ONLINE APPENDIX

Latent Factor Analysis in Short Panels

Alain-Philippe Fortin, Patrick Gagliardini, and Olivier Scaillet

We give proofs of Propositions 1-4 of the paper in Section B. We gather the results of our Monte Carlo experiments in Section C. We prove Lemmas 1-4 of the paper in Section D. We provide additional theory in Appendix E, namely the characterization of the pseudo likelihood and the PML estimator (E.1), the conditions for global identification and consistency (E.2), the asymptotic expansions for the FA estimators (E.3), the local analysis of the first-order conditions of FA estimators (E.4), the asymptotic normality of FA estimators (E.5), the definition of invariant tests (E.6), and proofs of additional lemmas (E.7). We give numerical checks of Inequalities (6) of Proposition 4 in Appendix F. Finally, we collect the maximum value of k as a function of T in Appendix G.

B Proofs of Propositions 1-4

Proof of Proposition 1: (a) The proof of this part is made in three steps. (i) We first establish the link between the LR statistic and the norm of matrix $\hat{S} = \hat{V}_\varepsilon^{-1/2} M_{\hat{F},\hat{V}_\varepsilon}(\hat{V}_y - \hat{V}_\varepsilon) M'_{\hat{F},\hat{V}_\varepsilon} \hat{V}_\varepsilon^{-1/2}$, namely we prove $LR(k) = \frac{n}{2}\|\hat{S}\|^2 + o_p(1)$. The next lemma is instrumental to this step.

Lemma 1 *Under Assumption 1, (a) the eigenvalues of matrix $\hat{S}$ are: $\hat{\gamma}_j$, for $j = k+1, ..., T$, and 0, with multiplicity k, where $1 + \hat{\gamma}_j$ for $j = k+1, ..., T$ are the $T-k$ smallest eigenvalues of $\hat{V}_y \hat{V}_\varepsilon^{-1}$, (b) the squared Frobenius norm is $\|\hat{S}\|^2 = \sum_{j=k+1}^T \hat{\gamma}_j^2$, and (c) $diag(\hat{S}) = 0$.*

Then, we apply a second-order expansion of the log function in the RHS of (2). The first-order term vanishes because $\sum_{j=k+1}^T \hat{\gamma}_j = tr(\hat{S}) = 0$ by Lemma 1 a) and c). The second-order term equals $\frac{n}{2}\|\hat{S}\|^2$ by Lemma 1 b). The remainder (third-order) term is $o_p(1)$ because we have $\sqrt{n}\hat{\gamma}_j = O_p(1)$

for $j = k+1, ..., T$. This bound results from the expansion of the sample covariance:

$$\hat{V}_y = \tilde{V}_y + \frac{1}{\sqrt{n}}\Psi_y + o_p(\frac{1}{\sqrt{n}}) = \tilde{V}_y + O_p(\frac{1}{\sqrt{n}}), \tag{B.1}$$

where $\tilde{V}_y := FF' + \tilde{V}_\varepsilon$ and $\Psi_y := \frac{1}{\sqrt{n}}(\varepsilon\beta F' + F\beta'\varepsilon') + \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right)$, see Equation (E.2) and Lemma 6 in Appendix E.2, and $\hat{V}_\varepsilon = \tilde{V}_\varepsilon + O_p(\frac{1}{\sqrt{n}})$, see Equation (E.20). Then, $\hat{V}_y\hat{V}_\varepsilon^{-1} = \tilde{V}_y\tilde{V}_\varepsilon^{-1} + O_p(\frac{1}{\sqrt{n}})$, matrix $\tilde{V}_y\tilde{V}_\varepsilon^{-1}$ has unit eigenvalues for order $j = k+1, ..., T$, and the eigenvalues of matrices $\hat{V}_y\hat{V}_\varepsilon^{-1}$ and $\tilde{V}_y\tilde{V}_\varepsilon^{-1}$ differ by quantities of order $O_p(\frac{1}{\sqrt{n}})$ by Weyl's inequalities.

(ii) Next, let us establish the asymptotic expansion of $n\|\hat{S}\|^2$ in order to show equation (3). Since $\hat{G}\hat{G}'\hat{V}_\varepsilon^{-1} = M_{\hat{F},\hat{V}_\varepsilon}$, we have $\hat{S} = \hat{V}_\varepsilon^{-1/2}\hat{G}\hat{S}^*\hat{G}'\hat{V}_\varepsilon^{-1/2}$, where $\hat{S}^* = \hat{G}'\hat{V}_\varepsilon^{-1}(\hat{V}_y - \hat{V}_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}$. Besides, we have $0 = diag(\hat{S})$ (see Lemma 1 (c)). Therefore, $0 = diag(\hat{S}) = \hat{V}_\varepsilon^{-1}diag(\hat{G}\hat{S}^*\hat{G}') = 2\hat{V}_\varepsilon^{-1}\hat{\boldsymbol{X}}'vech(\hat{S}^*)$, i.e., $vech(\hat{S}^*)$ is in the orthogonal complement of the range of $\hat{\boldsymbol{X}}$. [40] It follows from the local identification assumption A.5 that $M_{\hat{\boldsymbol{X}}}$ is well-defined and thus $vech(\hat{S}^*) = M_{\hat{\boldsymbol{X}}}vech(\hat{S}^*)$.[41] Next, we have

$$M_{\hat{\boldsymbol{X}}}vech(\hat{S}^*) = M_{\hat{\boldsymbol{X}}}vech(\hat{G}'\hat{V}_\varepsilon^{-1}(\hat{V}_y - \hat{V}_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}) = M_{\hat{\boldsymbol{X}}}vech(\hat{G}'\hat{V}_\varepsilon^{-1}(\hat{V}_y - \tilde{V}_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}), \tag{B.2}$$

