariance matrices with the second power:

$$\begin{aligned}
cov[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}_{2t-j}(\boldsymbol{\theta})] &= (\mathbf{1}_{K\times K} \otimes \boldsymbol{\Gamma}_j) \odot (\boldsymbol{\Gamma}_j \otimes \mathbf{1}_{K\times K})\\
&\quad + \left(\boldsymbol{\ell}_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right) \odot \left(\boldsymbol{\ell}'_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K\right),\\
cov[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}_{3t-j}(\boldsymbol{\theta})] &= \mathbf{0},
\end{aligned} \tag{A4}$$

$$\begin{aligned}
cov[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}_{4t-j}(\boldsymbol{\theta})] &= 4\left[\boldsymbol{\ell}_{K^2} \otimes vec'(\boldsymbol{\Gamma}_0)\right] \odot cov[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}_{2t-j}(\boldsymbol{\theta})]\\
&\quad +2\left[\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}'_K \otimes vecd'(\boldsymbol{\Gamma}_0)\right] \odot \left(\boldsymbol{\ell}_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right) \odot (\boldsymbol{\Gamma}_j \otimes \mathbf{1}_{K\times K})\\
&\quad +2\left[\boldsymbol{\ell}_{K^2} \otimes vecd'(\boldsymbol{\Gamma}_0) \otimes \boldsymbol{\ell}'_K\right] \odot (\mathbf{1}_{K\times K} \otimes \boldsymbol{\Gamma}_j) \odot \left(\boldsymbol{\ell}'_K \otimes \boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}_K\right),
\end{aligned} \tag{A5}$$

iii) Covariance matrices with the third power:

$$
\begin{aligned}
cov\left[\mathbf{m}_{3t}(\boldsymbol{\theta}),\mathbf{m}_{3t-j}(\boldsymbol{\theta})\right] = {} & [vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}_K\otimes\boldsymbol{\ell}'_{K^2}]\odot\{\boldsymbol{\ell}_K\otimes cov\left[\mathbf{m}_{1t}(\boldsymbol{\theta}),\mathbf{m}_{3t-j}(\boldsymbol{\theta})\right]\}\\
& +2\left(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j\right)\odot[(\boldsymbol{\Gamma}_j\odot\boldsymbol{\Gamma}_j)\otimes\mathbf{1}_{K\times K}]\\
& +2[vec(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot[\boldsymbol{\ell}_{K^2}\otimes vecd'(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_K]\odot(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K)\\
& +4[vec(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot[\boldsymbol{\ell}_{K^2}\otimes vec'(\boldsymbol{\Gamma}_0)]\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K})\\
& +4(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K)\odot(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K)\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K}),
\end{aligned}
$$

$$
cov\left[\mathbf{m}_{3t}(\boldsymbol{\theta}),\mathbf{m}_{4t-j}(\boldsymbol{\theta})\right]=\mathbf{0}, \tag{A6}
$$

iv) Covariance matrix of the fourth power:

$$
\begin{aligned}
cov\left[\mathbf{m}_{4t}(\boldsymbol{\theta}),\mathbf{m}_{4t-j}(\boldsymbol{\theta})\right] = {} & 4cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}),\mathbf{m}_{2t-j}(\boldsymbol{\theta})\right]\odot cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}),\mathbf{m}_{2t-j}(\boldsymbol{\theta})\right]\\
& +4[vec(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}),\mathbf{m}_{4t-j}(\boldsymbol{\theta})\right]\\
& +2[\boldsymbol{\ell}_K\otimes vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes\boldsymbol{\ell}_K\otimes vecd'(\boldsymbol{\Gamma}_0)\right]\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K})\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K})\\
& +2[\boldsymbol{\ell}_K\otimes vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vecd'(\boldsymbol{\Gamma}_0\otimes\boldsymbol{\ell}_K)\right]\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right)\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right)\\
& +2[vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}_K\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes\boldsymbol{\ell}_K\otimes vecd'(\boldsymbol{\Gamma}_0)\right]\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right)\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right)\\
& +2[vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}_K\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vecd'(\boldsymbol{\Gamma}_0\otimes\boldsymbol{\ell}_K)\right]\odot(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j)\odot(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j)\\
& +8[\boldsymbol{\ell}_K\otimes vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vec'(\boldsymbol{\Gamma}_0)\right]\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right)\odot(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K})\\
& +8[vecd(\boldsymbol{\Gamma}_0)\otimes\boldsymbol{\ell}_K\otimes\boldsymbol{\ell}'_{K^2}]\odot\left[\boldsymbol{\ell}_{K^2}\otimes vec'(\boldsymbol{\Gamma}_0)\right]\odot(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j)\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right)\\
& +8\left(\boldsymbol{\Gamma}_j\otimes\mathbf{1}_{K\times K}\right)\odot\left(\mathbf{1}_{K\times K}\otimes\boldsymbol{\Gamma}_j\right)\odot\left(\boldsymbol{\ell}'_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}_K\right)\odot\left(\boldsymbol{\ell}_K\otimes\boldsymbol{\Gamma}_j\otimes\boldsymbol{\ell}'_K\right).
\end{aligned}
$$

Then, we can show the asymptotic independence of the kurtosis and skewness components by noticing that

$$
\begin{aligned}
cov\left[\mathbf{s}_{\mathrm{st}|\infty}(\boldsymbol{\theta}),s_{\mathrm{k}t-j|\infty}(\boldsymbol{\theta})\right] &= \mathbf{b}'_1 cov\left[\mathbf{m}_{1t}(\boldsymbol{\theta}),\mathbf{m}_{2t-j}(\boldsymbol{\theta})\right]\mathbf{b}_2\\
&\quad+\mathbf{b}'_1 cov\left[\mathbf{m}_{1t}(\boldsymbol{\theta}),\mathbf{m}_{4t-j}(\boldsymbol{\theta})\right]\mathbf{b}_4\\
&\quad+\mathbf{b}'_3 cov\left[\mathbf{m}_{3t}(\boldsymbol{\theta}),\mathbf{m}_{2t-j}(\boldsymbol{\theta})\right]\mathbf{b}_2\\
&\quad+\mathbf{b}'_3 cov\left[\mathbf{m}_{3t}(\boldsymbol{\theta}),\mathbf{m}_{4t-j}(\boldsymbol{\theta})\right]\mathbf{b}_4\\
&=\mathbf{0},
\end{aligned}
$$

where the last equality follows from (A1), (A3), (A4) and (A6). Moreover, we can simplify even further the relevant expressions by exploiting the cancellation of cross-terms within the variance formulas,

$$
cov\left[\mathbf{m}_{1t}(\boldsymbol{\theta}),\mathbf{s}_{\mathrm{s}t-j}(\boldsymbol{\theta})\right]=\mathbf{0},\ \text{ and }\ cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}),s_{\mathrm{k}t-j|\infty}(\boldsymbol{\theta})\right]=\mathbf{0}. \tag{A7}
$$

For the sake of brevity, we prove the above equalities for the case when $R=K$; the proof for the case $R<K$ is similar, but more tedious.

To show the first equality in (A7), notice that for any j, we obtain

$$
cov\left[\mathbf{m}_{1t}(\boldsymbol{\theta}),\mathbf{m}_{1t-j}(\boldsymbol{\theta})\right]\mathbf{b}_1=-\left[2\boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_0+tr(\boldsymbol{\Gamma}_0)\boldsymbol{\Gamma}_j\right]
$$

because $\boldsymbol{\Omega}_\infty = \mathbf{I}_K - \boldsymbol{\Gamma}_0$ and $\mathbf{b}_1 = -tr(\boldsymbol{\Gamma}_0)\mathbf{I}_K - \boldsymbol{\Gamma}_0$. The remaining part follows from exploiting the following equalities:

$$\boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_0 = \left\{\left[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)\right] \odot \left(\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right)\right\}(\boldsymbol{\ell}_K \otimes \mathbf{I}_K) \tag{A8}$$

and

$$tr(\boldsymbol{\Gamma}_0)\boldsymbol{\Gamma}_j = \left\{\left[vecd'(\boldsymbol{\Gamma}_0) \otimes \mathbf{1}_{K\times K}\right] \odot \left(\boldsymbol{\ell}'_K \otimes \Gamma_j\right)\right\}(\boldsymbol{\ell}_K \otimes \mathbf{I}_K). \tag{A9}$$

For instance, to show (A8), define

$$\mathbf{T}_K = \begin{bmatrix}\mathbf{e}_1\mathbf{e}'_1 & \dots & \mathbf{e}_K\mathbf{e}'_K\end{bmatrix},$$

with $(\mathbf{e}_1|\dots|\mathbf{e}_K) = \mathbf{I}_K$, as the unique $K \times K^2$ "diagonalization" matrix that transforms $vec(\mathbf{A})$ into $vecd(\mathbf{A})$ as $vecd(\mathbf{A}) = \mathbf{T}'_K vec(\mathbf{A})$ (see Magnus (1988)). Similarly, let

$$\mathbf{T}_{K^2} = \begin{bmatrix}(\mathbf{e}_1\mathbf{e}'_1 \otimes \mathbf{e}_1\mathbf{e}'_1) & (\mathbf{e}_1\mathbf{e}'_2 \otimes \mathbf{e}_1\mathbf{e}'_2) & \dots & (\mathbf{e}_K\mathbf{e}'_{K-1} \otimes \mathbf{e}_K\mathbf{e}'_{K-1}) & (\mathbf{e}_K\mathbf{e}'_K \otimes \mathbf{e}_K\mathbf{e}'_K)\end{bmatrix},$$

which is $K^2 \times K^4$. Some straightforward algebra delivers the following key identities:

$$\begin{aligned}
\mathbf{e}'_i\mathbf{T}_K &= (\mathbf{e}_i \otimes \mathbf{e}_i)', \\
(\mathbf{e}_i \otimes \mathbf{e}_i)'\mathbf{T}_{K^2} &= (\mathbf{e}_i \otimes \mathbf{e}_i \otimes \mathbf{e}_i \otimes \mathbf{e}_i)', \\
\mathbf{T}'_{K^2}(\boldsymbol{\ell}_K \otimes \mathbf{I}_K)\mathbf{e}_i &= (\mathbf{I}_K \otimes \mathbf{e}_i \otimes \mathbf{I}_K \otimes \mathbf{e}_i)vec(\mathbf{I}_K), \\
\mathbf{T}'_{K^2}\boldsymbol{\ell}_{K^2} &= vec(\mathbf{I}_{K^2}),
\end{aligned}$$

for all $i = 1, \dots, K$. Moreover, $\mathbf{T}_K$ and $\mathbf{T}_{K^2}$ have the important property that

$$(\mathbf{A} \odot \mathbf{B}) = \mathbf{T}_K(\mathbf{A} \otimes \mathbf{B})\mathbf{T}'_{K^2}$$

for any pair of $K \times K^2$ matrices $\mathbf{A}$ and $\mathbf{B}$. As a consequence, we have that for any pair of indices $i_1, i_2 = 1, \dots, K$,

$$\begin{aligned}
\mathbf{e}'_{i_1}\{[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)] \odot (\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K)\}(\boldsymbol{\ell}_K \otimes \mathbf{I}_K)\mathbf{e}_{i_2} &= \mathbf{e}'_{i_1}\mathbf{T}_K\{[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)] \otimes (\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K)\} \\
&\quad \times \mathbf{T}'_{K^2}(\boldsymbol{\ell}_K \otimes \mathbf{I}_K)\mathbf{e}_{i_2} \\
&= (\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_1})'\{[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)] \otimes (\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K)\} \times (\mathbf{I}_K \otimes \mathbf{e}_{i_2} \otimes \mathbf{I}_K \otimes \mathbf{e}_{i_2})vec(\mathbf{I}_K) \\
&= \{\mathbf{e}'_{i_1}[\boldsymbol{\ell}_K \otimes vec'(\boldsymbol{\Gamma}_0)](\mathbf{I}_K \otimes \mathbf{e}_{i_2}) \otimes \mathbf{e}'_{i_1}(\boldsymbol{\Gamma}_j \otimes \boldsymbol{\ell}'_K)(\mathbf{I}_K \otimes \mathbf{e}_{i_2})\} \times vec(\mathbf{I}_K) \\
&= (\mathbf{e}'_{i_2}\boldsymbol{\Gamma}_0 \otimes \mathbf{e}'_{i_1}\boldsymbol{\Gamma}_j)vec(\mathbf{I}_K) = \mathbf{e}'_{i_1}\boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_0\mathbf{e}_{i_2}.
\end{aligned}$$

