Online Supplementary Material to "Instrumental variables inference in a small-dimensional VAR model with dynamic latent factors"

Federico Carlini[†] Patrick Gagliardini[‡]

In Appendix C we provide the proofs of Lemmas 1-4, Propositions 4 and 6-10. In Appendix D we first discuss general identification results for ABCD state space representations from the transfer function and from the spectral density function (Appendix D.1). We then apply these identification results for the state space model (2.1)-(2.2) when $p = q = 1$ to prove Proposition 5 (Appendix D.2). In Appendix E we give additional theoretical results concerning unidentifiability of VARMA representations (Section E.1), identification of the variance-covariance matrices Σ_u and Σ_v (Section E.2), and their estimation (Section E.3). We also provide results on parameter identification when some latent factors are White Noise (Section E.4), a closed-form expression for the asymptotic variance of our estimator when errors are Gaussian (Section E.5), and an algorithm to compute the asymptotic variance of the ML estimator in our state space model (Section E.6). In Appendix F we present additional tables and figures for the empirical analysis. Finally, in Appendix G we give the results of our Monte Carlo experiments. We use $|\cdot|$ to denote absolute value, Euclidean norm or Frobenius norm of a scalar, vector or matrix, respectively.

[†]Department of Economics and Finance, LUISS University Rome, Italy. fcarlini@luiss.it

[‡]Corresponding author. Faculty of Economics, Università della Svizzera italiana, Lugano, Switzerland, and SFI. patrick.gagliardini@usi.ch

C Proofs of Lemmas 1-4 and Propositions 4, 6-10

C.1 Proof of Lemma 1

From the state-space equations (2.1)-(2.2) we get:

$$
\begin{aligned}
f_t &= \bar{B}'(C(L)Y_t - u_t), && \text{(C.1)}\\
C(L)Y_t &= B[I_K - \Phi(L)]f_t + Bv_t + u_t. && \text{(C.2)}
\end{aligned}
$$

By plugging (C.1) into (C.2) we get:

$$
\left(I_n - B[I_K - \Phi(L)]\bar{B}'\right) C(L)Y_t = Bv_t + \left(I_n - B[I_K - \Phi(L)]\bar{B}'\right) u_t. \quad \text{(C.3)}
$$

The matrix polynomial on the LHS $A(L) := \left(I_n - B[I_K - \Phi(L)]\bar{B}'\right) C(L) = I_n - \sum_{j=1}^{p+q} A_j L^j$ has order $p+q$, and the matrix coefficients are:

$$
A_j = C_j + B\Phi_j\bar{B}' - \sum_{i=1}^{j-1} B\Phi_i\bar{B}'C_{j-i}, \quad j = 1, ..., p+q, \quad \text{(C.4)}
$$

where we set $C_j \equiv 0$ for $j > p$ and $\Phi_j \equiv 0$ for $j > q$. The process $Bv_t + \left(I_n - B[I_K - \Phi(L)]\bar{B}'\right) u_t = Bv_t + u_t - \sum_{j=1}^{q} B\Phi_j\bar{B}'u_{t-j}$ on the RHS has vanishing autocovariances at any order larger than q, and therefore admits a MA(q) representation: $Bv_t + u_t - \sum_{j=1}^{q} B\Phi_j\bar{B}'u_{t-j} = w_t + \sum_{j=1}^{q} \Psi_j w_{t-j} =: \Psi(L)w_t$ with $w_t \sim WN(0, \Sigma_w)$. The MA coefficients Ψ_j and the variance Σ_w satisfy:

$$
\begin{aligned}
\Sigma_w + \sum_{j=1}^{q} \Psi_j\Sigma_w\Psi_j' &= B\Sigma_v B' + \Sigma_u + \sum_{j=1}^{q} B\Phi_j\bar{B}'\Sigma_u\bar{B}\Phi_j'B', && \text{(C.5)}\\
\Psi_i\Sigma_w + \sum_{j=i+1}^{q} \Psi_j\Sigma_w\Psi_{j-i}' &= -B\Phi_i\bar{B}'\Sigma_u + \sum_{j=i+1}^{q} B\Phi_j\bar{B}'\Sigma_u\bar{B}\Phi_{j-i}'B', \quad i = 1, ..., q. && \text{(C.6)}
\end{aligned}
$$

These equations imply $\Psi_j = B\nu_j'$ for some $n \times K$ matrices ν_j, $j = 1, ..., q$. We get the VARMA($p+q$,q) representation $A(L)Y_t = \Psi(L)w_t$.

The determinant of the AR matrix polynomial in the VARMA($p+q$,q) representation is equal to:

$$\begin{aligned} detA(z) &= det\left(I_n - B[I_K - \Phi(z)]\bar{B}'\right) detC(z) \\ &= det[(\bar{B}:\bar{B}_\perp)(B:B_\perp)']det\left(I_n - B[I_K - \Phi(z)]\bar{B}'\right) detC(z) \\ &= det[(B:B_\perp)'\left(I_n - B[I_K - \Phi(z)]\bar{B}'\right)(\bar{B}:\bar{B}_\perp)]detC(z) \\ &= det\begin{bmatrix} (B'B)\Phi(z)(B'B)^{-1} & 0 \\ 0 & I_{n-K} \end{bmatrix} detC(z) = det\Phi(z)detC(z). \end{aligned}$$

Therefore, under Assumptions M.1 (iii) and M.2 (ii), the roots of $A(z)$ are outside the unit circle, and the VARMA($p+q$,q) representation in (3.11) is stationary and causal.

C.2 Proof of Lemma 2

We have to show:

$$\frac{1}{T}\sum_{t=1}^{T} Y_t Y_{t-i}' \overset{a.s.}{\to} \Gamma(i), \text{ as } T \to \infty, \tag{C.7}$$

for any $i \geq 0$. We use Theorem 4.1.1. in Hannan and Deistler (1988). Under Assumptions M.1 and M.2, process $\{(\boldsymbol{Y}_t', \boldsymbol{f}_t')'\}$ is a stable VAR(1) process, and $\{Y_t\}$ admits a MA(∞) representation $Y_t = \sum_{j=0}^{\infty} \Psi_j z_{t-j}$, where $z_t = (u_t', v_t')' \sim WN(0, \Sigma_z)$. Then, by Theorem 4.1.1. in Hannan and Deistler (1988), the strong consistency in (C.7) holds if

$$\frac{1}{T}\sum_{t=1}^{T} z_t z_{t-i}' \overset{a.s.}{\to} \Sigma_z \delta_{i,0}, \text{ as } T \to \infty, \tag{C.8}$$

for any $i \geq 0$, where $\delta_{i,0} = 1$ if $i = 0$, and $\delta_{i,0} = 0$ otherwise. Under Assumptions LS.1 (i) and (iii), process $\{X_t\}$ with $X_t := z_t z_{t-i}' - \Sigma_z \delta_{0,i}$ is a L^p-mixingale, for $p = \beta > 1$. Indeed, $|E[X_t|\mathcal{Z}_{t-m}]|_p \leq c_t \zeta_m$ for any $m \geq 0$, where $\mathcal{Z}_{t-m} = \sigma(z_{t-j}, j \geq m)$, $c_t = E[|z_t|^{2\beta}]^{1/\beta} + |\Sigma_z|$ is independent of t, and $\zeta_m = 0$ for $m \geq 1$. Moreover, $T^{-1}\sum_{t=1}^{T} c_t^p / t^p < \infty$. Then, by Corollary 2.8 in Mc Leish (1975), we have $T^{-1}\sum_{t=1}^{T} X_t \overset{a.s.}{\to} 0$ and $T \to \infty$, which yields (C.8).

C.3 Proof of Lemma 3

Condition (a) holds because $E(|u_t|^{2\beta}) < \infty$ and $E(|v_t|^{2\beta}) < \infty$ for some $\beta > 2$ from Assumption LS.1 (iii), and the components of vector ψ_t are quadratic functions of the elements of vector Y_t and its lags.

To check condition (b), let us consider process $X_t := (Y_t', f_t')'$. Vector X_t follows a VAR(p^*) process $P(L)X_t = \zeta_t$ with order $p^* = \max\{p, q\}$ and lag polynomial $P(z) = I_{n+K} - \sum_{j=1}^{p^*} P_j z^j$, where

$$P_j = \begin{pmatrix} C_j & B\Phi_j \\ 0 & \Phi_j \end{pmatrix} \qquad \zeta_t = \begin{pmatrix} u_t + Bv_t \\ v_t \end{pmatrix},$$

and we set $C_j = 0$ for $j > p$ and $\Phi_j = 0$ for $j > q$. Under Assumptions M.1 (iii) and M.2 (ii), the roots of $\det P(z) = \det C(z) \det \Phi(z)$ lie outside the complex unit circle. Moreover, under Assumption LS.1 (ii), process ζ_t is i.i.d. with positive density function $f(\zeta) = f_u(\zeta_1 - B\zeta_2) f_v(\zeta_2) > 0$. From Theorem 1 in Mokkadem (1988) it follows that X_t is geometrically strongly mixing. Process ψ_t is a measurable function of a finite number of lags of process X_t, so it is geometrically strongly mixing as well.

To check condition (c), we use that ψ_t is a stationary process, which implies

$$\frac{1}{T} V \left(\sum_{t=1}^{T} \psi_t \right) = \frac{1}{T} \sum_{j=-T+1}^{T-1} (T - |j|) cov(\psi_t, \psi_{t-j}).$$

Given that ψ_t is a geometrically strongly mixing process with α-coefficients $\alpha(j) = O(\rho^j)$ where $\rho < 1$ (Condition b)), and it has finite moments of order $\beta > 2$ (Condition a)), we can apply Serfling's inequality and we can bound $|cov(\psi_t, \psi_{t-j})| \leq \bar{K} \times \rho^{(1-2/\beta)j} |\psi_t|_\beta^2$, where $\bar{K}$ is a strictly positive constant and $|\psi_t|_\beta := E[|\psi_t|^\beta]^{1/\beta}$. The series $j\rho^{(1-2/\beta)j}$ is summable, and thus

$$\lim_{T\to\infty} \frac{1}{T} V \left(\sum_{t=1}^{T} \psi_t \right) = \lim_{T\to\infty} \sum_{j=-T+1}^{T-1} cov(\psi_t, \psi_{t-j}) = \sum_{j=-\infty}^{\infty} cov(\psi_t, \psi_{t-j}) = V_\psi.$$

C.4 Proof of Lemma 4

We have to show the asymptotic normality:

$$\sqrt{T} vec(\hat{B}_\perp(p)' \hat{\boldsymbol{A}}^\star) \xrightarrow{d} N(0, \Sigma_{22}), \tag{C.9}$$

as $T \to \infty$, and derive the asymptotic variance matrix Σ_{22}. From Theorem 4 (c), we have $\hat{K}(p) = K_0$ w.p.a. 1, so that we can neglect asymptotically the effect of estimation of K. From similar arguments as in Proposition 7, equation (B.4), the asymptotic expansion of $\sqrt{T}vec(\hat{\boldsymbol{A}}^{\star} - \boldsymbol{A}_0^{\star})$ is given by

$$\sqrt{T}vec(\hat{\boldsymbol{A}}^{\star} - \boldsymbol{A}_0^{\star}) = [E(\tilde{X}_{t-p-2}\tilde{X}_{t-p-2}')^{-1} \otimes I_n]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{X}_{t-p-2} \otimes u_t^{\star}\right) + o_p(1),$$

where $u_t^{\star} = Y_t - EL(Y_t|Y_{t-1}, \ldots, Y_{t-p^{\star}})$ and $\tilde{X}_{t-p-2} = X_{t-p-2} - EL(X_{t-p-2}|Y_{t-1}, \ldots, Y_{t-p-1})$ and $X_{t-p-2} := (Y_{t-p-2}', ..., Y_{t-p^{\star}}')'$, while from equation (C.43) in the Online Appendix and Theorem 2 (a) we have the asymptotic expansion:

$$\sqrt{T}vec[(\hat{B}_{\perp}(p) - B_{0\perp})'] = -A_{0\perp}'\sqrt{T}(\hat{b} - b_0) = -A_{0\perp}'S_{b_1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{Y}_{t-p-1} \otimes u_t^{*} + o_p(1),$$

where $u_t^{*} = Y_t - EL(Y_t|Y_{t-1}, \ldots, Y_{t-p-1})$, $\tilde{Y}_{t-p-1} = Y_{t-p-1} - EL(Y_{t-p-1}|Y_{t-1}, \ldots, Y_{t-p})$ and $A_{0\perp} = [0_{K\times(n-K)} : I_K] \otimes I_{n-K}$. Thus, we get:

$$\begin{aligned}
\sqrt{T}vec(\hat{B}_{\perp}(p)'\hat{\boldsymbol{A}}^{\star}) &= \sqrt{T}vec\{(\hat{B}_{\perp} - B_{0\perp} + B_{0\perp})'(\hat{\boldsymbol{A}}^{\star} - \boldsymbol{A}_0^{\star} + \boldsymbol{A}_0^{\star})\} - \sqrt{T}vec(B_{0\perp}'\boldsymbol{A}_0^{\star}) \\
&= \sqrt{T}vec\{(\hat{B}_{\perp} - B_{0\perp})'\boldsymbol{A}_0^{\star}\} + \sqrt{T}vec\{B_{0\perp}'(\hat{\boldsymbol{A}}^{\star} - \boldsymbol{A}_0^{\star})\} + o_p(1) \\
&= (\boldsymbol{A}_0^{\star\prime} \otimes I_{n-K_0})\sqrt{T}vec[(\hat{B}_{\perp} - B_{\perp 0})'] + (I_{(p^{\star}-p-1)n} \otimes B_{0\perp}')\sqrt{T}vec(\hat{\boldsymbol{A}}^{\star} - \boldsymbol{A}_0^{\star}) + o_p(1) \\
&= -([\boldsymbol{A}_0^{\star\prime}[0_{K_0\times(n-K_0)} : I_K]'] \otimes I_{n-K_0})\sqrt{T}(\hat{b} - b_0) + (I_{(p^{\star}-p-1)n} \otimes B_{\perp 0}')\sqrt{T}vec(\hat{\boldsymbol{A}}^{\star} - \boldsymbol{A}_0^{\star}) + o_p(1) \\
&= -(\tilde{\boldsymbol{A}}_0^{\star\prime} \otimes I_{n-K_0})\sqrt{T}(\hat{b} - b_0) + (I_{(p^{\star}-p-1)n} \otimes B_{\perp 0}')\sqrt{T}vec(\hat{\boldsymbol{A}}^{\star} - \boldsymbol{A}_0^{\star}) + o_p(1) \\
&= [\mathcal{M}_1 : \mathcal{M}_2]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} \tilde{Y}_{t-p-1} \otimes u_t^{*} \\ \tilde{X}_{t-p-2} \otimes u_t^{\star} \end{bmatrix} + o_p(1) = \mathcal{M}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_{2t} + o_p(1),
\end{aligned}$$

where $\tilde{\boldsymbol{A}}_0^{\star} = [A_{p+2,0,[K_0]}^{\star} : \cdots : A_{p^{\star},0,[K_0]}^{\star}]$ and $A_{i,0,[K_0]}^{\star}$ is the matrix with the last K_0 columns of $A_{i,0}^{\star}$, $\mathcal{M} = [\mathcal{M}_1 : \mathcal{M}_2]$ with $\mathcal{M}_1 = -(\tilde{\boldsymbol{A}}_0^{\star\prime} \otimes I_{n-K_0})S_{b_1}$ and $\mathcal{M}_2 = (I_{(p^{\star}-p-1)n} \otimes B_{0\perp}')\{E(\tilde{X}_{t-p-2}\tilde{X}_{t-p-2}')^{-1} \otimes I_n\}$, and $\psi_{2,t} = [(\tilde{Y}_{t-p-1} \otimes u_t^{*})', (\tilde{X}_{t-p-2} \otimes u_t^{\star})']'$. We apply the CLT to $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_{2t}$ using the same arguments of Section B.4. Hence, (C.9) follows with $\Sigma_{22} = \mathcal{M}V_{\psi,22}\mathcal{M}'$ and $V_{\psi,22} = \sum_{j=-\infty}^{\infty}Cov(\psi_{2,t}, \psi_{2,t-j})$.

C.5 Proof of Proposition 4

We check the three Assumptions ID.1, ID.2, ID.3 separately.

C.5.1 Assumption ID.1

With $p = q = 1$ we have:

$$EL(Y_t|Y_{t-1}, Y_{t-2}) = CY_{t-1} + B\Phi EL(f_{t-1}|Y_{t-1}, Y_{t-2}) = (C + B\Phi P_1')Y_{t-1} + B\Phi P_2' Y_{t-1}$$

where $EL(f_t|Y_t, Y_{t-1}) = P_1'Y_t + P_2'Y_{t-1}$. Thus, we have $A_2^* = B\Phi P_2'$. Let us write P_2 in terms of model parameters. We have

$$\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = V[\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix}]^{-1} Cov[\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix}, f_t] = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix} \begin{pmatrix} \Pi \\ \Pi\Phi' \end{pmatrix}$$

where $\Pi = Cov(Y_t, f_t)$, and the $\Sigma^{i,j}$ denote the blocks in the inverse variance-covariance matrix of Y_t, Y_{t-1}. Thus, we get:

$$\begin{aligned} P_2 &= \Sigma^{21}\Pi + \Sigma^{22}\Pi\Phi' = \Sigma^{22}((\Sigma^{22})^{-1}\Sigma^{21}\Pi + \Pi\Phi') = \Sigma^{22}(-\Sigma_{21}(\Sigma_{11})^{-1}\Pi + \Pi\Phi') \\ &= -\Sigma^{22}\left(\Gamma(1)'\Gamma(0)^{-1}\Pi - \Pi\Phi'\right), \end{aligned}$$

where we use $(\Sigma^{22})^{-1}\Sigma^{21} = -\Sigma_{21}(\Sigma_{11})^{-1} = \Gamma(1)'\Gamma(0)^{-1}$. Then we get

$$A_2^* = -B\Phi\left(\Gamma(1)'\Gamma(0)^{-1}\Pi - \Pi\Phi'\right)'\Sigma^{22}. \tag{C.10}$$

Let us now focus on matrix $\Gamma(1)'\Gamma(0)^{-1}\Pi - \Pi\Phi'$. We have:

$$\Gamma(1) = Cov(Y_t, Y_{t-1}) = C\Gamma(0) + B\Phi\Pi'$$

which yields $\Gamma(1)'\Gamma(0)^{-1} = \Gamma(0)C'\Gamma(0)^{-1} + \Pi\Phi'B'\Gamma(0)^{-1}$. Hence:

$$\Gamma(1)'\Gamma(0)^{-1}\Pi - \Pi\Phi' = \Gamma(0)C'\Gamma(0)^{-1}\Pi - \Pi\Phi'[I_K - B'\Gamma(0)^{-1}\Pi]. \tag{C.11}$$

Thus:

$$A_2^* = -B\Phi\left(\Pi'\Gamma(0)^{-1}C\Gamma(0) - [I_K - \Pi'\Gamma(0)^{-1}B]\Phi\Pi'\right)\Sigma^{22}. \tag{C.12}$$

The matrix Π is such that:

$$\Pi = Cov(Y_t, f_t) = C\Pi\Phi' + B\Sigma_f. \tag{C.13}$$

Moreover, the matrix $\Gamma(0)$ is such that:

$$\Gamma(0) \quad = \quad C\Gamma(0)C' + B\Sigma_f B' + \Sigma_u + C\Pi\Phi' B' + B\Phi\Pi' C'. \tag{C.14}$$

To further investigate condition (C.12), we consider the case $K = 1$, i.e. a single factor with variance σ_f^2 and autoregression coefficient ϕ. Then we have $\Pi = \sigma_f^2(I_n - \phi C)^{-1}B$. Then, equation (C.12) becomes:

$$A_2^* = \phi\sigma_f^2 BB' \left(\phi[1 - \sigma_f^2 B'\Gamma(0)^{-1}(I_n - \phi C)^{-1}B] - (I_n - \phi C')^{-1}\Gamma(0)^{-1}C\Gamma(0)(I_n - \phi C')\right)(I_n - \phi C')^{-1}\Sigma^{22},$$

which yields equation (3.3) using $\Sigma^{22} = (\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1}$.

Then, Assumption ID.1 is equivalent to $\phi \neq 0$ and

$$(I - \phi C)\Gamma(0)C'\Gamma(0)^{-1}(I - \phi C)^{-1}B \neq \phi[1 - \sigma_f^2 B'\Gamma(0)^{-1}(I - \phi C)^{-1}B]B.$$

The latter condition is equivalent to

$$C'\Gamma(0)^{-1}(I - \phi C)^{-1}B \neq \phi[1 - \sigma_f^2 B'\Gamma(0)^{-1}(I - \phi C)^{-1}B]\Gamma(0)^{-1}(I - \phi C)^{-1}B, \tag{C.15}$$

namely vector $\Gamma(0)^{-1}(I - \phi C)^{-1}B$ is *not* an eigenvector of matrix C' to eigenvalue $\phi[1 - \sigma_f^2 B'\Gamma(0)^{-1}(I - \phi C)^{-1}B]$.

Let us focus on DGP with $n = 2$ and $C = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$ and $\Sigma_u = \begin{pmatrix} \sigma_{u,1}^2 & 0 \\ 0 & \sigma_{u,2}^2 \end{pmatrix}$. Then, condition (C.15) is *not* met if either:

$$\text{Case A:} \quad \Gamma(0)^{-1}(I - \phi C)^{-1}B = \begin{pmatrix} * \\ 0 \end{pmatrix} \quad \text{and} \quad c_1 = \phi[1 - \sigma_f^2 B'\Gamma(0)^{-1}(I - \phi C)^{-1}B],$$

or

$$\text{Case B:} \quad \Gamma(0)^{-1}(I - \phi C)^{-1}B = \begin{pmatrix} 0 \\ * \end{pmatrix} \quad \text{and} \quad c_2 = \phi[1 - \sigma_f^2 B'\Gamma(0)^{-1}(I - \phi C)^{-1}B],$$

where $*$ denotes a non-zero entry.

To work out the details of the two cases, we consider the normalization $B = \begin{pmatrix} b \\ 1 \end{pmatrix}$ and write

$$\Gamma(0) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \qquad \Gamma(0)^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

To compute σ_1^2, σ_2^2 and ρ we use

$$\Pi = \sigma_f^2(I-\phi C)^{-1}B = \sigma_f^2 \begin{pmatrix} \frac{b}{1-\phi c_1} \\ \frac{1}{1-\phi c_2} \end{pmatrix}$$

and

$$C\Pi\Phi'B' = \phi\sigma_f^2 \begin{pmatrix} \frac{c_1 b^2}{1-\phi c_1} & \frac{c_1 b}{1-\phi c_1} \\ \frac{c_2 b}{1-\phi c_2} & \frac{c_2}{1-\phi c_2} \end{pmatrix}.$$

Using that the unconditional variance $\Gamma(0)$ satiesfies $\Gamma(0) = C\Gamma(0)C' + B\Sigma_f B' + \Sigma_u + C\Pi\Phi'B' + B\Phi\Pi'C'$, we get:

$$\begin{aligned} \begin{pmatrix} \sigma_1^2(1-c_1^2) & \rho\sigma_1\sigma_2(1-c_1c_2) \\ \rho\sigma_1\sigma_2(1-c_1c_2) & \sigma_2^2(1-c_2^2) \end{pmatrix} &= \sigma_f^2 \begin{pmatrix} b^2 & b \\ b & 1 \end{pmatrix} + \begin{pmatrix} \sigma_{u,1}^2 & 0 \\ 0 & \sigma_{u,2}^2 \end{pmatrix} \\ &\quad + \phi\sigma_f^2 \begin{pmatrix} \frac{2c_1 b^2}{1-\phi c_1} & \frac{c_1 b}{1-\phi c_1} + \frac{c_2 b}{1-\phi c_2} \\ \frac{c_1 b}{1-\phi c_1} + \frac{c_2 b}{1-\phi c_2} & \frac{2c_2}{1-\phi c_2} \end{pmatrix}, \end{aligned}$$

which yields:

$$\sigma_1^2 = \frac{1}{1-c_1^2}\left(\sigma_{u,1}^2 + \frac{1+\phi c_1}{1-\phi c_1} b^2 \sigma_f^2\right) \tag{C.16}$$

$$\sigma_2^2 = \frac{1}{1-c_2^2}\left(\sigma_{u,2}^2 + \frac{1+\phi c_2}{1-\phi c_2} \sigma_f^2\right) \tag{C.17}$$

$$\rho\sigma_1\sigma_2 = \frac{1}{1-c_1c_2} b\sigma_f^2 \left(\frac{1}{1-\phi c_1} + \frac{\phi c_2}{1-\phi c_2}\right). \tag{C.18}$$

Case A

We have

$$\Gamma(0)^{-1}(I-\phi C)^{-1}B = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \frac{\sigma_2^2}{1-\phi c_1}b - \frac{\rho\sigma_1\sigma_2}{1-\phi c_2} \\ -\frac{\rho\sigma_1\sigma_2}{1-\phi c_1}b + \frac{\sigma_1^2}{1-\phi c_2} \end{pmatrix}. \tag{C.19}$$

Hence a condition for Case A is that the second element of this vector vanishes, i.e.

$$\frac{\rho\sigma_1\sigma_2}{1-\phi c_1}b = \frac{\sigma_1^2}{1-\phi c_2}$$

which yields $b \neq 0$ the equation

$$\frac{1-\phi c_1}{1-\phi c_2} = \frac{1-c_1^2}{1-c_1c_2}\frac{\frac{1}{1-\phi c_1} + \frac{\phi c_2}{1-\phi c_2}}{\frac{1+\phi c_1}{1-\phi c_1} + \lambda_1},$$

where $\lambda_1 = \frac{\sigma_{u,1}^2}{b^2\sigma_f^2}$, which can be rewritten as:

$$[1+\phi c_1 + \lambda_1(1-\phi c_1)](1-\phi c_1)(1-c_1c_2) = (1-\phi^2 c_1c_2)(1-c_1^2). \tag{C.20}$$

The first element of vector $\Gamma(0)^{-1}(I-\phi C)^{-1}B$ becomes:

$$\begin{aligned} \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\left(\frac{\sigma_2^2 b}{1-\phi c_1} - \frac{\rho\sigma_1\sigma_2}{1-\phi c_2}\right) &= \frac{b}{\sigma_1^2(1-\phi c_1)} \\ &= \frac{(1-c_1^2)b}{(1-\phi c_1)\sigma_{u,1}^2 + (1+\phi c_1)b^2\sigma_f^2}, \end{aligned}$$

and we get:

$$\Gamma(0)^{-1}(I-\phi C)^{-1}B = \frac{(1-c_1^2)b}{(1-\phi c_1)\sigma_{u,1}^2 + (1+\phi c_1)b^2\sigma_f^2}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Thus, the condition $c_1 = \phi[1-\sigma_f^2 B'\Gamma(0)^{-1}(I-\phi C)^{-1}B$ becomes:

$$[(1-\phi c_1)\lambda_1 + 1 + \phi c_1](\phi - c_1) = \phi(1-c_1^2). \tag{C.21}$$

Thus, Case A is characterized by equations (C.20) and (C.21).

Now, from equation (C.21) we have $(1-\phi c_1)\lambda_1 + 1 + \phi c_1 = \phi(1-c_1^2)/(\phi - c_1)$, which we plug into equation (C.20) and get:

$$(1-\phi c_1)(1-c_1c_2) = (1-\phi^2 c_1c_2)(1-c_1/\phi).$$

This linear equation in c_2 admits the solution $c_2 = 1/\phi$. However, by the stationarity conditions, this equation cannot hold since both ϕ and c_2 are smaller than 1 in modulus. We deduce that no admissible parameter vector matches equations (C.20) and (C.21).

Case B

A condition for Case B is that the first element of vector $\Gamma(0)^{-1}(I-\phi C)^{-1}B$ vanishes, i.e.

$$\frac{\rho\sigma_1\sigma_2}{1-\phi c_2} = \frac{\sigma_2^2}{1-\phi c_1}b$$

which yields the equation

$$\frac{1-\phi c_2}{1-\phi c_1} = \frac{1-c_2^2}{1-c_1c_2}\frac{\frac{1}{1-\phi c_1}+\frac{\phi c_2}{1-\phi c_2}}{\frac{1+\phi c_2}{1-\phi c_2}+\lambda_2},$$

where $\lambda_2 = \frac{\sigma_{u,2}^2}{\sigma_f^2}$, which can be rewritten as:

$$[(1+\phi c_2)+\lambda_2(1-\phi c_2)](1-\phi c_2)(1-c_1c_2) = (1-\phi^2c_1c_2)(1-c_2^2). \tag{C.22}$$

The second element of vector $\Gamma(0)^{-1}(I-\phi C)^{-1}B$ becomes:

$$\begin{aligned}\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\left(-\frac{\rho\sigma_1\sigma_2}{1-\phi c_1}b+\frac{\sigma_1^2}{1-\phi c_2}\right) &= \frac{1}{\sigma_2^2(1-\phi c_2)}b \\ &= \frac{1-c_2^2}{(1-\phi c_2)\sigma_{u,2}^2+(1+\phi c_2)\sigma_f^2},\end{aligned}$$

and we get:

$$\Gamma(0)^{-1}(I-\phi C)^{-1}B = \frac{1-c_2^2}{(1-\phi c_2)\sigma_{u,2}^2+(1+\phi c_2)\sigma_f^2}\begin{pmatrix}0\\1\end{pmatrix}.$$

Thus, the condition $c_2 = \phi[1-\sigma_f^2 B'\Gamma(0)^{-1}(I-\phi C)^{-1}B]$ becomes:

$$[(1-\phi c_2)\lambda_2+(1+\phi c_2)](\phi-c_2) = \phi(1-c_2^2). \tag{C.23}$$

Thus, Case B is characterized by equations (C.22) and (C.23). They correspond to equations (C.20) and (C.21) after interchanging c_1 with c_2. By the above arguments, no admissible parameter vector matches equations (C.22) and (C.23).

Thus we have shown that, in the considered example, Assumption ID.1 is met if, and only if, $\phi \neq 0$.

C.5.2 Assumption ID.2

With $p = q = K = 1$, Assumption ID.2 holds if, and only if, B is not an eigenvector of matrix $\Sigma_u C' \Sigma_u^{-1}$. In our setting, the latter matrix is equal to $C = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$. Vector $B = \begin{pmatrix} b \\ 1 \end{pmatrix}$ is an eigenvector of C iff either $b = 0$, or $c_1 = c_2$. Thus, Assumption ID.2 is met if, and only if, $b \neq 0$ and $c_1 \neq c_2$.

C.5.3 Assumption ID.3

Assumption ID.3 is the condition $\phi \neq 0$. The result follows.