because the kernel of $M_{\hat{\boldsymbol{X}}}$ is $\{vech(\hat{G}'D\hat{G}) : D \text{ diagonal}\}$. Besides, we have the expansion $\sqrt{n}vech(\hat{G}'\hat{V}_\varepsilon^{-1}(\hat{V}_y - \tilde{V}_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}) = vech(\hat{Z}_n^*) + o_p(1)$, where $\hat{Z}_n^* = \hat{G}'\hat{V}_\varepsilon^{-1}Z_n\hat{V}_\varepsilon^{-1}\hat{G}$. It is because expansion (B.1) and $\hat{G}'\hat{V}_\varepsilon^{-1}F = \hat{G}'\hat{V}_\varepsilon^{-1}M_{\hat{F},\hat{V}_\varepsilon}F = O_p(\frac{1}{\sqrt{n}})$ by the root-n consistency of FA estimators (see Appendix E.5.1). Using $\|\hat{S}\|^2 = \|\hat{S}^*\|^2$, it follows that

$$\frac{n}{2}\|\hat{S}\|^2 = nvech(\hat{S}^*)'vech(\hat{S}^*) = nvech(\hat{S}^*)'M_{\hat{\boldsymbol{X}}}vech(\hat{S}^*) = vech(\hat{Z}_n^*)'M_{\hat{\boldsymbol{X}}}vech(\hat{Z}_n^*) + o_p(1). \tag{B.3}$$

[40] To see this step, write $\hat{G} = (\hat{g}_{t,i}) = [\hat{g}_1 : \cdots : \hat{g}_{T-k}]$. By definition of the $vech$ operator, $vech(\hat{G}'E_{t,t}\hat{G}) = \left[\frac{1}{\sqrt{2}}\hat{g}_{t,1}^2 : \cdots : \frac{1}{\sqrt{2}}\hat{g}_{t,T-k}^2 : \{\hat{g}_{t,i}\hat{g}_{t,j}\}_{i<j}\right]'$. Therefore, $\hat{\boldsymbol{X}}' = [\frac{1}{\sqrt{2}}\hat{g}_1 \odot \hat{g}_1 : \cdots : \frac{1}{\sqrt{2}}\hat{g}_{T-k} \odot \hat{g}_{T-k} : \{\hat{g}_i \odot \hat{g}_j\}_{i<j}]$. Thus, for any $(T-k)\times(T-k)$ symmetric matrix $A = (a_{i,j})$, $diag(\hat{G}A\hat{G}') = \sum_{i=1}^{T-k} a_{i,i}diag(\hat{g}_i\hat{g}_i') + 2\sum_{i<j} a_{i,j}diag(\hat{g}_i\hat{g}_j') = \sum_{i=1}^{T-k} a_{i,i}(\hat{g}_i \odot \hat{g}_i) + 2\sum_{i<j} a_{i,j}(\hat{g}_i \odot \hat{g}_j) = 2\hat{\boldsymbol{X}}'vech(A)$.

[41] Assumption A.5 is equivalent to $\boldsymbol{X}$ having full column rank by Lemma 7 in Appendix E.4. Besides, from Proposition 8 in Appendix E.6 and the fact that $\hat{G}\hat{O} = G + o_p(1)$ for some rotation matrix $\hat{O}$ (see below), we have $\hat{\mathscr{R}}\hat{\boldsymbol{X}} = \boldsymbol{X} + o_p(1)$ from some orthogonal matrix $\hat{\mathscr{R}}$. Hence, $\hat{\boldsymbol{X}}$ is invertible with probability approaching 1.

From $M_{\hat{F},\hat{V}_\varepsilon} = M_{F,V_\varepsilon} + o_p(1)$, we have $\hat{G}\hat{O} = G + o_p(1)$ for some (possibly data-dependent) $(T-k)\times(T-k)$ orthogonal matrix $\hat{O}$. Since $vech(\hat{Z}_n^*)' M_{\hat{\boldsymbol{X}}} vech(\hat{Z}_n^*)$ is invariant to post-multiplication of $\hat{G}$ by an orthogonal matrix (see Proposition 8 in Appendix E.6), from (B.3) we get $\frac{n}{2}\|\hat{S}\| = vech(Z_n^*)' M_{\boldsymbol{X}} vech(Z_n^*) + o_p(1)$, which - together with step (i) - yields asymptotic expansion (3).

(iii) Let us now establish the asymptotic normality of $vech(Z_n^*)$. For any integer m, we let A_m denote the unique $m^2 \times \frac{m(m+1)}{2}$ matrix satisfying $vec(S) = A_m vech(S)$ for any $m \times m$ symmetric matrix S.[42] Duplication matrix A_m satisfies $A_m' A_m = 2I_{\frac{m(m+1)}{2}}$, $A_m A_m' = I_{m^2} + K_{m,m}$, and $K_{m,m}A_m = A_m$, where $K_{m,m}$ is the commutation matrix (see also Magnus, Neudecker (2007) Theorem 12 in Chapter 2.8). Then, we have $vech(Z_n^*) = \boldsymbol{R}' vech(\mathscr{Z}_n)$, where $\mathscr{Z}_n = V_\varepsilon^{-1/2} Z_n V_\varepsilon^{-1/2}$, $\boldsymbol{R} = \frac{1}{2} A_T'(Q \otimes Q) A_{T-k}$, and $Q = V_\varepsilon^{-1/2} G$. Matrix $\boldsymbol{R}$ satisfies $\boldsymbol{R}'\boldsymbol{R} = I_p$. The next lemma establishes the asymptotic normality of $vech(\mathscr{Z}_n)$.