But since i_1, i_2 are arbitrary, we can conclude that (A8) holds. Analogous calculations allow us to show (A9). Therefore (A8) and (A9), together with the fact that $\mathbf{b}_3 = \boldsymbol{\ell}_K \otimes \mathbf{I}_K$ and (A2), imply that

$$cov\left[\mathbf{m}_{1t}(\boldsymbol{\theta}), \mathbf{s}_{st-j|\infty}(\boldsymbol{\theta})\right] = cov[\mathbf{m}_{1t}(\boldsymbol{\theta}), \mathbf{m}'_{1t}(\boldsymbol{\theta})]\mathbf{b}'_1 + cov[\mathbf{m}_{1t}(\boldsymbol{\theta}), \mathbf{m}'_{3t}(\boldsymbol{\theta})]\mathbf{b}'_3 = \mathbf{0}.$$

As for the second equality in (A7), again given that $\mathbf{\Omega}_\infty = \mathbf{I}_K - \mathbf{\Gamma}_0$ and

$$\mathbf{b}_2 = -\frac{1}{2}tr(\mathbf{\Gamma}_0)vec(\mathbf{I}_K) - vec(\mathbf{\Gamma}_0),$$

we can then use the same tedious but straightforward arguments as before to show that

$$cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}_{2t-j}(\boldsymbol{\theta})\right] vec(\mathbf{\Gamma}_0) = \{[\boldsymbol{\ell}_{K^2} \otimes vec'(\mathbf{\Gamma}_0)] \odot cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}_{2t-j}(\boldsymbol{\theta})\right]\}\boldsymbol{\ell}_{K^2},$$
$$tr(\mathbf{\Gamma}_0)cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}_{2t-j}(\boldsymbol{\theta})\right] vec(\mathbf{I}_K) = \{\left[\boldsymbol{\ell}_{K^2} \otimes \boldsymbol{\ell}'_K \otimes vecd'(\mathbf{\Gamma}_0)\right] \odot \left(\boldsymbol{\ell}_K \otimes \mathbf{\Gamma}_j \otimes \boldsymbol{\ell}'_K\right)$$
$$\odot (\mathbf{\Gamma}_j \otimes \mathbf{1}_{K\times K})\}\boldsymbol{\ell}_{K^2} + \{[\boldsymbol{\ell}_{K^2} \otimes vecd'(\mathbf{\Gamma}_0) \otimes \boldsymbol{\ell}'_K] \odot (\mathbf{1}_{K\times K} \otimes \mathbf{\Gamma}_j) \odot \boldsymbol{\ell}'_K \otimes \mathbf{\Gamma}_j \otimes \boldsymbol{\ell}_K)\}\boldsymbol{\ell}_{K^2},$$

which, together with the fact that $\mathbf{b}_4 = \boldsymbol{\ell}_{K^2}/4$ and (A5), imply that

$$cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}), s_{\mathrm{k}t-j|\infty}(\boldsymbol{\theta})\right] = cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}'_{2t}(\boldsymbol{\theta})\right]\mathbf{b}'_2 + cov\left[\mathbf{m}_{2t}(\boldsymbol{\theta}), \mathbf{m}'_{4t}(\boldsymbol{\theta})\right]\mathbf{b}'_4 = \mathbf{0},$$

as desired. This allows us to write

$$\lim_{T\to\infty} V\begin{bmatrix}\sqrt{T}\bar{s}_{\mathrm{k}T}(\boldsymbol{\theta}) \\ \sqrt{T}\bar{\mathbf{s}}_{\mathrm{s}T}(\boldsymbol{\theta})\end{bmatrix} = \begin{bmatrix}\mathcal{C}_{\mathrm{k}}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{C}_{\mathrm{s}}(\boldsymbol{\theta})\end{bmatrix}$$

where the expressions for $\mathcal{C}_{\mathrm{k}}(\boldsymbol{\theta})$ and $\mathcal{C}_{\mathrm{s}}(\boldsymbol{\theta})$ can be found in the statement of the Lemma. □

Lemma 7 *Let $\bar{\mathbf{s}}_{\mathrm{MV}T}(\boldsymbol{\theta})$ denote the Gaussian ML score with respect to the conditional mean and variance parameters $\boldsymbol{\theta}$. Then,*

$$i)\ \lim_{T\to\infty} cov[\sqrt{T}\bar{\mathbf{s}}_{\mathrm{MV}T}(\boldsymbol{\theta}), \sqrt{T}\bar{s}_{\mathrm{k}T}(\boldsymbol{\theta})|\boldsymbol{\theta}] = \mathbf{0},$$
$$ii)\ \lim_{T\to\infty} cov[\sqrt{T}\bar{\mathbf{s}}_{\mathrm{MV}T}(\boldsymbol{\theta}), \sqrt{T}\bar{\mathbf{s}}_{\mathrm{s}T}(\boldsymbol{\theta})|\boldsymbol{\theta}] = \mathbf{0}.$$

Proof. As shown in Mencía and Sentana (2012), the score of the latent model with respect to the mean-variance parameter vector $\boldsymbol{\theta}$ converges to the Gaussian score as we approach the null hypothesis along any of the possible directions through which the GH distribution approaches Gaussianity. This observation combined with the EM principle provides a very convenient way of studying explicitly the score with respect to $\boldsymbol{\theta}$. For ease of exposition assume $\boldsymbol{\xi}_0 = \mathbf{0}$. Then,

$$\mathbf{Y}_T = [\ell_T \otimes \boldsymbol{\pi}(\boldsymbol{\theta})] + [\mathbf{I}_T \otimes \mathbf{H}(\boldsymbol{\theta})]\{\mathbf{I}_{MT} - [\mathbf{C}_T \otimes \mathbf{F}(\boldsymbol{\theta})]\}^{-1}[\mathbf{I}_T \otimes \mathbf{M}(\boldsymbol{\theta})]\mathbf{E}_T$$
$$\equiv \mathbf{\Pi}_T(\boldsymbol{\theta}) + \mathbf{D}_T(\boldsymbol{\theta})\mathbf{E}_T.$$

where we have defined

$$\mathbf{C}_T \equiv \begin{bmatrix}\mathbf{0} & \mathbf{I}_{T-1} \\ \mathbf{0} & \mathbf{0}\end{bmatrix},$$
$$\mathbf{\Pi}_T(\boldsymbol{\theta}) \equiv \ell_T \otimes \boldsymbol{\pi}(\boldsymbol{\theta}),$$

and

$$\mathbf{D}_T(\boldsymbol{\theta}) \equiv [\mathbf{I}_T \otimes \mathbf{H}(\boldsymbol{\theta})]\{\mathbf{I}_{MT} - [\mathbf{C}_T \otimes \mathbf{F}(\boldsymbol{\theta})]\}^{-1}[\mathbf{I}_T \otimes \mathbf{M}(\boldsymbol{\theta})]. \quad \text{(A10)}$$

Under our assumption that no linear combination of $\mathbf{Y}_T$ has zero variance, the matrix $\mathbf{D}_T(\boldsymbol{\theta})$ has full row-rank. Let $\mathbf{D}^*_T(\boldsymbol{\theta})$ be a $(K-N)T \times KT$ matrix of differentiable functions such that

$$\tilde{\mathbf{D}}_T(\boldsymbol{\theta}) = [\mathbf{D}'_T(\boldsymbol{\theta}), (\mathbf{D}^*_T(\boldsymbol{\theta}))']' \quad \text{(A11)}$$

is non-singular. Let $\tilde{\mathbf{\Pi}}_T(\boldsymbol{\theta}) = [\mathbf{\Pi}'_T(\boldsymbol{\theta}), \mathbf{0}']'$ and define

$$\tilde{\mathbf{Y}}_T \equiv \tilde{\mathbf{\Pi}}_T(\boldsymbol{\theta}) + \tilde{\mathbf{D}}_T(\boldsymbol{\theta})\mathbf{E}_T.$$

This reasoning delivers the following alternative state space representation under the null,

$$\mathbf{Y}_T = \mathbf{S}_{NT,KT}\tilde{\mathbf{Y}}_T, \quad \text{with} \quad \tilde{\mathbf{Y}}_T \sim N[\tilde{\mathbf{\Pi}}_T(\boldsymbol{\theta}), \tilde{\mathbf{D}}'_T(\boldsymbol{\theta})\tilde{\mathbf{D}}_T(\boldsymbol{\theta})].$$

We now apply the EM principle to the previous representation noting that the measurement equation contains no unknown parameters. The score of the latent model is

$$\begin{aligned}\frac{\partial \ln f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial \tilde{\mathbf{\Pi}}'_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\tilde{\mathbf{D}}_T(\boldsymbol{\theta})\mathbf{E}_T + \frac{1}{2}\frac{\partial vec'[\tilde{\mathbf{D}}'_T(\boldsymbol{\theta})\tilde{\mathbf{D}}_T(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}[\tilde{\mathbf{D}}'_T(\boldsymbol{\theta}) \otimes \tilde{\mathbf{D}}'_T(\boldsymbol{\theta})]vec\left(\mathbf{E}_T\mathbf{E}'_T - \mathbf{I}_{KT}\right), \\ &\equiv \mathbf{b}_{\mathrm{MV}1,T}(\boldsymbol{\theta})\mathbf{E}_T + \mathbf{b}_{\mathrm{MV}2,T}(\boldsymbol{\theta})vec\left(\mathbf{E}_T\mathbf{E}'_T - \mathbf{I}_{KT}\right).\end{aligned}$$

where we have used $\mathbf{E}_T = \tilde{\mathbf{D}}_T^{-1}(\boldsymbol{\theta})[\tilde{\mathbf{Y}}_T - \tilde{\mathbf{\Pi}}_T(\boldsymbol{\theta})]$, with $\mathbf{b}_{\mathrm{MV}1,T}(\boldsymbol{\theta})$ and $\mathbf{b}_{\mathrm{MV}2,T}(\boldsymbol{\theta})$ defined in the obvious way. Smoothing the score above we obtain

$$\begin{aligned}\bar{\mathbf{s}}_{\mathrm{MV}T}(\boldsymbol{\theta}) &\equiv \frac{1}{T}\frac{\partial \ln f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \frac{1}{T}\mathbf{b}_{\mathrm{MV}1,T}(\boldsymbol{\theta})E[\mathbf{E}_T|\mathbf{Y}_T, \boldsymbol{\theta}] + \frac{1}{T}\mathbf{b}_{\mathrm{MV}2,T}(\boldsymbol{\theta})vec\{E[\mathbf{E}_T\mathbf{E}'_T|\mathbf{Y}_T, \boldsymbol{\theta}] - \mathbf{I}_{KT}\}.\end{aligned}$$

But since

$$E[\mathbf{E}_T\mathbf{E}'_T|\mathbf{Y}_T, \boldsymbol{\theta}] = V[\mathbf{E}_T|\mathbf{Y}_T, \boldsymbol{\theta}] + E[\mathbf{E}_T|\mathbf{Y}_T, \boldsymbol{\theta}]E[\mathbf{E}'_T|\mathbf{Y}_T, \boldsymbol{\theta}]$$

and $V[\mathbf{E}_T|\mathbf{Y}_T, \boldsymbol{\theta}]$ does not depend on $\mathbf{Y}_T$, it is clear that $\bar{\mathbf{s}}_{\mathrm{MV}T}(\boldsymbol{\theta})$ is a linear combination of $\mathbf{E}_{T|T}(\boldsymbol{\theta}) \equiv E[\mathbf{E}_T|\mathbf{Y}_T, \boldsymbol{\theta}]$ and $vec[\mathbf{E}_{T|T}(\boldsymbol{\theta})\mathbf{E}'_{T|T}(\boldsymbol{\theta})]$ (with coefficients possibly varying with T). If we then replace Kalman smoothed variables by their Wiener-Kolmogorov counterparts and the coefficients of the linear combination by their limits as $T \to \infty$, we obtain

$$\begin{aligned}\bar{\mathbf{s}}_{\mathrm{MV}T}(\boldsymbol{\theta}) = \; & \mathbf{b}_{\mathrm{MV}0}(\boldsymbol{\theta}) + \mathbf{b}_{\mathrm{MV}1}(\boldsymbol{\theta})'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta}) \\ & + \sum_{\ell=0}^{T-1}\left\{\mathbf{b}_{\mathrm{MV}2}(\boldsymbol{\theta})'\frac{1}{T}\sum_{t=\ell+1}^{T} vec\left[\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta})\boldsymbol{\varepsilon}'_{t-\ell|\infty}(\boldsymbol{\theta})\right]\right\}.\end{aligned}$$