C.6 Proof of Proposition 6

The eigenvalue-eigenvector equations for matrix $\hat{R}$ corresponding to the K largest eigenvalues are:

$$\hat{R}\hat{U}_j = \hat{\lambda}_j \hat{U}_j, \qquad j = 1, ..., K. \tag{C.24}$$

The LHS of (C.24) is equal to:

$$\begin{aligned} \hat{R}\hat{U}_j &= RU_j + R(\hat{U}_j - U_j) + (\hat{R} - R)U_j + (\hat{R} - R)(\hat{U}_j - U_j) \\ &= \lambda_j U_j + R(\hat{U}_j - U_j) + (\hat{R} - R)U_j + (\hat{R} - R)(\hat{U}_j - U_j). \end{aligned} \tag{C.25}$$

The RHS of (C.24) is:

$$\hat{\lambda}_j \hat{U}_j = \lambda_j U_j + (\hat{\lambda}_j - \lambda_j)U_j + \lambda_j(\hat{U}_j - U_j) + (\hat{\lambda}_j - \lambda_j)(\hat{U}_j - U_j). \tag{C.26}$$

Thus, from (C.24)- (C.26) we get:

$$\begin{aligned} & R(\hat{U}_j - U_j) + (\hat{R} - R)U_j + (\hat{R} - R)(\hat{U}_j - U_j) \\ = \; & (\hat{\lambda}_j - \lambda_j)U_j + \lambda_j(\hat{U}_j - U_j) + (\hat{\lambda}_j - \lambda_j)(\hat{U}_j - U_j). \end{aligned} \tag{C.27}$$

Let us now derive expansions for eigenvalues and eigenvectors from this equation. i) Let us take the scalar product of both sides of (C.27) with U_j. We get:

$$\begin{aligned} & \lambda_j U_j'(\hat{U}_j - U_j) + U_j'(\hat{R} - R)U_j + U_j'(\hat{R} - R)(\hat{U}_j - U_j) \\ = \quad & \hat{\lambda}_j - \lambda_j + \lambda_j U_j'(\hat{U}_j - U_j) + (\hat{\lambda}_j - \lambda_j)U_j'(\hat{U}_j - U_j), \end{aligned}$$

which yields after rearranging terms:

$$\hat{\lambda}_j - \lambda_j = U_j'(\hat{R} - R)U_j + U_j'(\hat{R} - R)(\hat{U}_j - U_j) - (\hat{\lambda}_j - \lambda_j)U_j'(\hat{U}_j - U_j), \tag{C.28}$$

for any $j = 1, ..., K$.

ii) Let us now left-multiply both sides of (C.27) times P_l for $l \neq j$, and use that $P_l R = R P_l = \lambda_l P_l$. We get:

$$\begin{aligned} & \lambda_l P_l(\hat{U}_j - U_j) + P_l(\hat{R} - R)U_j + P_l(\hat{R} - R)(\hat{U}_j - U_j) \\ = \quad & \lambda_j P_l(\hat{U}_j - U_j) + (\hat{\lambda}_j - \lambda_j)P_l(\hat{U}_j - U_j), \end{aligned}$$

which yields using the property of distinct eigenvalues:

$$\begin{aligned} P_l(\hat{U}_j - U_j) = \; & \frac{1}{\lambda_j - \lambda_l}P_l(\hat{R} - R)U_j + \frac{1}{\lambda_j - \lambda_l}P_l(\hat{R} - R)(\hat{U}_j - U_j) \\ & - \frac{1}{\lambda_j - \lambda_l}(\hat{\lambda}_j - \lambda_j)P_l(\hat{U}_j - U_j), \end{aligned} \tag{C.29}$$

for $l \neq j$.

iii) By the normalization property of the eigenvectors we have:

$$1 = \hat{U}_j'\hat{U}_j = 1 + 2U_j'(\hat{U}_j - U_j) + (\hat{U}_j - U_j)'(\hat{U}_j - U_j),$$

which yields:

$$P_j(\hat{U}_j - U_j) = -\frac{1}{2}U_j|\hat{U}_j - U_j|^2, \tag{C.30}$$

for any $j = 1, ..., K$.

By using the property $\sum_{l=0}^{K} P_l = I_n$ and equations (C.29) and (C.30) we get:

$$\begin{aligned}
\hat{U}_j - U_j &= \sum_{l=0}^{K} P_l(\hat{U}_j - U_j) \\
&= \sum_{l=0, l\neq j}^{K} \frac{1}{\lambda_j - \lambda_l} P_l(\hat{R} - R)U_j + \sum_{l=0, l\neq j}^{K} \frac{1}{\lambda_j - \lambda_l} P_l(\hat{R} - R)(\hat{U}_j - U_j) \\
&\quad - \sum_{l=0, l\neq j}^{K} \frac{1}{\lambda_j - \lambda_l} (\hat{\lambda}_j - \lambda_j) P_l(\hat{U}_j - U_j) - \frac{1}{2} U_j |\hat{U}_j - U_j|^2, \qquad \text{(C.31)}
\end{aligned}$$

for $j = 1, ..., K$. Equations (C.28) and (C.31) provide expansions for eigenvalues and eigenvectors.

Let us now bound the second-order terms. We use $|U_j| = 1$, $|P_j| = 1$, and $\sum_{l=0, l\neq j}^{K} |\lambda_j - \lambda_l|^{-1} \leq \rho$ for all $j = 1, ...K$, by definition of $\rho = \max_{j:0\leq j\leq K} \sum_{l=0, l\neq j}^{K} |\lambda_j - \lambda_l|^{-1}$. We rewrite (C.28) and (C.31) as

$$\begin{aligned}
\hat{\lambda}_j - \lambda_j &= U_j'(\hat{R} - R)U_j + \mathcal{R}_{\lambda,j}, \\
\hat{U}_j - U_j &= \sum_{l=0, l\neq j}^{K} \frac{1}{\lambda_j - \lambda_l} P_l(\hat{R} - R)U_j + \mathcal{R}_{U,j}, \qquad \text{(C.32)}
\end{aligned}$$

where the remainder terms are such that:

$$\mathcal{R}_{\lambda,j} \leq |\hat{R} - R||\hat{U}_j - U_j| + |\hat{\lambda}_j - \lambda_j||\hat{U}_j - U_j|, \qquad \text{(C.33)}$$

and:

$$\mathcal{R}_{U,j} \leq \rho \left(|\hat{R} - R||\hat{U}_j - U_j| + |\hat{\lambda}_j - \lambda_j||\hat{U}_j - U_j| \right) + \frac{1}{2}|\hat{U}_j - U_j|^2. \qquad \text{(C.34)}$$

Let us now show that the remainder terms are of order $|\hat{R} - R|^2$.

(i) By the Weilandt-Hoffmann inequality (see e.g. Tao (2012), p. 137), it holds $\sum_{j=1}^{n} |\hat{\lambda}_j - \lambda_j|^2 \leq |\hat{R} - R|^2$. Thus, we have in particular for the j-th eigenvalue:

$$|\hat{\lambda}_j - \lambda_j| \leq |\hat{R} - R|, \qquad \text{(C.35)}$$

for any $j = 1, ..., K$. From (C.32), (C.34), (C.35) and the triangular inequality, we get:

$$|\hat{U}_j - U_j| \leq \rho|\hat{R} - R| + 2\rho|\hat{R} - R||\hat{U}_j - U_j| + \frac{1}{2}|\hat{U}_j - U_j|^2. \qquad \text{(C.36)}$$

(ii) The standardized eigenvector $\hat{U}_j$ of matrix $\hat{R}$ is defined up to a sign change. Let us define $\hat{U}_j$ with the sign normalization such that $U_j'\hat{U}_j \geq 0$. Then, $|\hat{U}_j - U_j|^2 = 2 - 2U_j'\hat{U}_j \leq 2$, i.e.

$$|\hat{U}_j - U_j| \leq \sqrt{2}. \tag{C.37}$$

(iii) By using bound (C.37) in the RHS of (C.36) we get

$$|\hat{U}_j - U_j| \leq \rho|\hat{R} - R| + 2\sqrt{2}\rho|\hat{R} - R| + \frac{\sqrt{2}}{2}|\hat{U}_j - U_j|,$$

which yields:

$$|\hat{U}_j - U_j| \leq (6 + 5\sqrt{2})\rho|\hat{R} - R|, \tag{C.38}$$

for any $j = 1, ..., K$. (iv), Finally, from (C.33), (C.34), (C.35) and (C.38) we deduce

$$\mathcal{R}_{\lambda,j} \leq 2|\hat{R} - R||\hat{U}_j - U_j| \leq 2\rho(6 + 5\sqrt{2})|\hat{R} - R|^2,$$

and:

$$\mathcal{R}_{U,j} \quad \leq \quad 2\rho|\hat{R} - R||\hat{U}_j - U_j| + \frac{1}{2}|\hat{U}_j - U_j|^2 \leq \rho^2(55 + 40\sqrt{2})|\hat{R} - R|^2.$$

The conclusion follows.

C.7 Proof of Proposition 7

Let us start by deriving the asymptotic expansion of estimator $\hat{A}^*_{p+1}$. Let $X_t = (Y'_{t-1}, Y'_{t-2}, ..., Y'_{t-p-1})'$. The OLS estimator for A^*_j, with $j = 1, ..., p+1$, is given by

$$[\hat{A}^*_1 : \hat{A}^*_2 : \cdots : \hat{A}^*_{p+1}] = \left(\sum_{t=1}^{T} Y_t X_t'\right)\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1}.$$

From the results on partitioned regression, we have

$$\hat{A}^*_{p+1} - A^*_{p+1} = \left(\frac{1}{T}\sum_{t=1}^{T} u^*_t \hat{Y}'_{t-p-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} \hat{Y}_{t-p-1}\hat{Y}'_{t-p-1}\right)^{-1},$$

where $\hat{Y}_{t-p-1} = Y_{t-p-1} - \hat{\tilde{\mathbf{\Gamma}}}(p)'\hat{\mathbf{\Gamma}}(0)^{-1}\boldsymbol{Y}_{t-1}$ is the sample residual of the projection of Y_{t-p-1} onto $\boldsymbol{Y}_{t-1} = (Y'_{t-1}, ..., Y'_{t-p})'$ and $\hat{\mathbf{\Gamma}}(0) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Y}_t\boldsymbol{Y}'_t$ and $\hat{\tilde{\mathbf{\Gamma}}}(p) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Y}_t Y'_{t-p}$. By

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t^* X'_t = O_p(1), \tag{C.39}$$

and the results in Lemma 2, we get

$$\sqrt{T}(\hat{A}^*_{p+1} - A^*_{p+1}) = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t^*\tilde{Y}'_{t-p-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1}\right)^{-1} + o_p(1),$$

where $\tilde{Y}_{t-p-1} = Y_{t-p-1} - \tilde{\mathbf{\Gamma}}(p)'\mathbf{\Gamma}(0)^{-1}\boldsymbol{Y}_{t-1}$ is the population residual of the projection of Y_{t-p-1} onto $\boldsymbol{Y}_{t-1}$. Then, by the properties of the vec operator, we get:

$$\begin{aligned}\sqrt{T}vec(\hat{A}^*_{p+1} - A^*_{p+1}) &= \left[\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1}\right)^{-1} \otimes I_n\right] vec\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t^*\tilde{Y}'_{t-p-1}\right) + o_p(1) \\ &= \left[\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1}\right)^{-1} \otimes I_n\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{Y}_{t-p-1} \otimes u_t^* + o_p(1).\end{aligned}$$

By the convergence a.s. and in probability of $\frac{1}{T}\sum_{t=1}^{T}\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1}$ to $E(\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1}) = \Gamma(0) - \tilde{\mathbf{\Gamma}}(p)'\mathbf{\Gamma}(0)^{-1}\tilde{\mathbf{\Gamma}}(p)$ from Lemma 2, we get

$$\sqrt{T}vec(\hat{A}^*_{p+1} - A^*_{p+1}) = \left\{[E(\tilde{Y}_{t-p-1}\tilde{Y}'_{t-p-1})]^{-1} \otimes I_n\right\}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{Y}_{t-p-1} \otimes u_t^* + o_p(1). \tag{C.40}$$

Let us now derive the asymptotic expansion of estimator $\hat{R}$. We have:

$$\hat{R} = \hat{A}^*_{p+1}\hat{A}^{*\prime}_{p+1} = \left[A^*_{p+1} + (\hat{A}^*_{p+1} - A^*_{p+1})\right]\left[A^*_{p+1} + (\hat{A}^*_{p+1} - A^*_{p+1})\right]'.$$

By developing the product and using $\hat{A}^*_{p+1} - A^*_{p+1} = O_p(1/\sqrt{T})$ from (C.39) and (C.40), we get $\hat{R} = R + (\hat{A}^*_{p+1} - A^*_{p+1})A^{*\prime}_{p+1} + A^*_{p+1}(\hat{A}^*_{p+1} - A^*_{p+1})' + o_p(\frac{1}{\sqrt{T}})$. By applying the vec operator, we get the asymptotic expansion of $\hat{R}$:

$$\sqrt{T}vec(\hat{R} - R) = [(A^*_{p+1} \otimes I_n) + (I_n \otimes A^*_{p+1}) \cdot \mathcal{K}_{n,n}]\sqrt{T}vec(\hat{A}^*_{p+1} - A^*_{p+1}) + o_p(1),$$

where $\mathcal{K}_{n,n}$ is the commutation matrix such that $vec[(\hat{A}^*_{p+1} - A^*_{p+1})'] = \mathcal{K}_{n,n} vec(\hat{A}^*_{p+1} - A^*_{p+1})$.

The asymptotic expansion of $\hat{U}_j$ is obtained from (B.2) and the bound $\hat{R} - R = O_p(1/\sqrt{T})$, which yield:

$$\sqrt{T}(\hat{U}_j - U_j) = \sum_{i=0, i \neq j}^{K} \frac{1}{\lambda_j - \lambda_i} (U_j' \otimes P_i) \sqrt{T} vec(\hat{R} - R) + o_p(1), \tag{C.41}$$

for $j = 1, \ldots, K$. This yields the asymptotic expansion for $\hat{\boldsymbol{U}} = [\hat{U}_1 : \cdots : \hat{U}_K]$.

Finally, let us find the asymptotic expansion of $\hat{B}_1 = \hat{\boldsymbol{U}}_1 \hat{\boldsymbol{U}}_2^{-1}$. We have that

$$\begin{aligned} \hat{B}_1 - B_{1,0} &= \hat{\boldsymbol{U}}_1 \hat{\boldsymbol{U}}_2^{-1} - (B_{1,0}Q)(B_{2,0}Q)^{-1} = (\hat{\boldsymbol{U}}_1 - B_{1,0}Q)\hat{\boldsymbol{U}}_2^{-1} + (B_{1,0}Q)(\hat{\boldsymbol{U}}_2^{-1} - (B_{2,0}Q)^{-1}) \\ &= (\hat{\boldsymbol{U}}_1 - B_{1,0}Q)\hat{\boldsymbol{U}}_2^{-1} - (B_{1,0}Q)\hat{\boldsymbol{U}}_2^{-1}(\hat{\boldsymbol{U}}_2 - B_{2,0}Q)(B_{2,0}Q)^{-1}, \end{aligned}$$

where $B_{2,0} = I_K$ by IR.1. Using $\hat{\boldsymbol{U}}_1 - B_{1,0}Q = O_p(\frac{1}{\sqrt{T}})$ and $\hat{\boldsymbol{U}}_2 - B_{2,0}Q = O_p(\frac{1}{\sqrt{T}})$ from (C.41) and (B.1), and $B_{2,0} = I_K$, we get:

$$\begin{aligned} \sqrt{T}(\hat{B}_1 - B_{1,0}) &= \sqrt{T}(\hat{\boldsymbol{U}}_1 - B_{1,0}Q)(B_{2,0}Q)^{-1} - (B_{1,0}Q)(B_{2,0}Q)^{-1}\sqrt{T}(\hat{\boldsymbol{U}}_2 - B_{2,0}Q)(B_{2,0}Q)^{-1} + o_p(1) \\ &= \sqrt{T}(\hat{\boldsymbol{U}}_1 - B_{1,0}Q)Q^{-1} - B_{1,0}\sqrt{T}(\hat{\boldsymbol{U}}_2 - B_{2,0}Q)Q^{-1} + o_p(1). \end{aligned}$$

Now plug in

$$\begin{aligned} \hat{\boldsymbol{U}}_1 - B_{1,0}Q &= M_1(\hat{\boldsymbol{U}} - B_0Q), \\ \hat{\boldsymbol{U}}_2 - B_{2,0}Q &= M_2(\hat{\boldsymbol{U}} - B_0Q), \end{aligned}$$

where $M_1 = [I_{n-K} : 0_{(n-K)\times K}]$ and $M_2 = [0_{K\times(n-K)} : I_K]$, and get

$$\begin{aligned} \sqrt{T}(\hat{B}_1 - B_{1,0}) &= M_1\sqrt{T}(\hat{\boldsymbol{U}} - B_0Q)Q^{-1} - B_{1,0}M_2\sqrt{T}(\hat{\boldsymbol{U}} - B_0Q)Q^{-1} + o_p(1) \\ &= M\sqrt{T}(\hat{\boldsymbol{U}} - B_0Q)Q^{-1} + o_p(1), \end{aligned}$$

where $M = M_1 - B_{1,0}M_2 = [I_{n-K} : -B_{1,0}] = B_{0\perp}'$. By taking the vec, we get:

$$\sqrt{T} vec(\hat{B}_1 - B_{1,0}) = [(Q^{-1})' \otimes M)\sqrt{T}(\hat{\boldsymbol{U}} - B_0Q] + o_p(1).$$

Notice that $\sqrt{T} vec(\hat{\boldsymbol{U}} - B_0Q) = (\sqrt{T}(\hat{U}_1 - U_1)', \ldots, \sqrt{T}(\hat{U}_K - U_K)')'$ from (B.1).

C.8 Proof of Proposition 8

Let us consider the OLS estimator $\hat{\mathbf{\Delta}}$ of $\mathbf{\Delta}_0$ defined in Section 4.1:

$$\hat{\mathbf{\Delta}} = \hat{B}_{\perp}' \left(\sum_{t=1}^{T} Y_t \boldsymbol{Y}_{t-1}' \right) \left(\sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1}.$$

Using $\hat{B}_{\perp} = B_{0\perp} + (\hat{B}_{\perp} - B_{0\perp})$ and $B_{0\perp}' Y_t = \mathbf{\Delta}_0 \boldsymbol{Y}_{t-1} + \eta_t$, the latter equation yields

$$\hat{\mathbf{\Delta}} - \mathbf{\Delta}_0 = \left(\sum_{t=1}^{T} \eta_t \boldsymbol{Y}_{t-1}' \right) \left(\sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1} + (\hat{B}_{\perp} - B_{0\perp})' \left(\sum_{t=1}^{T} Y_t \boldsymbol{Y}_{t-1}' \right) \left(\sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1}.$$

Taking the vec operator, we get

$$\begin{aligned} \sqrt{T}(\hat{d} - d_0) &= \left[\left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1} \otimes I_{n-K} \right] \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \otimes \eta_t \\ &+ \left[\left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} Y_t' \right) \otimes I_{n-K} \right] \sqrt{T} vec[(\hat{B}_{\perp} - B_{0\perp})']. \end{aligned}$$

Now let us link $vec(B_{\perp}')$ to $b = vec(B_1)$. From the definition $B_{\perp} = [I_{n-K} : -B_1]'$ we get:

$$vec(B_{\perp}') = \begin{bmatrix} vec(I_{n-K}) \\ -b \end{bmatrix} = a_0 - A_{0\perp}' b, \tag{C.42}$$

where $a_0 := [vec(I_{n-K})', 0_{K(n-K)\times 1}']'$ and $A_{0\perp} = [0_{(n-K)K\times(n-K)^2} : I_{(n-K)K}] = [0_{K\times(n-K)} : I_K] \otimes I_{n-K}$. Thus:

$$vec[(\hat{B}_{\perp} - B_{0\perp})'] = -A_{0\perp}'(\hat{b} - b_0). \tag{C.43}$$

Hence, we get

$$\begin{aligned} \sqrt{T}(\hat{d} - d_0) &= \left[\left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1} \otimes I_{n-K} \right] \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \otimes \eta_t \\ &- \left[\left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} Y_{[K],t}' \right) \otimes I_{n-K} \right] \sqrt{T}(\hat{b} - b_0), \end{aligned}$$

where $Y_{[K],t} = (Y_{n-K+1,t}, ..., Y_{n,t})'$ is the lower K-dimensional subvector of Y_t. Using that sample moments of the process Y_t converge to the $\Gamma(j)$ (Lemma 2), we get the asymptotic expansion

$$\sqrt{T}(\hat{d} - d_0) = \left[\mathbf{\Gamma}(0)^{-1} \otimes I_{n-K}\right] \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{Y}_{t-1} \otimes \eta_t - \left[\mathbf{\Gamma}(0)^{-1} \tilde{\mathbf{\Gamma}}_{[K]}(-1) \otimes I_{n-K}\right] \sqrt{T}(\hat{b} - b_0) + o_p(1),$$

where $\tilde{\mathbf{\Gamma}}_{[K]}(j) = E[\boldsymbol{Y}_t Y'_{[K],t-j}]$.

C.9 Proof of Proposition 9

To derive the asymptotic expansion of estimator $\hat{\boldsymbol{C}}$, we first show that it can be obtained as the minimizer of a GMM objective function. Let us define:

$$\begin{aligned} H_t(\boldsymbol{C}, \tau) &= (Y_t - \boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Z}_t(\tau)' \\ &= (Y_t - \boldsymbol{C}\boldsymbol{Y}_{t-1})[Y'_{t-1}B_\perp - \boldsymbol{Y}'_{t-2}\boldsymbol{\Delta}' : \cdots : Y'_{t-M}B_\perp - \boldsymbol{Y}'_{t-M-1}\boldsymbol{\Delta}'] \end{aligned}$$

where $\boldsymbol{Z}_t(\tau) = (\eta_{t-1}(\tau)', ..., \eta_{t-M}(\tau)')'$, $\eta_t(\tau) = B'_\perp Y_t - \boldsymbol{\Delta}\boldsymbol{Y}_{t-1}$ and $\hat{\tau}$ is the estimator of $\tau = (b', d')'$ (dependence on b and d is via $B_\perp$ and $\boldsymbol{\Delta}$, respectively). We have

$$\frac{1}{T} \sum_{t=1}^{T} H_t(\boldsymbol{C}, \hat{\tau}) = \hat{Q}_{Y\boldsymbol{Z}} - \boldsymbol{C}\hat{Q}_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}.$$

Then, estimator $\hat{c} = vec(\hat{\boldsymbol{C}})$ is the minimizer of the GMM objective function:

$$\mathcal{Q}(c) = \left(\frac{1}{T} \sum_{t=1}^{T} h_t(c, \hat{\tau})\right)' \hat{\Omega}_c \left(\frac{1}{T} \sum_{t=1}^{T} h_t(c, \hat{\tau})\right)$$

where $h_t(c, \tau) \equiv vec(H_t(C, \tau))$.

If we take the first-order derivative with respect to parameter c, and multiply times $\sqrt{T}$, we get the first-order condition for estimator $\hat{c}$ given by

$$\left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial h_t(\hat{c}, \hat{\tau})}{\partial c'}\right)' \hat{\Omega}_c \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_t(\hat{c}, \hat{\tau})\right) = 0. \tag{C.44}$$

Now let us expand $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_t(\hat{c},\hat{\tau})$ at a first order w.r.t. $\hat{c}$ and $\hat{\tau}$. This yields

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_t(\hat{c},\hat{\tau}) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_t(c_0,\tau_0) \\ &+\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t(c_0,\tau_0)}{\partial c'}\sqrt{T}(\hat{c}-c_0) \\ &+\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t(c_0,\tau_0)}{\partial d'}\sqrt{T}(\hat{d}-d_0) \\ &+\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t(c_0,\tau_0)}{\partial b'}\sqrt{T}(\hat{b}-b_0)+o_p(1). \end{aligned} \tag{C.45}$$

To compute the derivatives on the RHS, the moment function can be written by vectorization as

$$\begin{aligned} h_t(c,\tau) &= vec\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Z}_t(\tau)'\right] \\ &= \boldsymbol{Z}_t(\tau)\otimes(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1}) \\ &= \boldsymbol{Z}_t(\tau)\otimes[Y_t-(\boldsymbol{Y}'_{t-1}\otimes I_n)c], \end{aligned} \tag{C.46}$$

and also as:

$$\begin{aligned} h_t(c,\tau) &= vec\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})[Y'_{t-1}B_\perp-\boldsymbol{Y}'_{t-2}\boldsymbol{\Delta}':\cdots:Y'_{t-M}B_\perp-\boldsymbol{Y}'_{t-M-1}\boldsymbol{\Delta}']\right] \\ &= \begin{pmatrix} vec\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{t-1}B_\perp\right] \\ \vdots \\ vec\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{t-M}B_\perp\right] \end{pmatrix} - \begin{pmatrix} vec\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-2}\boldsymbol{\Delta}'\right] \\ \vdots \\ vec\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-M-1}\boldsymbol{\Delta}'\right] \end{pmatrix} \\ &= \begin{pmatrix} I_{n-K}\otimes[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{t-1}] \\ \vdots \\ I_{n-K}\otimes[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{t-M}] \end{pmatrix} vec(B_\perp) - \begin{pmatrix} I_{n-K}\otimes[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-2}] \\ \vdots \\ I_{n-K}\otimes[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-M-1}] \end{pmatrix} vec(\boldsymbol{\Delta}') \\ &= \begin{pmatrix} \mathcal{K}_{n-K,n}\left([(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{t-1}]\otimes I_{n-K}\right) \\ \vdots \\ \mathcal{K}_{n-K,n}\left([(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{t-M}]\otimes I_{n-K}\right) \end{pmatrix} b_\perp \\ &\quad - \begin{pmatrix} \mathcal{K}_{n-K,n}\left([(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-2}]\otimes I_{n-K}\right) \\ \vdots \\ \mathcal{K}_{n-K,n}\left([(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-M-1}]\otimes I_{n-K}\right) \end{pmatrix} d, \end{aligned} \tag{C.47}$$

where $b_\perp := vec(B'_\perp) = a_0 - A'_{0\perp}b$ from (C.42), and $\mathcal{K}_{n-K,n}$ is the commutator matrix for orders

$n-K,n$. Then, from (C.46) we get the gradient of the moment function w.r.t. c:

$$\begin{aligned}\frac{\partial h_t(c,\tau)}{\partial c'} &= -\boldsymbol{Z}_t(\tau)\otimes(\boldsymbol{Y}'_{t-1}\otimes I_n),\\ &= -[\boldsymbol{Z}_t(\tau)\boldsymbol{Y}'_{t-1}]\otimes I_n,\end{aligned}$$

and from (C.47) the gradients w.r.t. d and b are:

$$\begin{aligned}\frac{\partial h_t(c,\tau)}{\partial d'} &= -\begin{pmatrix}\mathcal{K}_{n-K,n}\left(\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-2}\right]\otimes I_{n-K}\right)\\ \vdots\\ \mathcal{K}_{n-K,n}\left(\left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-M-1}\right]\otimes I_{n-K}\right)\end{pmatrix}\\ &= -(I_M\otimes\mathcal{K}_{n-K,n})\left[\begin{pmatrix}(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-2}\\ \vdots\\ (Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})\boldsymbol{Y}'_{t-M-1}\end{pmatrix}\otimes I_{n-K}\right],\end{aligned}$$

and

$$\frac{\partial h_t(c,\tau)}{\partial b'} = \frac{\partial h_t(c,\tau)}{\partial b'_\perp}\frac{\partial b_\perp}{\partial b'} = -(I_M\otimes\mathcal{K}_{n-K,n})\left[\begin{pmatrix}(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{[K],t-1}\\ \vdots\\ (Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})Y'_{[K],t-M}\end{pmatrix}\otimes I_{n-K}\right],$$

using $\frac{\partial b_\perp}{\partial b'} = [0_{K\times(n-K)} : I_K]'\otimes I_{n-K}$.

By evaluating at the true parameter values, and computing the expectations of the gradients, we get the Jacobian matrices:

$$J_c = E\left[\frac{\partial h_t(c_0,\tau_0)}{\partial c'}\right] = -Q'_{\boldsymbol{Y}_{-1}\boldsymbol{Z}}\otimes I_n,$$

and:

$$
\begin{aligned}
J_d &= E\left[\frac{\partial h_t(c_0,\tau_0)}{\partial d'}\right] = -(I_M \otimes \mathcal{K}_{n-K,n})\left[\begin{pmatrix} \tilde{\boldsymbol{\Gamma}}(-2)' - \boldsymbol{C}_0\boldsymbol{\Gamma}(1) \\ \vdots \\ \tilde{\boldsymbol{\Gamma}}(-M-1)' - \boldsymbol{C}_0\boldsymbol{\Gamma}(M) \end{pmatrix} \otimes I_{n-K}\right] \\
&= -(I_M \otimes \mathcal{K}_{n-K,n})\left[\begin{pmatrix} \Gamma_{\varepsilon Y}(2) & \cdots & \Gamma_{\varepsilon Y}(p+1) \\ \vdots & \cdots & \vdots \\ \Gamma_{\varepsilon Y}(M+1) & \cdots & \Gamma_{\varepsilon Y}(M+p) \end{pmatrix} \otimes I_{n-K},\right]
\end{aligned}
$$

and:

$$
\begin{aligned}
J_b &= E\left[\frac{\partial h_t(c_0,\tau_0)}{\partial b'}\right] = -(I_M \otimes \mathcal{K}_{n-K,n})\left[\begin{pmatrix} \Gamma_{[K]}(1) - \boldsymbol{C}_0\tilde{\boldsymbol{\Gamma}}_{[K]}(0) \\ \vdots \\ \Gamma_{[K]}(M) - \boldsymbol{C}_0\tilde{\boldsymbol{\Gamma}}_{[K]}(M-1) \end{pmatrix} \otimes I_{n-K}\right] \\
&= -(I_M \otimes \mathcal{K}_{n-K,n})\left[\begin{pmatrix} \Gamma_{\varepsilon Y_{[K]}}(1) \\ \vdots \\ \Gamma_{\varepsilon Y_{[K]}}(M) \end{pmatrix} \otimes I_{n-K}\right],
\end{aligned}
$$

where $\Gamma_{[K]}(j) = E(Y_t Y'_{[K],t-j})$ and $\tilde{\boldsymbol{\Gamma}}_{[K]}(j) = E(\boldsymbol{Y}_t Y'_{[K],t-j})$, and $\Gamma_{\varepsilon Y}(h) = E[\varepsilon_t Y'_{t-h}] = \Gamma(h) - \boldsymbol{C}_0\tilde{\boldsymbol{\Gamma}}(h-1)$, $\Gamma_{\varepsilon Y_{[K]}}(h) = E[\varepsilon_t Y'_{[K],t-h}] = \Gamma_{[K]}(h) - \boldsymbol{C}_0\tilde{\boldsymbol{\Gamma}}_{[K]}(h-1)$ with $\varepsilon_t = Y_t - \boldsymbol{C}_0\boldsymbol{Y}_{t-1}$.

From (C.45), taking the probability limits of the sample averages of partial derivatives, we get:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t(\hat{c},\hat{\tau}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t(c_0,\tau_0) + J_c\sqrt{T}(\hat{c}-c_0) + J_d\sqrt{T}(\hat{d}-d_0) + J_b\sqrt{T}(\hat{b}-b_0) + o_p(1).
$$

Plugging this expression and $\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t(\hat{c},\hat{\tau})}{\partial c'} = J_c + o_p(1)$ in (C.44), using $\hat{\Omega}_c = \Omega_c + o_p(1)$, and solving for $\sqrt{T}(\hat{c}-c_0)$ we get:

$$
\sqrt{T}(\hat{c}-c_0) = -(J_c'\Omega_c J_c)^{-1}J_c'\Omega_c\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t(c_0,\tau_0) + J_d\sqrt{T}(\hat{d}-d_0) + J_b\sqrt{T}(\hat{b}-b_0)\right\} + o_p(1).
$$

The conclusion follows.

C.10 Proof of Proposition 10

The proof is similar to the one of Proposition 9. Let us define:

$$L_t(\mathbf{\Phi}, \psi) = [\xi_t(\psi) - \mathbf{\Phi}\boldsymbol{\xi}_{t-1}(\psi)]\boldsymbol{W}_t(\psi)',$$

with $\xi_t(\psi) = \bar{B}'(Y_t - \boldsymbol{C}\boldsymbol{Y}_{t-1})$, $\bar{B} = B(B'B)^{-1}$, $\boldsymbol{\xi}_{t-1}(\psi) = (\xi_{t-1}(\psi)', ..., \xi_{t-q}(\psi)')'$, $\boldsymbol{W}_t(\psi) = (\xi_{t-q-1}(\psi)', ..., \xi_{t-q-L}(\psi)')'$, and $\psi = (b', c')'$. By introducing the argument $\phi = vec(\mathbf{\Phi})$, the estimator $\hat{\phi} = vec(\hat{\mathbf{\Phi}})$ is the minimizer of the GMM objective function:

$$\mathcal{P}(\phi) = \left(\frac{1}{T}\sum_{t=1}^{T} l_t(\phi, \hat{\psi})\right)' \hat{\Omega}_\phi \left(\frac{1}{T}\sum_{t=1}^{T} l_t(\phi, \hat{\psi})\right),$$

where $l_t(\phi, \psi) = vec[L_t(\mathbf{\Phi}, \psi)]$.