Lemma 2 *(a) Under Assumptions 1-2, A.2, A.6 (a)-(b), we have* $\Omega_n^{-1/2} vech(\mathscr{Z}_n) \Rightarrow N(0, I_{\frac{T(T+1)}{2}})$ *as* $n \to \infty$ *and* T *is fixed, where* $\Omega_n = D_n + \kappa_n I_{\frac{T(T+1)}{2}}$, *and* $\kappa_n = \frac{1}{n}\sum_{m=1}^{J_n}\left(\sum_{i \neq j \in I_m} \sigma_{ij}^2\right)$. *If additionally Assumption A.6 (c) holds, then* $vech(\mathscr{Z}_n) \Rightarrow N(0, \Omega)$, *with* $\Omega := D + \kappa I_{\frac{T(T+1)}{2}}$.

Lemma 2 yields the asymptotic normality of $vech(Z_n^*)$, namely $vech(Z_n^*) \Rightarrow N(0, \Omega_{Z^*})$, with $\Omega_{Z^*} = \boldsymbol{R}'\Omega\boldsymbol{R}$. Part (a) then follows from expansion (3) and the standard result on the distribution of idempotent quadratic forms of Gaussian vectors.

(b) We have $\hat{z}_{m,n}^* = \sum_{i \in I_m} \hat{G}'\hat{V}_\varepsilon^{-1}(\tilde{y}_i\tilde{y}_i')\hat{V}_\varepsilon^{-1}\hat{G}$ with $\tilde{y}_i = y_i - \bar{y}$, since $\hat{\varepsilon}_i = M_{\hat{F},\hat{V}_\varepsilon}\tilde{y}_i$ and $\hat{G}'\hat{V}_\varepsilon^{-1}M_{\hat{F},\hat{V}_\varepsilon} = \hat{G}'\hat{V}_\varepsilon^{-1}$. We get $\hat{z}_{m,n}^* = \sum_{i \in I_m} \hat{G}'\hat{V}_\varepsilon^{-1}(\tilde{\varepsilon}_i\tilde{\varepsilon}_i')\hat{V}_\varepsilon^{-1}\hat{G} + \sum_{i \in I_m} \hat{G}'\hat{V}_\varepsilon^{-1}(F\beta_i\beta_i'F')\hat{V}_\varepsilon^{-1}\hat{G}$ $+ \sum_{i \in I_m} \hat{G}'\hat{V}_\varepsilon^{-1}(F\beta_i\tilde{\varepsilon}_i' + \tilde{\varepsilon}_i\beta_i'F')\hat{V}_\varepsilon^{-1}\hat{G} =: \tilde{z}_{m,n}^* + z_{m,n,1}^* + z_{m,n,2}^*$, where $\tilde{\varepsilon}_i = \varepsilon_i - \bar{\varepsilon}$ by using $\tilde{y}_i\tilde{y}_i' = \tilde{\varepsilon}_i\tilde{\varepsilon}_i' + F\beta_i\beta_i'F' + F\beta_i\tilde{\varepsilon}_i' + \tilde{\varepsilon}_i\beta_i'F'$. Then, we can decompose $\hat{\Omega}_{Z^*}$ into a sum of a leading term and other terms, which are asymptotically negligible, so that $\hat{\Omega}_{Z^*} = \tilde{\Omega}_{Z^*} + o_p(1)$, with $\tilde{\Omega}_{Z^*} = \frac{1}{n}\sum_{m=1}^{J_n} vech(\bar{z}_{m,n}^*) vech(\bar{z}_{m,n}^*)'$, with $\bar{z}_{m,n}^*$ defined as $\tilde{z}_{m,n}^*$ after replacing

[42] The explicit form for A_m is $A_m = \left[\sqrt{2}(e_1 \otimes e_1) : \cdots : \sqrt{2}(e_m \otimes e_m) : \{e_i \otimes e_j + e_j \otimes e_i\}_{i<j}\right]$, with e_i being the ith unit vector of dimension m.