The rest of the proof follows from (A7) and from the fact that

$$cov\{vec[\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta})\boldsymbol{\varepsilon}'_{t-\ell|\infty}(\boldsymbol{\theta})], \mathbf{s}_{st-j}(\boldsymbol{\theta})\} = \mathbf{0} \text{ and } cov\{vec[\boldsymbol{\varepsilon}_{t|\infty}(\boldsymbol{\theta})\boldsymbol{\varepsilon}'_{t-\ell|\infty}(\boldsymbol{\theta})], s_{kt-j}(\boldsymbol{\theta})\} = \mathbf{0},$$

which can be established by an argument analogous to that of (A7). □

Proposition 1

To simplify the exposition, we focus on the case where $\boldsymbol{\theta}$ is fixed and known, so that and the task is to derive the scores with respect to the shape parameters $\boldsymbol{\varphi}$ only. We further assume that $\boldsymbol{\xi}_0 = \mathbf{0}$ and $\pi(\boldsymbol{\theta}) = \mathbf{0}$. These assumptions are not essential to the argument and may be removed at the cost of more notation. We can then write

$$\mathbf{Y}_T = \mathbf{D}_T(\boldsymbol{\theta})\mathbf{E}_T,$$

where $\mathbf{D}_T(\boldsymbol{\theta})$ is given in (A10). Note the $(NT \times KT)$ matrix $\mathbf{D}_T(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\varphi}$, although it depends on $\boldsymbol{\theta}$. Notice also that $f_{\mathbf{E}}(\mathbf{E}_T|\varphi)$ is continuous in $\mathbf{E}_T$ and differentiable in $\boldsymbol{\varphi}$ by construction because of the properties of the GH distribution $D(\mathbf{0}, \mathbf{I}_K, \boldsymbol{\varphi})$. Given that we are assuming $N \leq K$, we require the additional assumption that $\mathbf{D}_T(\boldsymbol{\theta})$ has full row rank. As in Lemma 7, we define $\mathbf{D}_T^*(\boldsymbol{\theta})$ such that (A11) is invertible. Similarly, we define the random vector $\mathbf{Y}_T^* = \mathbf{D}_T^*(\boldsymbol{\theta})\mathbf{E}_T$. Hence,

$$\tilde{\mathbf{Y}}_T \equiv \begin{bmatrix} \mathbf{Y}_T \\ \mathbf{Y}_T^* \end{bmatrix} = \tilde{\mathbf{D}}_T(\boldsymbol{\theta})\mathbf{E}_T$$

will be a KT-dimensional random vector with density $f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})$ with respect to Lebesgue measure on $\mathbb{R}^{KT}$ given by the usual change-of-variable formula,

$$f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi}) = \frac{f_{\mathbf{E}}(\tilde{\mathbf{D}}_T^{-1}(\boldsymbol{\theta})\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})}{|\det[\tilde{\mathbf{D}}_T(\boldsymbol{\theta})]|}.$$

Moreover, the density $f_{\tilde{\mathbf{Y}}}$ is continuous in $\tilde{\mathbf{Y}}_T$ and differentiable in $\boldsymbol{\varphi}$, and the marginal density

$$f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\varphi}) = \int_{\mathbb{R}^{(K-N)T}} f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi}) d\mathbf{Y}_T^*$$

is continuous in $\mathbf{Y}_T$ and differentiable in $\boldsymbol{\varphi}$ as well. Taking logs, differentiating with respect to $\boldsymbol{\varphi}$ on both sides of the foregoing equation, and exchanging the orders of the differentiation and integration operators on the right-hand side by virtue of theorem 16.8 in Billingsley (1995), we conclude that

$$\frac{\partial \ln f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} = \int_{\mathbb{R}^{(K-N)T}} \frac{\partial \ln f_{\mathbf{E}}(\tilde{\mathbf{D}}_T^{-1}(\boldsymbol{\theta})\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} \frac{f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})}{f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\varphi})} d\mathbf{Y}_T^*, \tag{A12}$$

for all $\mathbf{Y}_T$ and $\boldsymbol{\varphi}$ for which $f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\varphi}) > 0$ (and this holds for almost all $\boldsymbol{\varphi}$). The function $f_{\mathbf{Y}^*|\mathbf{Y}}(\mathbf{Y}_T^*|\mathbf{Y}_T, \boldsymbol{\varphi}) \equiv f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})/f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\varphi})$ is the conditional density of $\mathbf{Y}_T^*$ given $\mathbf{Y}_T$, which is a continuous density with respect to Lebesgue measure on $\mathbb{R}^{(K-N)T}$. In that precise sense, we write

$$E\left[\left.\frac{\partial \ln f_{\mathbf{E}}(\mathbf{E}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}}\right| \mathbf{Y}_T, \boldsymbol{\varphi}\right] = \int_{\mathbb{R}^{(K-N)T}} \frac{\partial \ln f_{\mathbf{E}}(\tilde{\mathbf{D}}_T^{-1}(\boldsymbol{\theta})\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} \frac{f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})}{f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\varphi})} d\mathbf{Y}_T^*.$$

Importantly, the value of the integral in (A12) is independent of the choice of $\mathbf{D}_T^*$. To see this, multiply both sides by $f_{\mathbf{Y}}(\mathbf{Y}_T|\boldsymbol{\varphi})$ and integrate with respect to $\mathbf{Y}_T$,

$$\begin{aligned} &\int_{\mathbb{R}^{NT}} \int_{\mathbb{R}^{(K-N)T}} \frac{\partial \ln f_{\mathbf{E}}(\tilde{\mathbf{D}}_T^{-1}(\boldsymbol{\theta})\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi}) d\mathbf{Y}_T^* d\mathbf{Y}_T \\ &= \int_{\mathbb{R}^{KT}} \frac{\partial \ln f_{\mathbf{E}}(\tilde{\mathbf{D}}_T^{-1}(\boldsymbol{\theta})\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{Y}}_T|\boldsymbol{\varphi}) d\tilde{\mathbf{Y}}_T = E\left[\left.\frac{\partial \ln f_{\mathbf{E}}(\mathbf{E}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}}\right| \boldsymbol{\varphi}\right] \end{aligned}$$

by Fubini's theorem. For all possible choices of $\mathbf{D}_T^*(\boldsymbol{\theta})$ we obtain a version of

$$E\left[\left.\frac{\partial \ln f_{\mathbf{E}}(\mathbf{E}_T|\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}}\right| \mathbf{Y}_T, \boldsymbol{\varphi}\right],$$

whose uniqueess follows from the a.s. equality of conditional expectations. Therefore, equation (5) holds. □

Proposition 2

It follows from Lemma 5 when $\boldsymbol{\beta} = \mathbf{0}$. □

Proposition 3

It follows from Lemma 6.i and 7.i. □

Proposition 4

It follows from Propositions 2 and 3. □

Proposition 5

It is a rewriting of Lemma 5. □

Proposition 6

It follows from Lemma 6.i, , 6.ii, 6.iii and 7.ii. □

Proposition 7

It follows by combining the arguments in the proof of Proposition 5 in Mencía and Sentana (2012) with the results in Propositions 5 and 6. □

B Asymptotic equivalence of smoothed scores sample moments

Consider the model (1)-(2) with $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{I}_K)$ and, to save notation, assume (i) $\boldsymbol{\pi} = \mathbf{0}$. To facilitate exposition we further assume that (ii) $\det(\mathbf{I}_M - \mathbf{F}z) = 0$ implies $|z| > 1$. This condition can be removed at the cost of considerably complicating the analysis.

Under these assumptions, the MA(∞) representation of $\{\mathbf{y}_t\}$ is

$$\mathbf{y}_t = \sum_{s=-\infty}^{\infty} \mathbf{D}(s)\boldsymbol{\varepsilon}_{t-s} \text{ for all } t,$$

where $\mathbf{D}(s) = \mathbf{H}\mathbf{F}^s\mathbf{M}$ for all $s \geq 0$, and $\mathbf{D}(s) = \mathbf{0}$ whenever $s < 0$.

Let $\mathcal{F}_T = \sigma(\{\mathbf{y}_t\}_{|t|\leq T})$ denote the σ-field generated by $\{\mathbf{y}_t\}_{|t|\leq T}$. Also, let $\mathcal{F}_\infty = \sigma(\cup_{T=0}^{\infty}\mathcal{F}_T)$. It is well known that the assumption of Gaussianity implies existence of sequences of $K \times N$

matrices $\{\mathbf{A}_{t|T}(\tau)\}$ for all t and T, and $\{\mathbf{A}(\tau)\}$ with $\mathbf{A}_{t|T}(\tau) = \mathbf{0}$ whenever $|t| > \tau$, such that

$$\boldsymbol{\varepsilon}_{t|T} = E(\boldsymbol{\varepsilon}_t|\mathcal{F}_T) = \sum_{\tau=-T}^{T} \mathbf{A}_{t|T}(\tau)\mathbf{y}_{t-\tau}, \text{ for all } t \text{ and } T,$$

$$\boldsymbol{\varepsilon}_{t|\infty} = E(\boldsymbol{\varepsilon}_t|\mathcal{F}_\infty) = \sum_{\tau=-\infty}^{\infty} \mathbf{A}(\tau)\mathbf{y}_{t-\tau}, \text{ for all } t.$$

For any real matrix $\mathbf{A}$, let $\|\mathbf{A}\| = \sqrt{tr(\mathbf{A}'\mathbf{A})}$ be its Frobenius norm.

The purpose of this appendix is to show that:

Proposition 8 *As* $T^* \equiv 2T + 1 \to \infty$,

$$\frac{1}{\sqrt{T^*}} \sum_{|t| \le T} (\boldsymbol{\varepsilon}_{t|\infty} - \boldsymbol{\varepsilon}_{t|T}) = o_{\mathrm{P}}(1).$$

In the proof of Proposition 8, we will make use of the following:

Lemma 8 *The following three properties hold:*

i) (L_2-optimality) Any $\mathcal{F}_T$-measurable function $\tilde{\boldsymbol{\varepsilon}}_T$ satisfies

$$E\left(\left\|\sum_{|t|\le T}(\boldsymbol{\varepsilon}_{t|\infty} - \boldsymbol{\varepsilon}_{t|T})\right\|^2\right) \le E\left(\left\|\sum_{|t|\le T}\boldsymbol{\varepsilon}_{t|\infty} - \tilde{\boldsymbol{\varepsilon}}_T\right\|^2\right) \quad \text{for all } T.$$

ii) (Geometric decay of $\mathbf{A}$) For some $\rho_\alpha \in (0,1)$, $C_\alpha > 0$ and all τ, $\|\mathbf{A}(\tau)\| \le C_\alpha \rho_\alpha^{|\tau|}$. Hence,

$$\sum_{\tau=-\infty}^{\infty} \|\mathbf{A}(\tau)\| < \infty.$$

iii)(Geometric decay of $\mathbf{D}$) For some $\rho_\delta \in (0,1)$, $C_\delta > 0$ and all s, $\|\mathbf{D}(s)\| \le C_\delta \rho_\delta^{|s|}$. Hence,

$$\sum_{s=-\infty}^{\infty} \|\mathbf{D}(s)\| < \infty.$$

Proof of Lemma 8.

Property (i) is a consequence of the fact that

$$\boldsymbol{\varepsilon}_{t|T} = E(\boldsymbol{\varepsilon}_{t|\infty}|\mathcal{F}_T) \text{ for all } t \text{ and } T$$

by virtue of the law of iterated expectations, and the standard result that an expectation conditional on $\mathcal{F}_T$ minimizes the L_2-distance to the set of $\mathcal{F}_T$-measurable functions.

In turn, Property (ii) follows from the fact that $\boldsymbol{\varepsilon}_{t|\infty}$ is a VARMA process. Hence,

$$\sum_{|\tau|>T} \|\mathbf{A}(\tau)\| \le 2C_\alpha \sum_{\tau=T+1}^{\infty} \rho_\alpha^\tau = \frac{2C_\alpha}{1-\rho_\alpha}\rho_\alpha^{T+1} \to 0 \text{ as } T \to \infty,$$

implying $\sum_{\tau=-\infty}^{\infty} \|\mathbf{A}(\tau)\| < \infty$.

To establish property (iii), note that $\|\mathbf{D}_s\| \leq \|\mathbf{H}\|\|\mathbf{F}\|^s\|\mathbf{M}\| \leq \sqrt{M}\|\mathbf{H}\|\|\mathbf{M}\||\lambda_{\mathbf{F}}|^s$, where we have denoted by $\lambda_{\mathbf{F}}$ the largest eigenvalue of $\mathbf{F}$. By assumption, $|\lambda_{\mathbf{F}}| < 1$, so $C_\delta = \sqrt{M}\|\mathbf{H}\|\|\mathbf{M}\|$ and $\rho_\delta = |\lambda_{\mathbf{F}}|$. Finally,

$$\sum_{|s|>S} \|\mathbf{D}(s)\| \leq 2C_\delta \sum_{s=S+1}^{\infty} \rho_\delta^s = \frac{2C_\delta}{1-\rho_\delta}\rho_\delta^{S+1} \to 0 \text{ as } S \to \infty,$$

implying $\sum_{s=-\infty}^{\infty} \|\mathbf{D}(s)\| < \infty$. □

Proof of Proposition 8.