By taking the first-order derivative with respect to ϕ we get the first order condition

$$\left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\hat{\phi}, \hat{\psi})}{\partial \phi'}\right)' \hat{\Omega}_\phi \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t(\hat{\phi}, \hat{\psi})\right) = 0. \quad \text{(C.48)}$$

Now let us expand $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t(\hat{\phi}, \hat{\psi})$ with a first-order Taylor expansion. We have:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t(\hat{\phi}, \hat{\psi}) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t(\phi_0, \psi_0) + \frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\phi_0, \psi_0)}{\partial \phi'}\sqrt{T}vec(\hat{\phi} - \phi_0) \\ &+ \frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\phi_0, \psi_0)}{\partial c'}\sqrt{T}(\hat{c} - c_0) + \frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\phi_0, \psi_0)}{\partial b'}\sqrt{T}(\hat{b} - b_0) + o_p(1). \quad \text{(C.49)}\end{aligned}$$

Let us define the following matrices:

$$\mathcal{J}_\phi = E\left[\frac{\partial l_t(\phi_0, \psi_0)}{\partial \phi'}\right], \qquad \mathcal{J}_c = E\left[\frac{\partial l_t(\phi_0, \psi_0)}{\partial c'}\right], \qquad \mathcal{J}_b = E\left[\frac{\partial l_t(\phi_0, \psi_0)}{\partial b'}\right].$$

From (C.49), taking the probability limits of the sample averages we get:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t(\hat{\phi}, \hat{\psi}) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t(\phi_0, \psi_0) + \mathcal{J}_\phi\sqrt{T}(\hat{\phi} - \phi_0) \\ &+ \mathcal{J}_c\sqrt{T}(\hat{c} - c_0) + \mathcal{J}_b\sqrt{T}(\hat{b} - b_0) + o_p(1).\end{aligned}$$

Plugging this expression in (C.48) and solving for $\hat{\phi}-\phi_0$ we get:

$$\sqrt{T}(\hat{\phi}-\phi_0) \;=\; -(\mathcal{J}_\phi'\Omega_\phi\mathcal{J}_\phi)^{-1}\mathcal{J}_\phi'\Omega_\phi\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}l_t(\phi_0,\psi_0)+\mathcal{J}_b\sqrt{T}(\hat{b}-b_0)+\mathcal{J}_c\sqrt{T}(\hat{c}-c_0)\right\}+o_p(1),$$

where Ω_ϕ is the probability limit of $\hat{\Omega}_\phi$.

To conclude the proof, let us compute explicitly matrices $\mathcal{J}_\phi$, $\mathcal{J}_c$ and $\mathcal{J}_b$. We can write the moment vector as:

$$l_t(\phi,\psi) \;=\; vec[(\xi_t(\psi)-\boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}(\psi))\boldsymbol{W}_t(\psi)'] = \boldsymbol{W}_t(\psi)\otimes[\xi_t(\psi)-\boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}(\psi)].$$

The term in square brackets can be written in different forms by vectorization:

$$\begin{aligned}
\xi_t(\psi)-\boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}(\psi) &= \xi_t(\psi)-(\boldsymbol{\xi}_{t-1}(\psi)'\otimes I_K)\phi \\
&= \bar{B}'Y_t-\boldsymbol{\Phi}(I_q\otimes\bar{B}')[Y_{t-1}',...,Y_{t-q}']' \\
&\quad -\left[\boldsymbol{Y}_{t-1}'\otimes\bar{B}'-\boldsymbol{\Phi}\left(\boldsymbol{Y}_{t-2}\otimes\bar{B}:\cdots:\boldsymbol{Y}_{t-q-1}\otimes\bar{B}\right)'\right]c \\
&= \left[(Y_t-\boldsymbol{C}\boldsymbol{Y}_{t-1})'\otimes I_K\right]\bar{b} \\
&\quad -\boldsymbol{\Phi}\left[(Y_{t-1}-\boldsymbol{C}\boldsymbol{Y}_{t-2})\otimes I_K:\cdots:(Y_{t-q}-\boldsymbol{C}\boldsymbol{Y}_{t-q-1})\otimes I_K\right]'\bar{b},
\end{aligned}$$

where $\bar{b}=vec(\bar{B}')$. Similarly for the instrument vector:

$$\begin{aligned}
\boldsymbol{W}_t(\psi) &= (I_S\otimes\bar{B}')(Y_{t-q-1}',...,Y_{t-q-L}')'-(\boldsymbol{Y}_{t-q-2}\otimes\bar{B}:\cdots:\boldsymbol{Y}_{t-q-L-1}\otimes\bar{B})'c \\
&= [(Y_{t-q-1}-\boldsymbol{C}\boldsymbol{Y}_{t-q-2})\otimes I_K:\cdots:(Y_{t-q-L}-\boldsymbol{C}\boldsymbol{Y}_{t-q-L-1})\otimes I_K]'\,\bar{b}.
\end{aligned}$$

Thus, we can write $l_t(\phi,\psi)$ in three different ways to highlight its dependence on parameter vectors ϕ, c and $\bar{b}$, respectively:

$$l_t(\phi,\psi)=\boldsymbol{W}_t(\psi)\otimes\left[\xi_t(\psi)-(\boldsymbol{\xi}_{t-1}(\psi)'\otimes I_K)\phi\right], \tag{C.50}$$

$$\begin{aligned}
l_t(\phi,\psi) &= \left[(I_S\otimes\bar{B}')(Y_{t-q-1}',...,Y_{t-q-L}')'-(\boldsymbol{Y}_{t-q-2}\otimes\bar{B}:\cdots:\boldsymbol{Y}_{t-q-L-1}\otimes\bar{B})'c\right] \\
&\quad\otimes\left\{\bar{B}'Y_t-\boldsymbol{\Phi}(I_q\otimes\bar{B}')[Y_{t-1}',...,Y_{t-q}']'\right. \\
&\quad\left.-\left[\boldsymbol{Y}_{t-1}'\otimes\bar{B}'-\boldsymbol{\Phi}\left(\boldsymbol{Y}_{t-2}\otimes\bar{B}:\cdots:\boldsymbol{Y}_{t-q-1}\otimes\bar{B}\right)'\right]c\right\},
\end{aligned} \tag{C.51}$$

and:

$$
\begin{aligned}
l_t(\phi,\psi) &= \left\{[(Y_{t-q-1} - \boldsymbol{C}\boldsymbol{Y}_{t-q-2}) \otimes I_K : \cdots : (Y_{t-q-L} - \boldsymbol{C}\boldsymbol{Y}_{t-q-L-1}) \otimes I_K]' \, \bar{b}\right\} \\
&\otimes \left\{ \left[(Y_t - \boldsymbol{C}\boldsymbol{Y}_{t-1})' \otimes I_K\right] \bar{b} \right. \\
&\left. -\boldsymbol{\Phi} \left[(Y_{t-1} - \boldsymbol{C}\boldsymbol{Y}_{t-2}) \otimes I_K : \cdots : (Y_{t-q} - \boldsymbol{C}\boldsymbol{Y}_{t-q-1}) \otimes I_K\right]' \bar{b} \right\}. \qquad \text{(C.52)}
\end{aligned}
$$

Then, we get the gradients of the moment vector w.r.t. ϕ, c and b:

$$
\frac{\partial l_t(\phi,\psi)}{\partial \phi'} = -\boldsymbol{W}_t(\psi) \otimes (\boldsymbol{\xi}_{t-1}(\psi)' \otimes I_K) = -[\boldsymbol{W}_t(\psi)\boldsymbol{\xi}_{t-1}(\psi)'] \otimes I_K,
$$

$$
\begin{aligned}
\frac{\partial l_t(\phi,\psi)}{\partial c'} &= -(\boldsymbol{Y}_{t-q-2} \otimes \bar{B} : \cdots : \boldsymbol{Y}_{t-q-L-1} \otimes \bar{B})' \otimes [\xi_t(\psi) - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}(\psi)] \\
&\quad -\boldsymbol{W}_t(\psi) \otimes \left[\boldsymbol{Y}'_{t-1} \otimes \bar{B}' - \boldsymbol{\Phi}\left(\boldsymbol{Y}_{t-2} \otimes \bar{B} : \cdots : \boldsymbol{Y}_{t-q-1} \otimes \bar{B}\right)'\right] \\
&= -\begin{bmatrix} [\xi_t(\psi) - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}(\psi)]\boldsymbol{Y}'_{t-q-2} \\ ...\boldsymbol{Y}'_{t-q-L-1} \end{bmatrix} \otimes \bar{B}' \\
&\quad -[\boldsymbol{W}_t(\psi)\boldsymbol{Y}'_{t-1}] \otimes \bar{B}' + \sum_{i=1}^{q} [\boldsymbol{W}_t(\psi)\boldsymbol{Y}'_{t-i-1}] \otimes (\Phi_i \bar{B}'),
\end{aligned}
$$

and:

$$
\frac{\partial l_t(\phi,\psi)}{\partial b'} = \frac{\partial l_t(\phi,\psi)}{\partial \bar{b}'} \frac{\partial \bar{b}(b)}{\partial b'},
$$

where:

$$
\begin{aligned}
\frac{\partial l_t(\phi,\psi)}{\partial \bar{b}'} &= [(Y_{t-q-1} - \boldsymbol{C}\boldsymbol{Y}_{t-q-2}) \otimes I_K : \cdots : (Y_{t-q-L} - \boldsymbol{C}\boldsymbol{Y}_{t-q-L-1}) \otimes I_K]' \otimes [\xi_t(\psi) - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}(\psi)] \\
&\quad +\boldsymbol{W}_t(\psi) \otimes \left\{(Y_t - \boldsymbol{C}\boldsymbol{Y}_{t-1})' \otimes I_K - \boldsymbol{\Phi}\left[(Y_{t-1} - \boldsymbol{C}\boldsymbol{Y}_{t-2}) \otimes I_K : \cdots : (Y_{t-q} - \boldsymbol{C}\boldsymbol{Y}_{t-q-1}) \otimes I_K\right]'\right\} \\
&= \begin{bmatrix} [\xi_t(\psi) - \boldsymbol{C}\boldsymbol{\xi}_{t-1}(\psi)](Y_{t-q-1} - \boldsymbol{C}\boldsymbol{Y}_{t-q-2})' \\ ...(Y_{t-q-L} - \boldsymbol{C}\boldsymbol{Y}_{t-q-L-1})' \end{bmatrix} \otimes I_K \\
&\quad +[\boldsymbol{W}_t(\psi)(Y_t - \boldsymbol{C}\boldsymbol{Y}_{t-1})'] \otimes I_K - \sum_{i=1}^{q} [\boldsymbol{W}_t(\psi)(Y_{t-i} - \boldsymbol{C}\boldsymbol{Y}_{t-i-1})'] \otimes \Phi_i.
\end{aligned}
$$

To compute the partial derivative $\partial\bar{b}/\partial b'$ at $b = b_0$, we use that for B in a neighborhood of B_0 we have:

$$B'B = (B_0'B_0)\left[I_K + (B_0'B_0)^{-1}(B - B_0)'B_0 + (B_0'B_0)^{-1}B_0'(B - B_0)\right] + O(|B - B_0|^2),$$

$$(B'B)^{-1} = \left[I_K - (B_0'B_0)^{-1}(B - B_0)'B_0 - (B_0'B_0)^{-1}B_0'(B - B_0)\right](B_0'B_0)^{-1} + O(|B - B_0|^2),$$

which implies:

$$\bar{B} = \bar{B}_0 + P_0(B - B_0)(B_0'B_0)^{-1} - \bar{B}_0(B - B_0)'\bar{B}_0 + O(|B - B_0|^2),$$

where $P_0 = I_n - B_0(B_0'B_0)^{-1}B_0'$. Now, we use that B and B_0 are normalized to have I_K as the lower $K \times K$ block. Hence:

$$\bar{B} = \bar{B}_0 + P_{0,[n-K]}(B_1 - B_{1,0})(B_0'B_0)^{-1} - \bar{B}_0(B_1 - B_{1,0})'B_{1,0}(B_0'B_0)^{-1} + O(|B_1 - B_{1,0}|^2),$$

where $P_{0,[n-K]}$ denotes the first $n - K$ columns of matrix P_0. By transposing and taking the vec we get:

$$\begin{aligned}\bar{b} &= \bar{b}_0 + \left(P_{0,[n-K]} \otimes (B_0'B_0)^{-1}\right) vec[(B_1 - B_{1,0})'] - \left(\bar{B}_0 \otimes [(B_0'B_0)^{-1}B_{1,0}']\right) vec(B_1 - B_{1,0}) + O(|b - b_0|^2) \\ &= \bar{b}_0 + \left[\left(P_{0,[n-K]} \otimes (B_0'B_0)^{-1}\right)\mathcal{K}_{n-K,K} - \bar{B}_0 \otimes \left((B_0'B_0)^{-1}B_{1,0}'\right)\right](b - b_0) + O(|b - b_0|^2).\end{aligned}$$

Thus, we get:

$$D_0 := \frac{\partial\bar{b}(b_0)}{\partial b'} = \left(P_{0,[n-K]} \otimes (B_0'B_0)^{-1}\right)\mathcal{K}_{n-K,K} - \bar{B}_0 \otimes \left((B_0'B_0)^{-1}B_{1,0}'\right).$$

By evaluating the derivatives at the true parameter values, and computing the expectations, we get:

$$\begin{aligned}\mathcal{J}_\phi &= -E(\boldsymbol{W}_t\boldsymbol{\xi}_{t-1}') \otimes I_K = -Q_{\boldsymbol{W}\boldsymbol{\xi}_{-1}} \otimes I_K \\ &= \begin{bmatrix} \Gamma_{\xi\xi}(q) & : \cdots : & \Gamma_{\xi\xi}(q+L-1) \\ \vdots & & \vdots \\ \Gamma_{\xi\xi}(1) & : \cdots : & \Gamma_{\xi\xi}(L) \end{bmatrix}' \otimes I_K,\end{aligned}$$

where $\Gamma_{\xi\xi}(h) := E[\xi_t \xi_{t-h}'] = \bar{B}_0' \left(\Gamma(h) - \boldsymbol{C}_0 \tilde{\boldsymbol{\Gamma}}(h-1) - \tilde{\boldsymbol{\Gamma}}(-h-1)' \boldsymbol{C}_0' + \boldsymbol{C}_0 \boldsymbol{\Gamma}(h) \boldsymbol{C}_0' \right) \bar{B}_0$, and

$$\begin{aligned} \mathcal{J}_c &= -\begin{bmatrix} E(e_t \boldsymbol{Y}_{t-q-2}') \\ \vdots \\ E(e_t \boldsymbol{Y}_{t-q-L-1}') \end{bmatrix} \otimes \bar{B}_0' \\ &\quad - E(\boldsymbol{W}_t \boldsymbol{Y}_{t-1}') \otimes \bar{B}_0' + \sum_{i=1}^{q} E(\boldsymbol{W}_t \boldsymbol{Y}_{t-i-1}') \otimes (\Phi_{i,0} \bar{B}_0'), \end{aligned}$$

and:

$$\mathcal{J}_b = \left\{ \begin{bmatrix} E(e_t \varepsilon_{t-q-1}') \\ \vdots \\ E(e_t \varepsilon_{t-q-L}') \end{bmatrix} \otimes I_K + E(\boldsymbol{W}_t \varepsilon_t') \otimes I_K - \sum_{i=1}^{q} E(\boldsymbol{W}_t \varepsilon_{t-i}') \otimes \Phi_{i,0} \right\} D_0,$$

where $e_t = \xi_t - \boldsymbol{\Phi}_0 \boldsymbol{\xi}_{t-1}$ and $\varepsilon_t = Y_t - \boldsymbol{C}_0 \boldsymbol{Y}_{t-1}$.

D Identification in ABCD state space representations

D.1 Review of identification of ABCD representations

This section reviews some results on the identification of linear state space systems provided in Glover and Willems (1974) and references therein. Consider a linear state space system in ABCD form

$$\begin{aligned} X_{t+1} &= \mathcal{A} X_t + \mathcal{B} W_t, \\ Y_t &= \mathcal{C} X_t + \mathcal{D} W_t, \end{aligned}$$

where $X_t \in \mathbb{R}^p$, $W_t \in \mathbb{R}^m$ and $Y_t \in \mathbb{R}^n$. The matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ are parameterized in terms of a parameter vector $\theta \in \mathbb{R}^q$. The latent input vector W_t is $WN(0, I)$. The eigenvalues of matrix $\mathcal{A}$ are assumed smaller than 1 in modulus, so that the system is covariance stationary.

By using $X_t = (I - \mathcal{A}L)^{-1} \mathcal{B} W_{t-1}$, we get

$$Y_t = \mathcal{C}(I - \mathcal{A}L)^{-1} \mathcal{B} W_{t-1} + \mathcal{D} W_t = G(L) W_t$$

where the transfer function is given by:

$$G(z) = \mathcal{C}(I - \mathcal{A}z)^{-1}\mathcal{B}z + \mathcal{D},$$

for complex argument z such that $z^{-1} \in \mathbb{C} \setminus \sigma(\mathcal{A})$, i.e. the complex plane excluding the spectrum of $\mathcal{A}$. Note that $G(z) = \mathcal{C}(Is - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}$ with $s = z^{-1}$, which yields the transfer function given in Glover and Willems (1974).

The spectral density function of the output process Y_t is:

$$f(\omega) = \frac{1}{2\pi}G(e^{i\omega})G(e^{i\omega})^* = \frac{1}{2\pi}G(e^{i\omega})G(e^{-i\omega})'$$

for the argument $\omega \in [-\pi, \pi]$, where $A^* = \bar{A}'$ is the Hermitian conjugate of complex matrix A. The spectral density function is in one-to-one relationship with the ACF. So, identification from the spectral density is tantamount to identification from the ACF. Moreover, since $f(-\omega) = \frac{1}{2\pi}G((e^{i\omega})^{-1})G((e^{i\omega})^{-1})^*$, we can equivalently consider identification from the function $\Phi(s) = G(s)G(\bar{s})'$, where

$$G(s) = \mathcal{C}(Is - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}.$$

We will refer to $G(s)$ and $\Phi(s)$ as transfer function and spectral function, respectively. These functions are well-defined for any complex argument s except the poles, which are located on the spectrum $\sigma(\mathcal{A})$ and lie inside the unit circle. Identification from those functions is equivalent to identification from the transfer and spectral functions defined in Section 3.4 of the paper.

D.1.1 Identification from the transfer function

Let $(\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1, \mathcal{D}_1)$ and $(\mathcal{A}_2, \mathcal{B}_2, \mathcal{C}_2, \mathcal{D}_2)$ be two sets of state space parameters, corresponding to structural parameters θ_1 and θ_2, respectively. Let $G_1(s)$ and $G_2(s)$ be the two corresponding transfer functions. The model is identifiable from the transfer function if, and only if,

$$G_1(s) = G_2(s), \; \forall s \in \mathbb{D} \qquad \Rightarrow \qquad \theta_1 = \theta_2,$$

where $\mathbb{D}$ is a domain in the complex plane where the transfer functions are well defined. Note that by complex analysis, an analytic function which equals zero on a path in $\mathbb{C}$, also vanishes on a larger domain outside the poles. Thus, the specification of $\mathbb{D}$ is not relevant for the purpose of identification.

The arguments used by Glover and Willems (1974) on p. 644 in the proof of their Proposition 1 yield the following lemma.

Lemma 5. *Suppose that the parameterization of the system is globally minimal, i.e., the dimension of the state vector is less than or equal to that of any other system having the same transfer function. Then, $G_1(s) = G_2(s)$ for any s in a complex domain if, and only if, there exists an invertible matrix T such that:*

$$\mathcal{A}_1 = T\mathcal{A}_2T^{-1}, \tag{D.1}$$

$$\mathcal{B}_1 = T\mathcal{B}_2, \tag{D.2}$$

$$\mathcal{C}_1 = \mathcal{C}_2T^{-1}, \tag{D.3}$$

$$\mathcal{D}_1 = \mathcal{D}_2. \tag{D.4}$$

Proof. Suppose we have $\mathcal{C}_1(Is - \mathcal{A}_1)^{-1}\mathcal{B}_1 + \mathcal{D}_1 = \mathcal{C}_2(Is - \mathcal{A}_2)^{-1}\mathcal{B}_2 + \mathcal{D}_2$ for all s. By evaluating at argument s with large modulus, it implies $\mathcal{D}_1 = \mathcal{D}_2$. Moreover, since the l.h.s. and r.h.s. must have the same poles, and the system is minimal, matrices $\mathcal{A}_1$ and $\mathcal{A}_2$ have the same dimensions and the same eigenvalues. Therefore, $\mathcal{A}_1 = T_1\Lambda T_1^{-1}$ and $\mathcal{A}_2 = T_2\Lambda T_2^{-1}$ for the diagonal matrix Λ of eigenvalues (we are assuming that $\mathcal{A}_1$ and $\mathcal{A}_2$ are diagonalizable to simplify). It then follows $\tilde{\mathcal{C}}_1(Is - \Lambda)^{-1}\tilde{\mathcal{B}}_1 = \tilde{\mathcal{C}}_2(Is - \Lambda)^{-1}\tilde{\mathcal{B}}_2$ for all s, where $\tilde{\mathcal{B}}_i = T_i^{-1}\mathcal{B}_i$ and $\tilde{\mathcal{C}}_i = \mathcal{C}_iT_i$. By writing $\Lambda = diag(\lambda_1, ..., \lambda_p)$, $\tilde{\mathcal{C}}_i = [c_{i,1} : \cdots : c_{i,p}]$ and $\tilde{\mathcal{B}}_i' = [b_{i,1} : \cdots : b_{i,p}]$, we get:

$$\sum_{j=1}^{p} \frac{1}{s - \lambda_j} c_{1,j} b_{1,j}' = \sum_{j=1}^{p} \frac{1}{s - \lambda_j} c_{2,j} b_{2,j}', \qquad \forall s,$$

which implies $c_{1,j}b_{1,j}' = c_{2,j}b_{2,j}'$ for $j = 1, ..., p$. Minimality of the parameterization excludes the possibility that some $b_{i,j}$ or $c_{i,j}$ are the zero vector. Therefore, $c_{1,j} = c_{2,j}r_j$ and $b_{1,j} = b_{2,j}r_j^{-1}$ for $r_j > 0$, and all j. Then, we get $\tilde{\mathcal{C}}_1 = \tilde{\mathcal{C}}_2R$ and $\tilde{\mathcal{B}}_1 = R^{-1}\tilde{\mathcal{B}}_2$ for $R = diag(r_1, ..., r_p)$. This yields $\mathcal{C}_1 = \mathcal{C}_2T_2RT_1^{-1}$ and $\mathcal{B}_1 = T_1R^{-1}T_2^{-1}\mathcal{B}_2$. The conclusion follows with $T = T_1R^{-1}T_2^{-1}$. □

The lemma yields the following corollary.

Corollary 1. *Suppose that the parameterization of the system is globally minimal. Then, the state space model is globally identified from the spectral function if, and only if, the validity of equations (D.1)-(D.4)* *for some invertible matrix T implies $\theta_1 = \theta_2$.*

D.1.2 Identification from the spectral function

The model is identifiable from the spectral density if, and only if,

$$\Phi_1(s) = \Phi_2(s),\ \forall s \in \mathbb{D} \qquad \Rightarrow \qquad \theta_1 = \theta_2.$$

Glover and Willems (1974) give the following Lemma.

Lemma 6. *Suppose that the parameterization of the system is globally minimal, i.e., the dimension of the state vector is less than or equal to that of any other system having the same output spectral density function. Then, $\Phi_1(s) = \Phi_2(s)$ for any s in a complex domain if, and only if, there exists an invertible matrix T and a symmetric matrix Q such that:*

$$\mathcal{A}_1 = T\mathcal{A}_2T^{-1}, \tag{D.5}$$

$$\mathcal{C}_1 = \mathcal{C}_2T^{-1}, \tag{D.6}$$

$$Q\mathcal{A}_1' + \mathcal{A}_1Q = -\mathcal{B}_1\mathcal{B}_1' + T\mathcal{B}_2\mathcal{B}_2'T', \tag{D.7}$$

$$Q\mathcal{C}_1' = -\mathcal{B}_1\mathcal{D}_1' + T\mathcal{B}_2\mathcal{D}_2', \tag{D.8}$$

$$\mathcal{D}_1\mathcal{D}_1' = \mathcal{D}_2\mathcal{D}_2'. \tag{D.9}$$

Proof. Consider the spectral function for a purely imaginary argument $s = it$, $t \in \mathbb{R}$. Equality of the spectral functions for two parameterizations on this line is equivalent to equality of the spectral functions on a larger domain excluding the poles. Then for imaginary argument s we have:

$$\begin{aligned}\Phi_i(s) \quad = \quad & \mathcal{C}_i(Is - \mathcal{A}_i)^{-1}\mathcal{B}_i\mathcal{B}_i'(-Is - \mathcal{A}_i')^{-1}\mathcal{C}_i' + \mathcal{C}_i(Is - \mathcal{A}_i)^{-1}\mathcal{B}_i\mathcal{D}_i' \\ & + \mathcal{D}_i\mathcal{B}_i'(-Is - \mathcal{A}_i')^{-1}\mathcal{C}_i' + \mathcal{D}_i\mathcal{D}_i'.\end{aligned}$$

Anderson (1969) on pages 138-139 states that minimality implies complete controllability, which in turn implies that for each $i = 1, 2$ there exists a unique positive definite matrix Q_i such that

$$-\mathcal{B}_i\mathcal{B}_i' = Q_i\mathcal{A}_i' + \mathcal{A}_iQ_i. \tag{D.10}$$

Then we have:

$$
\begin{aligned}
(Is-\mathcal{A}_i)^{-1}\mathcal{B}_i\mathcal{B}_i'(-Is-\mathcal{A}_i')^{-1} &= -(Is-\mathcal{A}_i)^{-1}(Q_i\mathcal{A}_i'+\mathcal{A}_iQ_i)(-Is-\mathcal{A}_i')^{-1}\\
&= (Is-\mathcal{A}_i)^{-1}Q_i+Q_i(-Is-\mathcal{A}_i')^{-1}.
\end{aligned}
$$

Thus, the spectral function becomes:

$$
\begin{aligned}
\Phi_i(s) &= \mathcal{C}_i(Is-\mathcal{A}_i)^{-1}Q_i\mathcal{C}_i'+\mathcal{C}_iQ_i(-Is-\mathcal{A}_i')^{-1}\mathcal{C}_i'+\mathcal{C}_i(Is-\mathcal{A}_i)^{-1}\mathcal{B}_i\mathcal{D}_i'\\
&\quad +\mathcal{D}_i\mathcal{B}_i'(-Is-\mathcal{A}_i')^{-1}\mathcal{C}_i'+\mathcal{D}_i\mathcal{D}_i'\\
&= \mathcal{C}_i(Is-\mathcal{A}_i)^{-1}(Q_i\mathcal{C}_i'+\mathcal{B}_i\mathcal{D}_i')\\
&\quad +(\mathcal{C}_iQ_i+\mathcal{D}_i\mathcal{B}_i')(-Is-\mathcal{A}_i')^{-1}\mathcal{C}_i'+\mathcal{D}_i\mathcal{D}_i'.
\end{aligned}
$$

Hence, the equation $\Phi_1(s)=\Phi_2(s)$ yields:

$$
\begin{aligned}
&\mathcal{C}_1(Is-\mathcal{A}_1)^{-1}(Q_1\mathcal{C}_1'+\mathcal{B}_1\mathcal{D}_1')+(\mathcal{C}_1Q_1+\mathcal{D}_1\mathcal{B}_1')(-Is-\mathcal{A}_1')^{-1}\mathcal{C}_1'+\mathcal{D}_1\mathcal{D}_1'\\
=\;&\mathcal{C}_2(Is-\mathcal{A}_2)^{-1}(Q_2\mathcal{C}_2'+\mathcal{B}_2\mathcal{D}_2')+(\mathcal{C}_2Q_2+\mathcal{D}_2\mathcal{B}_2')(-Is-\mathcal{A}_2')^{-1}\mathcal{C}_2'+\mathcal{D}_2\mathcal{D}_2'.
\end{aligned}
$$

By considering arguments s with large modulus, it follows (D.9). Moreover, by the arguments in the proof of Lemma 1, it follows equations (D.5), (D.6) and:

$$
Q_1\mathcal{C}_1'+\mathcal{B}_1\mathcal{D}_1'=T(Q_2\mathcal{C}_2'+\mathcal{B}_2\mathcal{D}_2'). \tag{D.11}
$$

Moreover, equation (D.6) implies $TQ_2\mathcal{C}_2'=TQ_2T'\mathcal{C}_1'$. Plugging the latter equation into (D.11) yields equation (D.8) with $Q=Q_1-TQ_2T'$. Finally, taking the difference between equation (D.10) for $i=1$, and the same equation for $i=2$ pre-multiplied by T and post-multiplied by T', we get equation (D.7).

□

This lemma yields the following corollary.

Corollary 2. *Suppose that the parameterization of the system is globally minimal. Then, the state space model is globally identified from the spectral function if, and only if, the validity of equations (D.5)-(D.9) for some T invertible and Q symmetric implies $\theta_1=\theta_2$.*

If system (D.5)-(D.9) admits solutions with $\theta_1 \neq \theta_2$, but system (D.1)-(D.4) implies $\theta_1 = \theta_2$, then the ABCD representation is identifiable from the transfer function but not from the spectral function.

D.1.3 Discussion of global minimality

A realization of a linear state space system is minimal if it is controllable and observable, see Glover and Willems (1974, Theorem 2.2.). We define the observability matrix as:

$$\mathcal{O}(\mathcal{C}, \mathcal{A}) = \begin{bmatrix} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \vdots \\ \mathcal{C}\mathcal{A}^{N-1} \end{bmatrix} \tag{D.12}$$

and the controllability matrix as

$$\mathfrak{C}(\mathcal{B}, \mathcal{A}) = \begin{bmatrix} \mathcal{B} & \mathcal{A}\mathcal{B} & \cdots & \mathcal{A}^{N-1}\mathcal{B} \end{bmatrix}. \tag{D.13}$$

A realization of the state space system is observable if the observability matrix $\mathcal{O}(\mathcal{C}, \mathcal{A})$ is of full rank. A realization of the state space system is controllable if the controllability matrix $\mathfrak{C}(\mathcal{B}, \mathcal{A})$ is of full rank.

D.2 Application of identification results for the state space (2.1)-(2.2) when $p = q = 1$ (proof of Proposition 5)

The ABC(D) representation of our model with state $X_t = (Y_t', f_{t+1}')'$ is

$$\mathcal{A} = \begin{bmatrix} C & B \\ 0 & \Phi \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} \Sigma_u^{1/2} & 0 \\ 0 & \Sigma_v^{1/2} \end{bmatrix}, \quad \mathcal{C} = \begin{bmatrix} I_n & 0_{n\times K} \end{bmatrix}, \quad \mathcal{D} = 0. \tag{D.14}$$

We first check that this representation is globally minimal. Then we derive the implications of Corollaries 2 and 4 for identification.