$\tilde{\varepsilon}_i$ with ε_i. Let us now show that $M_{\hat{\boldsymbol{X}}}\tilde{\Omega}_{Z^*}M_{\hat{\boldsymbol{X}}} = M_{\boldsymbol{X}}\Omega_{Z^*}M_{\boldsymbol{X}} + o_p(1)$ up to pre- and post-multiplication by a rotation matrix and its inverse. We have $M_{\hat{\boldsymbol{X}}}vech\left(\hat{G}'\hat{V}_\varepsilon^{-1}\left(\varepsilon_i\varepsilon_i'\right)\hat{V}_\varepsilon^{-1}\hat{G}\right) = M_{\hat{\boldsymbol{X}}}vech\left(\hat{G}'\hat{V}_\varepsilon^{-1}\left(\varepsilon_i\varepsilon_i' - \sigma_{ii}V_\varepsilon\right)\hat{V}_\varepsilon^{-1}\hat{G}\right)$, because of the kernel of $M_{\hat{\boldsymbol{X}}}$. Moreover, from the properties of matrix A_m introduced in part (a), we have: $vech\left(\hat{G}'\hat{V}_\varepsilon^{-1}(\varepsilon_i\varepsilon_i' - \sigma_{ii}V_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}\right)$ $= vech\left(\tilde{Q}'(e_ie_i' - \sigma_{ii}I_T)\tilde{Q}\right) = \hat{\boldsymbol{R}}'vech(e_ie_i' - \sigma_{ii}I_T)$, where $e_i = V_\varepsilon^{-1/2}\varepsilon_i$, and $\hat{\boldsymbol{R}} := \frac{1}{2}A_T'(\tilde{Q}\otimes\tilde{Q})A_{T-k}$ with $\tilde{Q} = V_\varepsilon^{1/2}\hat{V}_\varepsilon^{-1}\hat{G}$. We get $M_{\hat{\boldsymbol{X}}}vech(\bar{z}_{m,n}^*) = M_{\hat{\boldsymbol{X}}}\hat{\boldsymbol{R}}'vech(\zeta_{m,n})$, where $\zeta_{m,n} := \sum_{i\in I_m}(e_ie_i' - \sigma_{ii}I_T)$. Besides, $vech(\mathscr{Z}_n) = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n}vech(\zeta_{m,n})$. Then, $M_{\hat{\boldsymbol{X}}}\tilde{\Omega}_{Z^*}M_{\hat{\boldsymbol{X}}}$ $= M_{\hat{\boldsymbol{X}}}\hat{\boldsymbol{R}}'\tilde{\Omega}_n\hat{\boldsymbol{R}}M_{\hat{\boldsymbol{X}}}$ for $\tilde{\Omega}_n := \frac{1}{n}\sum_{m=1}^{J_n}vech(\zeta_{m,n})vech(\zeta_{m,n})'$. Further, $E[\tilde{\Omega}_n] = V[vech(\mathscr{Z}_n)]$ $= \Omega_n$. Moreover, $\tilde{\Omega}_n - E[\tilde{\Omega}_n] = o_p(1)$, by using $vec(\tilde{\Omega}_n) = \frac{1}{n}\sum_{m=1}^{J_n}vech(\zeta_{m,n})\otimes vech(\zeta_{m,n})$ and $\|V[vec(\tilde{\Omega}_n)]\| \leq C\frac{1}{n^2}\sum_{m=1}^{J_n}E\left[\|vech(\zeta_{m,n})\|^4\right] = o(1)$, where the latter bound is shown in the proof of Lemma 2 using Assumption 2 (d). Additionally, by Assumption A.6, we have $\Omega_n = \Omega + o(1)$. Thus, $\tilde{\Omega}_n = \Omega + o_p(1)$. Now, from the proof of part (a) we have $\hat{G}\hat{O} = G + o_p(1)$ for some $(T-k)\times(T-k)$ orthogonal matrix $\hat{O}$. Then, by Proposition 8 (e) in Appendix E.6, we have $\hat{\boldsymbol{R}}M_{\hat{\boldsymbol{X}}}\hat{\mathscr{R}}^{-1} = \boldsymbol{R}M_{\boldsymbol{X}} + o_p(1)$, for a p dimensional orthogonal matrix $\hat{\mathscr{R}} \equiv \mathscr{R}(\hat{O})$. We conclude that $\hat{\mathscr{R}}M_{\hat{\boldsymbol{X}}}\tilde{\Omega}_{Z^*}M_{\hat{\boldsymbol{X}}}\hat{\mathscr{R}}^{-1}$ is a consistent estimator of $M_{\boldsymbol{X}}\Omega_{Z^*}M_{\boldsymbol{X}}$ as $n\to\infty$ and T is fixed. Part (b) then follows from the continuity of eigenvalues for symmetric matrices, and their invariance under pre- and post-multiplication by an orthogonal matrix and its transpose.

(c) Under $H_1(k)$ and Assumption A.7 (a), we have $\hat{F}\xrightarrow{p}F^*$ and $\hat{V}_\varepsilon\xrightarrow{p}V_\varepsilon^*$. Then, $\hat{S}\xrightarrow{p}S^*$ with $S^* = (V_\varepsilon^*)^{-1/2}M_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*)M_{F^*,V_\varepsilon^*}'(V_\varepsilon^*)^{-1/2}\neq 0$. Indeed, if S^* were the null matrix, then we would have $M_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*)M_{F^*,V_\varepsilon^*}' = 0$, which implies $V_y - V_\varepsilon^* = P_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*) + (V_y - V_\varepsilon^*)P_{F^*,V_\varepsilon^*}' - P_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*)P_{F^*,V_\varepsilon^*}'$, with $P_{F^*,V_\varepsilon^*} = I_T - M_{F^*,V_\varepsilon^*}$. From the probability limits of Equation (FA2) for pseudo values, we have $P_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*) = (V_y - V_\varepsilon^*)P_{F^*,V_\varepsilon^*}' = P_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*)P_{F^*,V_\varepsilon^*}' = F^*(F^*)'$ (see proof of Lemma 1 (c)). Thus $V_y = F^*(F^*)' + V_\varepsilon^*$, in contradiction with Assumption A.7 (b). Thus, $n\|\hat{S}\|^2 \geq Cn$, w.p.a. 1, for a constant $C > 0$. Moreover, using $vech(\hat{z}_{m,n}^*) = vech(\hat{G}'\hat{V}_\varepsilon^{-1}(\sum_{i\in I_m}\tilde{y}_i\tilde{y}_i')\hat{V}_\varepsilon^{-1}\hat{G})$ and the conditions on Θ, we get $\|vech(\hat{z}_{m,n}^*)\| \leq C\sum_{i\in I_m}\|\tilde{y}_i\|^2$. Then, from Assumptions A.2 and A.3, $E[\|M_{\hat{\boldsymbol{X}}}\hat{\Omega}_{Z^*}M_{\hat{\boldsymbol{X}}}\|] \leq C\frac{1}{n}\sum_{m=1}^{J_n}b_{m,n}^2 = O(n\sum_{m=1}^{J_n}B_{m,n}^2)$. Moreover, $\sum_{m=1}^{J_n}B_{m,n}^2 = o(1)$. Indeed, Assumption 2

(d) implies $B_{m,n} \leq cn^{-\frac{\delta}{\delta+1}}$ uniformly in m, for any $c > 0$ and n large enough, and hence $\sum_{m=1}^{J_n} B_{m,n}^2 = cn^{-\frac{\delta}{\delta+1}} \sum_{m=1}^{J_n} B_{m,n} \leq c$, for any $c > 0$ and n large. Part (c) follows from the Lipschitz continuity of eigenvalues for symmetric matrices.