Fix some $\epsilon > 0$ and $k = 1, \ldots, K$ and define the event

$$\mathcal{E}_{k,T} \equiv \left\{ \left| \sum_{|t|\leq T} (\varepsilon_{k,t|\infty} - \varepsilon_{k,t|T}) \right| > \sqrt{T^*}\epsilon \right\}$$

By Chebyshev-Bienaymé's inequality,

$$\Pr(\mathcal{E}_{k,T}) \leq \frac{1}{T^*\epsilon^2} V\left[\sum_{|t|\leq T} (\varepsilon_{k,t|\infty} - \varepsilon_{k,t|T})\right] \leq \frac{1}{T^*\epsilon^2} E\left(\left\|\sum_{|t|\leq T} (\boldsymbol{\varepsilon}_{t|\infty} - \boldsymbol{\varepsilon}_{t|T})\right\|^2\right).$$

Further, Lemma 8.i implies that for any $\mathcal{F}_T$-measurable function $\tilde{\boldsymbol{\varepsilon}}_T$,

$$\Pr(\mathcal{E}_{k,T}) \leq \frac{1}{T^*\epsilon^2} E\left(\left\|\sum_{|t|\leq T} \boldsymbol{\varepsilon}_{t|\infty} - \tilde{\boldsymbol{\varepsilon}}_T\right\|^2\right).$$

Therefore, the proof will be completed if we establish that, for some suitable choice of $\tilde{\boldsymbol{\varepsilon}}_T$,

$$E\left(\left\|\sum_{|t|\leq T} \boldsymbol{\varepsilon}_{t|\infty} - \tilde{\boldsymbol{\varepsilon}}_T\right\|^2\right) = o(T).$$

To do so, consider the linear $\mathcal{F}_T$-measurable variable

$$\tilde{\boldsymbol{\varepsilon}}_T = \sum_{|t|\leq T} \sum_{|\tau|\leq T} \mathbf{A}(\tau)\mathbf{y}_{t-\tau}.$$

We have

$$\Delta_T = \sum_{|t|\leq T} \boldsymbol{\varepsilon}_{t|\infty} - \tilde{\boldsymbol{\varepsilon}}_T = \sum_{|t|\leq T} \sum_{|\tau|>T} \mathbf{A}(\tau)\mathbf{y}_{t-\tau} = \sum_{r=-\infty}^{\infty} \boldsymbol{\Phi}_T(r)\mathbf{y}_r,$$

where $\boldsymbol{\Phi}_T(0) = \mathbf{0}$,

$$\boldsymbol{\Phi}_T(r) = \sum_{j=\max\{1,r-2T\}}^{r} \mathbf{A}[-(T+j)], \text{ for } r > 0,$$

$$\boldsymbol{\Phi}_T(r) = \sum_{j=\max\{1,r-2T\}}^{r} \mathbf{A}(T+j), \text{ for } r < 0,$$

implying that

$$\|\boldsymbol{\Phi}_T(r)\| \leq C_\alpha \rho_\alpha^{T+1}/(1-\rho_\alpha), \text{ for } |r| \leq 2T+1, \text{ and}$$

$$\|\boldsymbol{\Phi}_T(r)\| \leq C_\alpha \rho_\alpha^{r-T}/(1-\rho_\alpha), \text{ for } |r| > 2T+1,$$

whence it follows immediately that $\sum_{r=-\infty}^{\infty} \|\mathbf{\Phi}_T(r)\| < C_\phi T \rho_\alpha^T$ for some constant $C_\phi > 0$. Finally,

$$\begin{aligned}
\sqrt{E(\|\Delta_T\|^2)} &= \sqrt{\sum_{r=-\infty}^{\infty} \left\| \sum_{s=-\infty}^{\infty} \mathbf{\Phi}_T(r)\mathbf{\Psi}(s-r) \right\|^2} \\
&\leq \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \|\mathbf{\Phi}_T(r)\| \|\mathbf{D}(s-r)\| \\
&\leq \left(\sum_{r=-\infty}^{\infty} \|\mathbf{\Phi}_T(r)\| \right) \left(\sum_{s=-\infty}^{\infty} \|\mathbf{D}(s)\| \right) < \infty,
\end{aligned}$$

where the last inequality follows from Lemma 8.ii and 8.iii. As a consequence of this, $E(\|\Delta_T\|^2) = o(T)$. □

C An algorithm for computing the asymptotic variance

Consider a VARMA process with scalar VAR part for the K_x-dimensional process $\mathbf{x}_t$,

$$\phi(L)\mathbf{x}_t = \mathbf{\Theta}(L)\mathbf{u}_t$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\mathbf{\Theta}(z) = \mathbf{\Theta}_0 + \mathbf{\Theta}_1 z + \ldots \mathbf{\Theta}_q z^q$. The error process $\mathbf{u}_t$ is assumed to be K-dimensional white noise, i.e. $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}$, $E(\mathbf{u}_t\mathbf{u}_{t-j}') = \mathbf{0}$ for $j \neq 0$. Next, write the VARMA process in companion VAR(1) form as

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{Q}\mathbf{u}_t,$$

where $\mathbf{X}_t = (\mathbf{x}_t, ..., \mathbf{x}_{t-p+1}, \mathbf{u}_t, ..., \mathbf{u}_{t-p+1})'$,

$$\mathbf{A} = \begin{pmatrix} \bar{\mathbf{\Phi}} \otimes \mathbf{I}_{K_x} & \mathbf{e}_1 \otimes \bar{\mathbf{\Theta}} \\ \mathbf{0} & \mathbf{J}_q \otimes \mathbf{I}_K \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \mathbf{\Theta}_0 \\ \mathbf{0} \\ \mathbf{I}_K \\ \mathbf{0} \end{pmatrix},$$

with $\mathbf{e}_1$ being the first vector of the canonical basis in $\mathbb{R}^p$,

$$\bar{\mathbf{\Phi}} = \begin{pmatrix} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{pmatrix}, \quad \bar{\mathbf{\Theta}} = \begin{pmatrix} \mathbf{\Theta}_1 & \cdots & \mathbf{\Theta}_q \end{pmatrix}, \text{ and } \mathbf{J}_q = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{q-1} & \mathbf{0} \end{pmatrix}.$$

Suppose we can find an invertible matrix $\mathbf{C}$ and a block diagonal matrix $\mathbf{\Lambda}$ (with Jordan blocks) such that $\mathbf{A} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1}$. Then, we can transform the original system by defining $\mathbf{Z}_t = \mathbf{C}^{-1}\mathbf{X}_t$, a possibly complex-valued stochastic process that satisfies

$$\mathbf{Z}_t = \mathbf{\Lambda}\mathbf{Z}_{t-1} + \boldsymbol{\eta}_t,$$

with $\boldsymbol{\eta}_t = \mathbf{C}^{-1}\mathbf{Q}\mathbf{u}_t$ being white-noise (and possibly complex-valued). Then, it can be shown that a computationally convenient decomposition of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1}, \tag{C13}$$

where

$$\mathbf{C} = \begin{pmatrix} \bar{\mathbf{C}} \otimes \mathbf{I}_{K_x} & -(\bar{\mathbf{\Phi}}^{-q} \otimes \mathbf{I}_{K_x})\mathbf{\Theta}^* \\ \mathbf{0} & \mathbf{I}_{Kq} \end{pmatrix}, \ \mathbf{\Lambda} = \begin{pmatrix} \bar{\mathbf{\Lambda}} \otimes \mathbf{I}_{K_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_q \otimes \mathbf{I}_K \end{pmatrix}$$

and

$$\mathbf{C}^{-1} = \begin{pmatrix} \bar{\mathbf{C}}^{-1} \otimes \mathbf{I}_{K_x} & (\bar{\mathbf{C}}^{-1}\bar{\mathbf{\Phi}}^{-q} \otimes \mathbf{I}_{K_x})\mathbf{\Theta}^* \\ \mathbf{0} & \mathbf{I}_{Kq} \end{pmatrix},$$

with $\bar{\mathbf{\Phi}} = \bar{\mathbf{C}}\bar{\mathbf{\Lambda}}\bar{\mathbf{C}}^{-1}$ providing the Jordan decomposition of $\bar{\mathbf{\Phi}}$, and

$$\mathbf{\Theta}^* = \sum_{h=1}^{q} (\bar{\mathbf{\Phi}}^{q-h}\mathbf{e}_1 \otimes \bar{\mathbf{\Theta}})(\mathbf{J}_q^{h-1} \otimes \mathbf{I}_K).$$

Notice that the decomposition outlined above is convenient to handle large systems because it reduces substantially the size of the matrices for which the Jordan decomposition needs to be performed.

We can also show that the autocovariance function of the Wiener-Kolmogorov filter derived in Lemma 4 is the autocovariance function of the stable solution to the difference equation embodied in its VARMA representation. For that reason, we decompose $\mathbf{A}$ as in (C13), with the absolute values of the eigenvalues in decreasing order. But since we have assumed no unit roots, we will have that $K_{\mathrm{S}} = K_x p + Kq - K_{\mathrm{U}}$, where K_{U} is the number of roots outside the unit circle and K_{S} the number of roots inside the unit circle.

Let $\mathbf{R} = \mathbf{C}\mathbf{Q}\overline{\mathbf{Q}}'\overline{\mathbf{C}}'$ denote the variance-covariance matrix of $\boldsymbol{\eta}_t$. We can partition the system into its unstable and stable parts as follows:

$$\mathbf{Z}_t = \begin{pmatrix} \mathbf{Z}_{\mathrm{U}t} \\ \mathbf{Z}_{\mathrm{S}t} \end{pmatrix}, \ \boldsymbol{\eta}_t = \begin{pmatrix} \boldsymbol{\eta}_{\mathrm{U}t} \\ \boldsymbol{\eta}_{\mathrm{S}t} \end{pmatrix}, \ \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_{\mathrm{UU}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{\mathrm{SS}} \end{pmatrix}, \text{ and } \mathbf{R} = \begin{pmatrix} \mathbf{R}_{\mathrm{UU}} & \mathbf{R}_{\mathrm{US}} \\ \mathbf{R}_{\mathrm{SU}} & \mathbf{R}_{\mathrm{SS}} \end{pmatrix}.$$

Next, if we write

$$\mathbf{Z}_{\mathrm{U}t} = \mathbf{\Lambda}_{\mathrm{UU}}^{-1}(\mathbf{Z}_{\mathrm{U}t+1} - \boldsymbol{\eta}_{\mathrm{U}t+1}) \quad \text{and} \quad \mathbf{Z}_{\mathrm{S}t} = \mathbf{\Lambda}_{\mathrm{SS}}\mathbf{Z}_{\mathrm{S}t-1} + \boldsymbol{\eta}_{\mathrm{S}t},$$

and partition

$$\mathbf{\Gamma}_{\mathbf{Z}}(j) = \begin{bmatrix} \mathbf{\Gamma}_{\mathrm{UU}}(j) & \mathbf{\Gamma}_{\mathrm{US}}(j) \\ \mathbf{\Gamma}_{\mathrm{SU}}(j) & \mathbf{\Gamma}_{\mathrm{SS}}(j) \end{bmatrix} = \begin{bmatrix} E(\mathbf{Z}_{\mathrm{U}t}\bar{\mathbf{Z}}'_{\mathrm{U}t-j}) & E(\mathbf{Z}_{\mathrm{U}t}\bar{\mathbf{Z}}'_{\mathrm{S}t-j}) \\ E(\mathbf{Z}_{\mathrm{S}t}\bar{\mathbf{Z}}'_{\mathrm{U}t-j}) & E(\mathbf{Z}_{\mathrm{S}t}\bar{\mathbf{Z}}'_{\mathrm{S}t-j}) \end{bmatrix},$$

we can show that the autocovariance function of $\mathbf{Z}_t$ can be computed from

$$\begin{aligned}
vec[\mathbf{\Gamma}_{\mathrm{UU}}(0)] &= [\mathbf{I}_{K_{\mathrm{U}}^2} - (\mathbf{\Lambda}_{\mathrm{UU}}^{-1} \otimes \mathbf{\Lambda}_{\mathrm{UU}}^{-1})]^{-1} vec[\mathbf{\Lambda}_{\mathrm{UU}}^{-1}\mathbf{R}_{\mathrm{UU}}(\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-1})'], \\
\mathbf{\Gamma}_{\mathrm{UU}}(j) &= \mathbf{\Gamma}_{\mathrm{UU}}(0)(\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-j})', \quad \text{for } j > 0 \\
\mathbf{\Gamma}_{\mathrm{UU}}(j) &= \overline{\mathbf{\Gamma}}'_{\mathrm{UU}}(-j), \quad \text{for } j < 0 \\
vec[\mathbf{\Gamma}_{\mathrm{SS}}(0)] &= [\mathbf{I}_{K_{\mathrm{S}}^2} - (\mathbf{\Lambda}_{\mathrm{SS}} \otimes \mathbf{\Lambda}_{\mathrm{SS}})]^{-1} vec\,(\mathbf{R}_{\mathrm{SS}}), \\
\mathbf{\Gamma}_{\mathrm{SS}}(j) &= \mathbf{\Lambda}_{\mathrm{SS}}^{j}\mathbf{\Gamma}_{\mathrm{SS}}(0), \quad \text{for } j > 0 \\
\mathbf{\Gamma}_{\mathrm{SS}}(j) &= \overline{\mathbf{\Gamma}}'_{\mathrm{SS}}(-j), \quad \text{for } j < 0 \\
\mathbf{\Gamma}_{\mathrm{SU}}(j) &= -\sum_{h=1}^{j} (\overline{\mathbf{\Lambda}}_{\mathrm{SS}}^{j-h})'\mathbf{R}_{\mathrm{SU}}(\overline{\mathbf{\Lambda}}_{\mathrm{UU}}^{-h})', \quad \text{for } j > 0 \\
\mathbf{\Gamma}_{\mathrm{SU}}(j) &= \mathbf{0}, \quad \text{for } j \leq 0, \quad \text{and} \\
\mathbf{\Gamma}_{\mathrm{US}}(j) &= \overline{\mathbf{\Gamma}}'_{\mathrm{SU}}(-j).
\end{aligned}$$

Finally, we can recover the autocovariance function of $\mathbf{X}_t$ from

$$\mathbf{\Gamma_X}(j) = E[\mathbf{X}_t\mathbf{X}'_{t-j}] = E[(\mathbf{C}\mathbf{Z}_t)(\overline{\mathbf{C}}\overline{\mathbf{Z}}_{t-j})'] = \mathbf{C}\Gamma_{\mathbf{Z}}(j)\overline{\mathbf{C}}'.$$

Obviously, the autocovariance function of $\mathbf{x}_t$ is the first block of $\mathbf{\Gamma_X}$.