D.2.1 Minimality

In the ABC(D) representation of our model, the system is always controllable, i.e. $\mathfrak{C}(\mathcal{B}, \mathcal{A})$ is full rank because $\mathcal{B}$ is of full rank. Moreover, the system is observable. In fact, the observability matrix

$$\mathcal{O}(\mathcal{C},\mathcal{A}) = \begin{bmatrix} I_n & 0 \\ C & B \\ \vdots & \vdots \end{bmatrix} \tag{D.15}$$

is full rank because B is full rank. Therefore, the system is minimal.

D.2.2 Identification from the transfer function

Let us consider Equations (D.1)-(D.4) for the ABCD representation (D.14). Equation (D.4) is trivially satisfied. Equation (D.3) implies $\mathcal{C}_2 = \mathcal{C}_1 T$, which yields:

$$T = \begin{bmatrix} I_n & 0 \\ T_{21} & T_{22} \end{bmatrix}, \qquad T^{-1} = \begin{bmatrix} I_n & 0 \\ -(T_{22})^{-1}T_{21} & (T_{22})^{-1} \end{bmatrix},$$

with T_{22} invertible $K \times K$ and T_{21} of dimensions $K \times n$. Equation (D.2) yields $T_{21} = 0$, $\Sigma_{u,1} = \Sigma_{u,2}$ and $\Sigma_{v,1}^{1/2} = T_{22}\Sigma_{v,2}^{1/2}$.

Let us now focus on equation (D.1). We have:

$$\begin{aligned} T\mathcal{A}_2T^{-1} &= \begin{bmatrix} I_n & 0 \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} C_2 & B_2 \\ 0 & \Phi_2 \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & (T_{22})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} C_2 & B_2(T_{22})^{-1} \\ 0 & T_{22}\Phi_2(T_{22})^{-1} \end{bmatrix}. \end{aligned}$$

Hence equation (D.1) is equivalent to the system:

$$\begin{cases} C_2 &= C_1 \\ B_2(T_{22})^{-1} &= B_1 \\ T_{22}\Phi_2(T_{22})^{-1} &= \Phi_1. \end{cases}$$

With the normalization of the lower $K \times K$ block of matrix B being the identity I_K, the second equation in the system implies $T_{22} = I_K$, and hence $B_2 = B_1$. Then, from the third equation we get $\Phi_1 = \Phi_2$.

The above arguments show that the state space parameters are identifiable from the transfer function. This yields in Proposition 5 (a).

D.2.3 Identification from the spectral function

Let us derive the implications of Equations (D.5)-(D.9). Equation (D.9) is trivially satisfied. Equation (D.8) becomes $Q\mathcal{C}_1' = 0$, which yields:

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & Q_{22} \end{bmatrix},$$

with Q_{22} a symmetric $K \times K$ matrix. Further, as for the identification from the transfer function, Equation (D.6) implies $\mathcal{C}_2 = \mathcal{C}_1 T$, which yields:

$$T = \begin{bmatrix} I_n & 0 \\ T_{21} & T_{22} \end{bmatrix}, \qquad T^{-1} = \begin{bmatrix} I_n & 0 \\ -(T_{22})^{-1}T_{21} & (T_{22})^{-1} \end{bmatrix},$$

with T_{22} invertible $K \times K$ and T_{21} of dimensions $K \times n$.

Let us now focus on equation (D.5). We have:

$$\begin{aligned} T\mathcal{A}_2T^{-1} &= \begin{bmatrix} I_n & 0 \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} C_2 & B_2 \\ 0 & \Phi_2 \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -(T_{22})^{-1}T_{21} & (T_{22})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} C_2 - B_2(T_{22})^{-1}T_{21} & B_2(T_{22})^{-1} \\ T_{21}C_2 - T_{21}B_2(T_{22})^{-1}T_{21} - T_{22}\Phi_2(T_{22})^{-1}T_{21} & T_{21}B_2(T_{22})^{-1} + T_{22}\Phi_2(T_{22})^{-1} \end{bmatrix}. \end{aligned}$$

Hence equation (D.5) is equivalent to the system:

$$\left\{ \begin{aligned} C_2 - B_2(T_{22})^{-1}T_{21} &= C_1 \\ B_2(T_{22})^{-1} &= B_1 \\ T_{21}C_2 - T_{21}B_2(T_{22})^{-1}T_{21} - T_{22}\Phi_2(T_{22})^{-1}T_{21} &= 0 \\ T_{21}B_2(T_{22})^{-1} + T_{22}\Phi_2(T_{22})^{-1} &= \Phi_1 \end{aligned} \right.$$

i.e. to the system:

$$\begin{cases} C_2 &= C_1 + B_2(T_{22})^{-1}T_{21} \\ B_2(T_{22})^{-1} &= B_1 \\ T_{21}C_1 &= T_{22}\Phi_2(T_{22})^{-1}T_{21} + T_{21}B_2(T_{22})^{-1}T_{21} \\ T_{21}B_2(T_{22})^{-1} + T_{22}\Phi_2(T_{22})^{-1} &= \Phi_1. \end{cases}$$

Let us finally consider equation (D.7). We have:

$$Q\mathcal{A}_1' + \mathcal{A}_1 Q = \begin{bmatrix} 0 & B_1 Q_{22} \\ Q_{22}B_1' & Q_{22}\Phi_1' + \Phi_1 Q_{22} \end{bmatrix},$$

$$\mathcal{B}_1\mathcal{B}_1' = \begin{bmatrix} \Sigma_{u,1} & 0 \\ 0 & \Sigma_{v,1} \end{bmatrix},$$

and:

$$T\mathcal{B}_2\mathcal{B}_2'T' \quad = \quad \begin{bmatrix} \Sigma_{u,2} & \Sigma_{u,2}(T_{21})' \\ T_{21}\Sigma_{u,2} & T_{21}\Sigma_{u,2}(T_{21})' + T_{22}\Sigma_{v,2}(T_{22})' \end{bmatrix}.$$

Therefore, equation (D.7) is equivalent to the system:

$$\begin{cases} 0 &= -\Sigma_{u,1} + \Sigma_{u,2} \\ B_1 Q_{22} &= \Sigma_{u,2}(T_{21})' \\ Q_{22}\Phi_1' + \Phi_1 Q_{22} &= -\Sigma_{v,1} + T_{21}\Sigma_{u,2}(T_{21})' + T_{22}\Sigma_{v,2}(T_{22})'. \end{cases}$$

We conclude that for our state space system equations (D.5)-(D.9) are equivalent to:

$$
\begin{aligned}
C_2 &= C_1 + B_2(T_{22})^{-1}T_{21} && \text{(D.16)}\\
B_2(T_{22})^{-1} &= B_1 && \text{(D.17)}\\
T_{21}C_1 &= T_{22}\Phi_2(T_{22})^{-1}T_{21} + T_{21}B_2(T_{22})^{-1}T_{21} && \text{(D.18)}\\
\Phi_1 &= T_{21}B_2(T_{22})^{-1} + T_{22}\Phi_2(T_{22})^{-1} && \text{(D.19)}\\
\Sigma_{u,1} &= \Sigma_{u,2} && \text{(D.20)}\\
B_1Q_{22} &= \Sigma_{u,2}(T_{21})' && \text{(D.21)}\\
Q_{22}\Phi_1' + \Phi_1Q_{22} &= -\Sigma_{v,1} + T_{21}\Sigma_{u,2}(T_{21})' + T_{22}\Sigma_{v,2}(T_{22})'. && \text{(D.22)}
\end{aligned}
$$

We now discuss global identification from these equations. We use the identification restriction that the lower $K \times K$ block of matrix B is the identity I_K. Then, equation (D.17) implies $T_{22} = I_K$ and hence $B_1 = B_2$. So the system becomes:

$$
\begin{aligned}
C_2 &= C_1 + B_2T_{21} && \text{(D.23)}\\
B_2 &= B_1 && \text{(D.24)}\\
T_{21}C_1 &= \Phi_2T_{21} + T_{21}B_2T_{21} && \text{(D.25)}\\
\Phi_1 &= T_{21}B_2 + \Phi_2 && \text{(D.26)}\\
\Sigma_{u,1} &= \Sigma_{u,2} && \text{(D.27)}\\
B_1Q_{22} &= \Sigma_{u,2}(T_{21})' && \text{(D.28)}\\
Q_{22}\Phi_1' + \Phi_1Q_{22} &= -\Sigma_{v,1} + T_{21}\Sigma_{u,2}(T_{21})' + \Sigma_{v,2}. && \text{(D.29)}
\end{aligned}
$$

Thus, we get

$$
\begin{aligned}
C_2 &= C_1 + B_2T_{21}\\
B_2 &= B_1\\
\Phi_2 &= \Phi_1 - T_{21}B_1\\
\Sigma_{u,2} &= \Sigma_{u,1}\\
\Sigma_{v,2} &= \Sigma_{v,1} + Q_{22}\Phi_1' + \Phi_1Q_{22} - T_{21}\Sigma_{u,1}(T_{21})',
\end{aligned}
$$

where matrices Q_{22} and T_{21} solve

$$\begin{aligned} T_{21}C_1 &= \Phi_1 T_{21}, && \text{(D.30)} \\ B_1 Q_{22} &= \Sigma_{u,1}(T_{21})'. && \text{(D.31)} \end{aligned}$$

Therefore, the ABCD model is identifiable by the spectral function if equations (D.30)-(D.31) admit $T_{21} = 0$ and $Q_{22} = 0$ as the only solution. Now, from (D.31) we have $T_{21} = Q_{22}B_1'\Sigma_{u,1}^{-1}$, and plugging this into (D.30) we get:

$$Q_{22}B_1'\Sigma_{u,1}^{-1}C_1 = \Phi_1 Q_{22}B_1'\Sigma_{u,1}^{-1}.$$

By right-multiplying times $\Sigma_{u,1}$ and transposing this equation, we get:

$$\Sigma_{u,1}C_1'\Sigma_{u,1}^{-1}B_1 Q_{22} = B_1 Q_{22}\Phi_1'.$$

This yields in Proposition 5 (b).

E Additional results

E.1 Proof of the unidentifiability of the (unrestricted) VARMA representation

In this section we show that parameters of the VARMA($p+q$,q) representation are not identifiable in general, unless the structural relationships in equations (A.2)-(A.3) are fully imposed. We present our argument in the case $p = q = 1$. Then, the VARMA(2,1) representation reads:

$$Y_t = (B\Phi\bar{B}' + C)Y_{t-1} - (B\Phi\bar{B}'C)Y_{t-2} + w_t + B\nu' w_{t-1}, \quad w_t \sim WN(0, \Sigma_w). \tag{E.1}$$

The matrices ν and Σ_w are such that $\Sigma_w + B\nu'\Sigma_w\nu B' = B\Sigma_v B' + \Sigma_u + B\Phi\bar{B}'\Sigma_u\bar{B}\Phi' B'$ and $\nu'\Sigma_w = \Phi\bar{B}'\Sigma_u$. Let us suppose that these two restrictions are not imposed on the parameters of (E.1), and investigate identifiability of B, C, Φ, ν, Σ_w.

Proposition 11. *Assume the matrix equation*

$$aB_0'a + C_0'a - a(B_0'B_0)\Phi_0' = 0 \tag{E.2}$$

admits non-trivial solutions for the $n \times K$ matrix a. Then, given the DGP in the state-space model (2.1)-(2.2) *with $p = q = 1$, the parameters B_0, C_0, Φ_0, ν_0 and $\Sigma_{w,0}$ are not identifiable from the VARMA(2,1) representation* (E.1).

Equation (E.2) is a so-called non-symmetric algebraic Riccati equation. Dilip and Pillai (2016) show that in general, this equation has non-trivial real-valued solutions. As an example, consider the single-factor case $K = 1$. Then, equation (E.2) is an eigenvalue-eigenvector equation $C_0' a = ((B_0' B_0)\Phi_0 - B_0' a)a$ for matrix C_0' (notice Φ_0 is a scalar here). Suppose there exists an eigenvector v of C_0' which is not in $\mathcal{R}(B_0)^\perp$, and let λ be its eigenvalue, with $\lambda \neq \Phi_0$. Then $a = \frac{1}{B_0' v}(\Phi_0 - \lambda)v$ is a solution of equation (E.2).

Proof of Proposition 11: Let us consider the VARMA(2,1) representation in (E.1) written for the true parameter values:

$$[I_n - (B_0 \Phi_0 \bar{B}_0' + C_0)L + (B_0 \Phi_0 \bar{B}_0' C_0)L^2]Y_t = (I_n + B_0 \nu_0' L)w_t.$$

Let us multiply both sides by $I_n - B_0 \varsigma' B_{0\perp}' L$, where ς is a $(n-K) \times K$ matrix. Then we get:

$$Y_t = [B_0(\Phi_0 B_0' + \varsigma' B_{0\perp}') + C_0]Y_{t-1} - [B_0(\varsigma' B_{0\perp}' + \Phi_0 B_0')C_0]Y_{t-2} + w_t + B_0(\nu_0 - B_{0\perp}\varsigma)' w_{t-1}. \quad \text{(E.3)}$$

This yields another VARMA(2,1) representation as (E.1) with same parameter B_0 but different parameters C, Φ, ν, if we can find matrices C, Φ, ς such that

$$B_0(\Phi_0 B_0' + \varsigma' B_{0\perp}') + C_0 = B_0 \Phi B_0' + C \quad \text{(E.4)}$$

$$B_0(\varsigma' B_{0\perp}' + \Phi_0 B_0')C_0 = B_0 \Phi B_0' C, \quad \text{(E.5)}$$

and either $C \neq C_0$, or $\Phi \neq \Phi_0$, or $\varsigma \neq 0$. For this purpose, let us transform equation (E.4). Pre-multiplying equation (E.4) by $\bar{B}_0'$ we get

$$\Phi_0 B_0' + \varsigma' B_{0\perp}' + \bar{B}_0' C_0 = \Phi B_0' + \bar{B}_0' C. \quad \text{(E.6)}$$

Post-multiplying equation (E.6) by $\bar{B}_0$ we get $\Phi_0 + \bar{B}_0' C_0 \bar{B}_0 = \Phi + \bar{B}_0' C \bar{B}_0$, i.e.

$$\Phi - \Phi_0 = -\bar{B}_0'(C - C_0)\bar{B}_0. \quad \text{(E.7)}$$

Post-multiplying equation (E.6) by $\bar{B}_{0\perp}$ we get $\varsigma' + \bar{B}_0' C_0 \bar{B}_{0\perp} = \bar{B}_0' C \bar{B}_{0\perp}$, i.e.

$$\varsigma' = \bar{B}_0'(C - C_0)\bar{B}_{0\perp}. \tag{E.8}$$

Now, pre-multiplying (E.4) by $B_{0\perp}'$ we get $B_{0\perp}'(C - C_0) = 0$, which is equivalent to

$$C - C_0 = B_0 a' \tag{E.9}$$

where a is an $n \times K$ matrix. The equation (E.7) and (E.9) imply $\Phi - \Phi_0 = -a'\bar{B}_0$, while equations (E.8) and (E.9) imply $\varsigma' = a'\bar{B}_{0\perp}$. Therefore, equation (E.4) is satisfied if, and only if,

$$C = C_0 + B_0 a', \quad \Phi = \Phi_0 - a'\bar{B}_0, \quad \varsigma' = a'\bar{B}_{0\perp}. \tag{E.10}$$

By plugging these equations into (E.5), using that B_0 is full rank, we get:

$$\begin{aligned} (\varsigma' B_{0\perp}' + \Phi_0 B_0')C_0 &= \Phi B_0' C, \\ a'\bar{B}_{0\perp} B_{0\perp}' C_0 + \Phi_0 B_0' C_0 &= (\Phi_0 - a'\bar{B}_0) B_0'(C_0 + B_0 a'), \\ a'\bar{B}_{0\perp} B_{0\perp}' C_0 - \Phi_0 (B_0' B_0) a' + a'\bar{B}_0 B_0' C_0 + a' B_0 a' &= 0. \end{aligned}$$

Hence, by using $\bar{B}_{0\perp} B_{0\perp}' + \bar{B}_0 B_0' = I_n$, we get

$$a' C_0 - \Phi_0 (B_0' B_0) a' + a' B_0 a' = 0. \tag{E.11}$$

In Proposition 11 we assume that the matrix equation (E.11) admits non-trivial solutions for $n \times K$ matrix a. Then, we have solutions of equations (E.4)-(E.5) with $C \neq C_0$, $\Phi \neq \Phi_0$ and $\varsigma \neq 0$.

E.2 Identification of variance-covariance matrices Σ_u and Σ_v

Some transformations of the variance-covariance matrix Σ_u are semi-parametrically identifiable from second-order moments of processes η_t and ξ_t. Indeed, we have $V(\eta_t) = B_{0\perp}' \Sigma_u B_{0\perp}$ and $Cov(\eta_t, \xi_t) = B_{0\perp}' \Sigma_u \bar{B}_0$. For the identification of the complementary transformation $\bar{B}_0' \Sigma_u \bar{B}_0$ and of variance Σ_v, we use the ACF of process ξ_t and the VAR(q) dynamics of the factor. Indeed, from the Yule-Walker equations of process (f_t) we get:

$$\Gamma_f(0) = \sum_{j=1}^{q} \Phi_j \Gamma_f(-j) + \Sigma_v, \tag{E.12}$$

$$\Gamma_f(h) = \sum_{j=1}^{q} \Phi_j \Gamma_f(h-j), \qquad 1 \leq h \leq q, \tag{E.13}$$

where $\Gamma_f(h) = Cov(f_t, f_{t-h})$. Since the autoregressive matrices Φ_j and the autocovariances $\Gamma_f(h) = Cov(\xi_t, \xi_{t-h})$ for $h \neq 0$ are identified (Section 3.3), system (E.13) yields a linear matrix equation to identify the factor variance $\Gamma_f(0) = V(f_t)$. If we just consider the equation for $h = q$ in (E.13), we get $\Phi_q \Gamma_f(0) = Cov(\xi_t, \xi_{t-q}) - \sum_{j=1}^{q-1} \Phi_j Cov(\xi_t, \xi_{t-q+j})$. Under Assumption ID.3, the latter equation admits a unique solution and $\Gamma_f(0)$ is identified. From $V(\xi_t) = \Gamma_f(0) + \bar{B}_0' \Sigma_u \bar{B}_0$, the identification of $\bar{B}_0' \Sigma_u \bar{B}_0$ follows immediately. Finally, equation (E.12) yields the identification of Σ_v.

Proposition 12. *Under Assumptions M.1, M.2, IR.1, ID.1, ID.2 and ID.3, matrices Σ_u and Σ_v are identifiable from the autocovariances of process $\{Y_t\}$.*

We can link Proposition 12 to specifications (i)-(iii) for Σ_u considered in Section 2.1. Differently from a static factor model, we can identify the variance-covariance matrix Σ_u of the innovation u_t without imposing a diagonal structure on it (case (i)), because we rely on the factor dynamics to identify parameters $B_0, \boldsymbol{C}_0$. Once Σ_u is identified, matrices Λ and D in parameterization (ii) are identifiable under the identification conditions for static factor models (e.g. Anderson and Rubin (1956), Lawley and Maxwell (1971)). In case (iii), identification of matrices Q_0 and D requires e.g. exclusion restrictions on the elements of Q_0 as in the literature on Structural VAR models (see e.g. Stock and Watson (2016) for a recent survey).

E.3 Estimators of Σ_u and Σ_v

We estimate transformations $\widehat{B_{0\perp}' \Sigma_u B_{0\perp}} = \hat{Q}_{\eta\eta}$, $\widehat{B_{0\perp}' \Sigma_u \bar{B}_0} = \hat{Q}_{\eta\xi}$ and $\widehat{\bar{B}_0' \Sigma_u \bar{B}_0} = \hat{Q}_{\xi\xi} - \hat{\Gamma}_f(0)$, where $\hat{Q}_{\eta\eta}$, $\hat{Q}_{\xi\xi}$ and $\hat{Q}_{\eta\xi}$ denote the sample variances of $\hat{\eta}_t$ and $\hat{\xi}_t$, and their sample covariance matrix, respectively. We estimate the factor variance $\hat{\Gamma}_f(0)$ by the symmetric part of the matrix which solves approximately in least square sense the linear system (E.13) after replacing $\boldsymbol{\Phi}$ with $\hat{\boldsymbol{\Phi}}$, and $\Gamma_f(h)$ with the sample ACF $\hat{\Gamma}_\xi(h)$ of $\hat{\xi}_t$ for $h \neq 0$. The symmetric part of matrix A is $\{A\}_s := \frac{1}{2}(A + A')$. Finally, $\hat{\Sigma}_v = \hat{\Gamma}_f(0) - \{\sum_{j=1}^{q} \hat{\Phi}_j \hat{\Gamma}_\xi(-j)\}_s$. This procedure does not ensure positive-definiteness of estimates

$\hat{\Sigma}_u$ and $\hat{\Sigma}_v$. Positive-semidefiniteness can be imposed by some regularization procedure, e.g. replacing negative eigenvalues with zero.

E.4 Identification of model parameters when some of the factors are static

In Section 3, we have investigated the identification strategy of the model parameters for the state space model (2.1)-(2.2) when the autoregressive matrix $\mathbf{\Phi}_0$ is assumed to be a full rank matrix (Section 3.1), and matrix $\Phi_{q,0}$ is non-singular (Section 3.3). If $\mathbf{\Phi}_0$ is a full rank matrix, then all the elements of the factor vector f_t and their linear combinations show a certain dynamics. In this section we analyse a model in which the factors have some directions driven by a pure shock and discuss identification. To simplify, we assume $p = q = 1$ throughout this section.

Let us assume that the rank of matrix Φ is $r < K$. Upon rotation of the factors, this model can be written as

$$Y_t = CY_{t-1} + \Lambda_g g_t + \Lambda_h h_t + u_t \tag{E.14}$$

$$\begin{bmatrix} g_t \\ h_t \end{bmatrix} = \begin{bmatrix} \Phi_{gg} & \Phi_{gh} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} g_{t-1} \\ h_{t-1} \end{bmatrix} + \begin{bmatrix} v_t^g \\ v_t^h \end{bmatrix}, \tag{E.15}$$

where we split the factor vector $f_t = (g_t', h_t')'$ such that g_t and h_t are respectively r and $(K - r)$ dimensional vectors, Λ_g and Λ_h are respectively $n \times r$ and $n \times (K - r)$ matrices of loadings, and $[\Phi_{gg} : \Phi_{gh}]$ has full row rank. The unobservable factors h_t are driven by a pure shock, instead the unobservable factors g_t evolve dynamically. More precisely, process (h_t) is White Noise and process (g_t) follows a VARMA(1,1) model.

For model (E.14)-(E.15) we have

$$EL(Y_t|Y_{t-1}, Y_{t-2}) = C_0 Y_{t-1} + \Lambda_g EL(g_t|Y_{t-1}, Y_{t-2}) = (C_0 + \Lambda_g F_{g,1}^{(2)})Y_{t-1} + \Lambda_g F_{g,2}^{(2)} Y_{t-2},$$

where $F_{g,1}^{(2)}, F_{g,2}^{(2)}$ are the coefficients of the linear projection of g_t on Y_{t-1}, Y_{t-2}. To identify the parameters we should impose identification restrictions on Λ_g as in Section 3 for the loadings B, as well as a rank condition on $F_{g,2}^{(2)}$.

Assumption IR.1. *Matrix Λ_g is such that*

$$\Lambda_g = \begin{bmatrix} \Lambda_{g,1} \\ I_r \end{bmatrix}$$

where $\Lambda_{g,1}$ is the upper $(n-r) \times r$ block.

Assumption RR.1. *The matrix $F_{g,2}^{(2)}$ has full rank.*

If Identification Restriction IR.1 and Assumption RR.1 hold, then Λ_g is identifiable with the same argument for the identification of B in Section 3.1 relying on the spectrum of matrix $A_2^* A_2^{*\prime}$. In particular, the number of non-zero eigenvalues of this matrix identifies the number r of dynamic factors.

A second identification restriction for the identification of $\Lambda_{g,\perp}$ has to be defined.

Assumption IR.2. *Matrix $\Lambda_{g,\perp}$ is such that $\Lambda_{g,\perp} = [I_{n-r} \,:\, -\Lambda_{g,1}]'$.*

For the identification of matrix C, let us consider the following moment conditions:

$$E[(\Lambda_g g_t + \Lambda_h h_t + u_t)(\Lambda_{g,\perp}'(\Lambda_h h_{t+j} + u_{t+j}))'] = 0, \quad j \geq 1.$$

These conditions hold because we have $E(g_t u_{t+j}') = 0$, $E(g_t h_{t+j}') = 0$, $E(h_t u_{t+j}') = 0$, $E(h_t h_{t+j}') = 0$, $E(u_t u_{t+j}') = 0$, $E(u_t h_{t+j}') = 0$, for $j \geq 1$. We can notice that $\Lambda_g g_t + \Lambda_h h_t + u_t = Y_t - C_0 Y_{t-1}$ and $\Lambda_{g,\perp}'(\Lambda_h h_{t+j} + u_{t+j}) = \Lambda_{g,\perp}'(Y_{t+j} - C_0 Y_{t+j-1})$. Moreover, the parameter $\delta_0' = \Lambda_{g,\perp}' C_0$ is identifiable as in Section 3.2 by regressing $\Lambda_{g,\perp}' Y_t$ on Y_{t-1}. We get the moment conditions

$$E[Y_t(Y_{t+j}'\Lambda_{g,\perp} - Y_{t+j-1}'\delta_0)] = CE[Y_{t-1}(Y_{t+j}'\Lambda_{g,\perp} - Y_{t+j-1}'\delta_0)],$$

for $j = 1, \ldots, M$. They correspond to the restrictions in (3.7), using a negative index j in (3.6). Under similar rank conditions as in Assumption ID.2 matrix C is identifiable.

To identify the loading matrix Λ_h, we have to consider the model given by

$$Y_t - C_0 Y_{t-1} = [\Lambda_{g,0} : \Lambda_h] \begin{bmatrix} g_t \\ h_t \end{bmatrix} + u_t, \tag{E.16}$$

as a static factor model for observable data $Y_t - C_0 Y_{t-1}$. As noticed by Lawley and Maxwell (1971), the identification of the loadings can be obtained if we impose some structure on the variance of u_t, i.e. Σ_u is diagonal. Other conditions need to be specified, but necessary and sufficient restrictions are not provided in closed form, see e.g. Lawley and Maxwell (1971), Anderson and Rubin (1956). Williams (2020) shows that the identifiability of the parameters of a factor model does not require that the variance-covariance matrix Σ_u is diagonal. Therefore, for the identification of Λ_h we state the following assumption.

Assumption RR.2. *The model in equation* (E.16) *is such that the loading matrix* $[\Lambda_g : \Lambda_h]$ *is identifiable.*

Then, given that Λ_g is identifiable, we can recover Λ_h. Finally, the identification of matrices Φ_{gg} and Φ_{gh} are obtained with a system of equations with IV variables as in Section 3.3.

E.4.1 A special case

Under Assumption RR.3 introduced below regarding the factor autoregressive matrix, the dynamic component of the factor vector follows a VAR model without MA component, which simplifies identification. Write $\Phi = \Psi R'$ with Ψ and R some $K \times r$ full rank matrices.

Assumption RR.3. *Matrix* $R'\Psi$ *is non-singular.*

Note that matrix $R'\Psi$ is invariant to transformations of the latent factor. [1] Let $\Psi_\perp$ be a $K \times (K - r)$ matrix spanning the orthogonal complement to the range of Ψ. Under Assumption RR.3, matrix $S = [\Psi_\perp : R]$ is non-singular. Indeed, $S'[\Psi : \Psi_\perp] = \begin{bmatrix} 0 & \Psi_\perp'\Psi_\perp \\ R'\Psi & R'\Psi_\perp \end{bmatrix}$ is invertible, if $R'\Psi$ is non-singular.

Consider the transformation $\tilde{f}_t = S' f_t$. We have:

$$
\begin{aligned}
\tilde{f}_{1t} &= \Psi_\perp' f_t = \Psi_\perp' v_t \equiv \tilde{v}_{1t}, \\
\tilde{f}_{2t} &= R' f_t = (R'\Psi) R' f_{t-1} + R' v_t = \tilde{\Phi} \tilde{f}_{2,t-1} + \tilde{v}_{2t},
\end{aligned}
$$

where $\tilde{\Phi} = R'\Psi$. Then, with $\tilde{B} = B(S')^{-1}$ the model writes

$$
\begin{aligned}
y_t &= C y_{t-1} + \tilde{B}_2 \tilde{f}_{2t} + \tilde{u}_t, \\
\tilde{f}_{2t} &= \tilde{\Phi} \tilde{f}_{2,t-1} + \tilde{v}_{2t},
\end{aligned}
\tag{E.17}
$$

where $\tilde{u}_t = u_t + \tilde{B}_1 \tilde{v}_{1,t}$. This is a model with $\tilde{K} = r$ dynamic factors in vector $\tilde{f}_{2t}$ following a VAR model with autoregressive matrix $\tilde{\Phi}$ of full-rank. Hence, Assumption ID.3 is met here. Notice however

[1] In fact, if we transform $f_t \to A f_t$, with A non-singular, then we have $\Phi \to A\Phi A^{-1}$. Thus, $\Psi \to A\Psi$ and $R \to (A^{-1})' R$, and $R'\Psi$ is invariant to this transformation.

that $\tilde{u}_t$ and $\tilde{v}_{2t}$ are correlated when $\Psi'_{\perp}\Sigma_v R \neq 0$. In fact, $\tilde{u}_t$ contains $K - r$ static factors $\tilde{v}_{1t}$ possibly correlated with the shocks of the dynamic factors.

When the rank of matrix $R'\Psi$ is reduced, the latent factors involve components with non-Markov dynamics as seen above.

Let us briefly discuss the identification in model (E.17) under Assumption RR.3. Suppose that $p = q = 1$ is known to the econometrician. We apply the ideas in the paper accounting for the possible correlation between $\tilde{u}_t$ and $\tilde{v}_{2t}$. (i) We first use the pseudo-model $Y_t = A_1^* Y_{t-1} + A_2^* Y_{t-2} + u_t^*$. Under the assumption $rank(A_2^*) = r$, from the SVD of $A_2^*(A_2^*)'$ we identify $\tilde{K} = r$ and the range of $\tilde{B}_2$. (ii) The orthogonality restrictions $E[\varepsilon_t \eta'_{t-j}] = 0$ are valid for $j \leq -1$, where $\varepsilon_t = Y_t - CY_{t-1}$ and the instrument is $\eta_t = \tilde{B}'_{2,\perp}(Y_t - CY_{t-1})$. So we can identify C by the IV condition. (iii) Finally we can identify $\tilde{\Phi}$ from the VARMA(1,1) dynamics of $\xi_t = \bar{\tilde{B}}'_2(Y_t - CY_{t-1})$ using instruments.

E.5 Asymptotic variance of our estimator with Gaussian innovations

Consider our state space model with $p = q = 1$:

$$\begin{aligned} Y_t &= CY_{t-1} + Bf_t + u_t \\ f_t &= \Phi f_{t-1} + v_t \end{aligned}$$

where $u_t \sim WN(0, \Sigma_u)$ and $v_t \sim WN(0, \Sigma_v)$ are uncorrelated white noises, and matrix B is normalized such that the lower $K \times K$ block is the identity.