Proof of Proposition 2: We have $LR(k) = \frac{n}{2}\|\hat{S}\|^2 + o_p(1) = vech(Z_n^*)' M_{\boldsymbol{X}} vech(Z_n^*) + o_p(1)$ from expansion (3). Moreover, the kernel of matrix $M_{\boldsymbol{X}}$ implies that $M_{\boldsymbol{X}} vech(Z_n^*) = A(F, V_\varepsilon) z_n^{AD}$, where vector z_n^{AD} stacks the $T(T-1)/2$ above-diagonal elements of matrix Z_n and $A(F, V_\varepsilon)$ is a deterministic matrix whose elements only depend on F, V_ε. From Conditions (a) and (b) of Proposition 2, and Lemma 2, we have $z_n^{AD} \Rightarrow N(0, \Omega_z)$, where the diagonal matrix Ω_z is the same as if the errors were independent normally distributed - up to replacing q with $q + \kappa$.

Proof of Proposition 3: Let us first get the asymptotic expansion of $\hat{V}_y - \tilde{V}_\varepsilon = \frac{1}{n}\tilde{Y}\tilde{Y}' - \tilde{V}_\varepsilon$. With the drifting DGP $Y = \mu 1_n' + F\beta' + F_{k+1}\beta_{loc}' + \varepsilon$, and using $\bar{\beta} = 0$, $\bar{\beta}_{loc} = 0$, $\frac{1}{n}[\beta : \beta_{loc}]'[\beta : \beta_{loc}] = I_{k+1}$ and Lemma 6 (a) in Appendix D, we get $\hat{V}_y - \tilde{V}_\varepsilon = FF' + \frac{1}{\sqrt{n}}\Psi_{y,loc} + R_y$, where

$$\Psi_{y,loc} = c_{k+1}\rho_{k+1}\rho_{k+1}' + \frac{1}{\sqrt{n}}(\varepsilon\beta F' + F\beta'\varepsilon') + \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right), \tag{B.4}$$

and $R_y = \frac{1}{n}(\varepsilon\beta_{loc}F_{k+1}' + F_{k+1}\beta_{loc}'\varepsilon') + [F_{k+1}F_{k+1}' - n^{-1/2}c_{k+1}\rho_{k+1}\rho_{k+1}'] + o_p(\frac{1}{\sqrt{n}})$. Using $F_{k+1} = \sqrt{\gamma_{k+1}}\rho_{k+1}$ and $\sqrt{n}\gamma_{k+1} = c_{k+1} + o(1)$, we get $R_y = o_p(1/\sqrt{n})$. Subsituting the expansion for $\hat{V}_y - \tilde{V}_\varepsilon$ into (B.2), and repeating the arguments leading to expansion (3) yields expansion (5). From Lemma 2, we get $vech(Z_{n,loc}^*) \Rightarrow N(c_{k+1} vech(\xi_{k+1}\xi_{k+1}'), \Omega_{Z^*})$ as $n \to \infty$. The result then follows from the standard result on the distribution of idempotent quadratic forms of non-central Gaussian vectors.

Proof of Proposition 4: The proof of part (a) is in three steps. (i) The testing problem asymptotically simplifies to the null hypothesis $H_0 : \lambda_1 = ... = \lambda_{df} = 0$ vs. the alternative hypothesis $H_1 : \exists \lambda_j > 0, j = 1, ..., df$. Let us define $\boldsymbol{\lambda}_0 = (0, ..., 0)'$ for the null hypothesis and pick a given vector $\boldsymbol{\lambda}_1 = (\lambda_1, ..., \lambda_{df})'$ in the alternative hypothesis, and consider the test of $\boldsymbol{\lambda}_0$ versus $\boldsymbol{\lambda}_1$ (simple hypothesis). By Neyman-Pearson Lemma, the most powerful test for $\boldsymbol{\lambda}_0$ versus $\boldsymbol{\lambda}_1$ rejects the null hypothesis when $f(z; \lambda_1, ..., \lambda_{df})/f(z; 0, ..., 0)$ is large, i.e., the test function is $\phi(z) = \mathbf{1}\left\{\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)} \geq C\right\}$ for a constant $C > 0$ set to ensure the correct asymptotic size.

(ii) Let us now show that the density ratio $\frac{f(z;\lambda_1,\ldots,\lambda_{df})}{f(z;0,\ldots,0)}$ is an increasing function of z. To show this, we can rely on an expansion of the density of $\sum_{j=1}^{df}\mu_j\chi^2(1,\lambda_j^2)$ in terms of central chi-square densities (Kotz, Johnson, and Boyd (1967) Equations (144) and (151)):

$$f(z;\lambda_1,...,\lambda_{df})=\sum_{k=0}^{\infty}\bar{c}_k(\lambda_1,...,\lambda_{df})g(z;df+2k,0), \tag{B.5}$$