D A Gibbs sampler algorithm for the common trend model with asymmetric Student t innovations

In this section, we develop a Gibbs sampler for the model we use in the empirical application in section 7.3, namely

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{H}\boldsymbol{\xi}_t,\\
\boldsymbol{\xi}_t &= \mathbf{c}(\boldsymbol{\theta}) + \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1} + \mathbf{M}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t,\\
\boldsymbol{\varepsilon}_t &= \boldsymbol{\alpha}(\boldsymbol{\varphi}) + \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})\boldsymbol{\beta} + \zeta_t^{-1/2}\boldsymbol{\Upsilon}^{1/2}(\boldsymbol{\varphi})\mathbf{z}_t,\\
\zeta_t|\boldsymbol{\theta},\boldsymbol{\varphi} &\sim iid\ \Gamma(\nu/2, 1/2),\\
\mathbf{z}_t|\boldsymbol{\theta},\boldsymbol{\varphi} &\sim iid\ N(\mathbf{0}, \mathbf{I}_K),
\end{aligned}$$

where $\boldsymbol{\theta}$ are mean-variance parameters and $\boldsymbol{\varphi} = (\nu, \boldsymbol{\beta}')'$ are shape parameters describing the asymmetric Student t distribution (a member of the GH family of distributions). More specifically, $\boldsymbol{\theta} = (\mu, \delta, \rho_x, \rho_{\epsilon_E}, \rho_{\epsilon_I}, \sigma_x^2, \sigma_{v_E}^2, \sigma_{v_I}^2)'$,

$$\mathbf{y}_t = \begin{pmatrix} y_{Et} \\ y_{It} \end{pmatrix}, \boldsymbol{\xi}_t = \begin{pmatrix} x_t \\ x_{t-1} \\ \epsilon_{Et} \\ \epsilon_{It} \end{pmatrix}, \boldsymbol{\varepsilon}_t = \begin{pmatrix} f_t \\ v_{Et} \\ v_{It} \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{c}(\boldsymbol{\theta}) = \begin{pmatrix} (1-\rho_x)\mu \\ 0 \\ (1-\rho_{\epsilon_E})\delta/2 \\ -(1-\rho_{\epsilon_I})\delta/2 \end{pmatrix}, \mathbf{F}(\boldsymbol{\theta}) = \begin{pmatrix} 1+\rho_x & -\rho_x & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \rho_{\epsilon_E} & 0 \\ 0 & 0 & 0 & \rho_{\epsilon_E} \end{pmatrix}, \mathbf{M}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma_x & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma_{v_E} & 0 \\ 0 & 0 & \sigma_{v_I} \end{pmatrix},$$

$$\boldsymbol{\alpha}(\boldsymbol{\varphi}) = -a(\boldsymbol{\varphi})\boldsymbol{\beta}, \text{ and } \boldsymbol{\Upsilon}(\boldsymbol{\varphi}) = (\nu-2)\left\{\mathbf{I}_K + \frac{[a(\boldsymbol{\varphi})-1]}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right\}, \text{ with}$$

$$a(\boldsymbol{\varphi}) = \frac{-(\nu-4)+\sqrt{(\nu-4)^2+8(\nu-4)\boldsymbol{\beta}'\boldsymbol{\beta}}}{4\boldsymbol{\beta}'\boldsymbol{\beta}}.$$

We produce draws from the posterior distribution by means of a Gibbs sampler in which we augment the original parameter space, consisting of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, with the state variables $\boldsymbol{\xi}_{0:T} = \{\boldsymbol{\xi}_t\}_{t=0}^T$ and the mixing variables $\zeta_{1:T} = \{\zeta_t\}_{t=1}^T$. Throughout, we implicitly assume prior independence between $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$.

Given $\mathbf{y}_{1:T}$ and initial values $(\boldsymbol{\theta}^0, \boldsymbol{\varphi}^0, \boldsymbol{\xi}_{0:T}^0)$, we draw, for $s = 1, \ldots, S$, in the following way:

Block I: $\zeta_{1:T}^s \sim p(\zeta_{1:T}|\boldsymbol{\theta}^{s-1}, \boldsymbol{\xi}_{0:T}^{s-1}, \boldsymbol{\varphi}^{s-1}, \mathbf{y}_{1:T})$, which is given by

$$\begin{aligned}\zeta_t|\boldsymbol{\theta}, \boldsymbol{\varphi}, \boldsymbol{\xi}_{0:T}, \mathbf{y}_{1:T} \sim GIG\left(\frac{K+\nu}{2}, \sqrt{(\nu-2)a(\boldsymbol{\varphi})\boldsymbol{\beta}'\boldsymbol{\beta}}, \sqrt{\mathrm{q}_t+1}\right),\\ \mathrm{q}_t = \mathbf{p}_t'\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\varphi})\mathbf{p}_t,\\ \mathbf{p}_t = [\mathbf{M}'(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta})]^{-1}\mathbf{M}'(\boldsymbol{\theta})[\boldsymbol{\xi}_t - \mathbf{c}(\boldsymbol{\theta}) - \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1}] + a(\boldsymbol{\varphi})\boldsymbol{\beta}.\end{aligned}$$

Dapugnar (1989) developed a generator of GIG pseudo-random numbers based on the ratio-of-uniforms method. In our practical implementation, we switch to a generator of gamma pseudo-random numbers whenever the norm of $\boldsymbol{\beta}$ is below the square root of $\beta_{\text{tolerance}} = 10^{-3}$ as the generator may become inefficient and unstable when the GIG distribution approaches the gamma. We also set $a(\boldsymbol{\varphi}) = 1$ and $\boldsymbol{\Upsilon}(\boldsymbol{\varphi}) = (\nu-2)\mathbf{I}_K$ for small values of the norm of $\boldsymbol{\beta}$.

Block II: $\boldsymbol{\xi}_{0:T}^s \sim p(\boldsymbol{\xi}_{0:T}|\boldsymbol{\theta}^{s-1}, \boldsymbol{\varphi}^{s-1}, \zeta_{1:T}^s, \mathbf{y}_{1:T})$, which is obtained from a modified version of the simulation smoother in Durbin and Koopman (2002) (see also Koopman and Durbin (1998)). Specifically, we proceed as follows. First of all, we note that, conditional on $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$ and $\zeta_{1:T}$, the system above admits the following representation as a Gaussian linear state space model:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{H}\boldsymbol{\xi}_t,\\ \boldsymbol{\xi}_t &= \mathbf{c}_t(\boldsymbol{\theta}, \boldsymbol{\varphi}) + \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1} + \mathbf{M}_t(\boldsymbol{\theta}, \boldsymbol{\varphi})\mathbf{z}_t,\end{aligned}$$

where

$$\begin{aligned}\mathbf{c}_t(\boldsymbol{\theta}, \boldsymbol{\varphi}) &= \mathbf{c}(\boldsymbol{\theta}) + \mathbf{M}(\boldsymbol{\theta})[\boldsymbol{\alpha}(\boldsymbol{\varphi}) + \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})\boldsymbol{\beta}],\\ \mathbf{M}_t(\boldsymbol{\theta}, \boldsymbol{\varphi}) &= \zeta_t^{-1/2}\mathbf{M}(\boldsymbol{\theta})\boldsymbol{\Upsilon}^{1/2}(\boldsymbol{\varphi}).\end{aligned}$$

The algorithm has three parts:

1. We draw $\{\mathbf{z}_t^+\}_{t=1}^T$ from $\mathbf{z}_t^+ \sim iid\ N(\mathbf{0}, \mathbf{I}_K)$ and $\boldsymbol{\xi}_0^+ \sim N(\boldsymbol{\xi}_{0|0}, \mathbf{P}_{0|0})$. We compute $\{\mathbf{y}_t^+\}_{t=1}^T$ and $\{\boldsymbol{\xi}_t^+\}_{t=1}^T$ by means of the recursion

$$\begin{aligned}\boldsymbol{\xi}_t^+ &= \mathbf{c}_t(\boldsymbol{\theta}, \boldsymbol{\varphi}) + \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1}^+ + \mathbf{M}_t(\boldsymbol{\theta}, \boldsymbol{\varphi})\mathbf{z}_t^+,\\ \mathbf{y}_t^+ &= \mathbf{y}_t - \mathbf{H}\boldsymbol{\xi}_t^+.\end{aligned}$$

2. We run the Kalman filter followed by the Kalman smoother, storing the sequence of smoothed states $\{\hat{\boldsymbol{\xi}}_t\}_{t=0}^T$, where we denote $\hat{\boldsymbol{\xi}}_t = \boldsymbol{\xi}_{t|T}$. Specifically, for $t = 1, \ldots, T$ we first compute

$$\begin{aligned}\mathbf{K}_t &= \mathbf{P}_{t|t-1}\mathbf{H}'(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}')^{-1},\\ \mathbf{P}_{t|t} &= (\mathbf{I}_M - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1},\\ \mathbf{P}_{t+1|t} &= \mathbf{F}(\boldsymbol{\theta})\mathbf{P}_{t|t}\mathbf{F}(\boldsymbol{\theta})' + \mathbf{M}_{t+1}(\boldsymbol{\theta}, \boldsymbol{\varphi})\mathbf{M}_{t+1}'(\boldsymbol{\theta}, \boldsymbol{\varphi}),\\ \boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t^+ - \mathbf{H}\boldsymbol{\xi}_{t|t-1}),\\ \boldsymbol{\xi}_{t+1|t} &= \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t|t}.\end{aligned}$$

Then, for $\tau = 1, \ldots, T-1$ we compute

$$\mathbf{J}_{T-\tau} = \mathbf{P}_{T-\tau|T-\tau}\mathbf{F}(\boldsymbol{\theta})'\mathbf{P}^{-1}_{T-\tau+1|T-\tau},$$
$$\hat{\boldsymbol{\xi}}_{T-\tau} = \boldsymbol{\xi}_{T-\tau|T-\tau} + \mathbf{J}_{T-\tau}(\hat{\boldsymbol{\xi}}_{T-\tau+1} - \boldsymbol{\xi}_{T-\tau+1|T-\tau+1}).$$

Notice that we have neglected the time-varying constants in the state-transition equation (see Jarocinski (2015) for details).

3. We compute $\{\boldsymbol{\xi}^*_t\}_{t=0}^T$ as $\boldsymbol{\xi}^*_t = \boldsymbol{\xi}^+_t + \hat{\boldsymbol{\xi}}_t$ for $t = 0, \ldots, T$.

It turns out $\boldsymbol{\xi}^*_{0:T}$ is a draw from $p(\boldsymbol{\xi}_{0:T}|\boldsymbol{\theta}, \boldsymbol{\varphi}, \zeta_{1:T}, \mathbf{y}_{1:T})$ as desired.