Let $b = vec(B_1)$, $c = vec(C)$ and $\phi = vec(\Phi)$. The asymptotic expansion of our estimator is (see Theorem 2):

$$\sqrt{T}\begin{pmatrix} \hat{b} - b_0 \\ \hat{c} - c_0 \\ \hat{\phi} - \phi_0 \end{pmatrix} = \begin{bmatrix} S_{b1} & 0 & 0 & 0 \\ S_{c1} & S_{c2} & S_{c3} & 0 \\ S_{\phi 1} & S_{\phi 2} & S_{\phi 3} & S_{\phi 4} \end{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \tilde{Y}_{t-2} \otimes u_t^* \\ Y_{t-1} \otimes \eta_t \\ \mathbf{Z}_t \otimes \varepsilon_t \\ \mathbf{W}_t \otimes e_t \end{pmatrix} + o_p(1) = S \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_t + o_p(1)$$

where $\tilde{Y}_{t-2} = Y_{t-2} - EL(Y_{t-2}|Y_{t-1})$, $u_t^* = Y_t - EL(Y_t|Y_{t-1}, Y_{t-2})$, $\varepsilon_t = Y_t - CY_{t-1}$, $e_t = \xi_t - \Phi\xi_{t-1}$, $\mathbf{Z}_t = (\eta'_{t-1}, ..., \eta'_{t-M})'$, $\mathbf{W}_t = (\xi'_{t-2}, ..., \xi'_{t-1-L})'$, with $\eta_t = B'_{\perp}(Y_t - CY_{t-1}) = B'_{\perp}u_t$ and $\xi_t = \bar{B}'(Y_t - CY_{t-1}) = f_t + \bar{B}'u_t$, where $\bar{B} = B(B'B)^{-1}$. The blocks of matrix S are defined in (B.7).

Let us now derive results necessary to compute the asymptotic variance.

E.5.1 Computing the blocks of matrix S

i) The ACF

To compute the ACF $\Gamma(j)$ of vector Y_t we use the ABC(D) representation written as

$$\begin{aligned} X_t &= \mathcal{A}X_{t-1} + \mathcal{B}w_t \\ Y_t &= [I_n \ : \ 0]X_t, \end{aligned}$$

where $X_t = (Y_t', f_t')'$ and $w_t = (u_t', v_t')' \sim N(0, \Sigma_w)$ and

$$\mathcal{A} = \begin{pmatrix} C & B\Phi \\ 0 & \Phi \end{pmatrix}, \qquad \mathcal{B} = \begin{pmatrix} I_n & B \\ 0 & I_K \end{pmatrix}, \qquad \Sigma_w = V(w_t) = \begin{pmatrix} \Sigma_u & 0 \\ 0 & \Sigma_v \end{pmatrix}.$$

Then, the ACF $V_X(j) = Cov(X_t, X_{t-j})$ of process X_t is given by

$$V_X(j) = \mathcal{A}^j V_X(0)$$

for any $j \geq 0$, where the variance $V_X(0)$ is such that

$$vec(V_X(0)) = (I_{(n+K)^2} - \mathcal{A} \otimes \mathcal{A})^{-1}(\mathcal{B} \otimes \mathcal{B})vec(\Sigma_w).$$

The autocovariances $\Gamma(j)$ of process Y_t are obtained by the upper-left $n \times n$ block of matrix $V_X(j)$, and using $\Gamma(-j) = \Gamma(j)'$. The autocovariances $\Gamma_f(j)$ of process f_t are obtained by the lower-right $K \times K$ block of matrix $V_X(j)$.

ii) Pseudo-true values

We have $EL(Y_t|Y_{t-1}, Y_{t-2}) = A_1^* Y_{t-1} + A_2^* Y_{t-2}$ with

$$\begin{aligned} \begin{bmatrix} A_1^* & A_2^* \end{bmatrix} &= Cov(Y_t, \mathbf{Y}_{t-1})V(\mathbf{Y}_{t-1})^{-1} \\ &= \begin{bmatrix} \Gamma(1) & \Gamma(2) \end{bmatrix} \begin{bmatrix} \Gamma(0) & \Gamma(1) \\ \Gamma(1)' & \Gamma(0) \end{bmatrix}^{-1} \end{aligned}$$

where $\mathbf{Y}_{t-1} = (Y'_{t-1}, Y'_{t-2})'$. In particular, by the formula of the partitioned inverse:

$$A_2^* = \left(\Gamma(2) - \Gamma(1)\Gamma(0)^{-1}\Gamma(1)\right)\left(\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1)\right)^{-1}. \tag{E.18}$$

Alternatively, we have $A_2^* = B\Phi P_2'$ where P_2 is the coefficient matrix in the linear projection $EL(f_{t-1}|Y_{t-1}, Y_{t-2}) = P_1'Y_{t-1} + P_2'Y_{t-2}$ that is given by

$$\begin{aligned} \left[\begin{array}{cc} P_1' & P_2' \end{array}\right] &= Cov(f_{t-1}, \mathbf{Y}_{t-1})V(\mathbf{Y}_{t-1})^{-1} \\ &= \left[\begin{array}{cc} \Pi' & \Phi\Pi' \end{array}\right]\left[\begin{array}{cc} \Gamma(0) & \Gamma(1) \\ \Gamma(1)' & \Gamma(0) \end{array}\right]^{-1} \end{aligned}$$

where $\Pi = Cov(Y_t, f_t)$ is obtained by the upper-right block of $V_X(0)$. Then:

$$A_2^* = B\Phi\left(\Phi\Pi' - \Pi'\Gamma(0)^{-1}\Gamma(1)\right)\left(\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1)\right)^{-1}. \tag{E.19}$$

iii) Matrix S_{b1}

From (B.7) we have

$$\begin{aligned} S_{b1} &= \left(Q \otimes B_\perp'\right)\left(\sum_{i=0, i\neq 1}^{K} \frac{1}{\lambda_1 - \lambda_i}(U_1 \otimes P_i) \;:\; \cdots \;:\; \sum_{i=0, i\neq K}^{K} \frac{1}{\lambda_K - \lambda_i}(U_K \otimes P_i)\right)' \\ &\quad \times (I_{n^2} + \mathcal{K}_{n,n})\left[\left(A_2^*(\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1}\right) \otimes I_n\right] \end{aligned}$$

where U_1, ... , U_K are the standardized eigenvectors of matrix $R = A_2^*(A_2^*)'$ associated with the non-zero eigenvalues $\lambda_1 > \cdots > \lambda_K > 0$, matrices $P_i = U_i U_i'$ are the corresponding eigenprojectors, and $P_0 = I_n - B(B'B)^{-1}B'$ is the eigenprojector for eigenvalue $\lambda_0 = 0$; the $K \times K$ matrix Q is such that $Q'(B'B)Q = I_K$.

Now, write $Q = [Q_1 \;:\; \cdots \;:\; Q_K]$, where the Q_k are the columns of matrix Q and are equal to $Q_k = \frac{1}{\sqrt{\mu_k}}E_k$, where the E_k and μ_k are standardized eigenvectors and eigenvalues of $B'B$. We have

$U_k = BQ_k$ from (B.1). Then, $B'_\perp P_i = 0$ for $i = 1, ..., K$, while $B'_\perp P_0 = B'_\perp$. Moreover:

$$
\begin{aligned}
& \left(Q \otimes B'_\perp\right) \left(\sum_{i=0, i\neq 1}^{K} \frac{1}{\lambda_1 - \lambda_i} (U_1 \otimes P_i) \ : \ \cdots \ : \ \sum_{i=0, i\neq K}^{K} \frac{1}{\lambda_K - \lambda_i} (U_K \otimes P_i) \right)' \\
= & \left[\begin{array}{ccc} Q_1 \otimes B'_\perp & \cdots & Q_K \otimes B'_\perp \end{array} \right] \left[\begin{array}{c} U'_1 \otimes (\sum_{i=0,i\neq 1}^{K} \frac{1}{\lambda_1 - \lambda_i} P_i) \\ \vdots \\ U'_K \otimes (\sum_{i=0,i\neq K}^{K} \frac{1}{\lambda_K - \lambda_i} P_i) \end{array} \right] \\
= & \sum_{k=1}^{K} \frac{1}{\lambda_k} (Q_k U'_k) \otimes B'_\perp = [\sum_{k=1}^{K} \frac{1}{\lambda_k} (Q_k Q'_k) B'] \otimes B'_\perp = [(\sum_{k=1}^{K} \frac{1}{\lambda_k \mu_k} E_k E'_k) B'] \otimes B'_\perp .
\end{aligned}
$$

Thus we get

$$
S_{b1} = \left([(\sum_{k=1}^{K} \frac{1}{\lambda_k \mu_k} E_k E'_k) B'] \otimes B'_\perp \right) (I_{n^2} + \mathcal{K}_{n,n}) \left[\left(A_2^* (\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1} \right) \otimes I_n \right] .
$$

Now, we use that by the properties of the commutation matrix we have

$$
\mathcal{K}_{n,n} \left[\left(A_2^* (\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1} \right) \otimes I_n \right] = \left[I_n \otimes \left(A_2^* (\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1} \right) \right] \mathcal{K}_{n,n},
$$

and $A_2^* = BD'$ for $D' = \bar{B}' A_2^* = \Phi \left(\Phi \Pi' - \Pi' \Gamma(0)^{-1} \Gamma(1) \right) \left(\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1) \right)^{-1}$ from (E.18) and (E.19). Thus, because $B'_\perp A_2^* = 0$ we get:

$$
\begin{aligned}
S_{b1} &= \left([(\sum_{k=1}^{K} \frac{1}{\lambda_k \mu_k} E_k E'_k) B'] \otimes B'_\perp \right) \left[\left(A_2^* (\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1} \right) \otimes I_n \right] \\
&= [(\sum_{k=1}^{K} \frac{1}{\lambda_k \mu_k} E_k E'_k) B' A_2^* (\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1}] \otimes B'_\perp \\
&= [(\sum_{k=1}^{K} \frac{1}{\lambda_k} E_k E'_k) D' (\Gamma(0) - \Gamma(1)'\Gamma(0)^{-1}\Gamma(1))^{-1}] \otimes B'_\perp .
\end{aligned}
$$

iv) Matrix $[S_{c1} \ : \ S_{c2} \ : \ S_{c3}]$

From (B.7) with $p = 1$ we know that:

$$[S_{c1} : S_{c2} : S_{c3}] = -(J_c'\Omega_c J_c)^{-1} J_c'\Omega_c \left[\begin{array}{ccc} \left(J_b - J_d(\Gamma(0)^{-1}\Gamma_{[K]}(-1) \otimes I_{n-K})\right) S_{b1} & J_d(\Gamma(0)^{-1} \otimes I_{n-K}) & I_{n(n-K)M} \end{array} \right]$$

where $\Gamma_{[K]}(-1)$ denotes the K right columns of matrix $\Gamma(-1) = \Gamma(1)'$. Let us now compute matrices J_c, J_b and J_d.

- We have $J_c = -Q'_{\mathbf{Y}\_\mathbf{Z}} \otimes I_n$ and

$$Q_{\mathbf{Y}\_\mathbf{Z}} = \left(\begin{array}{cccc} \Gamma(0) - \Gamma(1)C' & \Gamma(1) - \Gamma(2)C' & \cdots & \Gamma(M-1) - \Gamma(M)C' \end{array} \right) (I_M \otimes B_\perp).$$

  Equivalently, by using $E(Y_{t-1}\eta'_{t-j}) = E(Y_{t-1}u'_{t-j})B_\perp = C^{j-1}\Sigma_u B_\perp$, for $j \geq 1$, which yields:

$$Q_{\mathbf{Y}\_\mathbf{Z}} = \left(\begin{array}{cccc} I_n & C & \cdots & C^{M-1} \end{array} \right) [I_M \otimes (\Sigma_u B_\perp)].$$

- We have

$$J_d = -(I_M \otimes \mathcal{K}_{n-K,n})[\left(\begin{array}{c} \Gamma(2) - C\Gamma(1) \\ \vdots \\ \Gamma(M+1) - C\Gamma(M) \end{array} \right) \otimes I_{n-K}].$$

- Moreover:

$$J_b = -(I_M \otimes \mathcal{K}_{n-K,n})[\left(\begin{array}{c} \Gamma(1) - C\Gamma(0) \\ \vdots \\ \Gamma(M) - C\Gamma(M-1) \end{array} \right)_{[n-K+1:n]} \otimes I_{n-K}],$$

  where $(\cdot)_{[n-K+1:n]}$ denotes the K right columns of a matrix.

v) Matrix $[S_{\phi 1} : S_{\phi 2} : S_{\phi 3} : S_{\phi 4}]$

From (B.7) we have

$$[S_{\phi 1} : S_{\phi 2} : S_{\phi 3} : S_{\phi 4}] = -(\mathcal{J}_\phi'\Omega_\phi \mathcal{J}_\phi)^{-1}\mathcal{J}_\phi'\Omega_\phi \left[\begin{array}{cccc} \mathcal{J}_b S_{b1} + \mathcal{J}_c S_{c1} & \mathcal{J}_c S_{c2} & \mathcal{J}_c S_{c3} & I_{K^2 L} \end{array} \right].$$

We now compute matrices $\mathcal{J}_\phi$, $\mathcal{J}_b$ and $\mathcal{J}_c$.

- We have $\mathcal{J}_\phi = -Q_{\mathbf{W}\xi_-} \otimes I_K$ with

$$Q_{\mathbf{W}\xi_-} = \begin{pmatrix} \Gamma_f(0)\Phi' \\ \Gamma_f(0)(\Phi')^2 \\ \vdots \\ \Gamma_f(0)(\Phi')^L \end{pmatrix}. \tag{E.20}$$

- With $q = 1$ we have

$$\begin{aligned} \mathcal{J}_c &= -\begin{pmatrix} E[e_t Y_{t-3}'] \\ \vdots \\ E[e_t Y_{t-L-2}] \end{pmatrix} \otimes \bar{B}' - E(\mathbf{W}_t Y_{t-1}') \otimes \bar{B}' + E(\mathbf{W}_t Y_{t-2}') \otimes (\Phi \bar{B}') \\ &= -\begin{pmatrix} \bar{B}'[\Gamma(-1) - C\Gamma(-2)] \\ \vdots \\ \bar{B}'[\Gamma(-L) - C\Gamma(-L-1)] \end{pmatrix} \otimes \bar{B}' + \begin{pmatrix} \bar{B}'[\Gamma(0) - C\Gamma(-1)] \\ \vdots \\ \bar{B}'[\Gamma(1-L) - C\Gamma(-L)] \end{pmatrix} \otimes (\Phi \bar{B}'). \end{aligned}$$

- Finally:

$$\begin{aligned} \mathcal{J}_b &= \left\{ \begin{pmatrix} E(e_t \varepsilon_{t-2}') \\ \vdots \\ E(e_t \varepsilon_{t-1-L}) \end{pmatrix} \otimes I_K + E(\mathbf{W}_t \varepsilon_t') \otimes I_K - E(\mathbf{W}_t \varepsilon_{t-1}') \otimes \Phi \right\} D_0 \\ &= \left\{ \begin{pmatrix} \Gamma_f(0)(\Phi')^2 B' \\ \vdots \\ \Gamma_f(0)(\Phi')^{L+1} B' \end{pmatrix} \otimes I_K - \begin{pmatrix} \Gamma_f(0)\Phi' B' \\ \vdots \\ \Gamma_f(0)(\Phi')^L B' \end{pmatrix} \otimes \Phi \right\} D_0 \\ &= \left\{ (Q_{\mathbf{W}\xi_-} \Phi' B') \otimes I_K - (Q_{\mathbf{W}\xi_-} B') \otimes \Phi \right\} D_0 \\ &= Q_{\mathbf{W}\xi_-} \otimes (\Phi (B'B)^{-1} B_1') - (Q_{\mathbf{W}\xi_-} \Phi') \otimes ((B'B)^{-1} B_1'), \end{aligned}$$

for $Q_{\mathbf{W}\xi_-}$ given in (E.20), where $D_0 = (P_{0,[n-K]} \otimes (B'B)^{-1})\mathcal{K}_{n-K,K} - \bar{B} \otimes ((B'B)^{-1} B_1')$, $P_{0,[n-K]}$ denotes the first $n-K$ columns of $P_0 = I_n - B(B'B)^{-1}B'$ and B_1 is the upper block of B.

E.5.2 Computing the asymptotic variance of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t$

The asymptotic variance is $\sum_{j=-\infty}^{\infty} Cov(\psi_t, \psi_{t-j})$. We can compute in closed form the autocovariances $Cov(\psi_t, \psi_{t-j})$ when the WN processes u_t and v_t are Gaussian. Then, any block in ψ_t is the Kronecker product of two Gaussian zero-mean vectors. Thus, what we need are covariances of the type

$$Cov(x_1 \otimes x_2, x_3 \otimes x_4)$$

where $x_1, ..., x_4$ are zero-mean jointly Gaussian vectors. We use the next Lemma.

Lemma 7. *Let x_1, x_2, x_3, x_4 be four zero-mean jointly Gaussian vectors of dimensions N, m, r, s, respectively, such that $X = (x_1', x_3')'$ and $Y = (x_2', x_4')'$ have variances $V(X) = \Omega_{XX}$ and $V(Y) = \Omega_{YY}$ and covariance $Cov(X, Y) = \Omega_{XY}$. Then:*

$$Cov(x_1 \otimes x_2, x_3 \otimes x_4) = A_1 \left[\Omega_{XX} \otimes \Omega_{YY} + \mathcal{K}_{N+r,m+s}(\Omega_{XY}' \otimes \Omega_{XY})\right] A_2', \tag{E.21}$$

where $A_1 = \left[\begin{array}{cc} I_N \otimes [I_m \ : \ O_{m\times s}] & O_{Nm\times r(m+s)} \end{array}\right]$ *and* $A_2 = \left[\begin{array}{cc} O_{rs\times N(m+s)} & I_r \otimes [O_{s\times m} \ : \ I_s] \end{array}\right]$.

Proof. The result follows from writing $x_1 \otimes x_2 = A_1(X \otimes Y)$ and $x_3 \otimes x_4 = A_2(X \otimes Y)$, and $V(X \otimes Y) = \Omega_{XX} \otimes \Omega_{YY} + \mathcal{K}_{N+r,m+s}(\Omega_{XY}' \otimes \Omega_{XY})$ (see e.g. Magnus and Neudecker's book). □

The derivation of the asymptotic variance $\sum_{j=-\infty}^{\infty} Cov(\psi_t, \psi_{t-j})$ proceeds as follow. In Step 1, we use Lemma 1 to rewrite this asymptotic variance in terms of variances and covariances of Gaussian vectors. Indeed, let us define the zero-mean Gaussian vectors:

$$x_1 = \begin{pmatrix} \tilde{Y}_{t-2} \\ Y_{t-1} \\ \mathbf{Z}_t \\ \mathbf{W}_t \end{pmatrix}, \quad x_2 = \begin{pmatrix} u_t^* \\ \eta_t \\ \varepsilon_t \\ e_t \end{pmatrix}, \quad x_3 = \begin{pmatrix} \tilde{Y}_{t-2-j} \\ Y_{t-1-j} \\ \mathbf{Z}_{t-j} \\ \mathbf{W}_{t-j} \end{pmatrix}, \quad x_4 = \begin{pmatrix} u_{t-j}^* \\ \eta_{t-j} \\ \varepsilon_{t-j} \\ e_{t-j} \end{pmatrix}.$$

The dimensions N, m, r and s of these four vectors are $N = r = 2n+(n-K)M+KL$ and $m = s = 3n$. In order to write vector ψ_t as a linear transformation of the Kronecker product $x_1 \otimes x_2$, let us partition the identity matrices I_N and I_m as:

$$I_N = \begin{pmatrix} \underset{n\times N}{\mathcal{S}_{11}} \\ \underset{n\times N}{\mathcal{S}_{21}} \\ \underset{(n-K)M\times N}{\mathcal{S}_{31}} \\ \underset{KL\times N}{\mathcal{S}_{41}} \end{pmatrix}, \qquad I_m = \begin{pmatrix} \underset{n\times m}{\mathcal{S}_{12}} \\ \underset{(n-K)\times m}{\mathcal{S}_{22}} \\ \underset{n\times m}{\mathcal{S}_{32}} \\ \underset{k\times m}{\mathcal{S}_{42}} \end{pmatrix},$$

reflecting the blocks in x_1 and x_2. Then, we have for instance $\tilde{Y}_{t-2} \otimes u_t^* = (\mathcal{S}_{11}x_1) \otimes (\mathcal{S}_{12}x_2) = (\mathcal{S}_{11} \otimes \mathcal{S}_{12})(x_1 \otimes x_2)$, and similarly for the other Kronecker products in vector ψ_t. Thus, $\psi_t = \mathcal{S}(x_1 \otimes x_2)$, where

$$\mathcal{S} = \begin{pmatrix} \mathcal{S}_{11} \otimes \mathcal{S}_{12} \\ \mathcal{S}_{21} \otimes \mathcal{S}_{22} \\ \mathcal{S}_{31} \otimes \mathcal{S}_{32} \\ \mathcal{S}_{41} \otimes \mathcal{S}_{42} \end{pmatrix}, \tag{E.22}$$

and $\psi_{t-j} = \mathcal{S}(x_3 \otimes x_4)$. Hence

$$Cov(\psi_t, \psi_{t-j}) = \mathcal{S} Cov(x_1 \otimes x_2, x_3 \otimes x_4)\mathcal{S}'.$$

Now, by Lemma 1 we have $Cov(x_1 \otimes x_2, x_3 \otimes x_4) = A_1[\Omega_{XX} \otimes \Omega_{YY} + \mathcal{K}_{2N,2m}(\Omega_{XY}' \otimes \Omega_{XY})]A_2'$, where Ω_{XX}, Ω_{YY} and Ω_{XY} are the variance matrices, and the covariance, of $X = (x_1', x_3')'$, $Y = (x_2', x_4')'$, and

$$A_1 = \begin{bmatrix} I_N \otimes [I_m \ : \ O_{m\times m}] & O_{Nm\times 2Nm} \end{bmatrix}, \tag{E.23}$$
$$A_2 = \begin{bmatrix} O_{Nm\times 2Nm} & I_N \otimes [O_{m\times m} \ : \ I_m] \end{bmatrix}. \tag{E.24}$$

In Step 2, we write vectors x_1, x_2, x_3 and x_4 as linear transformations of current and lagged realizations of process Y_t. Indeed, by using $\tilde{Y}_{t-2} = Y_{t-2} - \Gamma(1)'\Gamma(0)^{-1}Y_{t-1}$ and

$$\mathbf{Z}_t = \begin{pmatrix} B_\perp'(Y_{t-1} - CY_{t-2}) \\ \vdots \\ B_\perp'(Y_{t-M} - CY_{t-M-1}) \end{pmatrix}, \quad \mathbf{W}_t = \begin{pmatrix} \bar{B}'(Y_{t-2} - CY_{t-3}) \\ \vdots \\ \bar{B}'(Y_{t-L-1} - CY_{t-L-2}) \end{pmatrix},$$

and denoting $L_{max} = \max\{M+1, L+2\}$ we have

$$x_1 = \begin{pmatrix} -\Gamma(1)'\Gamma(0)^{-1} & I_n & O_{n\times n} & O_{n\times n} & \cdots & O_{n\times n} \\ I_n & O_{n\times n} & O_{n\times n} & O_{n\times n} & \cdots & O_{n\times n} \\ B'_\perp & -B'_\perp C & O_{n\times n} & O_{n\times n} & \cdots & O_{n\times n} \\ O_{n\times n} & B'_\perp & -B'_\perp C & O_{n\times n} & \cdots & O_{n\times n} \\ & \ddots & \ddots & & & \\ O_{n\times n} & \bar{B}' & -\bar{B}'C & O_{n\times n} & \cdots & O_{n\times n} \\ O_{n\times n} & O_{n\times n} & \bar{B}' & -\bar{B}'C & \cdots & O_{n\times n} \\ & & \ddots & \ddots & & \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ Y_{t-3} \\ Y_{t-4} \\ \vdots \\ Y_{t-L_{max}} \end{pmatrix} \equiv M_1 \mathbf{Y}_{t-1:t-L_{max}} \tag{E.25}$$

where in vector $\mathbf{Y}_{t-1:t-L_{max}}$ we stack the lags $t-1$ up to $t-L_{\max}$ of vector Y, and matrix M_1 is $L \times L_{max}n$. Similarly, using $u_t^* = Y_t - A_1^* Y_{t-1} - A_2^* Y_{t-2}$, $\eta_t = B'_\perp(Y_t - CY_{t-1})$, $\varepsilon_t = Y_t - CY_{t-1}$ and $e_t = \bar{B}'(Y_t - CY_{t-1}) - \Phi\bar{B}'(Y_{t-1} - CY_{t-2}) = \bar{B}'Y_t - (\bar{B}'C + \Phi\bar{B}')Y_{t-1} + \Phi\bar{B}'CY_{t-2}$, we have

$$x_2 = \begin{pmatrix} I_n & -A_1^* & -A_2^* \\ B'_\perp & -B'_\perp C & O_{n\times n} \\ I_n & -C & O_{n\times n} \\ \bar{B}' & -(\bar{B}'C + \Phi\bar{B}') & \Phi\bar{B}'C \end{pmatrix} \begin{pmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \end{pmatrix} \equiv M_2 \mathbf{Y}_{t:t-2}, \tag{E.26}$$

where matrix M_2 is $m \times 3n$. Thus, $x_1 = M_1\mathbf{Y}_{t-1:t-L_{max}}$, $x_2 = M_2\mathbf{Y}_{t:t-2}$, $x_3 = M_1\mathbf{Y}_{t-1-j:t-L_{max}-j}$, and $x_4 = M_2\mathbf{Y}_{t-j:t-2-j}$.

In Step 3, we use the writing of $x_1, ..., x_4$ as linear transformations of Y_t and its lags to derive the matrices Ω_{XX}, Ω_{YY} and Ω_{XY}. We have for $j \geq 0$:

$$X = \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \begin{pmatrix} M_1 & : & O_{L\times jn} \\ O_{L\times jn} & : & M_1 \end{pmatrix} \mathbf{Y}_{t-1:t-L_{max}-j} \equiv M_X(j)\mathbf{Y}_{t-1:t-L_{max}-j},$$

and

$$Y = \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} M_2 & : & O_{m\times jn} \\ O_{m\times jn} & : & M_2 \end{pmatrix} \mathbf{Y}_{t:t-2-j} \equiv M_Y(j)\mathbf{Y}_{t:t-2-j}.$$

Note that in matrix $M_X(j)$ the blocks in the upper and lower parts are not vertically aligned, and similarly for $M_Y(j)$. We have

$$V(\mathbf{Y}_{t-1:t-L_{max}-j}) = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \cdots & \cdots & \Gamma(L_{max}+j-1) \\ \Gamma(-1) & \Gamma(0) & \Gamma(1) & \cdots & \Gamma(L_{max}+j-2) \\ & \ddots & \ddots & & \\ \Gamma(-L_{max}-j+1) & \Gamma(-L_{max}-j+2) & \cdots & \cdots & \Gamma(0) \end{pmatrix} \equiv \mathbb{T}_1(j),$$

$$V(\mathbf{Y}_{t:t-2-j}) = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \cdots & \cdots & \Gamma(j+2) \\ \Gamma(-1) & \Gamma(0) & \Gamma(1) & \cdots & \Gamma(j+1) \\ & \ddots & \ddots & & \\ \Gamma(-j-2) & \Gamma(-j-1) & \cdots & \cdots & \Gamma(0) \end{pmatrix} \equiv \mathbb{T}_2(j),$$

and

$$\begin{aligned} Cov(\mathbf{Y}_{t-1:t-L_{max}-j}, \mathbf{Y}_{t:t-2-j}) &= \begin{pmatrix} \Gamma(-1) & \Gamma(0) & \cdots & \cdots & \Gamma(j+1) \\ \Gamma(-2) & \Gamma(-1) & \Gamma(0) & \cdots & \Gamma(j) \\ & \ddots & \ddots & & \\ \Gamma(-L_{max}-j) & \Gamma(-L_{max}-j+1) & \cdots & \cdots & \Gamma(-L_{max}+2) \end{pmatrix} \\ &\equiv \mathbb{T}_{12}(j). \end{aligned}$$

Thus, we get $\Omega_{XX} = M_X(j)\mathbb{T}_1(j)M_X(j)'$, $\Omega_{YY} = M_Y(j)\mathbb{T}_2(j)M_Y(j)'$ and $\Omega_{XY} = M_X(j)\mathbb{T}_{12}(j)M_Y(j)'$.

To conclude, we summarize our result. The asymptotic variance is

$$\sum_{j=-\infty}^{\infty} Cov(\psi_t, \psi_{t-j}) = \Gamma_\psi(0) + \sum_{j=1}^{\infty}[\Gamma_\psi(j) + \Gamma_\psi(j)']$$

where:

$$\Gamma_\psi(j) = \mathcal{S}A_1\left[\Omega_{XX}(j)\otimes\Omega_{YY}(j) + \mathcal{K}_{2N,2m}(\Omega_{XY}(j)'\otimes\Omega_{XY}(j))\right]A_2'\mathcal{S}'$$

for $j \geq 0$, the matrices A_1 and A_2 are given in (E.23), the matrix $\mathcal{S}$ is defined in (E.22), and

$$\begin{aligned} \Omega_{XX}(j) &= M_X(j)\mathbb{T}_1(j)M_X(j)', \\ \Omega_{YY}(j) &= M_Y(j)\mathbb{T}_2(j)M_Y(j)', \\ \Omega_{XY}(j) &= M_X(j)\mathbb{T}_{12}(j)M_Y(j)', \end{aligned}$$

and

$$M_X(j) = \begin{pmatrix} M_1 & : & O_{L\times jn} \\ O_{L\times jn} & : & M_1 \end{pmatrix}, \quad M_Y(j) = \begin{pmatrix} M_2 & : & O_{m\times jn} \\ O_{m\times jn} & : & M_2 \end{pmatrix}$$

and the matrices M_1 and M_2 are given in (E.25) and (E.26).

E.6 Asymptotic variance of MLE in the state space model

Consider the state space model

$$\begin{aligned} Y_t &= CY_{t-1} + Bf_t + u_t \\ f_t &= \Phi f_{t-1} + v_t \end{aligned}$$

where $u_t \sim WN(0, \Sigma_u)$ and $v_t \sim WN(0, \Sigma_v)$ are uncorrelated white noises, and matrix B is such that the lower $K \times K$ block is the identity. We collect in vector θ the set of unknown parameters.

Let us start with the MLE. The errors u_t and v_t are assumed Gaussian. Under regularity conditions (including global and local identification), the MLE $\hat{\theta}$ is asymptotically normal as $T \to \infty$, i.e.

$$\sqrt{T}(\hat{\theta} - \theta_0) \;\Rightarrow\; N(0, I_0^{-1}),$$

where the asymptotic variance is given by the inverse information matrix, that is defined as

$$\begin{aligned} I_0 &= \lim_{T\to\infty} E[-\frac{1}{T}\frac{\partial^2 \mathcal{L}_T(\theta_0)}{\partial\theta\partial\theta'}] \\ &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left\{E[\frac{\partial m_t(\theta_0)'}{\partial\theta}\Sigma_t(\theta_0)^{-1}\frac{\partial m_t(\theta_0)}{\partial\theta'}] + \frac{1}{2}E[\frac{\partial vec\Sigma_t(\theta_0)'}{\partial\theta}(\Sigma_t(\theta_0)^{-1}\otimes\Sigma_t(\theta_0)^{-1})\frac{\partial vec\Sigma_t(\theta_0)}{\partial\theta'}]\right\} \end{aligned}$$

where $m_t(\theta_0) = EL[Y_t|\mathcal{Y}^{t-1}]$ with $\mathcal{Y}^{t-1} = (Y_{t-1}, ..., Y_1)$ and $\Sigma_t(\theta_0) = V[Y_t - m_t(\theta_0)]$. Note that under the Gaussian assumption, EL coincides with conditional expectation, and $\Sigma_t(\theta_0)$ coincides with the conditional variance of Y_t given $\mathcal{Y}^{t-1}$ (which is independent of $\mathcal{Y}^{t-1}$).

E.6.1 Derivation of the information matrix formula

See Lutkepohl (2007), equation (18.4.7).