where the coefficients $\bar{c}_k(\lambda_1,...,\lambda_{df}) = Ae^{-\sum_{j=1}^{df}\lambda_j^2/2}E[Q(\lambda_1,...,\lambda_{df})^k]/k!$ involve moments of the quadratic form $Q(\lambda_1,...,\lambda_{df}) = (1/2)\sum_{j=1}^{df}\left(\nu_j^{1/2}X_j+\lambda_j(1-\nu_j)^{1/2}\right)^2$ of the mutually independent variables $X_j \sim N(0,1)$, $A = \prod_{j=1}^{df}\mu_j^{-1/2}$, and $\nu_j = 1-\frac{1}{\mu_j}\min_\ell \mu_\ell$. Without loss of generality for checking the monotonicity, we have rescaled the density so that $\min_j \mu_j = 1$. Then, from (B.5), we get the ratio: $\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)} = \frac{\sum_{k=0}^{\infty}\bar{c}_k(\lambda_1,...,\lambda_{df})g(z;df+2k,0)}{\sum_{k=0}^{\infty}\bar{c}_k(0,...,0)g(z;df+2k,0)}$. By dividing both the numerator and the denominator by the central chi-square density $g(z;df,0)$, we get $\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)} = e^{-\sum_{j=1}^{df}\lambda_j^2/2}\frac{\sum_{k=0}^{\infty}c_k(\lambda_1,...,\lambda_{df})\psi_k(z)}{\sum_{k=0}^{\infty}c_k(0,...,0)\psi_k(z)} =: e^{-\sum_{j=1}^{df}\lambda_j^2/2}\Psi(z;\lambda_1,...,\lambda_{df})$, where $\psi_k(z) := g(z;df+2k,0)/g(z;df,0) = \frac{\Gamma(\frac{df}{2})}{2^k\Gamma(\frac{df}{2}+k)}z^k$ is the ratio of central chi-square distributions with $df+2k$ and df degrees of freedom, and $c_k(\lambda_1,...,\lambda_{df}) = E[Q(\lambda_1,...,\lambda_{df})^k]/k!$. We use the short notation $c_k(\lambda) := c_k(\lambda_1,...,\lambda_{df})$ and $c_k(0) := c_k(0,...,0)$. The factor $e^{-\sum_{j=1}^{df}\lambda_j^2/2}$ does not impact on the monotonicity of the density ratio. We take the derivative of $\Psi(z;\lambda_1,...,\lambda_{df})$ with respect to argument z and get $\partial_z\Psi(z;\lambda_1,...,\lambda_{df}) = \frac{\left(\sum_{k=1}^{\infty}c_k(\lambda)\psi_k'(z)\right)\left(1+\sum_{k=1}^{\infty}c_k(0)\psi_k(z)\right)}{\left(\sum_{k=0}^{\infty}c_k(0)\psi_k(z)\right)^2} - \frac{\left(1+\sum_{k=1}^{\infty}c_k(\lambda)\psi_k(z)\right)\left(\sum_{k=1}^{\infty}c_k(0)\psi_k'(z)\right)}{\left(\sum_{k=0}^{\infty}c_k(0)\psi_k(z)\right)^2}$. The sign is given by the difference of the numerators, which is $\sum_{k=1}^{\infty}[c_k(\lambda)-c_k(0)]\psi_k'(z)+\sum_{k,l=1,k\neq l}^{\infty}c_k(\lambda)c_l(0)[\psi_k'(z)\psi_l(z)-\psi_k(z)\psi_l'(z)] = \sum_{k=1}^{\infty}[c_k(\lambda)-c_k(0)]\psi_k'(z)+\sum_{k,l=1,k>l}^{\infty}[c_k(\lambda)c_l(0)-c_l(\lambda)c_k(0)][\psi_k'(z)\psi_l(z)-\psi_k(z)\psi_l'(z)]$. We use $\psi_k'(z) = \frac{\Gamma(\frac{d}{2})k}{2^k\Gamma(\frac{d}{2}+k)}z^{k-1}$ and $\psi_k'(z)\psi_l(z)-\psi_k(z)\psi_l'(z) = (k-l)\frac{\Gamma(\frac{d}{2})^2}{2^{k+l}\Gamma(\frac{d}{2}+k)\Gamma(\frac{d}{2}+l)}z^{k+l-1}$ for $k>l$ and $z\geq 0$. The difference of the numerators in the derivative of the density ratio becomes: $\frac{1}{2}\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+1)}[c_1(\lambda)-c_1(0)] + \frac{1}{2^2}\frac{2\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+2)}[c_2(\lambda)-c_2(0)]z + \sum_{m=3}^{\infty}\frac{1}{2^m}\Big(m\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+m)}[c_m(\lambda)-c_m(0)]$ $+\sum_{k>l\geq 1,k+l=m}\frac{(k-l)\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d}{2}+k)\Gamma(\frac{d}{2}+l)}[c_k(\lambda)c_l(0)-c_l(\lambda)c_k(0)]\Big)z^{m-1} = \sum_{m=1}^{\infty}\frac{1}{2^m}\kappa_m z^{m-1}$, with $\kappa_m := \sum_{k>l\geq 0,k+l=m}(k-l)\frac{\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d}{2}+k)\Gamma(\frac{d}{2}+l)}[c_k(\lambda)c_l(0)-c_l(\lambda)c_k(0)]$. A direct calculation shows that $\kappa_1,\kappa_2\geq 0$. Hence, a sufficient condition for monotonicity of the density ratio is $\kappa_m\geq 0$, for all $m\geq 3$, i.e., Inequalities (6). Thus, the test rejects for large values of the argument, i.e., $\phi(z) = \mathbf{1}\{z \geq \bar{C}\}$, where the constant $\bar{C}$ is determined by fixing the asymptotic size under the null hypothesis.

(iii) Since the test function ϕ does not depend on $\boldsymbol{\lambda}_1$, it is AUMPI in the class of hypothesis tests based on the LR statistic (or the squared norm statistic). It yields part (a).

Let us now turn to the proof of part (b). From the definition of the κ_m coefficients written as $\kappa_m = \sum_{j>l\geq 0, j+l=m} \frac{(j-l)\Gamma(\frac{df}{2})^2}{\Gamma(\frac{df}{2}+j)\Gamma(\frac{df}{2}+l)} c_j(0)c_l(0)[\frac{c_j(\lambda)}{c_j(0)} - \frac{c_l(\lambda)}{c_l(0)}]$, it is sufficient to get $\kappa_m \geq 0$, for all m, that sequence $\frac{c_j(\lambda)}{c_j(0)}$, for $j = 0, 1, ...$, is increasing. To prove that, we link the coefficients $c_j(\lambda)$ to the complete exponential Bell's polynomials (Bell (1934)) and establish the following recurrence.