Block III: $\boldsymbol{\varphi}^s \sim p(\boldsymbol{\varphi}|\boldsymbol{\theta}^{s-1}, \boldsymbol{\xi}^s_{0:T}, \zeta^s_{1:T}, \mathbf{y}_{1:T})$, which we obtain by implementing an Adaptive Rejection Metropolis Sampler (ARMS, see Gilks and Wild (1992) and Gilks, Best, and Tan (1995)). We note that $\boldsymbol{\varepsilon}^{s-1}_{1:T} = \{\boldsymbol{\varepsilon}^{s-1}_t\}_{t=1}^T$, where

$$\boldsymbol{\varepsilon}^{s-1}_t = [\mathbf{M}'(\boldsymbol{\theta}^{s-1})\mathbf{M}(\boldsymbol{\theta}^{s-1})]^{-1}\mathbf{M}'(\boldsymbol{\theta}^{s-1})[\boldsymbol{\xi}^{s-1}_t - \mathbf{c}(\boldsymbol{\theta}^{s-1}) - \mathbf{F}(\boldsymbol{\theta}^{s-1})\boldsymbol{\xi}^{s-1}_{t-1}],$$

has the sufficiency property $\boldsymbol{\varphi}|\boldsymbol{\theta}^{s-1}, \boldsymbol{\xi}^{s-1}_{0:T}, \zeta^s_{1:T}, \mathbf{y}_{1:T} \sim \boldsymbol{\varphi}|\boldsymbol{\varepsilon}^{s-1}_{1:T}, \zeta^s_{1:T}$. In addition,

$$p(\boldsymbol{\varphi}|\boldsymbol{\varepsilon}_{1:T}, \zeta_{1:T}) \propto \left[\prod_{t=1}^T p(\boldsymbol{\varepsilon}_t|\boldsymbol{\varepsilon}_{1:t-1}, \boldsymbol{\varphi}, \zeta_{1:t})p(\zeta_t|\boldsymbol{\varepsilon}_{1:t-1}, \boldsymbol{\varphi}, \zeta_{1:t-1})\right] p(\boldsymbol{\varphi}),$$
$$\boldsymbol{\varepsilon}_t|\boldsymbol{\varepsilon}_{1:t-1}, \boldsymbol{\varphi}, \zeta_{1:t} \sim N[\boldsymbol{\alpha}(\boldsymbol{\varphi}) + \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})\boldsymbol{\beta}, \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})],$$
$$\zeta_t|\boldsymbol{\varepsilon}_{1:t-1}, \boldsymbol{\varphi}, \zeta_{1:t-1} \sim \Gamma\left(\nu/2, 1/2\right).$$

Thus, the log-likelihood we employ (up to an additive term constant in $\boldsymbol{\varphi}$) is

$$\mathcal{L}(\boldsymbol{\varphi}) = -\frac{T}{2}\log\{\det[\boldsymbol{\Upsilon}(\boldsymbol{\varphi})]\} - \frac{1}{2}\sum_{t=1}^T \tilde{\boldsymbol{\varepsilon}}_t'\tilde{\boldsymbol{\varepsilon}}_t - T\left[\frac{\nu}{2}\log(2) + \log\Gamma\left(\frac{\nu}{2}\right)\right] + \frac{\nu}{2}\sum_{t=1}^T \log(\zeta_t),$$

where

$$\tilde{\boldsymbol{\varepsilon}}_t \equiv \zeta_t^{1/2}\boldsymbol{\Upsilon}^{-1/2}(\boldsymbol{\varphi})[\boldsymbol{\varepsilon}_t - \boldsymbol{\alpha}(\boldsymbol{\varphi}) - \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})\boldsymbol{\beta}].$$

We apply ARMS to each parameter in turn. Let ϑ be the result of applying a certain transformation to the specific entry of $\boldsymbol{\varphi}$ being updated. In particular, for the parameter ν we let $\vartheta = \nu_{\min}/\nu$ (we take $\nu_{\min} = 4$) while for β_j we use $\vartheta = [1 + \exp(-\beta_j)]^{-1}$, $j = x, 1, 2$. The transformation is chosen in all cases to ensure $\vartheta \in [0, 1]$.

Let ϑ^0 be the starting value and $\mathcal{L}^0$ its log-posterior. ARMS updates ϑ^0 to ϑ^1 as follows:

1. Construct a grid $\vartheta_1, \ldots, \vartheta_{n_{\text{ARMS}}}$ and compute their log-posteriors $\mathcal{L}_1, \ldots, \mathcal{L}_{n_{\text{ARMS}}}$.

2. Form the piecewise-linear function h given by

$$h(\vartheta) = \max\left\{\mathrm{L}_j(\vartheta), \min\left[\mathrm{L}_{j-1}(\vartheta), \mathrm{L}_{j+1}(\vartheta)\right]\right\}, \qquad \vartheta_j < \vartheta \leq \vartheta_{j+1},$$
$$\mathrm{L}_j(\vartheta) = \mathcal{L}_j + \mathcal{L}_{j+1}\frac{(\vartheta - \vartheta_j)}{(\vartheta_{j+1} - \vartheta_j)}.$$

Next, draw ϑ^* from the piecewise exponencial distribution with density proportional to $\exp[h(\vartheta)]$. In other words, draw first a sub-interval and, conditioning on it, from a scaled truncated exponential distribution. Compute the associated log-posterior $\mathcal{L}^*$.

3. Draw $u_{\mathrm{ARS}} \sim \mathrm{U}[0,1]$. If $\log(u_{\mathrm{ARS}}) > \mathcal{L}^* - h(\vartheta^*)$, augment the grid of ϑ by ϑ^* and that of $\mathcal{L}$ by $\mathcal{L}^*$ and go back to 2. Otherwise, move on to 4.

4. Draw $u_{\mathrm{MH}} \sim \mathrm{U}[0,1]$. If $\log(u_{\mathrm{MH}}) > \mathcal{L}^* - \mathcal{L}^0$, set $\vartheta^1 = \vartheta^0$. Otherwise, set $\vartheta^1 = \vartheta^*$.

In the implementation, each draw $\boldsymbol{\varphi}^s$ is obtained by repeating the algorithm above n_{MH} times before proceeding with the Gibbs sampler.

We have also considered Slice Sampling (SS, see Neal (2003)) as an alternative method to update ϑ^0 to ϑ^1. The alternative sampling is done as follows:

1. Draw $e \sim \exp(1)$ (so that $y = \exp(\mathcal{L}^0 - e) \sim \mathrm{U}[0, \exp(\mathcal{L}^0)]$).

2. Given a positive real number w, draw $u \sim \mathrm{U}[0,1]$ and let $\vartheta_{\mathrm{L}} = \max\{\vartheta^0 - uw, 0\}$ and $\vartheta_{\mathrm{R}} = \min\{\mathrm{L} + w, 1\}$. Let $\mathcal{L}_{\mathrm{L}}$ and $\mathcal{L}_{\mathrm{R}}$ be their respective log-posteriors.
Given an integer m_{SS}, draw $v \sim \mathrm{U}[0,1]$ and form $m_{\mathrm{L}} = v m_{\mathrm{SS}}$ and $m_{\mathrm{R}} = m_{\mathrm{SS}} - 1 - m_{\mathrm{L}}$. While $\mathcal{L}_{\mathrm{L}} > \mathcal{L}^0 - e$ and $m_{\mathrm{L}} > 0$ update ϑ_{L} to $\max\{\vartheta_{\mathrm{L}} - w, 0\}$ (recomputing $\mathcal{L}_{\mathrm{L}}$) and m_{L} to $m_{\mathrm{L}} - 1$. Likewise, update ϑ_{R} to $\min\{\vartheta_{\mathrm{R}} + w, 1\}$ (recomputing $\mathcal{L}_{\mathrm{R}}$) and m_{R} to $m_{\mathrm{R}} - 1$ while $\mathcal{L}_{\mathrm{R}} > \mathcal{L}^0 - e$ and $m_{\mathrm{R}} > 0$.

3. Draw $\vartheta^* \sim \mathrm{U}[\vartheta_{\mathrm{L}}, \vartheta_{\mathrm{R}}]$ and let $\mathcal{L}^*$ be its log-posterior. While $\mathcal{L}^* < \mathcal{L}^0 - e$, either $\vartheta^* < \vartheta^0$, in which case update ϑ_{L} to ϑ^*, or $\vartheta^* \geq \vartheta^0$, in which case update ϑ_{R} to ϑ^*. Re-draw $\vartheta^* \sim \mathrm{U}[\vartheta_{\mathrm{L}}, \vartheta_{\mathrm{R}}]$ and re-compute $\mathcal{L}^*$. When this process terminates, set $\vartheta^1 = \vartheta^*$.

We report the output of the algorithm based on ARMS but our results are robust to the sampling method.

Block IV: $\boldsymbol{\theta}^s \sim p(\boldsymbol{\theta}|\boldsymbol{\xi}^s_{0:T}, \boldsymbol{\varphi}^s, \zeta^s_{1:T}, \mathbf{y}_{1:T})$, which is obtained in blocks. First, we note the sufficiency property $\boldsymbol{\theta}|\boldsymbol{\xi}^s_{0:T}, \boldsymbol{\varphi}^s, \zeta^s_{1:T}, \mathbf{y}_{1:T} \sim \boldsymbol{\theta}|\boldsymbol{\xi}^s_{0:T}, \boldsymbol{\varphi}^s, \zeta^s_{1:T}$. Next, we partition $\boldsymbol{\theta} = (\boldsymbol{\theta}'_c, \boldsymbol{\theta}'_\rho, \boldsymbol{\theta}'_\gamma)'$, with $\boldsymbol{\theta}_c = (\mu, \delta)'$, $\boldsymbol{\theta}_\rho = (\rho_x, \rho_{\epsilon_E}, \rho_{\epsilon_I})'$ and $\boldsymbol{\theta}_\sigma = (\sigma^2_x, \sigma^2_{v_E}, \sigma^2_{v_I})'$. We proceed as follows:

1. We set a Gaussian prior on $\boldsymbol{\theta}_c$ given by $\boldsymbol{\theta}_c \sim N(\underline{\mathbf{c}}, \underline{\mathbf{S}}_c)$ and we draw from the posterior $\boldsymbol{\theta}_c|\boldsymbol{\theta}^{s-1}_\rho, \boldsymbol{\theta}^{s-1}_\sigma, \boldsymbol{\xi}^s_{0:T}, \boldsymbol{\varphi}^s, \zeta^s_{1:T}$, which is

$$\boldsymbol{\theta}_c|\boldsymbol{\theta}_\rho, \boldsymbol{\theta}_\sigma, \boldsymbol{\xi}_{0:T}, \boldsymbol{\varphi}, \zeta_{1:T} \sim N\left(\bar{\mathbf{c}}, \bar{\mathbf{S}}_c\right), \text{ with } \bar{\mathbf{c}} = \bar{\mathbf{S}}_c(\underline{\mathbf{S}}_c^{-1}\underline{\mathbf{c}} + \hat{\mathbf{S}}_c^{-1}\hat{\mathbf{c}}) \text{ and } \bar{\mathbf{S}}_c = (\underline{\mathbf{S}}_c^{-1} + \hat{\mathbf{S}}_c^{-1})^{-1},$$

where

$$\hat{\mathbf{S}}_c = \left[\sum_{t=1}^T \zeta_t\right]^{-1} [\mathbf{D}_c \mathbf{M}(\boldsymbol{\theta}) \boldsymbol{\Upsilon}(\boldsymbol{\varphi}) \mathbf{M}(\boldsymbol{\theta}) \mathbf{D}_c], \text{ and } \hat{\mathbf{c}} = \left[\sum_{t=1}^T \zeta_t\right]^{-1} \left[\sum_{t=1}^T \sqrt{\zeta_t} \mathbf{Y}_{ct}\right],$$

with

$$\mathbf{D}_c = \begin{bmatrix} 1-\rho_x & 0 & 0 & 0 \\ 0 & 0 & 1-\rho_{\epsilon_E} & -(1-\rho_{\epsilon_I}) \end{bmatrix}$$

and

$$\mathbf{Y}_{ct} = \mathbf{D}_c\{\boldsymbol{\xi}_t - \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1} - \mathbf{M}(\boldsymbol{\theta})[\boldsymbol{\alpha}(\boldsymbol{\varphi}) + \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})\boldsymbol{\beta}]\}.$$

2. We set a Gaussian prior on $\boldsymbol{\theta}_\rho$ given by $\boldsymbol{\theta}_\rho \sim N(\underline{\boldsymbol{\rho}}, \underline{\mathbf{S}}_\rho)$ and we draw from the posterior $\boldsymbol{\theta}_\rho|\boldsymbol{\theta}_c^s, \boldsymbol{\theta}_\sigma^{s-1}, \boldsymbol{\xi}_{0:T}^s, \boldsymbol{\varphi}^s, \zeta_{1:T}^s$, which is