E.6.2 Numerical computation of the information matrix

The numerical computation of the information matrix I_0 uses the Kalman filter. We have the following equations from the standard Kalman filter:

$$
\begin{aligned}
f_{t|t-1} &= \Phi f_{t-1|t-1} && \text{(E.27)} \\
Y_{t|t-1} &= B f_{t|t-1} + C Y_{t-1} && \text{(E.28)} \\
V_{t|t-1} &= \Phi V_{t-1|t-1} \Phi' + \Sigma_v && \text{(E.29)} \\
\Sigma_{t|t-1} &= B V_{t|t-1} B' + \Sigma_u && \text{(E.30)} \\
f_{t|t} &= f_{t|t-1} + V_{t|t-1} B' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) && \text{(E.31)} \\
V_{t|t} &= V_{t|t-1} - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B V_{t|t-1} && \text{(E.32)}
\end{aligned}
$$

where $f_{t|t-1}$ and $f_{t|t}$ are the predicted and filtered values of the factor, $V_{t|t-1}$ and $V_{t|t}$ are the corresponding variances of the prediction and filtering errors, respectively, and similarly $Y_{t|t-1}$, $Y_{t|t}$, $\Sigma_{t|t-1}$ and $\Sigma_{t|t}$ for the endogenous variables. All these quantities depend on θ, but we omit this dependence in the notation for expository purpose. Note that $Y_{t|t-1}$ corresponds to $m_t(\theta)$, and $\Sigma_{t|t-1}$ to $\Sigma_t(\theta)$ introduced above.

Thus, the equations

$$
\begin{aligned}
Y_{t|t-1} &= B \Phi f_{t-1|t-1} + C Y_{t-1} && \text{(E.33)} \\
\Sigma_{t|t-1} &= B(\Phi V_{t-1|t-1} \Phi' + \Sigma_v) B' + \Sigma_u && \text{(E.34)}
\end{aligned}
$$

link the prediction vector, and prediction error variance, of Y_t to the lagged filtered value of the factor, and the associated variance of the filtering error. The latter satisfy the recursions

$$
\begin{aligned}
f_{t|t} &= \Phi f_{t-1|t-1} + V_{t|t-1} B' (B V_{t|t-1} B' + \Sigma_u)^{-1} (Y_t - B \Phi f_{t-1|t-1} - C Y_{t-1}) \\
V_{t|t} &= V_{t|t-1} - V_{t|t-1} B' (B V_{t|t-1} B' + \Sigma_u)^{-1} B V_{t|t-1}
\end{aligned}
$$

where

$$
V_{t|t-1} = \Phi V_{t-1|t-1} \Phi' + \Sigma_v .
$$

By differentiating these equations w.r.t. θ we get recursive relations for the gradients, that we now derive.

One-factor model

We consider the case of a single factor model ($K = 1$), in which the derivations are a bit simpler. Here, the normalization of the loading vector is $B = (b', 1)'$. Then, the vector of the parameters is

$$\theta = (c', b', \phi, S_u', \sigma_v^2)',$$

where $c = vec(C)$ and $S_u = vech(\Sigma_u)$.

The recursive equations are:

$$Y_{t|t-1} = B\phi f_{t-1|t-1} + CY_{t-1} \tag{E.35}$$

$$\Sigma_{t|t-1} = [\phi^2 V_{t-1|t-1} + \sigma_v^2]BB' + \Sigma_u \tag{E.36}$$

$$\begin{aligned} f_{t|t} &= \phi f_{t-1|t-1} + V_{t|t-1}B'(V_{t|t-1}BB' + \Sigma_u)^{-1}(Y_t - B\phi f_{t-1|t-1} - CY_{t-1}) \\ V_{t|t} &= V_{t|t-1} - V_{t|t-1}^2 B'(V_{t|t-1}BB' + \Sigma_u)^{-1}B \end{aligned}$$

$$V_{t|t-1} = \phi^2 V_{t-1|t-1} + \sigma_v^2.$$

By writing $Y_{t|t-1} = B\phi f_{t-1|t-1} + CY_{t-1} = B\phi f_{t-1|t-1} + (Y_{t-1}' \otimes I_n)c$, we have:

$$\begin{aligned} \frac{\partial Y_{t|t-1}}{\partial c'} &= B\phi \frac{\partial f_{t-1|t-1}}{\partial c'} + Y_{t-1}' \otimes I_n \\ \frac{\partial Y_{t|t-1}}{\partial B'} &= B\phi \frac{\partial f_{t-1|t-1}}{\partial B'} + I_n \phi f_{t-1|t-1} \\ \frac{\partial Y_{t|t-1}}{\partial \phi} &= B\phi \frac{\partial f_{t-1|t-1}}{\partial \phi} + B f_{t-1|t-1} \\ \frac{\partial Y_{t|t-1}}{\partial S_u'} &= B\phi \frac{\partial f_{t-1|t-1}}{\partial S_u'} \\ \frac{\partial Y_{t|t-1}}{\partial \sigma_v^2} &= B\phi \frac{\partial f_{t-1|t-1}}{\partial \sigma_v^2}, \end{aligned}$$

where the gradient w.r.t. b is obtained from the gradient w.r.t. B by deleting the last row. Thus we get:

$$\frac{\partial m_t(\theta_0)}{\partial \theta'} = B_0 \phi_0 \frac{\partial f_{t-1|t-1}}{\partial \theta'} + \left[\begin{array}{ccccc} Y_{t-1}' \otimes I_n & I_{1:n-1}\phi_0 f_{t-1|t-1} & B_0 f_{t-1|t-1} & 0_{n \times \frac{n(n+1)}{2}} & 0_{n \times 1} \end{array} \right], \tag{E.37}$$

where $I_{1:n-1}$ is the $n \times (n-1)$ matrix that is obtained from the identity matrix I_n by deleting the last column.

By using $vec(\Sigma_{t|t-1}) = (\phi^2 V_{t-1|t-1} + \sigma_v^2) B \otimes B + M S_u$, where M is a $n^2 \times \frac{n(n+1)}{2}$ matrix such that $vec(\Sigma_u) = M vech(\Sigma_u)$, we get

$$\begin{aligned}
\frac{\partial vec(\Sigma_{t|t-1})}{\partial c'} &= (B \otimes B)\phi^2 \frac{\partial V_{t-1|t-1}}{\partial c'} \\
\frac{\partial vec(\Sigma_{t|t-1})}{\partial B'} &= (B \otimes B)\phi^2 \frac{\partial V_{t-1|t-1}}{\partial B'} + (\phi^2 V_{t-1|t-1} + \sigma_v^2)\frac{\partial B \otimes B}{\partial B'} \\
\frac{\partial vec(\Sigma_{t|t-1})}{\partial \phi} &= (B \otimes B)\phi^2 \frac{\partial V_{t-1|t-1}}{\partial \phi} + 2\phi V_{t-1|t-1} B \otimes B \\
\frac{\partial vec(\Sigma_{t|t-1})}{\partial S_u'} &= (B \otimes B)\phi^2 \frac{\partial V_{t-1|t-1}}{\partial S_u'} + M \\
\frac{\partial vec(\Sigma_{t|t-1})}{\partial \sigma_v^2} &= (B \otimes B)\phi^2 \frac{\partial V_{t-1|t-1}}{\partial \sigma_v^2} + B \otimes B.
\end{aligned}$$

Note that $\frac{\partial B \otimes B}{\partial B_k} = e_k \otimes B + B \otimes e_k$, where e_k is the kth column of the identity matrix I_n. Thus we get:

$$\begin{aligned}
\frac{\partial vec \Sigma_t(\theta_0)}{\partial \theta'} &= (B_0 \otimes B_0)\phi_0^2 \tfrac{\partial V_{t-1|t-1}}{\partial \theta'} \\
&+ \left[\begin{array}{ccccc} 0_{n^2 \times n} & (\phi_0^2 V_{t-1|t-1} + \sigma_{v,0}^2)\frac{\partial B_0 \otimes B_0}{\partial b_0'} & 2\phi_0 V_{t-1|t-1} B_0 \otimes B_0 & M & B_0 \otimes B_0 \end{array}\right]
\end{aligned}$$

where $\frac{\partial B_0 \otimes B_0}{\partial b_0'} = I_{1:n-1} \otimes B_0 + B_0 \otimes I_{1:n-1}$.

Let us now obtain the recursive equations for the gradient of $f_{t|t}$.

(i) Derivative of $f_{t|t}$ w.r.t. c. We have

$$\begin{aligned}
\frac{\partial f_{t|t}}{\partial c'} &= \phi \frac{\partial f_{t-1|t-1}}{\partial c'} + B' \Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial c'} \\
&\quad - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B B' \Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial c'} \\
&\quad - V_{t|t-1} B' \Sigma_{t|t-1}^{-1}[B\phi \frac{\partial f_{t-1|t-1}}{\partial c'} + (Y_{t-1}' \otimes I_n)].
\end{aligned}$$

Using $\frac{\partial V_{t|t-1}}{\partial c} = \phi^2 \frac{\partial V_{t-1|t-1}}{\partial c}$ and rearranging terms, we get:

$$\begin{aligned}
\frac{\partial f_{t|t}}{\partial c} &= \phi \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) \frac{\partial f_{t-1|t-1}}{\partial c} \\
&\quad + \phi^2 \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) B' \Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t-1|t-1}}{\partial c} - V_{t|t-1}[Y_{t-1} \otimes (\Sigma_{t|t-1}^{-1} B)].
\end{aligned}$$

(ii) Derivative of $f_{t|t}$ w.r.t. b. We have

$$\begin{aligned}
\frac{\partial f_{t|t}}{\partial B'} &= \phi\frac{\partial f_{t-1|t-1}}{\partial B'} + B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial B'} + V_{t|t-1}(Y_t - Y_{t|t-1})'\Sigma_{t|t-1}^{-1} \\
&\quad -V_{t|t-1}B'\Sigma_{t|t-1}^{-1}BB'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial B'} \\
&\quad -V_{t|t-1}^2\left(B'\Sigma_{t|t-1}^{-1}\frac{\partial BB'}{\partial B_k}\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\right)_{k=1,\ldots,n} \\
&\quad -V_{t|t-1}B'\Sigma_{t|t-1}^{-1}[B\phi\frac{\partial f_{t-1|t-1}}{\partial B'} + \phi f_{t-1|t-1}I_n].
\end{aligned}$$

We use

$$\begin{aligned}
B'\Sigma_{t|t-1}^{-1}\left(\frac{\partial BB'}{\partial B_k}\right)\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1}) &= \left\{[(Y_t - Y_{t|t-1})'\Sigma_{t|t-1}^{-1}] \otimes [B'\Sigma_{t|t-1}^{-1}]\right\} vec\left(\frac{\partial BB'}{\partial B_k}\right) \\
&= \left\{[(Y_t - Y_{t|t-1})'\Sigma_{t|t-1}^{-1}] \otimes [B'\Sigma_{t|t-1}^{-1}]\right\}(e_k \otimes B + B \otimes e_k) \\
&= (Y_t - Y_{t|t-1})'\Sigma_{t|t-1}^{-1}e_k(B'\Sigma_{t|t-1}^{-1}B) \\
&\quad +(Y_t - Y_{t|t-1})'\Sigma_{t|t-1}^{-1}BB'\Sigma_{t|t-1}^{-1}e_k,
\end{aligned}$$

where e_k is the kth column of I_n, and $\frac{\partial V_{t|t-1}}{\partial B} = \phi^2\frac{\partial V_{t-1|t-1}}{\partial B}$ to get:

$$\begin{aligned}
\frac{\partial f_{t|t}}{\partial b} &= \phi\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)\frac{\partial f_{t-1|t-1}}{\partial b} \\
&\quad +\phi^2\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t-1|t-1}}{\partial b} \\
&\quad +V_{t|t-1}[\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1} - B\phi f_{t-1|t-1})]_{1:n-1} \\
&\quad -V_{t|t-1}^2(B'\Sigma_{t|t-1}^{-1}B)[\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})]_{1:n-1} \\
&\quad -V_{t|t-1}^2B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})[\Sigma_{t|t-1}^{-1}B]_{1:n-1} \\
&= \phi\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)\frac{\partial f_{t-1|t-1}}{\partial b} \\
&\quad +\phi^2\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t-1|t-1}}{\partial b} \\
&\quad +V_{t|t-1}\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)[\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})]_{1:n-1} - V_{t|t-1}f_{t|t}[\Sigma_{t|t-1}^{-1}B]_{1:n-1}.
\end{aligned}$$

(iii) Derivative of $f_{t|t}$ w.r.t. ϕ. We have

$$
\begin{aligned}
\frac{\partial f_{t|t}}{\partial \phi} &= \phi \frac{\partial f_{t-1|t-1}}{\partial \phi} + f_{t-1|t-1} + B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial \phi} \\
&\quad - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}BB'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial \phi} \\
&\quad - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}[B\phi\frac{\partial f_{t-1|t-1}}{\partial \phi} + Bf_{t-1|t-1}] \\
&= \phi\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)\frac{\partial f_{t-1|t-1}}{\partial \phi} \\
&\quad + \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial \phi} + \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)f_{t-1|t-1}.
\end{aligned}
$$

Using $\frac{\partial V_{t|t-1}}{\partial \phi} = \phi^2 \frac{\partial V_{t-1|t-1}}{\partial \phi} + 2\phi V_{t-1|t-1}$ we get:

$$
\begin{aligned}
\frac{\partial f_{t|t}}{\partial \phi} &= \phi\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)\frac{\partial f_{t-1|t-1}}{\partial \phi} \\
&\quad + \phi^2\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t-1|t-1}}{\partial \phi} \\
&\quad + \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)f_{t-1|t-1} + 2\phi V_{t-1|t-1}\left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1}).
\end{aligned}
$$

(iv) Derivative of $f_{t|t}$ w.r.t. S_u. We have

$$
\begin{aligned}
\frac{\partial f_{t|t}}{\partial S_u'} &= \phi\frac{\partial f_{t-1|t-1}}{\partial S_u'} + B'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial S_u'} \\
&\quad - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}BB'\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\frac{\partial V_{t|t-1}}{\partial S_u'} \\
&\quad - V_{t|t-1}\left(B'\Sigma_{t|t-1}^{-1}\frac{\partial \Sigma_u}{\partial S_{u,k}}\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})\right)_{k=1,\ldots,\frac{n(n+1)}{2}} \\
&\quad - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}[B\phi\frac{\partial f_{t-1|t-1}}{\partial S_u'}].
\end{aligned}
$$

Using

$$
B'\Sigma_{t|t-1}^{-1}\frac{\partial \Sigma_u}{\partial S_{u,k}}\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1}) = \left\{[(Y_t - Y_{t|t-1})'\Sigma_{t|t-1}^{-1}] \otimes [B'\Sigma_{t|t-1}^{-1}]\right\}\frac{\partial vec\Sigma_u}{\partial S_{u,k}}
$$

and $\frac{\partial V_{t|t-1}}{\partial S_u} = \phi^2 \frac{\partial V_{t-1|t-1}}{\partial S_u}$ and rearranging terms, we get:

$$
\begin{aligned}
\frac{\partial f_{t|t}}{\partial S_u} &= \phi \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) \frac{\partial f_{t-1|t-1}}{\partial S_u} \\
&\quad + \phi^2 \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) B' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \frac{\partial V_{t-1|t-1}}{\partial S_u} \\
&\quad - V_{t|t-1} M' \left\{ [\Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1})] \otimes [\Sigma_{t|t-1}^{-1} B] \right\}.
\end{aligned}
$$

(v) Derivative of $f_{t|t}$ w.r.t. σ_v^2. We have

$$
\begin{aligned}
\frac{\partial f_{t|t}}{\partial \sigma_v^2} &= \phi \frac{\partial f_{t-1|t-1}}{\partial \sigma_v^2} + B' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \frac{\partial V_{t|t-1}}{\partial \sigma_v^2} \\
&\quad - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B B' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \frac{\partial V_{t|t-1}}{\partial \sigma_v^2} \\
&\quad - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B \phi \frac{\partial f_{t-1|t-1}}{\partial \sigma_v^2}
\end{aligned}
$$

and $\frac{\partial V_{t|t-1}}{\partial \sigma_v^2} = \phi^2 \frac{\partial V_{t-1|t-1}}{\partial \sigma_v^2} + 1$ which yields

$$
\begin{aligned}
\frac{\partial f_{t|t}}{\partial \sigma_v^2} &= \phi \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) \frac{\partial f_{t-1|t-1}}{\partial \sigma_v^2} \\
&\quad + \phi^2 \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) B' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \frac{\partial V_{t-1|t-1}}{\partial \sigma_v^2} \\
&\quad + \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) B' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}).
\end{aligned}
$$

Putting these equations together and evaluating the parameters at the true values, we get

$$
\frac{\partial f_{t|t}}{\partial \theta} = \phi_0 \gamma_t \frac{\partial f_{t-1|t-1}}{\partial \theta} + \phi_0^2 \gamma_t B_0' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \frac{\partial V_{t-1|t-1}}{\partial \theta} + A_t \tag{E.38}
$$

where

$$
\gamma_t = 1 - V_{t|t-1} B_0' \Sigma_{t|t-1}^{-1} B_0
$$

and

$$A_t = \begin{bmatrix} -V_{t|t-1}[Y_{t-1} \otimes (\Sigma_{t|t-1}^{-1} B_0)] \\ V_{t|t-1}\gamma_t[\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})]_{1:n-1} - V_{t|t-1} f_{t|t}[\Sigma_{t|t-1}^{-1} B_0]_{1:n-1} \\ \gamma_t \left(f_{t-1|t-1} + 2\phi_0 V_{t-1|t-1} B_0' \Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1}) \right) \\ -V_{t|t-1} M' \left\{ [\Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1})] \otimes [\Sigma_{t|t-1}^{-1} B_0] \right\} \\ \gamma_t B_0' \Sigma_{t|t-1}^{-1}(Y_t - Y_{t|t-1}) \end{bmatrix}.$$

Let us finally obtain the recursive equations for the gradient of $V_{t|t}$.

(i) Derivative of $V_{t|t}$ w.r.t. c. We have

$$\begin{aligned} \frac{\partial V_{t|t}}{\partial c'} &= \frac{\partial V_{t|t-1}}{\partial c'} - 2V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B \frac{\partial V_{t|t-1}}{\partial c'} \\ &\quad + V_{t|t-1}^2 B' \Sigma_{t|t-1}^{-1} B B' \Sigma_{t|t-1}^{-1} B \frac{\partial V_{t|t-1}}{\partial c'} \\ &= \phi^2 \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right)^2 \frac{\partial V_{t-1|t-1}}{\partial c'}. \end{aligned}$$

(ii) Derivative of $V_{t|t}$ w.r.t. b. We have

$$\begin{aligned} \frac{\partial V_{t|t}}{\partial B'} &= \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right)^2 \frac{\partial V_{t|t-1}}{\partial B'} \\ &\quad + V_{t|t-1}^3 \left(B' \Sigma_{t|t-1}^{-1} \frac{\partial B B'}{\partial B_k} \Sigma_{t|t-1}^{-1} B \right)_{k=1,\dots,n} - 2V_{t|t-1}^2 B' \Sigma_{t|t-1}^{-1}. \end{aligned}$$

By using $B' \Sigma_{t|t-1}^{-1} \frac{\partial B B'}{\partial B_k} \Sigma_{t|t-1}^{-1} B = 2(B' \Sigma_{t|t-1}^{-1} B) B' \Sigma_{t|t-1}^{-1} e_k$ and $\frac{\partial V_{t|t-1}}{\partial b} = \phi^2 \frac{\partial V_{t-1|t-1}}{\partial b}$ we get

$$\begin{aligned} \frac{\partial V_{t|t}}{\partial b} &= \phi^2 \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right)^2 \frac{\partial V_{t-1|t-1}}{\partial b} \\ &\quad - 2V_{t|t-1}^2 \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right) [\Sigma_{t|t-1}^{-1} B]_{1:n-1}. \end{aligned}$$

(iii) Derivative of $V_{t|t}$ w.r.t. ϕ. We have

$$\begin{aligned} \frac{\partial V_{t|t}}{\partial \phi} &= \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right)^2 \frac{\partial V_{t|t-1}}{\partial \phi} \\ &= \phi^2 \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right)^2 \frac{\partial V_{t-1|t-1}}{\partial \phi} + 2\phi V_{t-1|t-1} \left(1 - V_{t|t-1} B' \Sigma_{t|t-1}^{-1} B\right)^2. \end{aligned}$$

(iv) Derivative of $V_{t|t}$ w.r.t. S_u. We have

$$\begin{aligned}\frac{\partial V_{t|t}}{\partial S_u'} &= \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)^2 \frac{\partial V_{t|t-1}}{\partial S_u'} \\ &\quad + V_{t|t-1}^2 \left(B'\Sigma_{t|t-1}^{-1}\frac{\partial \Sigma_u}{\partial S_{u,k}}\Sigma_{t|t-1}^{-1}B\right)_{k=1,\ldots,n}.\end{aligned}$$

By using $\left(B'\Sigma_{t|t-1}^{-1}\frac{\partial \Sigma_u}{\partial S_{u,k}}\Sigma_{t|t-1}^{-1}B\right)_{k=1,\ldots,\frac{n(n+1)}{2}} = \left[(B'\Sigma_{t|t-1}^{-1}) \otimes (B'\Sigma_{t|t-1}^{-1})\right] M$ and $\frac{\partial V_{t|t-1}}{\partial S_u} = \phi^2 \frac{\partial V_{t-1|t-1}}{\partial S_u}$ we get

$$\begin{aligned}\frac{\partial V_{t|t}}{\partial S_u} &= \phi^2 \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)^2 \frac{\partial V_{t-1|t-1}}{\partial S_u} \\ &\quad + V_{t|t-1}^2 M' \left[(\Sigma_{t|t-1}^{-1}B) \otimes (\Sigma_{t|t-1}^{-1}B)\right].\end{aligned}$$

(v) Derivative of $V_{t|t}$ w.r.t. σ_v^2. We have

$$\begin{aligned}\frac{\partial V_{t|t}}{\partial \sigma_v^2} &= \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)^2 \frac{\partial V_{t|t-1}}{\partial \sigma_v^2} \\ &= \phi^2 \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)^2 \frac{\partial V_{t-1|t-1}}{\partial \sigma_v^2} + \left(1 - V_{t|t-1}B'\Sigma_{t|t-1}^{-1}B\right)^2.\end{aligned}$$

Putting these equations together and evaluating the parameters at the true values, we get

$$\frac{\partial V_{t|t}}{\partial \theta} = \phi^2 \gamma_t^2 \frac{\partial V_{t-1|t-1}}{\partial \theta} + D_t g \tag{E.39}$$

where

$$D_t = \begin{bmatrix} 0_{n^2 \times 1} \\ -2V_{t|t-1}^2 \gamma_t [\Sigma_{t|t-1}^{-1} B_0]_{1:n-1} \\ 2\phi_0 V_{t-1|t-1} \gamma_t^2 \\ V_{t|t-1}^2 M' \left[(\Sigma_{t|t-1}^{-1} B_0) \otimes (\Sigma_{t|t-1}^{-1} B_0)\right] \\ \gamma_t^2 \end{bmatrix}.$$

To summarize, we can compute the input for the information matrix by augmenting the Kalman filter with the equations for the gradients:

$$
\begin{aligned}
f_{t|t-1} &= \phi_0 f_{t-1|t-1} \\
Y_{t|t-1} &= B_0 f_{t|t-1} + C_0 Y_{t-1} \\
V_{t|t-1} &= \phi_0^2 V_{t-1|t-1} + \sigma_{v,0}^2 \\
\Sigma_{t|t-1} &= V_{t|t-1} B_0 B_0' + \Sigma_{u,0} \\
\gamma_t &= 1 - V_{t|t-1} B_0' \Sigma_{t|t-1}^{-1} B_0 \\
f_{t|t} &= f_{t|t-1} + V_{t|t-1} B_0' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \\
V_{t|t} &= \gamma_t V_{t|t-1} \\
\frac{\partial f_{t|t}}{\partial \theta} &= \phi_0 \gamma_t \frac{\partial f_{t-1|t-1}}{\partial \theta} + \phi_0^2 \gamma_t B_0' \Sigma_{t|t-1}^{-1} (Y_t - Y_{t|t-1}) \frac{\partial V_{t-1|t-1}}{\partial \theta} + A_t \\
\frac{\partial V_{t|t}}{\partial \theta} &= \phi_0^2 \gamma_t^2 \frac{\partial V_{t-1|t-1}}{\partial \theta} + D_t \\
\frac{\partial m_t(\theta_0)}{\partial \theta'} &= B_0 \phi_0 \frac{\partial f_{t-1|t-1}}{\partial \theta'} + \left[\begin{array}{ccccc} Y_{t-1}' \otimes I_n & I_{1:n-1} \phi_0 f_{t-1|t-1} & B_0 f_{t-1|t-1} & 0_{n \times \frac{n(n+1)}{2}} & 0_{n \times 1} \end{array} \right] \\
\frac{\partial vec \Sigma_t(\theta_0)}{\partial \theta'} &= (B_0 \otimes B_0) \phi_0^2 \frac{\partial V_{t-1|t-1}}{\partial \theta'} \\
&\quad + \left[\begin{array}{ccccc} 0_{n^2 \times n} & (\phi_0^2 V_{t-1|t-1} + \sigma_{v,0}^2) \frac{\partial B_0 \otimes B_0}{\partial b_0'} & 2\phi_0 V_{t-1|t-1} B_0 \otimes B_0 & M & B_0 \otimes B_0 \end{array} \right].
\end{aligned}
$$

The iteration is started with $f_{0|0} = f_{1|0} = 0$, $V_{0|0} = V_{1|0} = \frac{\sigma_{v,0}^2}{1-\phi_0^2}$, $Y_{1|0} = Y_0 = 0$, $\Sigma_{1|0} = \Omega$, where $vec(\Omega) = (I_{n^2} - C \otimes C)^{-1} (\frac{\sigma_{v,0}^2}{1-\phi_0^2} B_0 B_0' + \Sigma_{u,0})$, and $\frac{\partial f_{0|0}}{\partial \theta} = 0$ and $\frac{\partial V_{0|0}}{\partial \theta} = 0$.

F Additional empirical results

In Table 2 and Figure 4 we provide summary statistics and the auto- and cross-correlation functions (ACF and XCF) for the four series of log 5-minute realized volatilities of the CAC 40, OMX Stockholm 30, IBEX 35 and DAX 30 indices used in the empirical illustration of Section 78. Figure 5 provides the ACF and XCF of $\hat{\xi}_t$ and $\hat{\eta}_t = (\hat{\eta}_{1,t}, \hat{\eta}_{2,t}, \hat{\eta}_{3,t})'$.

	Mean	St.Dv.	5% quant.	Median	95% quant.	Skew	Kurt.
CAC 40	-9.6483	0.9036	-11.0429	-9.6860	-8.1353	0.4145	3.6503
OMX Stockholm 30	-10.1363	0.9073	-11.4314	-10.2404	-8.5189	0.7201	4.2395
IBEX 35	-9.2547	0.8441	-10.5308	-9.2937	-7.8389	0.4377	3.6228
DAX 30	-9.6486	0.9058	-11.0877	-9.6844	-8.1321	0.3427	3.5243

Table 2: This table summarizes some descriptive statistics of daily log 5-minute realized volatilities for European financial indices for the period from June 1, 2009 to November 19, 2021.

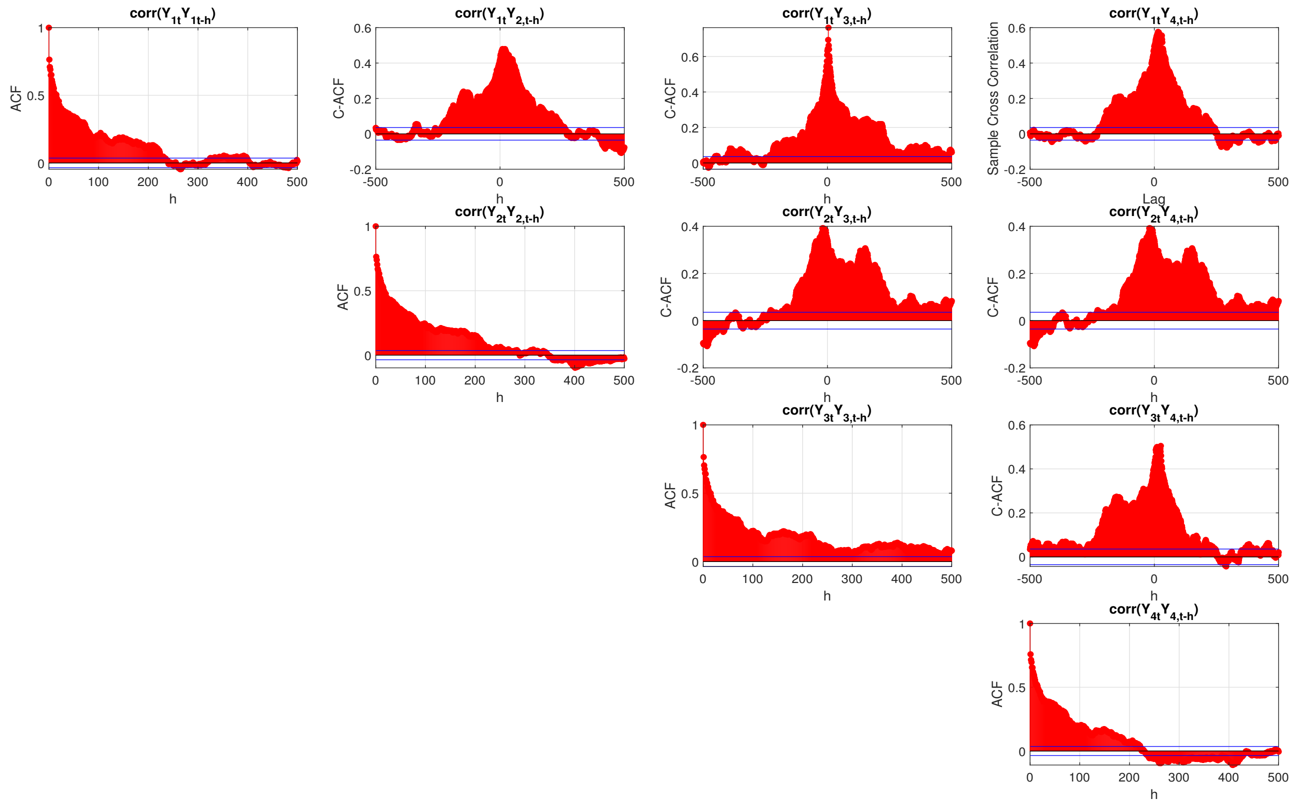

Figure 4: This figure plots the autocorrelation and cross-correlation functions of the log realized volatilities. The series Y_{1t}, Y_{2t}, Y_{3t} and Y_{4t} are the demeaned log realized volatilities of the CAC 40, OMX Stockholm 30, IBEX 35 and DAX 30 indices, respectively. The plots on the diagonal are the autocorrelation functions of Y_{it}, for $i = 1, 2, 3, 4$. The plots that are in out-of-diagonal position refer to cross-correlation functions of Y_{it} and Y_{jt} for $i \neq j$. In each panel, the horizontal axis represents the lag h.