Lemma 3 *We have* $c_{l+1}(\lambda) = \frac{1}{l+1} \sum_{i=0}^{l} \left(\frac{1}{2} \sum_{j=1}^{df} \nu_j^i \left[\nu_j + (i+1)(1-\nu_j)\lambda_j^2 \right] \right) c_{l-i}(\lambda)$, *for* $l \geq 0$.

We use $\frac{c_l(\lambda)}{c_l(0)} = \frac{\tilde{c}_l(\lambda)}{\bar{\gamma}_l}$, where we obtain the sequences $\bar{\gamma}_l := c_l(0)\nu_{df}^{-l}$ and $\tilde{c}_l(\lambda) := c_l(\lambda)\nu_{df}^{-l}$ by standardization with ν_{df}^{-l}. From Lemma 3, we have $\bar{\gamma}_{l+1} = \frac{1}{l+1} \sum_{i=0}^{l} \frac{1}{2} \left(1 + \sum_{j=2}^{df-1} \rho_j^{i+1} \right) \bar{\gamma}_{l-i}$ with $\bar{\gamma}_0 = 1$, and $\tilde{c}_{l+1}(\lambda) = \frac{1}{l+1} \sum_{i=0}^{l} \left(\frac{1}{2} \sum_{j=1}^{df} \rho_j^i \left[\rho_j + \frac{i+1}{\nu_{df}}(1-\nu_j)\lambda_j^2 \right] \right) \tilde{c}_{l-i}(\lambda)$ with $\tilde{c}_0(\lambda) = 1$ (note that $\rho_1 = 0$ and $\rho_{df} = 1$). To prove that sequence $\frac{\tilde{c}_l(\lambda)}{\bar{\gamma}_l}$ is increasing, the next lemma provides a sufficient condition from "separation" of the coefficients that define the recursive relations.

Lemma 4 *Let* (a_i) *be a real sequence, and let* $b_i = \frac{1}{2} \left(1 + \sum_{j=2}^{df-1} \rho_j^i \right)$, *for* $i \geq 1$, *where* $0 \leq \rho_j \leq 1$. *Let sequences* (g_l) *and* (c_l) *be defined recursively by* $g_{l+1} = \frac{1}{l}(b_1 g_l + b_2 g_{l-1} + ... + b_l)$ *and* $c_{l+1} = \frac{1}{l}(a_1 c_l + a_2 c_{l-1} + ... + a_l)$, *with* $g_1 = c_1 = 1$. *Suppose that* $a_i \geq \max\{\frac{df-1}{2}, 1\}$, *for all* i *(separation condition). Then, sequence* $(\frac{c_l}{g_l})$ *is increasing.*

We apply Lemma 4 to sequences $\tilde{c}_l(\lambda)$ and $\bar{\gamma}_l$. We detail the case $df \geq 3$ (for $df = 2$ the analysis is simpler). The separation condition $\frac{1}{2} \sum_{j=1}^{df} \rho_j^i \left[\rho_j + \frac{i+1}{\nu_{df}}(1-\nu_j)\lambda_j^2 \right] \geq \frac{df-1}{2}$, for $i = 0$, yields $\lambda_1^2 + \sum_{j=2}^{df}(1-\nu_j)\lambda_j^2 \geq \nu_{df} \left(df - 2 - \sum_{j=2}^{df-1} \rho_j \right)$, and, for $i \geq 1$, it yields $\sum_{j=2}^{df-1} \rho_j^i (1 - \nu_j)\lambda_j^2 + (1-\nu_{df})\lambda_{df}^2 \geq \frac{\nu_{df}}{i+1} \left(df - 2 - \sum_{j=2}^{df-1} \rho_j^{i+1} \right)$. Inequalities (7) follow.

C Monte Carlo experiments

This appendix gives a Monte Carlo assessment of size and power and selection procedure for the number of factors for the LR test under non-Gaussian errors. Let us start with a description of the DGP we use in our simulations. In the DGP, the betas are $\beta_i \overset{i.i.d.}{\sim} N(0, I_k)$, with $k = 3$, and the matrix of factor values is $F = V_\varepsilon^{1/2} U \Gamma^{1/2}$, where $U = \tilde{F}(\tilde{F}'\tilde{F})^{-1/2}$ and $vec(\tilde{F}) \sim N(0, I_{Tk})$. We generate the diagonal elements of $V_\varepsilon = diag(h_1, ..., h_T)$ through a common time-varying component in idiosyncratic volatilities (Renault, Van Der Heijden and Werker (2023)) via the ARCH $h_t = 0.6 + 0.5 h_{t-1} z_{t-1}^2$, with $z_t \sim IIN(0, 1)$. This common component induces a deviation from spherical errors. The diagonal matrix $\Gamma = T diag(3, 2, n^{-\bar{\kappa}})$ yields $\frac{1}{T} F' V_\varepsilon^{-1} F = diag(3, 2, n^{-\bar{\kappa}})$, i.e., the "signal-to-noise" ratios equal 3, 2 and $n^{-\bar{\kappa}}$ for the three factors. We take $\bar{\kappa} = \infty$ to study the size of $LR(2)$. To study the power of $LR(2)$, we take $\bar{\kappa} = 0$ to get a global alternative and $\bar{\kappa} = 1/2$ to get a local alternative (weak factor). We generate the idiosyncratic errors by $\varepsilon_{i,t} = h_t^{1/2} h_{i,t}^{1/2} z_{i,t}$, where $h_{i,t} = c_i + \alpha_i h_{i,t-1} z_{i,t-1}^2$, with $z_{i,t} \sim IIN(0, 1)$ mutually independent of z_t. We use the constraint $c_i = \sigma_{ii}(1 - \alpha_i)$ with uniform draws for the idiosyncratic variances $V[\varepsilon_{i,t}] = \sigma_{ii} \overset{i.i.d.}{\sim} U[1, 4]$, so that $V[\varepsilon_{i,t}/h_t^{1/2}] = \frac{c_i}{1-\alpha_i} = \sigma_{ii}$. Such a setting allows for cross-sectional heterogeneity in the variances of the scaled $\varepsilon_{i,t}/h_t^{1/2}$. The ARCH parameters are uniform draws $\alpha_i \overset{i.i.d.}{\sim} U[0.2, 0.5]$ with an upper boundary of the interval ensuring existence of fourth-order moments. We generate $5,000$ panels of returns of size $n \times T$ for each of the 100 draws of the $T \times k$ factor matrix F and common ARCH process h_t, $t = 1, ..., T$, in order to keep the factor values constant within repetitions, but also to study the potential heterogeneity of size and power results across different factor paths. The factor betas β_i, idiosyncratic variances σ_{ii}, and individual ARCH parameters α_i are the same across all repetitions in all designs of the section. We use three different cross-sectional sizes $n = 500, 1000, 5000$, and three values of time-series dimension $T = 6, 12, 24$. The variance matrix $\hat{\Omega}_{\bar{Z}^*}$ is computed using the parametric structure of Lemma 9. We get the $T - 1$ estimated parameters by least squares, as detailed in OA Section E.5.3 i). The p-values are computed over $5,000$ draws.