   $\boldsymbol{\theta}_\rho|\boldsymbol{\theta}_c, \boldsymbol{\theta}_\sigma, \boldsymbol{\xi}_{0:T}, \boldsymbol{\varphi}, \zeta_{1:T} \sim N\left(\bar{\boldsymbol{\rho}}, \bar{\mathbf{S}}_\rho\right)$, with $\bar{\boldsymbol{\rho}} = \bar{\mathbf{S}}_\rho(\underline{\mathbf{S}}_\rho^{-1}\underline{\boldsymbol{\rho}} + \hat{\mathbf{S}}_\rho^{-1}\hat{\boldsymbol{\rho}})$ and $\bar{\mathbf{S}}_\rho = (\underline{\mathbf{S}}_\rho^{-1} + \hat{\mathbf{S}}_\rho^{-1})^{-1}$,

   where

$$\hat{\mathbf{S}}_\rho = \left[\sum_{t=1}^T \mathbf{X}_{\rho t}^2\right]^{-1} \text{ and } \hat{\boldsymbol{\rho}} = \left[\sum_{t=1}^T \mathbf{X}_{\rho t}^2\right]^{-1}\left[\sum_{t=1}^T \mathbf{X}_{\rho t}\mathbf{Y}_{\rho t}\right],$$

   with

$$\begin{aligned} \mathbf{X}_{\rho t} &= \zeta_t^{1/2}\mathrm{diag}[\mathbf{W}_\rho\mathbf{D}_\rho\{\boldsymbol{\xi}_{t-1} - [\mathbf{I}_K - \mathbf{D}_\rho\mathbf{F}(\boldsymbol{\theta})]^{-1}\mathbf{C}(\boldsymbol{\theta})\}], \\ \mathbf{Y}_{\rho t} &= \zeta_t^{1/2}\mathbf{W}_\rho\mathbf{D}_\rho\{\boldsymbol{\xi}_t - [\mathbf{I}_K - \mathbf{D}_\rho\mathbf{F}(\boldsymbol{\theta})]^{-1}\mathbf{C}(\boldsymbol{\theta}) - \mathbf{M}(\boldsymbol{\theta})[\boldsymbol{\alpha}(\boldsymbol{\varphi}) + \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})\boldsymbol{\beta}]\}, \end{aligned}$$

$$\mathbf{W}_\rho = [\mathbf{D}_\rho\mathbf{M}(\boldsymbol{\theta})\boldsymbol{\Upsilon}^{1/2}(\boldsymbol{\varphi})]^{-1} \text{ and } \mathbf{D}_\rho \equiv \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

3. We set an inverse gamma prior on $\boldsymbol{\theta}_\sigma$ given by $\sigma_j^{-2} \sim \Gamma(\nu_j/2, \varsigma_j/2)$, for $j = x, v_E, v_I$, with these parameters being prior-independent across j. However, for the purposes of generating draws from the posterior distribution $\boldsymbol{\theta}_\sigma|\boldsymbol{\theta}_c^s, \boldsymbol{\theta}_\rho^s, \boldsymbol{\xi}_{0:T}^s, \boldsymbol{\varphi}^s, \zeta_{1:T}^s$, we need to consider two separate cases.

   If $\boldsymbol{\beta} = \mathbf{0}$, the three parameters are posterior-independent and direct sampling can be implemented because the prior conjugates with the likelihood. More formally,

$$\sigma_j^{-2}|\boldsymbol{\theta}_c, \boldsymbol{\theta}_\rho, \boldsymbol{\xi}_{0:T}, \boldsymbol{\varphi}, \zeta_{1:T} \sim \Gamma\left[\frac{T+\nu_j}{2}, \frac{1}{2}\left(\sum_{t=1}^T \eta_{jt}^2 + \varsigma_j\right)\right], \text{ for } j = x, v_E, v_I,$$

   where

$$\boldsymbol{\eta}_t = \begin{bmatrix} \eta_{xt} \\ \eta_{v_E t} \\ \eta_{v_I t} \end{bmatrix} = \zeta_t^{1/2}\boldsymbol{\Upsilon}^{-1/2}(\boldsymbol{\theta})\mathbf{D}_\rho[\boldsymbol{\xi}_t - \mathbf{C}(\boldsymbol{\theta}) - \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1}].$$

   On the other hand, if $\boldsymbol{\beta} \neq \mathbf{0}$, direct sampling is not available. In this case, we generate draws from the posterior distribution by componentwise application of ARMS. The log-likelihood that we employ (up to an additive term constant in $\boldsymbol{\theta}_\sigma$) is

$$\mathcal{L}(\boldsymbol{\theta}_\sigma) = -\frac{T}{2}[\log(\sigma_x^2) + \log(\sigma_{v_E}^2) + \log(\sigma_{v_I}^2)] - \frac{1}{2}\sum_{t=1}^T \tilde{\boldsymbol{\eta}}_t'\tilde{\boldsymbol{\eta}}_t,$$

   with

$$\begin{aligned} \tilde{\boldsymbol{\eta}}_t &= \zeta_t^{1/2}\boldsymbol{\Upsilon}^{-1/2}(\boldsymbol{\varphi})\mathrm{diag}(\sigma_x^{-1}, \sigma_{v_E}^{-1}, \sigma_{v_I}^{-1})\mathbf{D}_\rho \\ &\quad \times[\boldsymbol{\xi}_t - \mathbf{C}(\boldsymbol{\theta}) - \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\xi}_{t-1} - \boldsymbol{\alpha}(\boldsymbol{\varphi}) - \zeta_t^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\varphi})\boldsymbol{\beta}]. \end{aligned}$$

The procedure is exactly as explained above. Again, we also employed slice sampling, with our results being robust to this variation.

E Additional Monte Carlo results (HAC)

Table E1: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the bivariate cointegrated, dynamic single factor model ($T = 100$)

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%)					
					Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_J	Kt	0.17	1.71	4.56	24.76	2.29	15.76	30.25	2.94	19.88
	Sk	2.75	9.55	16.99	8.60	9.67	8.55	20.60	12.29	15.96
	GH	2.20	7.98	14.20	15.89	8.31	12.47	34.06	10.92	23.73
$H_{S_{\mathbf{f}}}$	Kt	0.17	1.67	4.62	5.14	3.75	2.13	9.13	4.50	2.94
	Sk	1.35	6.37	12.77	6.16	6.34	6.13	11.00	12.46	6.39
	GH	0.75	4.06	9.02	6.07	5.31	4.34	13.72	11.78	5.19
$H_{S_{\mathbf{v}}}$	Kt	0.17	1.64	4.77	18.89	1.73	17.50	24.15	2.00	20.95
	Sk	1.67	7.55	13.97	6.65	7.33	7.06	14.73	7.91	17.73
	GH	1.17	5.80	11.44	12.19	5.74	11.72	25.37	5.83	27.46
Red	Kt	0.25	1.89	5.13	24.17	2.51	13.68	28.05	3.40	17.73
	Sk	1.52	6.57	13.12	6.80	7.07	6.41	15.39	8.46	6.81
	GH	1.00	5.21	10.87	14.35	5.94	9.68	28.67	7.84	12.08

Notes: Results based on 10,000 samples of size $T = 100$ from model (16) with $\rho_x = .5$, $\rho_{\epsilon_E} = .2$, $\rho_{\epsilon_I} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_E = 2$ and $q_I = .5$, where $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2$ represents the signal-to-noise ratio for y_{it} for $i = E, I$. The column labels J, S_f, $S_{\mathbf{v}}$ refer to the alternative $\boldsymbol{\varepsilon}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$ (i.e. $R = 3$), $f_t \sim GH(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ ($R = 1$) and $\mathbf{v}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ ($R = 2$), respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f, and $S_{\mathbf{v}}$ alternative hypotheses, while Red denotes the reduced form tests discussed in section 5.4.2. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table E2: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the bivariate cointegrated, dynamic single factor model ($T = 250$)

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%)					
					Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_J	Kt	0.12	1.79	5.15	61.57	3.57	42.37	65.21	5.22	49.08
	Sk	1.61	7.05	12.87	6.15	7.45	6.23	43.14	14.99	29.48
	GH	1.17	5.77	10.80	27.25	6.48	18.65	68.37	14.98	52.18
$H_{S_{\mathbf{f}}}$	Kt	0.13	1.73	4.97	11.68	6.51	2.35	20.10	8.69	4.21
	Sk	1.13	6.06	12.15	5.64	5.42	5.90	18.32	27.79	7.48
	GH	0.56	3.94	8.70	8.23	5.77	4.21	27.72	28.67	6.26
$H_{S_{\mathbf{v}}}$	Kt	0.12	1.48	4.86	50.02	1.99	46.26	58.78	2.45	52.70
	Sk	1.31	5.94	11.84	4.88	6.36	4.98	32.34	6.23	39.44
	GH	0.95	4.38	9.29	22.91	4.98	21.18	59.80	5.03	63.81
Red	Kt	0.15	1.85	5.65	59.11	3.96	34.63	61.54	5.98	43.07
	Sk	1.29	6.14	11.61	5.63	6.12	5.25	35.12	12.29	6.82
	GH	0.85	4.70	9.70	28.61	5.47	16.30	62.18	12.54	24.03

Notes: Results based on 10,000 samples of size $T = 250$ from model (16) with $\rho_x = .5$, $\rho_{\epsilon_E} = .2$, $\rho_{\epsilon_I} = .8$, $\sigma_f^2 = 1$ and $\sigma_{v_i}^2$ chosen such that $q_E = 2$ and $q_I = .5$, where $q_i = \sigma_x^2/\sigma_{\epsilon_i}^2$ represents the signal-to-noise ratio for y_{it} for $i = E, I$. The column labels J, S_f, $S_{\mathbf{v}}$ refer to the alternative $\boldsymbol{\varepsilon}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$ (i.e. $R = 3$), $f_t \sim GH(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ ($R = 1$) and $\mathbf{v}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ ($R = 2$), respectively. The row labels H_J, H_{S_f}, and $H_{S_{\mathbf{v}}}$ refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f, and $S_{\mathbf{v}}$ alternative hypotheses, while Red denotes the reduced form tests discussed in section 5.4.2. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table E3: Monte Carlo rejection rates (in %) under the null and alternative hypotheses for the local-level model

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%)					
					Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	S_v	J	S_f	S_v
H_J	Kt	0.08	1.52	4.77	23.84	7.06	3.41	35.64	11.68	6.77
	Sk	1.33	6.15	12.01	4.90	5.70	5.38	41.94	24.15	11.18
	GH	0.73	4.65	9.57	11.60	6.60	4.98	57.08	29.14	12.57
H_{S_f}	Kt	0.10	1.60	5.08	17.91	8.72	2.24	27.32	12.57	3.53
	Sk	0.96	6.01	11.76	5.29	5.49	5.44	47.23	33.42	5.09
	GH	0.52	3.64	8.25	10.89	6.60	3.79	57.18	36.83	4.49
H_{S_v}	Kt	0.17	1.67	4.78	14.25	2.95	4.71	31.21	5.41	8.68
	Sk	1.03	5.41	10.78	3.94	5.18	4.61	24.47	4.22	15.97
	GH	0.51	3.44	7.72	8.27	3.95	4.41	41.34	4.47	18.39
Red	Kt	0.05	1.46	5.20	22.68	6.98	2.85	33.28	11.69	5.45
	Sk	1.07	5.75	11.45	4.65	5.50	5.01	55.45	28.16	6.84
	GH	0.53	3.48	8.12	12.35	6.38	4.02	64.79	31.23	6.66

Notes: Results based on 10,000 samples of size $T = 250$ from the local-level model discussed in section 5.3 in which the signal-to-noise ratio $q = \sigma_f^2/\sigma_v^2$ is set to 2. The column labels J, S_f, $S_{\mathbf{v}}$ refer to the alternative $\boldsymbol{\varepsilon}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$ $(R = 2)$, $f_t \sim GH(\eta, \psi, \beta)$, $v_t \sim N(0, 1)$ $(R = 1)$ and $v_t \sim GH(\eta, \psi, \beta)$, $f_t \sim N(0, 1)$ $(R = 1)$, respectively. The row labels H_J, H_{S_f}, and H_{S_v} refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f, and S_v alternative hypotheses, Red denotes the reduced form tests discussed in section 5.4.2, while HK denotes the original Harvey and Koopman (1992) tests discussed in section 5.4.1. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Table E4: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the multivariate local-level model