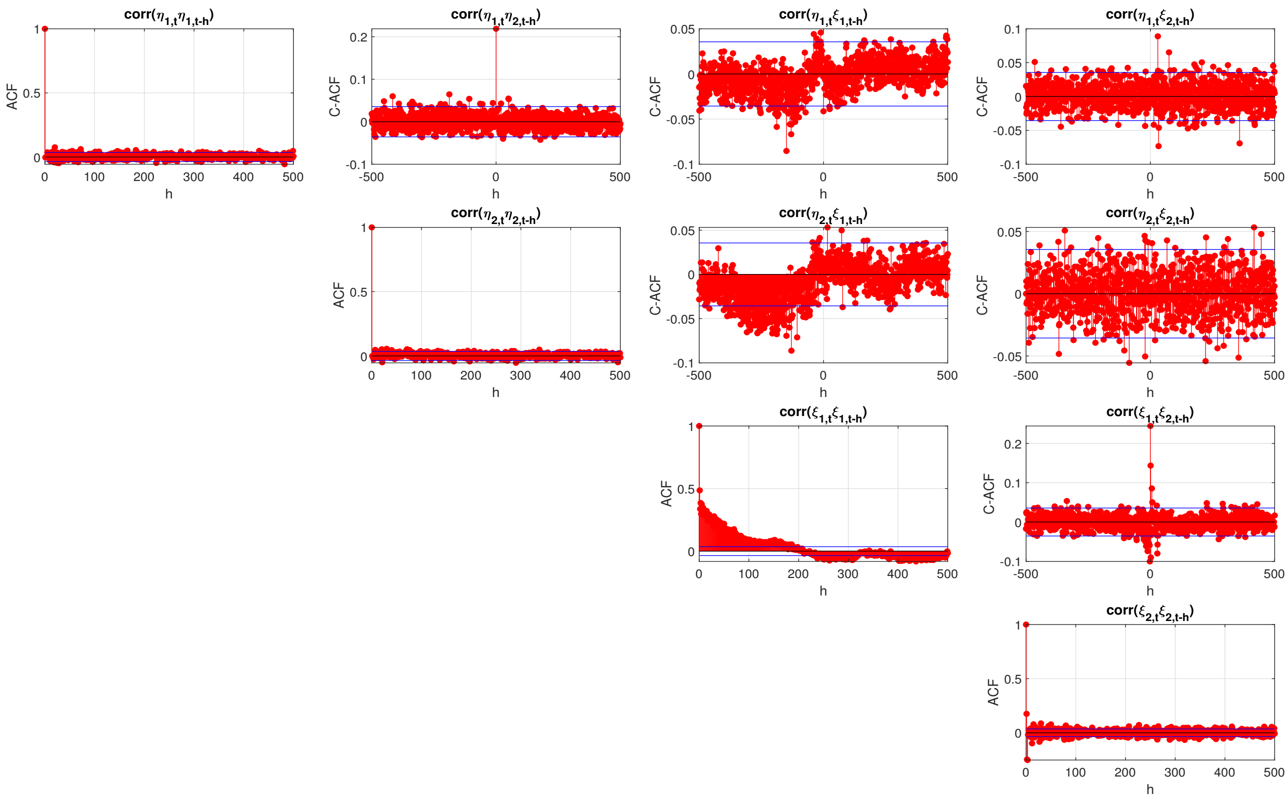

Figure 5: This figure plots the estimated autocorrelation and cross-correlation functions of the two persistent components ξ_{it}, $i = 1, 2$ and the two White Noise components η_{jt}, $j = 1, 2$. The plots on the diagonal are the autocorrelation functions of ξ_{it} and η_{jt}, for $i, j = 1, 2$. The plots that are in out-of-diagonal position refer to cross-correlation functions.

G Monte Carlo results

This section reports the results of the Monte Carlo experiments. The finite-sample bias and variance of the model parameters' estimators are provided in Section G.1. The finite-sample size and power of the test statistic for the number of common factors are presented in Section G.2. The finite-sample percentages of selected model orders K, p, q are displayed in Section G.3.

G.1 Bias and standard deviations of estimators

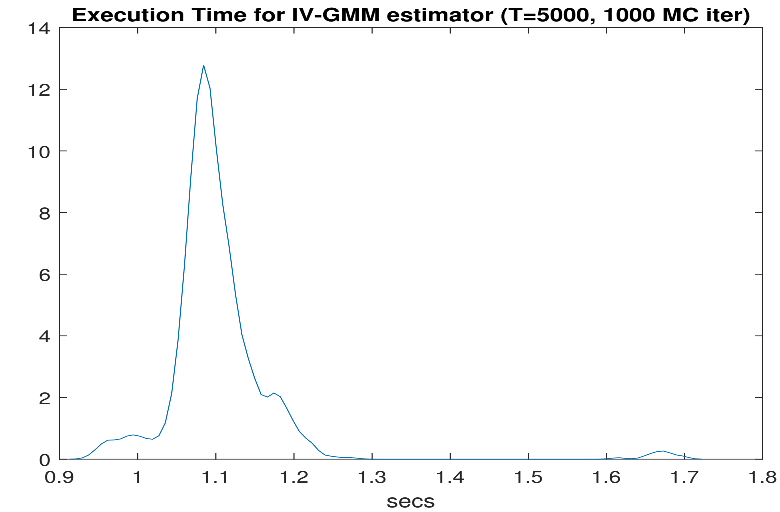

Figure 6: This figure plots the kernel density of the execution times (in seconds) for our estimator in 1000 Monte Carlo samples. Data are generated according to DGP 1, with $T = 5000$.

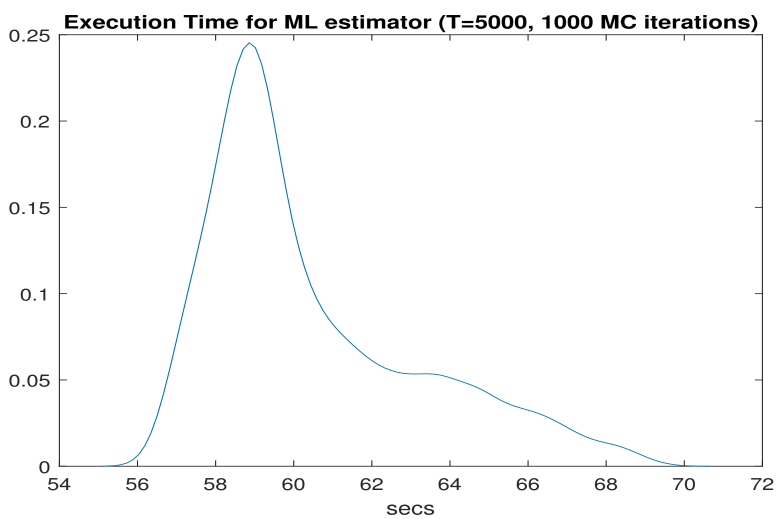

Figure 7: This figure plots the kernel density of the execution times (in seconds) for the ML estimator in 1000 Monte Carlo samples. Data are generated according to DGP 1, with $T = 5000$.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_1	0.5000	0.0105	0.1680	0.0086	0.1162	0.0020	0.0491
b_2	0.5000	-0.0014	0.1681	-0.0057	0.1116	0.0013	0.0496
b_3	0.5000	0.0088	0.1730	0.0065	0.1191	0.0019	0.0519
c_{11}	-0.2158	0.0400	0.0766	0.0140	0.0634	-0.0008	0.0279
c_{21}	0.2860	0.0503	0.0748	0.0265	0.0555	0.0019	0.0174
c_{31}	0.3860	0.0420	0.0698	0.0170	0.0530	0.0007	0.0251
c_{41}	0.4719	0.0864	0.1096	0.0407	0.0839	0.0032	0.0343
c_{12}	-0.3768	0.0156	0.0482	0.0088	0.0342	0.0008	0.0131
c_{22}	-0.6122	0.0058	0.0485	0.0036	0.0351	0.0000	0.0140
c_{32}	-0.0122	0.0180	0.0475	0.0109	0.0349	0.0014	0.0130
c_{42}	-0.0244	0.0275	0.0659	0.0161	0.0470	0.0012	0.0170
c_{13}	0.3280	0.0730	0.0781	0.0318	0.0668	0.0021	0.0286
c_{23}	0.2537	0.0677	0.0794	0.0360	0.0593	0.0040	0.0244
c_{33}	0.4537	0.0634	0.0810	0.0282	0.0664	0.0013	0.0302
c_{43}	-0.6927	0.1395	0.1290	0.0700	0.1039	0.0070	0.0418
c_{14}	-0.4669	0.0621	0.0836	0.0263	0.0652	0.0007	0.0238
c_{24}	0.2445	0.0586	0.0814	0.0319	0.0594	0.0025	0.0183
c_{34}	0.3445	0.0608	0.0781	0.0275	0.0571	0.0023	0.0224
c_{44}	0.2890	0.1130	0.1337	0.0557	0.0955	0.0035	0.0345
ϕ	0.8000	-0.1184	0.2496	-0.0301	0.1150	-0.0030	0.0176

Table 3: Bias and standard deviation for our estimator in DGP 1 ($n = 4$, $K = 1$, Gaussian noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_1	0.5000	-0.0002	0.0745	0.0007	0.0502	0.0043	0.0228
b_2	0.5000	-0.0161	0.0748	-0.0124	0.0520	-0.0046	0.0253
b_3	0.5000	-0.0872	0.0659	-0.0816	0.0518	-0.0623	0.0417
c_{11}	-0.2158	-0.0161	0.0471	-0.0181	0.0341	-0.0236	0.0158
c_{21}	0.2860	-0.0088	0.0417	-0.0132	0.0296	-0.0212	0.0146
c_{31}	0.3860	0.0080	0.0435	0.0038	0.0328	-0.0061	0.0186
c_{41}	0.4719	-0.0233	0.0595	-0.0313	0.0434	-0.0425	0.0212
c_{12}	-0.3768	-0.0052	0.0370	-0.0062	0.0267	-0.0043	0.0117
c_{22}	-0.6122	-0.0048	0.0404	-0.0048	0.0288	-0.0030	0.0121
c_{32}	-0.0122	0.0019	0.0386	0.0004	0.0288	0.0018	0.0128
c_{42}	-0.0244	-0.0094	0.0465	-0.0116	0.0326	-0.0080	0.0156
c_{13}	0.3280	0.0294	0.0588	0.0245	0.0466	0.0091	0.0284
c_{23}	0.2537	0.0337	0.0536	0.0275	0.0438	0.0113	0.0284
c_{33}	0.4537	0.0485	0.0582	0.0447	0.0485	0.0275	0.0350
c_{43}	-0.6927	0.0611	0.0981	0.0504	0.0818	0.0208	0.0534
c_{14}	-0.4669	-0.0357	0.0342	-0.0382	0.0244	-0.0379	0.0130
c_{24}	0.2445	-0.0291	0.0316	-0.0329	0.0226	-0.0352	0.0110
c_{34}	0.3445	-0.0077	0.0309	-0.0123	0.0226	-0.0181	0.0100
c_{44}	0.2890	-0.0701	0.0415	-0.0747	0.0301	-0.0729	0.0175
ϕ	0.8000	-0.1224	0.0934	-0.1102	0.0716	-0.0826	0.0401

Table 4: Bias and standard deviation for the ML estimator in DGP 1 ($n = 4$, $K = 1$, Gaussian noises). The second column reports the true values of the parameters.

		True		*Random*	
Param.	θ_0	Bias	Std.	Bias	Std.
b_1	0.5000	0.0011	0.0563	0.2149	0.2223
b_2	0.5000	0.0019	0.0498	0.1795	0.2242
b_3	0.5000	0.0029	0.0640	0.5522	0.1048
c_{11}	-0.2158	-0.0015	0.0320	0.3495	0.0842
c_{21}	0.2860	0.0002	0.0272	0.2997	0.0833
c_{31}	0.3860	-0.0003	0.0333	0.3464	0.0374
c_{41}	0.4719	0.0011	0.0383	0.5584	0.0575
c_{12}	-0.3768	-0.0011	0.0262	0.0090	0.0605
c_{22}	-0.6122	-0.0009	0.0286	-0.0043	0.0370
c_{32}	-0.0122	-0.0008	0.0275	0.0306	0.0828
c_{42}	-0.0244	0.0005	0.0326	0.0269	0.0766
c_{13}	0.3280	-0.0005	0.0402	0.4666	0.0607
c_{23}	0.2537	0.0000	0.0333	0.3940	0.0731
c_{33}	0.4537	-0.0023	0.0466	0.5240	0.0523
c_{43}	-0.6927	0.0013	0.0582	0.7081	0.0542
c_{14}	-0.4669	-0.0001	0.0226	0.1120	0.0673
c_{24}	0.2445	0.0007	0.0215	0.1088	0.0583
c_{34}	0.3445	0.0015	0.0217	0.1389	0.0639
c_{44}	0.2890	0.0004	0.0249	0.2163	0.0669
ϕ	0.8000	-0.0061	0.0340	0.1414	0.0571

Table 5: Bias and standard deviation for the ML estimator in DGP 1 ($n = 4$, $K = 1$, Gaussian noises, $T = 5000$) with different initial values in the iterative maximization algorithm. The second column reports the true values of the parameters. The third and fourth columns are the results when the initial values are the true parameter values. The fifth and sixth columns are the results when the initial values are the true parameter values plus a random draw from a uniform variate $U[-0.5, 0.5]$.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_1	0.5000	0.0205	0.1833	0.0109	0.1184	0.0011	0.0498
b_2	0.5000	0.0077	0.1776	0.0025	0.1218	0.0009	0.0489
b_3	0.5000	0.0127	0.1747	0.0102	0.1144	0.0039	0.0516
c_{11}	-0.2158	0.0411	0.0789	0.0155	0.0576	-0.0014	0.0294
c_{21}	0.2860	0.0515	0.0797	0.0240	0.0511	0.0025	0.0164
c_{31}	0.3860	0.0457	0.0751	0.0178	0.0543	-0.0001	0.0264
c_{41}	0.4719	0.0913	0.1196	0.0421	0.0837	0.0028	0.0360
c_{12}	-0.3768	0.0137	0.0464	0.0091	0.0338	0.0009	0.0132
c_{22}	-0.6122	0.0051	0.0478	0.0014	0.0326	-0.0002	0.0142
c_{32}	-0.0122	0.0171	0.0485	0.0109	0.0345	0.0021	0.0130
c_{42}	-0.0244	0.0288	0.0699	0.0162	0.0466	0.0020	0.0175
c_{13}	0.3280	0.0724	0.0806	0.0341	0.0634	0.0006	0.0304
c_{23}	0.2537	0.0719	0.0758	0.0355	0.0596	0.0032	0.0236
c_{33}	0.4537	0.0659	0.0839	0.0301	0.0661	0.0009	0.0321
c_{43}	-0.6927	0.1440	0.1290	0.0709	0.1038	0.0061	0.0427
c_{14}	-0.4669	0.0635	0.0812	0.0280	0.0598	0.0006	0.0250
c_{24}	0.2445	0.0650	0.0810	0.0306	0.0564	0.0042	0.0179
c_{34}	0.3445	0.0647	0.0750	0.0303	0.0578	0.0022	0.0226
c_{44}	0.2890	0.1177	0.1326	0.0574	0.0965	0.0043	0.0361
ϕ	0.8000	-0.1174	0.2539	-0.0313	0.1260	-0.0014	0.0168

Table 6: Bias and standard deviation for our estimator in DGP 2 ($n = 4$, $K = 1$, Student noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_1	0.5000	0.0092	0.0733	0.0130	0.0513	0.0200	0.0219
b_2	0.5000	-0.0222	0.0697	-0.0154	0.0465	-0.0163	0.0199
b_3	0.5000	-0.0261	0.0768	-0.0079	0.0617	0.0293	0.0299
c_{11}	-0.2158	-0.0292	0.0465	-0.0311	0.0347	-0.0383	0.0149
c_{21}	0.2860	-0.0174	0.0395	-0.0214	0.0284	-0.0281	0.0126
c_{31}	0.3860	-0.0180	0.0468	-0.0239	0.0373	-0.0403	0.0170
c_{41}	0.4719	-0.0463	0.0576	-0.0525	0.0433	-0.0657	0.0180
c_{12}	-0.3768	-0.0015	0.0369	-0.0018	0.0265	0.0015	0.0116
c_{22}	-0.6122	0.0002	0.0394	-0.0004	0.0278	0.0038	0.0123
c_{32}	-0.0122	0.0024	0.0360	0.0000	0.0275	0.0003	0.0123
c_{42}	-0.0244	0.0007	0.0463	0.0014	0.0331	0.0083	0.0148
c_{13}	0.3280	-0.0130	0.0622	-0.0225	0.0494	-0.0435	0.0203
c_{23}	0.2537	-0.0001	0.0506	-0.0110	0.0401	-0.0289	0.0154
c_{33}	0.4537	-0.0053	0.0675	-0.0175	0.0573	-0.0470	0.0251
c_{43}	-0.6927	-0.0169	0.0960	-0.0344	0.0780	-0.0717	0.0292
c_{14}	-0.4669	-0.0306	0.0316	-0.0305	0.0223	-0.0283	0.0098
c_{24}	0.2445	-0.0219	0.0311	-0.0223	0.0234	-0.0189	0.0102
c_{34}	0.3445	-0.0185	0.0308	-0.0230	0.0225	-0.0301	0.0101
c_{44}	0.2890	-0.0575	0.0378	-0.0550	0.0280	-0.0474	0.0130
ϕ	0.8000	-0.0605	0.0701	-0.0443	0.0503	-0.0178	0.0170

Table 7: Bias and standard deviation for ML estimator in DGP 2 ($n = 4$, $K = 1$, Student noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_{11}	-1.5000	-0.0423	0.2599	-0.0184	0.1605	-0.0036	0.0664
b_{21}	-1.2000	-0.0332	0.2359	-0.0156	0.1475	-0.0030	0.0597
b_{31}	1.3000	0.0439	0.1995	0.0203	0.1232	0.0014	0.0476
b_{12}	1.6000	0.0163	0.1968	0.0137	0.1281	0.0017	0.0534
b_{22}	1.5000	0.0168	0.1729	0.0142	0.1124	0.0026	0.0477
b_{32}	-0.8000	-0.0158	0.1535	-0.0128	0.0956	-0.0011	0.0407
c_{11}	0.8000	0.0104	0.2621	-0.0286	0.2109	-0.0180	0.0899
c_{21}	0.0000	0.0106	0.2240	-0.0260	0.1771	-0.0162	0.0758
c_{31}	0.2000	-0.0425	0.1924	-0.0074	0.1513	0.0097	0.0731
c_{41}	0.0000	-0.0532	0.1228	-0.0282	0.1003	0.0013	0.0520
c_{51}	0.0000	-0.0409	0.0907	-0.0420	0.0776	-0.0101	0.0260
c_{12}	0.0000	-0.0540	0.3109	-0.0561	0.2572	-0.0221	0.1217
c_{22}	0.8000	-0.0743	0.2652	-0.0633	0.2171	-0.0199	0.0972
c_{32}	0.0000	0.0129	0.2314	0.0144	0.1825	0.0120	0.0977
c_{42}	0.4000	-0.0328	0.1439	-0.0218	0.1155	0.0013	0.0682
c_{52}	0.0000	-0.0690	0.1137	-0.0567	0.0939	-0.0125	0.0311
c_{13}	0.4000	0.0853	0.3549	0.0463	0.2628	0.0002	0.1055
c_{23}	0.0000	0.0786	0.3112	0.0451	0.2318	0.0010	0.0941
c_{33}	-0.6000	-0.0316	0.2514	-0.0221	0.1795	-0.0006	0.0698
c_{43}	0.0000	0.0309	0.1553	0.0142	0.1030	0.0030	0.0397
c_{53}	0.5000	0.0646	0.1542	0.0351	0.1248	0.0039	0.0558
c_{14}	0.0000	-0.0004	0.3586	0.0125	0.2844	-0.0058	0.1288
c_{24}	0.6000	-0.0074	0.3189	0.0073	0.2528	-0.0056	0.1145
c_{34}	0.0000	0.0159	0.2414	-0.0005	0.1833	0.0039	0.0802
c_{44}	-0.5000	0.0408	0.1297	0.0169	0.0949	0.0024	0.0403
c_{54}	0.0000	0.0183	0.1821	0.0156	0.1468	-0.0012	0.0690
c_{15}	0.0000	-0.0506	0.5149	-0.0402	0.3528	-0.0195	0.1403
c_{25}	0.0000	-0.0682	0.4395	-0.0464	0.2979	-0.0182	0.1138
c_{35}	0.0000	-0.0005	0.3862	-0.0012	0.2726	0.0087	0.1159
c_{45}	0.0000	-0.0215	0.2414	-0.0210	0.1688	-0.0004	0.0747
c_{55}	0.8000	-0.0906	0.1824	-0.0587	0.1239	-0.0142	0.0397
ϕ_{11}	0.9000	-0.2317	0.5046	-0.0754	0.3098	-0.0041	0.0549
ϕ_{21}	0.0000	-0.0866	0.6865	-0.1473	0.5237	-0.0529	0.2080
ϕ_{12}	0.0000	0.1728	0.5550	0.1214	0.4183	-0.0071	0.1985
ϕ_{22}	-0.9000	0.6333	0.6904	0.4847	0.6643	0.0645	0.2352

Table 8: Bias and standard deviation for our estimator in DGP 3 ($n = 5$, $K = 2$, $\phi_0 = 0.9$, $\sigma_{v,0} = 1$, Gaussian noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)
b_{11}	-1.5000	-0.0332	0.2494	-0.0037	0.1546	-0.0014	0.0655
b_{21}	-1.2000	-0.0272	0.2060	-0.0088	0.1407	-0.0019	0.0585
b_{31}	1.3000	0.0351	0.1888	0.0192	0.1161	0.0005	0.0491
b_{12}	1.6000	0.0104	0.1916	0.0095	0.1235	0.0014	0.0552
b_{22}	1.5000	0.0111	0.1677	0.0140	0.1110	0.0014	0.0474
b_{32}	-0.8000	-0.0157	0.1474	-0.0112	0.0915	-0.0010	0.0400
c_{11}	0.8000	0.0534	0.2083	0.0080	0.1842	-0.0067	0.0858
c_{21}	0.0000	0.0397	0.1854	0.0024	0.1558	-0.0065	0.0744
c_{31}	0.2000	-0.0699	0.1669	-0.0249	0.1400	0.0018	0.0709
c_{41}	0.0000	-0.0683	0.1113	-0.0387	0.0923	-0.0036	0.0481
c_{51}	0.0000	-0.0263	0.0670	-0.0266	0.0564	-0.0082	0.0207
c_{12}	0.0000	-0.0241	0.2640	-0.0247	0.2386	-0.0092	0.1104
c_{22}	0.8000	-0.0421	0.2131	-0.0311	0.1902	-0.0116	0.0895
c_{32}	0.0000	-0.0017	0.1962	-0.0007	0.1685	0.0023	0.0894
c_{42}	0.4000	-0.0309	0.1194	-0.0245	0.1068	-0.0027	0.0609
c_{52}	0.0000	-0.0517	0.0966	-0.0356	0.0664	-0.0096	0.0254
c_{13}	0.4000	0.0783	0.3411	0.0396	0.2489	-0.0045	0.1018
c_{23}	0.0000	0.0669	0.2916	0.0393	0.2203	0.0014	0.0925
c_{33}	-0.6000	-0.0279	0.2337	-0.0163	0.1673	0.0023	0.0686
c_{43}	0.0000	0.0348	0.1399	0.0174	0.0948	0.0030	0.0398
c_{53}	0.5000	0.0702	0.1494	0.0359	0.1210	0.0020	0.0542
c_{14}	0.0000	0.0003	0.3742	0.0144	0.2856	-0.0115	0.1297
c_{24}	0.6000	-0.0081	0.3331	0.0142	0.2558	-0.0136	0.1180
c_{34}	0.0000	0.0095	0.2461	-0.0065	0.1894	0.0064	0.0796
c_{44}	-0.5000	0.0339	0.1235	0.0139	0.0867	0.0016	0.0378
c_{54}	0.0000	0.0166	0.1927	0.0188	0.1478	-0.0032	0.0709
c_{15}	0.0000	-0.1017	0.4142	-0.0535	0.3173	-0.0303	0.1343
c_{25}	0.0000	-0.1062	0.3471	-0.0596	0.2722	-0.0211	0.1084
c_{35}	0.0000	0.0395	0.3259	0.0194	0.2493	0.0184	0.1115
c_{45}	0.0000	0.0004	0.2112	-0.0038	0.1565	0.0060	0.0699
c_{55}	0.8000	-0.0565	0.1292	-0.0362	0.0956	-0.0102	0.0327
ϕ_{11}	0.9000	-0.0773	0.2367	-0.0163	0.1253	-0.0049	0.0568
ϕ_{21}	0.0000	-0.1541	0.5634	-0.1990	0.4331	-0.0358	0.1944
ϕ_{12}	0.0000	0.1002	0.5331	0.0565	0.3776	-0.0149	0.1999
ϕ_{22}	-0.9000	0.3598	0.7433	0.1820	0.4277	0.0142	0.0481

Table 9: Difference between the median and true value, and Interquartile Range (IQR) divided by twice the 75% standard Normal quantile $z_{75\%}$, for our estimator in DGP 3 ($n = 5$, $K = 2$, $\phi_0 = 0.9$, $\sigma_{v,0} = 1$, Gaussian noises). The second column reports the true values of the parameters. The quantity $IQR/(2z_{75\%})$ equals the standard deviation in a Gaussian $N(0,1)$ distribution.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_{11}	-1.5000	-0.0662	0.3565	-0.0249	0.2126	-0.0045	0.0898
b_{21}	-1.2000	-0.0462	0.3202	-0.0183	0.1963	-0.0033	0.0802
b_{31}	1.3000	0.0588	0.2724	0.0231	0.1644	0.0005	0.0647
b_{12}	1.6000	0.0375	0.3042	0.0264	0.2022	0.0040	0.0833
b_{22}	1.5000	0.0393	0.2631	0.0275	0.1766	0.0062	0.0742
b_{32}	-0.8000	-0.0332	0.2328	-0.0237	0.1492	-0.0021	0.0641
c_{11}	0.8000	-0.0412	0.2108	-0.0401	0.1518	-0.0116	0.0614
c_{21}	0.0000	-0.0302	0.1795	-0.0328	0.1270	-0.0100	0.0503
c_{31}	0.2000	-0.0010	0.1599	0.0105	0.1136	0.0071	0.0511
c_{41}	0.0000	-0.0187	0.1210	-0.0016	0.0813	0.0040	0.0385
c_{51}	0.0000	-0.0304	0.0934	-0.0218	0.0593	-0.0037	0.0182
c_{12}	0.0000	-0.0382	0.2672	-0.0286	0.1868	-0.0120	0.0856
c_{22}	0.8000	-0.0577	0.2226	-0.0343	0.1555	-0.0105	0.0678
c_{32}	0.0000	-0.0010	0.2027	0.0015	0.1454	0.0071	0.0707
c_{42}	0.4000	-0.0235	0.1446	-0.0124	0.0997	0.0029	0.0527
c_{52}	0.0000	-0.0466	0.1209	-0.0286	0.0727	-0.0048	0.0221
c_{13}	0.4000	0.0406	0.2692	0.0195	0.1771	-0.0019	0.0724
c_{23}	0.0000	0.0366	0.2403	0.0180	0.1607	-0.0014	0.0649
c_{33}	-0.6000	-0.0249	0.1962	-0.0138	0.1241	0.0009	0.0488
c_{43}	0.0000	0.0109	0.1291	0.0018	0.0799	0.0025	0.0303
c_{53}	0.5000	0.0181	0.1601	0.0058	0.1008	0.0019	0.0407
c_{14}	0.0000	0.0249	0.2786	0.0168	0.2102	-0.0026	0.0858
c_{24}	0.6000	0.0121	0.2513	0.0098	0.1884	-0.0034	0.0744
c_{34}	0.0000	-0.0166	0.1948	-0.0135	0.1372	0.0014	0.0557
c_{44}	-0.5000	0.0142	0.1195	0.0021	0.0797	0.0019	0.0331
c_{54}	0.0000	0.0035	0.1670	0.0010	0.1201	-0.0006	0.0458
c_{15}	0.0000	-0.0118	0.3379	-0.0172	0.2243	-0.0127	0.0855
c_{25}	0.0000	-0.0308	0.2857	-0.0200	0.1880	-0.0115	0.0689
c_{35}	0.0000	-0.0247	0.2704	-0.0086	0.1861	0.0074	0.0730
c_{45}	0.0000	-0.0215	0.1731	-0.0110	0.1176	0.0034	0.0484
c_{55}	0.8000	-0.0571	0.1578	-0.0315	0.0924	-0.0058	0.0277
ϕ_{11}	0.9000	-0.1230	0.3577	-0.0291	0.1872	-0.0058	0.0498
ϕ_{21}	0.0000	-0.0158	0.6187	-0.0559	0.4155	-0.0226	0.1463
ϕ_{12}	0.0000	0.0506	0.5055	0.0065	0.3450	-0.0224	0.1635
ϕ_{22}	-0.9000	0.5260	0.6863	0.2929	0.5566	0.0156	0.1020

Table 10: Bias and standard deviation for our estimator in DGP 4 ($n = 5$, $K = 2$, $\phi_0 = 0.9$, $\sigma_{v,0} = 0.6$, Gaussian noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_{11}	-1.5000	-0.0630	0.4274	-0.0271	0.2465	-0.0042	0.0941
b_{21}	-1.2000	-0.0505	0.3910	-0.0221	0.2247	-0.0033	0.0846
b_{31}	1.3000	0.0640	0.3211	0.0279	0.1871	0.0004	0.0670
b_{12}	1.6000	0.0171	0.3039	0.0192	0.1873	0.0037	0.0758
b_{22}	1.5000	0.0189	0.2685	0.0206	0.1661	0.0049	0.0674
b_{32}	-0.8000	-0.0122	0.2392	-0.0172	0.1415	-0.0027	0.0572
c_{11}	0.8000	-0.0647	0.2670	-0.0664	0.2013	-0.0266	0.0850
c_{21}	0.0000	-0.0486	0.2271	-0.0545	0.1721	-0.0220	0.0698
c_{31}	0.2000	0.0164	0.2020	0.0284	0.1442	0.0197	0.0704
c_{41}	0.0000	-0.0103	0.1351	0.0073	0.1056	0.0132	0.0520
c_{51}	0.0000	-0.0408	0.1125	-0.0289	0.0755	-0.0048	0.0249
c_{12}	0.0000	-0.1184	0.3299	-0.0701	0.2342	-0.0286	0.1188
c_{22}	0.8000	-0.1235	0.2812	-0.0718	0.2009	-0.0242	0.0956
c_{32}	0.0000	0.0580	0.2410	0.0327	0.1801	0.0208	0.0971
c_{42}	0.4000	0.0055	0.1592	0.0044	0.1301	0.0129	0.0709
c_{52}	0.0000	-0.0674	0.1350	-0.0408	0.0966	-0.0060	0.0307
c_{13}	0.4000	-0.0831	0.3157	-0.0352	0.2397	-0.0179	0.1024
c_{23}	0.0000	-0.0646	0.2729	-0.0323	0.2174	-0.0141	0.0919
c_{33}	-0.6000	0.0718	0.2346	0.0276	0.1687	0.0107	0.0706
c_{43}	0.0000	0.0653	0.1615	0.0289	0.1127	0.0080	0.0421
c_{53}	0.5000	-0.0045	0.1507	-0.0069	0.1264	-0.0017	0.0559
c_{14}	0.0000	-0.0914	0.3107	-0.0359	0.2499	-0.0176	0.1189
c_{24}	0.6000	-0.0845	0.2760	-0.0377	0.2290	-0.0151	0.1048
c_{34}	0.0000	0.0744	0.2161	0.0327	0.1679	0.0114	0.0776
c_{44}	-0.5000	0.0610	0.1302	0.0298	0.1016	0.0068	0.0434
c_{54}	0.0000	-0.0154	0.1685	-0.0046	0.1445	-0.0031	0.0640
c_{15}	0.0000	-0.1226	0.4548	-0.0563	0.3014	-0.0292	0.1263
c_{25}	0.0000	-0.1212	0.3874	-0.0603	0.2547	-0.0252	0.1021
c_{35}	0.0000	0.0531	0.3487	0.0181	0.2388	0.0201	0.1069
c_{45}	0.0000	0.0194	0.2404	0.0054	0.1701	0.0116	0.0732
c_{55}	0.8000	-0.0903	0.1721	-0.0504	0.1167	-0.0082	0.0344
ϕ_{11}	0.7000	-0.1774	0.4976	-0.0079	0.2865	0.0446	0.0975
ϕ_{21}	0.0000	-0.0632	0.5958	-0.0443	0.4417	-0.0167	0.1779
ϕ_{12}	0.0000	-0.0061	0.5396	-0.0408	0.4203	-0.0781	0.2323
ϕ_{22}	-0.7000	0.5471	0.6191	0.3239	0.5297	0.0351	0.1296