		Panel A: Null hypothesis			Panel B: Alternative hypotheses (5%) Student t			asymmetric Student t		
		1%	5%	10%	J	S_f	$S_{\mathbf{v}}$	J	S_f	$S_{\mathbf{v}}$
H_J	Kt	0.23	3.15	7.37	96.21	6.69	95.78	93.94	10.63	95.13
	Sk	8.95	21.52	31.50	18.46	20.76	18.73	69.96	33.29	40.30
	GH	8.63	20.71	30.31	75.83	21.25	74.13	92.99	35.27	85.33
$H_{S_\mathbf{f}}$	Kt	0.06	1.78	5.56	32.25	30.09	1.68	35.83	34.12	2.19
	Sk	1.18	5.73	11.57	4.49	4.44	5.59	44.24	60.98	5.06
	GH	0.62	3.68	8.33	17.26	16.19	3.77	58.05	68.64	3.65
$H_{S_\mathbf{v}}$	Kt	0.29	2.82	7.06	95.86	3.07	95.80	95.52	2.90	95.60
	Sk	7.73	18.94	28.70	16.00	18.45	16.05	63.85	19.16	40.45
	GH	7.42	18.34	27.33	74.66	17.18	74.22	93.08	18.06	86.57
Red	Kt	0.25	3.13	7.36	96.71	6.96	95.77	93.87	11.11	95.14
	Sk	7.74	18.94	28.44	16.14	18.56	16.26	62.83	31.52	26.57
	GH	7.11	17.91	27.21	75.61	19.07	73.54	91.69	33.83	79.37

Notes: Results based on 10,000 samples of size $T = 250$ from a 10-variate version of the local-level model with $\boldsymbol{\pi} = \mathbf{0}$, $\mathbf{c} = \boldsymbol{\ell}_{10}$ and $\boldsymbol{\gamma} = q^{-1}\boldsymbol{\ell}_{10}$, where q reflects the signal-to-noise ratio, which we set to 2. The column labels J, S_f, $S_\mathbf{v}$ refer to the alternative $\boldsymbol{\varepsilon}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$ (i.e. $R = 11$), $f_t \sim GH(\eta, \psi, \beta)$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_N)$ ($R = 1$) and $\mathbf{v}_t \sim GH(\eta, \psi, \boldsymbol{\beta})$, $f_t \sim N(0, 1)$ ($R = 10$), respectively. The row labels H_J, H_{S_f}, and $H_{S_\mathbf{v}}$ refer to the score tests in Propositions 4 and 7 corresponding to the J, S_f, and $S_\mathbf{v}$ alternative hypotheses. In Panel B, Student t refers to the DGP for the GH being symmetric Student t with 8 degrees of freedom and, analogously, asymmetric Student t to the asymmetric Student t with 8 degrees of freedom and skewness vector $\boldsymbol{\beta} = -\boldsymbol{\ell}_R$. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

F Inferring real output from GDP and GDI over a long sample

Table F1: Parameter estimates and normality tests over the postwar period

Panel A: ML estimates

Param.	estimate	std. err.
μ	0.755	0.110
δ	0.304	0.031
α_x	0.493	0.059
α_{ϵ_E}	0.265	0.196
α_{ϵ_I}	0.939	0.024
σ_f^2	0.526	0.054
$\sigma_{v_E}^2$	0.076	0.021
$\sigma_{v_I}^2$	0.093	0.019

Panel B: Normality tests

		statistic	p-value
H_{S_f}	Kt	19.061	0.000
	Sk	1.161	0.281
	GH	20.221	0.000
H_{S_v}	Kt	6.537	0.005
	Sk	3.859	0.145
	GH	10.396	0.011
H_R	Kt	13.266	0.000
	Sk	1.232	0.540
	GH	14.498	0.002

Notes: Data: Quarterly real GDP and GDI from 1952Q1 to 2015Q2. Model: Bivariate cointegrated, dynamic single factor model (16); see section 7 for parameter definitions. In Panel A, estimates are Gaussian ML of the bivariate Gaussian likelihood of the stationary transformation $\Delta y_{Et} + \Delta y_{It}$ and $y_{Et} - y_{It}$ in the time domain. Standard errors are obtained from the asymptotic information matrix, which is computed using its frequency domain closed-form expression. In Panel B, the row labels H_{S_f} and H_{S_v} refer to the score tests in Propositions 4 and 7 corresponding to the S_f and S_v alternative hypotheses, respectively, while Red denotes the reduced form tests discussed in section 5.4.2. For each of those labels, Kt and Sk refer to the kurtosis and skewness components of the corresponding test statistics, while GH indicates the sum of the two.

Figure F1: Smoothed innovations and influence functions for the kurtosis and skewness tests: Sample 1952Q1 to 2015Q2.

Figure F1a: Smoothed innovations for the underlying factor

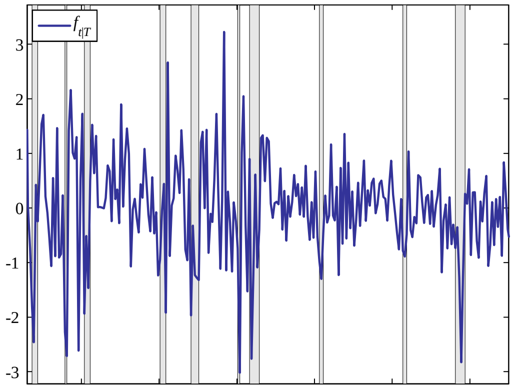


Figure F1b: Smoothed innovations for the measurement errors

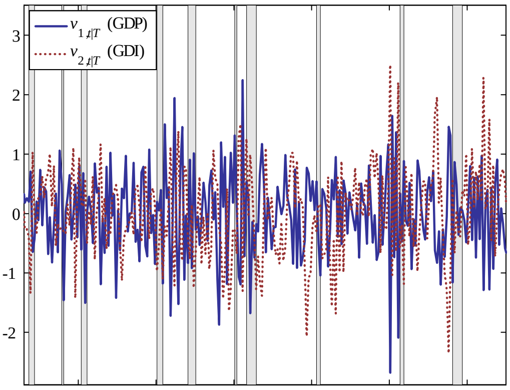


Figure F1c: Influence functions for the underlying factor (kurtosis)

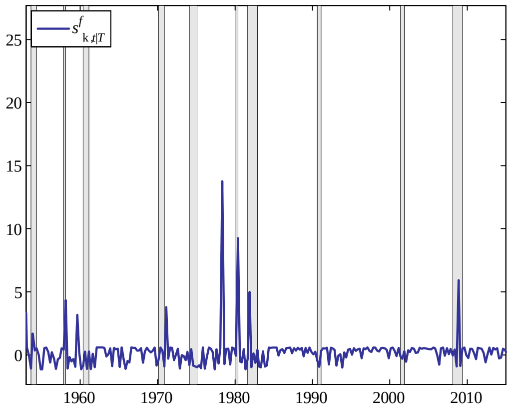


Figure F1d: Influence functions for the measurement errors (kurtosis)

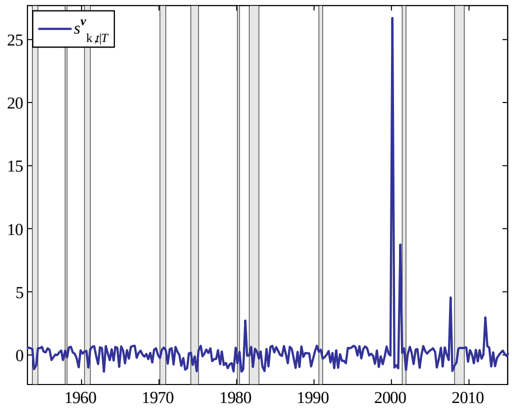


Figure F1e: Influence functions for the underlying factor (skewness)

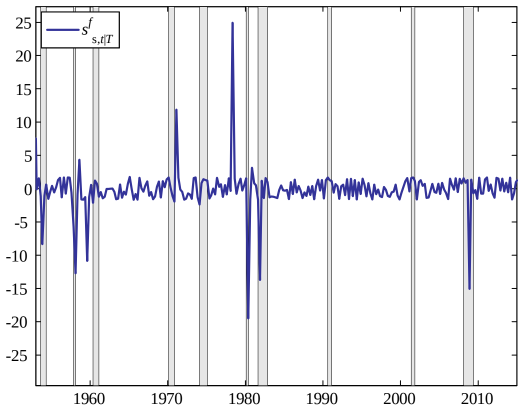
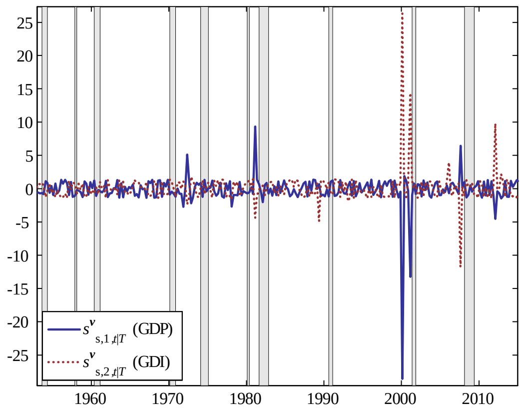


Figure F1f: Influence functions for the measurement errors (skewness)



Notes: Smoothed innovations and influence functions were obtained from fitting the bivariate cointegrated, dynamic single factor model (16) to the quarterly real GDP and GDI from 1952Q1 to 2015Q2; see Table F1 for parameter estimates. Shaded areas represent NBER recessions.

Figure F2: Posterior densities of shape parameters under the asymmetric Student t alternative: Sample 1952Q1 to 2015Q2

Figure F2a: η

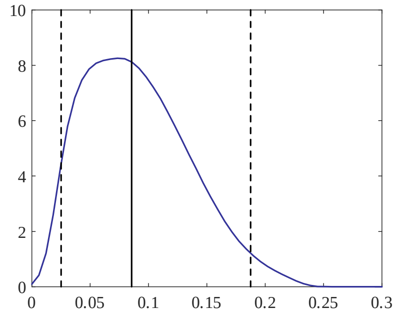


Figure F2b: β_x

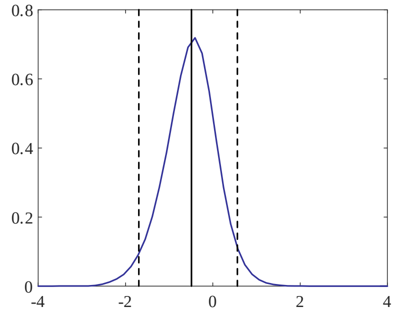


Figure F2c: β_{v_E}

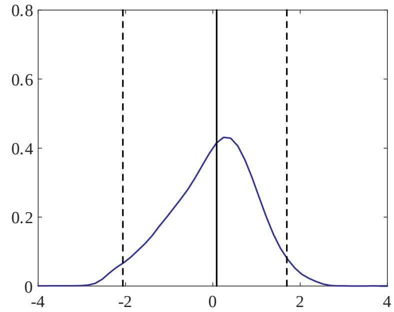
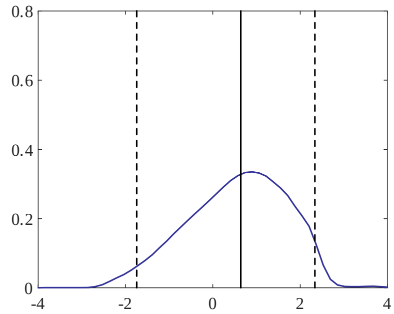


Figure F2d: β_{v_I}



Notes: Model: Bivariate cointegrated, dynamic single factor model (16) with multivariate asymmetric Student t innovations; see Section 7 for parameter definitions. η refers to the reciprocal of degrees of freedom while β_x (β_{v_E}) [β_{v_I}] refers to the skewness parameter of the "true GDP" (expenditure) [income] measure. Solid vertical lines refer to the median values while dashed lines report the 2.5% and 97.5% quantiles.

References

Billingsley, P. (1995): *Probability and measure*, 3^{rd} edition, John Wiley and Sons, New York.

Dapugnar, J. (1989): "An easily implemented, generalized inverse gaussian generator", *Communications in Statistics–Simulation* 18, 703–710.

Durbin, J. and S. J. Koopman (2002): "A simple and efficient simulation smoother for state space time series analysis", *Biometrika* 89 (3), 603–616.

Gilks, W. R., N. G. Best and K. K. C. Tan (1995): "Adaptive rejection Metropolis sampling within Gibbs sampling", *Journal of the Royal Statistical Society. Series C* 44 (4), 445–472.

Gilks, W. R., and P. Wild (1992): Adaptive rejection sampling for Gibbs sampling", *Applied Statistics* 41 (2), 337–348.

Gouriéroux C., A. Holly and A. Monfort (1980): "Kühn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters", Discussion Paper 770, Harvard Institute of Economic Research.

Jarocinski, M. (2015): "A note on implementing the Durbin and Koopman simulation smoother", European Central Bank WP No 1867.

Koopman, S. J. and J. Durbin (1998): "Fast filtering and smoothing for multivariate state space models", *Journal of Time Series Analysis* 21 (3), 281–296.

Magnus, J.R. (1988): *Linear structures*, Oxford University Press, New York.

Mencía J. and E. Sentana (2012): "Distributional tests in multivariate dynamic models with normal and Student t innovations", *Review of Economics and Statistics* 94, 133–152.

Neal, R. (2003): "Slice sampling", *The Annals of Statistics* 31 (3), 705–767.