Table 11: Bias and standard deviation for our estimator in DGP 5 ($n = 5$, $K = 2$, $\phi_0 = 0.7$, $\sigma_{v,0} = 0.6$, Gaussian noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_{11}	-1.5000	-0.1108	0.7914	-0.0488	0.3652	-0.0063	0.1393
b_{21}	-1.2000	-0.0909	0.7308	-0.0384	0.3371	-0.0046	0.1248
b_{31}	1.3000	0.1189	0.5583	0.0422	0.2788	-0.0011	0.0996
b_{12}	1.6000	0.0693	0.5472	0.0446	0.3134	0.0085	0.1238
b_{22}	1.5000	0.0630	0.4857	0.0443	0.2768	0.0112	0.1097
b_{32}	-0.8000	-0.0386	0.4125	-0.0355	0.2331	-0.0049	0.0941
c_{11}	0.8000	-0.0679	0.1997	-0.0504	0.1436	-0.0129	0.0522
c_{21}	0.0000	-0.0494	0.1719	-0.0383	0.1175	-0.0106	0.0422
c_{31}	0.2000	0.0195	0.1565	0.0184	0.1070	0.0088	0.0423
c_{41}	0.0000	0.0044	0.1221	0.0118	0.0801	0.0068	0.0332
c_{51}	0.0000	-0.0213	0.0956	-0.0150	0.0597	-0.0022	0.0180
c_{12}	0.0000	-0.0782	0.2475	-0.0489	0.1721	-0.0124	0.0677
c_{22}	0.8000	-0.0899	0.2173	-0.0501	0.1461	-0.0104	0.0540
c_{32}	0.0000	0.0366	0.1979	0.0183	0.1360	0.0085	0.0562
c_{42}	0.4000	0.0068	0.1373	0.0077	0.0946	0.0057	0.0426
c_{52}	0.0000	-0.0383	0.1174	-0.0227	0.0727	-0.0027	0.0214
c_{13}	0.4000	-0.0423	0.2181	-0.0191	0.1598	-0.0086	0.0654
c_{23}	0.0000	-0.0326	0.1947	-0.0180	0.1498	-0.0071	0.0587
c_{33}	-0.6000	0.0305	0.1742	0.0110	0.1199	0.0045	0.0452
c_{43}	0.0000	0.0315	0.1290	0.0112	0.0846	0.0036	0.0292
c_{53}	0.5000	-0.0104	0.1350	-0.0098	0.0987	-0.0008	0.0395
c_{14}	0.0000	-0.0442	0.2228	-0.0166	0.1758	-0.0073	0.0751
c_{24}	0.6000	-0.0464	0.2026	-0.0201	0.1635	-0.0070	0.0660
c_{34}	0.0000	0.0332	0.1645	0.0109	0.1249	0.0041	0.0495
c_{44}	-0.5000	0.0325	0.1147	0.0135	0.0830	0.0030	0.0308
c_{54}	0.0000	-0.0165	0.1418	-0.0079	0.1125	-0.0016	0.0433
c_{15}	0.0000	-0.0542	0.3052	-0.0340	0.1906	-0.0144	0.0747
c_{25}	0.0000	-0.0580	0.2576	-0.0361	0.1618	-0.0127	0.0603
c_{35}	0.0000	0.0073	0.2615	0.0053	0.1610	0.0095	0.0645
c_{45}	0.0000	0.0104	0.1753	0.0056	0.1109	0.0060	0.0448
c_{55}	0.8000	-0.0549	0.1409	-0.0305	0.0898	-0.0043	0.0259
ϕ_{11}	0.7000	-0.0982	0.4403	0.0172	0.2102	0.0178	0.0717
ϕ_{21}	0.0000	0.0228	0.6230	0.0116	0.3799	-0.0089	0.1484
ϕ_{12}	0.0000	-0.0602	0.5510	-0.0798	0.3708	-0.0438	0.1758
ϕ_{22}	-0.7000	0.4108	0.5974	0.2156	0.4617	0.0162	0.0800

Table 12: Bias and standard deviation for our estimator in DGP 6 ($n = 5$, $K = 2$, $\phi_0 = 0.7$, $\sigma_{v,0} = 0.6$, Gaussian noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_{11}	-1.5000	-2.6187	65.8725	-1.2791	25.5578	-0.0324	0.3360
b_{21}	-1.2000	-2.0605	51.0188	-1.7170	42.1275	-0.0285	0.3036
b_{31}	1.3000	1.5312	39.5820	1.9711	45.6743	0.0111	0.2436
b_{12}	1.6000	-2.7543	98.6036	-0.4349	16.6540	0.0308	0.2712
b_{22}	1.5000	-2.0407	76.4933	-0.8691	27.2747	0.0332	0.2405
b_{32}	-0.8000	1.7117	59.2185	0.9694	29.6995	-0.0179	0.2013
c_{11}	0.8000	-0.0823	0.2498	-0.0682	0.1801	-0.0126	0.0535
c_{21}	0.0000	-0.0493	0.2084	-0.0486	0.1477	-0.0099	0.0425
c_{31}	0.2000	0.0337	0.2238	0.0204	0.1228	0.0068	0.0407
c_{41}	0.0000	0.0086	0.1770	0.0132	0.0951	0.0070	0.0348
c_{51}	0.0000	-0.0169	0.1466	-0.0146	0.0884	-0.0025	0.0228
c_{12}	0.0000	-0.0853	0.2793	-0.0538	0.1943	-0.0115	0.0656
c_{22}	0.8000	-0.0985	0.2288	-0.0668	0.1742	-0.0103	0.0538
c_{32}	0.0000	0.0554	0.2239	0.0258	0.1498	0.0063	0.0546
c_{42}	0.4000	0.0079	0.1603	0.0084	0.1131	0.0057	0.0415
c_{52}	0.0000	-0.0260	0.1599	-0.0236	0.1099	-0.0033	0.0292
c_{13}	0.4000	-0.0994	0.2365	-0.0719	0.1702	-0.0162	0.0667
c_{23}	0.0000	-0.0736	0.1934	-0.0638	0.1511	-0.0125	0.0588
c_{33}	-0.6000	0.0673	0.1916	0.0537	0.1351	0.0073	0.0497
c_{43}	0.0000	0.0417	0.1541	0.0302	0.0922	0.0074	0.0359
c_{53}	0.5000	-0.0227	0.1518	-0.0222	0.0966	-0.0035	0.0445
c_{14}	0.0000	-0.0845	0.2148	-0.0550	0.1583	-0.0138	0.0694
c_{24}	0.6000	-0.0811	0.1873	-0.0579	0.1469	-0.0118	0.0599
c_{34}	0.0000	0.0622	0.1727	0.0446	0.1221	0.0072	0.0483
c_{44}	-0.5000	0.0439	0.1332	0.0312	0.0920	0.0063	0.0380
c_{54}	0.0000	-0.0254	0.1360	-0.0137	0.0989	-0.0024	0.0458
c_{15}	0.0000	-0.0850	0.2944	-0.0446	0.2082	-0.0145	0.0752
c_{25}	0.0000	-0.0783	0.2365	-0.0497	0.1784	-0.0132	0.0635
c_{35}	0.0000	0.0341	0.2503	0.0181	0.1680	0.0076	0.0669
c_{45}	0.0000	0.0219	0.1973	0.0181	0.1246	0.0076	0.0454
c_{55}	0.8000	-0.0573	0.1613	-0.0354	0.1090	-0.0057	0.0356
ϕ_{11}	0.4000	2.9343	94.0909	-0.1971	5.0240	0.0656	0.1747
ϕ_{21}	0.0000	-1.9473	61.8889	0.3173	8.0495	0.0193	0.2026
ϕ_{12}	0.0000	4.3236	143.7967	-0.1937	3.3216	-0.0552	0.2509
ϕ_{22}	-0.4000	-2.5941	94.1016	0.4482	5.2015	0.0296	0.2061

Table 13: Bias and standard deviation for our estimator in DGP 7 ($n = 5$, $K = 2$, $\phi_0 = 0.4$, $\sigma_{v,0} = 0.6$, Gaussian noises). The second column reports the true values of the parameters.

		$T=500$		$T=1000$		$T=5000$	
Param.	θ_0	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)
b_{11}	-1.5000	0.1375	1.4029	-0.0283	0.9184	0.0029	0.3051
b_{21}	-1.2000	0.1022	1.2850	-0.0326	0.8061	-0.0080	0.2706
b_{31}	1.3000	-0.0802	1.0326	0.0072	0.6914	-0.0184	0.2255
b_{12}	1.6000	-0.0101	1.0898	0.0113	0.6854	0.0081	0.2542
b_{22}	1.5000	-0.0070	0.9429	0.0233	0.6058	0.0059	0.2252
b_{32}	-0.8000	-0.0365	0.7987	-0.0273	0.4954	-0.0096	0.1838
c_{11}	0.8000	-0.0246	0.1662	-0.0240	0.1152	-0.0030	0.0401
c_{21}	0.0000	-0.0158	0.1359	-0.0145	0.0943	-0.0031	0.0333
c_{31}	0.2000	0.0003	0.1195	-0.0055	0.0836	0.0018	0.0309
c_{41}	0.0000	-0.0079	0.0913	-0.0012	0.0636	0.0021	0.0263
c_{51}	0.0000	-0.0119	0.0751	-0.0100	0.0535	-0.0022	0.0189
c_{12}	0.0000	-0.0447	0.1660	-0.0209	0.1126	-0.0046	0.0495
c_{22}	0.8000	-0.0480	0.1441	-0.0265	0.1101	-0.0037	0.0419
c_{32}	0.0000	0.0203	0.1264	0.0040	0.0899	0.0014	0.0418
c_{42}	0.4000	0.0003	0.0928	0.0012	0.0744	0.0017	0.0327
c_{52}	0.0000	-0.0189	0.0863	-0.0142	0.0563	-0.0025	0.0217
c_{13}	0.4000	-0.0892	0.1708	-0.0657	0.1281	-0.0144	0.0580
c_{23}	0.0000	-0.0660	0.1522	-0.0618	0.1159	-0.0124	0.0503
c_{43}	0.0000	0.0402	0.1043	0.0286	0.0724	0.0064	0.0289
c_{53}	0.5000	-0.0200	0.0976	-0.0141	0.0697	-0.0018	0.0329
c_{14}	0.0000	-0.0892	0.1585	-0.0468	0.1439	-0.0088	0.0576
c_{24}	0.6000	-0.0751	0.1422	-0.0482	0.1253	-0.0078	0.0530
c_{34}	0.0000	0.0638	0.1207	0.0413	0.1027	0.0051	0.0448
c_{44}	-0.5000	0.0373	0.0887	0.0274	0.0651	0.0047	0.0309
c_{54}	0.0000	-0.0233	0.1031	-0.0135	0.0775	-0.0010	0.0366
c_{15}	0.0000	-0.0680	0.1773	-0.0495	0.1355	-0.0144	0.0610
c_{25}	0.0000	-0.0642	0.1497	-0.0442	0.1173	-0.0125	0.0467
c_{35}	0.0000	0.0356	0.1310	0.0224	0.0995	0.0068	0.0473
c_{45}	0.0000	0.0250	0.1063	0.0201	0.0834	0.0047	0.0366
c_{55}	0.8000	-0.0324	0.1028	-0.0215	0.0654	-0.0036	0.0275
ϕ_{11}	0.4000	0.0096	0.6245	0.0671	0.4293	0.0503	0.1517
ϕ_{21}	0.0000	-0.0354	0.7005	-0.0012	0.5182	0.0277	0.2027
ϕ_{12}	0.0000	-0.0412	0.5356	-0.0523	0.4372	-0.0399	0.2004
ϕ_{22}	-0.4000	0.2939	0.8744	0.0927	0.5012	0.0002	0.1100

Table 14: Difference between the median and true value, and Interquartile Range (IQR) divided by twice the 75% standard Normal quantile $z_{75\%}$, for our estimator in DGP 7 ($n = 5$, $K = 2$, $\phi_0 = 0.4$, $\sigma_{v,0} = 0.6$, Gaussian noises). The second column reports the true values of the parameters. The quantity $IQR/(2z_{75\%})$ equals the standard deviation in a Gaussian $N(0, 1)$ distribution.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_{11}	-1.5000	-0.0402	0.2824	-0.0194	0.1685	-0.0021	0.0607
b_{21}	-1.2000	-0.0300	0.2524	-0.0140	0.1590	-0.0005	0.0554
b_{31}	1.3000	0.0436	0.2188	0.0208	0.1298	0.0004	0.0471
b_{12}	1.6000	0.0078	0.1928	0.0010	0.1144	-0.0007	0.0400
b_{22}	1.5000	0.0087	0.1681	0.0038	0.1042	0.0002	0.0362
b_{32}	-0.8000	-0.0071	0.1458	-0.0029	0.0881	0.0000	0.0297
c_{11}	0.8000	-0.0105	0.6686	-0.0685	0.6029	-0.0466	0.4284
c_{21}	0.0000	-0.0176	0.5780	-0.0670	0.5355	-0.0490	0.3863
c_{31}	0.2000	-0.0512	0.4903	0.0011	0.4167	0.0076	0.2743
c_{41}	0.0000	-0.0887	0.3233	-0.0535	0.2569	-0.0275	0.1562
c_{51}	0.0000	-0.0854	0.2580	-0.0893	0.2692	-0.0543	0.2226
c_{12}	0.0000	-0.1829	0.6892	-0.1233	0.6305	-0.0855	0.4248
c_{22}	0.8000	-0.1790	0.5873	-0.1251	0.5521	-0.0808	0.3714
c_{32}	0.0000	0.1194	0.5230	0.0658	0.4503	0.0419	0.2965
c_{42}	0.4000	0.0408	0.3588	0.0050	0.2942	-0.0032	0.1889
c_{52}	0.0000	-0.0825	0.2447	-0.0760	0.2650	-0.0574	0.1903
c_{13}	0.4000	-0.0784	0.6358	-0.0666	0.5244	-0.0358	0.3631
c_{23}	0.0000	-0.0560	0.5528	-0.0469	0.4588	-0.0243	0.3189
c_{33}	-0.6000	0.0940	0.4558	0.0732	0.3798	0.0394	0.2597
c_{43}	0.0000	0.1024	0.2954	0.0736	0.2417	0.0437	0.1699
c_{53}	0.5000	0.0329	0.2683	0.0191	0.2146	0.0182	0.1628
c_{14}	0.0000	-0.1491	0.8655	-0.0990	0.5411	-0.0391	0.3676
c_{24}	0.6000	-0.1318	0.7881	-0.0850	0.4808	-0.0302	0.3193
c_{34}	0.0000	0.1319	0.5172	0.0976	0.3610	0.0414	0.2564
c_{44}	-0.5000	0.1022	0.2454	0.0807	0.2069	0.0385	0.1664
c_{54}	0.0000	-0.0124	0.4666	0.0042	0.2574	0.0113	0.1661
c_{15}	0.0000	-0.3164	0.9029	-0.2446	1.0087	-0.1318	0.8330
c_{25}	0.0000	-0.2906	0.7769	-0.2229	0.8886	-0.1155	0.7589
c_{35}	0.0000	0.2038	0.6634	0.1535	0.6901	0.0852	0.5038
c_{45}	0.0000	0.0969	0.4443	0.0659	0.4169	0.0351	0.2561
c_{55}	0.8000	-0.1434	0.3529	-0.1086	0.4665	-0.0516	0.4582
ϕ_{11}	0.9000	-0.3410	0.5747	-0.2749	0.5382	-0.1077	0.3335
ϕ_{21}	0.0000	-0.1586	1.8350	-0.2519	2.1722	-0.1702	0.9468
ϕ_{12}	0.0000	0.0659	1.1689	0.1434	1.8705	0.1243	1.1423
ϕ_{22}	-0.9000	0.8966	0.6990	0.8258	0.7262	0.5325	0.6585

Table 15: Bias and standard deviation for our estimator in DGP 8 ($n = 5$, $K = 2$, ARCH noises). The second column reports the true values of the parameters.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)	med($\hat{\theta}$)-θ_0	IQR/($2z_{75\%}$)
b_{11}	-1.5000	-0.0222	0.2401	-0.0112	0.1456	-0.0015	0.0548
b_{21}	-1.2000	-0.0275	0.2134	-0.0140	0.1416	0.0007	0.0499
b_{31}	1.3000	0.0272	0.1961	0.0174	0.1201	-0.0035	0.0424
b_{12}	1.6000	0.0007	0.1658	0.0025	0.1011	-0.0007	0.0361
b_{22}	1.5000	0.0070	0.1545	-0.0012	0.0964	0.0015	0.0316
b_{32}	-0.8000	-0.0090	0.1245	-0.0034	0.0802	-0.0005	0.0263
c_{11}	0.8000	0.0491	0.3561	0.0116	0.3064	-0.0078	0.2156
c_{21}	0.0000	0.0338	0.2979	0.0098	0.2582	-0.0115	0.1805
c_{31}	0.2000	-0.0712	0.2593	-0.0311	0.2272	-0.0045	0.1620
c_{41}	0.0000	-0.0813	0.1471	-0.0577	0.1323	-0.0266	0.1022
c_{51}	0.0000	-0.0405	0.1077	-0.0353	0.0872	-0.0184	0.0522
c_{12}	0.0000	-0.1191	0.3948	-0.0723	0.3655	-0.0295	0.2717
c_{22}	0.8000	-0.1201	0.3562	-0.0725	0.3179	-0.0321	0.2249
c_{32}	0.0000	0.0680	0.2958	0.0340	0.2513	0.0092	0.2058
c_{42}	0.4000	0.0085	0.1643	-0.0097	0.1458	-0.0127	0.1249
c_{52}	0.0000	-0.0624	0.1396	-0.0472	0.1126	-0.0233	0.0695
c_{13}	0.4000	-0.0983	0.4252	-0.0785	0.3727	-0.0309	0.2544
c_{23}	0.0000	-0.0786	0.3529	-0.0637	0.3341	-0.0274	0.2199
c_{33}	-0.6000	0.1024	0.2984	0.0653	0.2595	0.0269	0.1702
c_{43}	0.0000	0.0814	0.1819	0.0581	0.1396	0.0238	0.0923
c_{53}	0.5000	0.0254	0.1665	0.0188	0.1597	0.0079	0.1038
c_{14}	0.0000	-0.1423	0.4545	-0.0829	0.4369	-0.0362	0.2828
c_{24}	0.6000	-0.1338	0.3955	-0.0693	0.3554	-0.0281	0.2557
c_{34}	0.0000	0.1165	0.2990	0.0703	0.2670	0.0308	0.1830
c_{44}	-0.5000	0.0847	0.1563	0.0565	0.1296	0.0166	0.0846
c_{54}	0.0000	-0.0214	0.1998	-0.0054	0.1851	0.0093	0.1252
c_{15}	0.0000	-0.2894	0.5842	-0.2127	0.5090	-0.1077	0.3178
c_{25}	0.0000	-0.2586	0.5108	-0.1838	0.4436	-0.0962	0.2631
c_{35}	0.0000	0.1957	0.4347	0.1382	0.3747	0.0703	0.2430
c_{45}	0.0000	0.0936	0.2480	0.0580	0.2186	0.0291	0.1546
c_{55}	0.8000	-0.1006	0.1910	-0.0755	0.1635	-0.0265	0.0882
ϕ_{11}	0.9000	-0.1596	0.3858	-0.1118	0.2690	-0.0355	0.1127
ϕ_{21}	0.0000	-0.1613	0.5686	-0.1731	0.5803	-0.1415	0.4241
ϕ_{12}	0.0000	0.0046	0.4973	0.0028	0.4808	0.0055	0.3031
ϕ_{22}	-0.9000	0.7900	0.9401	0.5671	1.0108	0.2207	0.6004

Table 16: Difference between the median and true value, and Interquartile Range (IQR) divided by twice the 75% standard Normal quantile $z_{75\%}$, for our estimator in DGP 8 ($n = 5$, $K = 2$, ARCH noises). The second column reports the true values of the parameters. The quantity $IQR/(2z_{75\%})$ equals the standard deviation in a Gaussian $N(0,1)$ distribution.

		$T = 500$		$T = 1000$		$T = 5000$	
Param.	θ_0	Bias	Std.	Bias	Std.	Bias	Std.
b_{11}	0.5000	0.0214	0.1850	0.0120	0.1144	0.0013	0.0444
b_{21}	0.5000	0.0115	0.1847	0.0050	0.1173	-0.0013	0.0456
b_{31}	0.5000	0.0187	0.1775	0.0124	0.1140	0.0019	0.0457
$c_{1,11}$	-0.2158	0.0338	0.1073	0.0171	0.0816	0.0039	0.0427
$c_{1,21}$	-0.3768	0.0390	0.1068	0.0252	0.0773	0.0067	0.0394
$c_{1,31}$	0.3280	0.0392	0.1066	0.0211	0.0803	0.0065	0.0390
$c_{1,41}$	-0.4669	0.0851	0.1590	0.0534	0.1190	0.0139	0.0665
$c_{1,12}$	0.2860	0.0478	0.1011	0.0294	0.0754	0.0087	0.0381
$c_{1,22}$	-0.6122	0.0438	0.1045	0.0276	0.0745	0.0077	0.0401
$c_{1,32}$	0.2537	0.0503	0.0949	0.0306	0.0673	0.0097	0.0336
$c_{1,42}$	0.2445	0.1044	0.1506	0.0690	0.1087	0.0204	0.0571
$c_{1,13}$	0.3860	0.0721	0.0980	0.0384	0.0740	0.0120	0.0358
$c_{1,23}$	-0.0122	0.0697	0.0957	0.0417	0.0699	0.0116	0.0336
$c_{1,33}$	0.4537	0.0614	0.1048	0.0308	0.0762	0.0088	0.0360
$c_{1,43}$	0.3445	0.1418	0.1461	0.0833	0.1064	0.0245	0.0477
$c_{1,14}$	0.4719	0.1087	0.1176	0.0683	0.0968	0.0202	0.0527
$c_{1,24}$	-0.0244	0.1088	0.1150	0.0715	0.0944	0.0208	0.0494
$c_{1,34}$	-0.6927	0.1118	0.1127	0.0706	0.0875	0.0211	0.0422
$c_{1,44}$	0.2890	0.2097	0.1840	0.1376	0.1496	0.0402	0.0762
$c_{2,11}$	-0.3000	0.0748	0.1100	0.0461	0.0849	0.0123	0.0451
$c_{2,21}$	0.0000	0.0730	0.1089	0.0473	0.0819	0.0141	0.0432
$c_{2,31}$	0.1000	0.0793	0.0973	0.0493	0.0759	0.0146	0.0366
$c_{2,41}$	0.0000	0.1530	0.1577	0.1002	0.1214	0.0276	0.0649
$c_{2,12}$	0.0000	-0.0377	0.0830	-0.0188	0.0598	-0.0051	0.0282
$c_{2,22}$	-0.3000	-0.0435	0.0851	-0.0264	0.0607	-0.0073	0.0267
$c_{2,32}$	0.0000	-0.0378	0.0896	-0.0190	0.0618	-0.0048	0.0279
$c_{2,42}$	0.2000	-0.0722	0.1253	-0.0406	0.0883	-0.0107	0.0443
$c_{2,13}$	0.1000	-0.1238	0.1517	-0.0741	0.1180	-0.0210	0.0592
$c_{2,23}$	0.0000	-0.1214	0.1472	-0.0783	0.1155	-0.0220	0.0546
$c_{2,33}$	0.3000	-0.1255	0.1501	-0.0749	0.1102	-0.0217	0.0488
$c_{2,43}$	0.0000	-0.2379	0.2441	-0.1517	0.1872	-0.0436	0.0868
$c_{2,14}$	0.0000	0.0039	0.1007	0.0022	0.0737	0.0010	0.0398
$c_{2,24}$	-0.2000	-0.0026	0.1040	-0.0041	0.0724	-0.0016	0.0394
$c_{2,34}$	0.0000	-0.0076	0.0990	-0.0065	0.0736	-0.0035	0.0386
$c_{2,44}$	-0.4000	-0.0042	0.1530	-0.0079	0.1151	-0.0023	0.0685
ϕ	0.8000	-0.1385	0.2651	-0.0142	0.1271	0.0096	0.0370

Table 17: Bias and standard deviation for our estimator in DGP 9 ($n = 4$, $K = 1$, $p = 2$, $\phi_0 = 0.8$, $\sigma_{v,0} = 1$, Gaussian noises). The second column reports the true values of the parameters.

G.2 Size and power of the test on the number of latent factors

	Nominal size:	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 0.5\%$
DGP 3	$T = 500$	0.1330	0.0700	0.0160	0.0040
	$T = 1000$	0.1270	0.0650	0.0110	0.0080
	$T = 5000$	0.0940	0.0420	0.0130	0.0040
DGP 4	$T = 500$	0.1000	0.0530	0.0090	0.0010
	$T = 1000$	0.1070	0.0570	0.0120	0.0040
	$T = 5000$	0.0950	0.0370	0.0140	0.0060
DGP 5	$T = 500$	0.0850	0.0450	0.0080	0.0010
	$T = 1000$	0.1000	0.0570	0.0090	0.0040
	$T = 5000$	0.0920	0.0430	0.0120	0.0040
DGP 6	$T = 500$	0.1040	0.0550	0.0130	0.0030
	$T = 1000$	0.1070	0.0530	0.0130	0.0050
	$T = 5000$	0.0990	0.0470	0.0110	0.0040
DGP 7	$T = 500$	0.0350	0.0110	0.0000	0.0000
	$T = 1000$	0.0680	0.0290	0.0010	0.0000
	$T = 5000$	0.0980	0.0400	0.0070	0.0030
DGP 8	$T = 500$	0.0990	0.0510	0.0170	0.0060
	$T = 1000$	0.1150	0.0530	0.0150	0.0040
	$T = 5000$	0.0970	0.0470	0.0090	0.0040

Table 18: Empirical size of the test for the null hypothesis $\mathcal{H}_0 : K = 2$ against the alternative $\mathcal{H}_a : K > 2$, in DGPs 3-8, for different nominal sizes. The number of Monte Carlo replications is $N_{rep} = 1000$.

	Nominal size:	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 0.5\%$
DGP 3	$T = 500$	0.9850	0.9560	0.7840	0.6830
	$T = 1000$	1.0000	1.0000	1.0000	0.9990
	$T = 5000$	1.0000	1.0000	1.0000	1.0000
DGP 4	$T = 500$	0.8230	0.6920	0.3800	0.2900
	$T = 1000$	0.9940	0.9830	0.9110	0.8430
	$T = 5000$	1.0000	1.0000	1.0000	1.0000
DGP 5	$T = 500$	0.7110	0.5590	0.2450	0.1670
	$T = 1000$	0.9840	0.9530	0.8000	0.6820
	$T = 5000$	1.0000	1.0000	1.0000	1.0000
DGP 6	$T = 500$	0.8940	0.7650	0.4460	0.3070
	$T = 1000$	1.0000	0.9960	0.9670	0.9240
	$T = 5000$	1.0000	1.0000	1.0000	1.0000
DGP 7	$T = 500$	0.2370	0.1140	0.0220	0.0100
	$T = 1000$	0.4700	0.3090	0.1020	0.0580
	$T = 5000$	1.0000	1.0000	0.9950	0.9890
DGP 8	$T = 500$	0.8700	0.7360	0.3650	0.2460
	$T = 1000$	0.9970	0.9900	0.9230	0.8500
	$T = 5000$	1.0000	0.9990	0.9970	0.9970

Table 19: Empirical power of the test for null hypothesis $\mathcal{H}_0 : K = 1$ against the alternative $\mathcal{H}_a : K > 1$, in DGPs 3-8, for different nominal sizes. The number of Monte Carlo replications is $N_{rep} = 1000$.

G.3 Selection of model orders K, p, q

$\alpha = 5\%$					
(p, K)	(1,1)	(1,2)	(2,1)	(2,2)	others
$T = 500$	0.6550	0.0700	0.2520	0.0110	0.0120
$T = 1000$	0.8870	0.0420	0.0640	0.0020	0.0050
$T = 5000$	0.9310	0.0150	0.0540	0.0000	0.0000
$\alpha = 1\%$					
(p, K)	(1,1)	(1,2)	(2,1)	(2,2)	others
$T = 500$	0.8210	0.0280	0.1240	0.0020	0.0250
$T = 1000$	0.9310	0.0150	0.0540	0.0000	0.0000
$T = 5000$	0.9680	0.0140	0.0180	0.0000	0.0000
$\alpha = 0.5\%$					
(p, K)	(1,1)	(1,2)	(2,1)	(2,2)	others
$T = 500$	0.8550	0.0160	0.0930	0.0010	0.0350
$T = 1000$	0.9530	0.0080	0.0390	0.0000	0.0000
$T = 5000$	0.9830	0.0050	0.0120	0.0000	0.0000
$\alpha = 0.1\%$					
(p, K)	(1,1)	(1,2)	(2,1)	(2,2)	others
$T = 500$	0.6880	0.0000	0.0050	0.0000	0.3070
$T = 1000$	0.9960	0.0000	0.0000	0.0000	0.0040
$T = 5000$	0.9970	0.0000	0.0000	0.0000	0.0030

Table 20: The table reports the percentage of times that model (p, K) is selected in 1000 Monte Carlo replications. Orders (p, K) are selected with the testing procedure described in Section 5, with $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$. Data are simulated according to DGP 1 ($K = p = q = 1$), with $T = 500, 1000, 5000$.

(p, K, q)	$\alpha = 0.01$ Percentage
$(2, 0, 0)$	4%
$(4, 0, 0)$	2%
$(1, 1, 1)$	66%
$(1, 1, 2)$	10%
$(1, 1, 3)$	12%
$(2, 1, 1)$	1%
$(2, 2, 1)$	2%
$(1, 2, 1)$	1%
$(1, 2, 2)$	0%
otherwise	0%

Table 21: The table reports the percentage of times that model (p, K, q) is selected in 1000 Monte Carlo replications. Orders (p, K) are selected with the testing procedure described in Section 5, with $\alpha = 0.01$. Order q is selected with BIC. Data are simulated according to DGP 1 ($K = p = q = 1$), with $T = 1000$.

References

Anderson, B. D. (1969). The inverse problem of stationary covariance generation. Journal of Statistical Physics 1.1, pages 133-147.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, Volume 5, pages 111-150.

Bartlett, M. (1938). The characteristic function of a conditional statistic. *Journal of the London Mathematical Society*, 1(1):62-67.

De Jong, R. and Davidson, J. (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. Econometrica, 68, 407-424.

De Schutter, B. (2000). Minimal state-space realization in linear system theory: an overview. *Journal of computational and applied mathematics*, 121(1-2):331-354.

Dilip, A. S. A. and Pillai, H. K. (2016). Characterization of solutions of non-symmetric algebraic Riccati equations. *Linear Algebra and its Applications*, 507:356-372.

Hannan, E. J. and Deistler, M. (1988). *The statistical theory of linear systems*, Volume 70. SIAM.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor analysis as a statistical method*. Butterworths.

Lutkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer.

Mc Leish, D. L. (1975). A maximal inequality and dependent strong laws. *The Annals of Probability*, 3(5): 829-839.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and their Applications*, 29(2):309-315.

Stock, J. and Watson, M. (2016). Factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. Prepared for the *Handbook of Macroeconomics*, Volume 2A, Chapter 8.

Tao, T. (2012). *Topics in Random Matrix Theory*. Graduate Studies in Mathematics.

Williams, B. (2020). Identification of the linear factor model. Econometric Reviews, 39(1):92-109.