We provide the size and power results in % in Table 1. Size of $LR(2)$ is close to its nominal level 5%, with size distortions smaller than 1%, except for the case $T = 24$ and $n = 500$. The impact of the factor values on size is small for T above 6. The labels global power and local power refer to $\bar{\kappa} = 0$ and $\bar{\kappa} = 1/2$, and power computation is not size adjusted. The global power is equal to 100%, while the local power ranges from 80% to 85% for $T = 6$, and is equal to 100% for $T = 12$ and $T = 24$. The approximate constancy of local power w.r.t. n, for large n, is coherent with theory implying convergence to asymptotic local power. In the last panel of Table 1, we provide the average of the estimated number $\hat{k}_{LR}$ of factors, obtained by sequential testing with $LR(k)$, for $k = 0, \dots, k_{max}$, with $k_{max} = 2, 7, 17$ for $T = 6, 12, 24$ (see Table 3 of OA). We follow the procedure described in Section 4.2, with size $\alpha_n = 10/n$. If we reject for all $k = 0, \dots, k_{max}$, then the estimated number of factors is set to $\hat{k}_{LR} = k_{max} + 1$. For all sample sizes $T = 6, 12, 24$, the average estimated number of factors is very close to the true number 2. We can conclude that our selection procedure for the number of factors works well in our simulations.

D Proofs of Lemmas 1-4

Proof of Lemma 1: Let $\hat{U}$ be the $T \times k$ matrix whose orthonormal columns are the eigenvectors for the k largest eigenvalues of matrix $\hat{V}_\varepsilon^{-1/2} \hat{V}_y \hat{V}_\varepsilon^{-1/2}$. Those eigenvalues are $1 + \hat{\gamma}_j$, $j = 1, ..., k$, while it holds $\hat{F} = \hat{V}_\varepsilon^{1/2} \hat{U} \hat{\Gamma}^{1/2}$, where $\hat{\Gamma} = diag(\hat{\gamma}_1, ..., \hat{\gamma}_k)$. We have $I_T - \hat{U}\hat{U}' = I_T - \hat{V}_\varepsilon^{-1/2} \hat{F} \hat{\Gamma}^{-1} \hat{F}' \hat{V}_\varepsilon^{-1/2} = I_T - \hat{V}_\varepsilon^{-1/2} \hat{F} (\hat{F}' \hat{V}_\varepsilon^{-1} \hat{F})^{-1} \hat{F}' \hat{V}_\varepsilon^{-1/2} = \hat{V}_\varepsilon^{-1/2} M_{\hat{F}, \hat{V}_\varepsilon} \hat{V}_\varepsilon^{1/2} = \hat{V}_\varepsilon^{1/2} M'_{\hat{F}, \hat{V}_\varepsilon} \hat{V}_\varepsilon^{-1/2}$. Thus, $\hat{S} = (I_T - \hat{U}\hat{U}') \left(\hat{V}_\varepsilon^{-1/2} \hat{V}_y \hat{V}_\varepsilon^{-1/2} - I_T \right) (I_T - \hat{U}\hat{U}')$. By the spectral decomposition of $\hat{V}_\varepsilon^{-1/2} \hat{V}_y \hat{V}_\varepsilon^{-1/2}$, we get $(I_T - \hat{U}\hat{U}') \left(\hat{V}_\varepsilon^{-1/2} \hat{V}_y \hat{V}_\varepsilon^{-1/2} - I_T \right) (I_T - \hat{U}\hat{U}') = \sum_{j=k+1}^{T} \hat{\gamma}_j \hat{P}_j$, where the $\hat{P}_j$ are the orthogonal projection matrices onto the eigenspaces for the $T - k$ smallest eigenvalues. Then, Part (a) follows. Part (b) is a consequence of the squared Frobenius norm of a symmetric matrix being equal to the sum of its squared eigenvalues. For part (c), let $P_{\hat{F}, \hat{V}_\varepsilon} = I_T - M_{\hat{F}, \hat{V}_\varepsilon}$ and note that $\hat{F}\hat{F}' = P_{\hat{F}, \hat{V}_\varepsilon} (\hat{V}_y - \hat{V}_\varepsilon) + (\hat{V}_y - \hat{V}_\varepsilon) P'_{\hat{F}, \hat{V}_\varepsilon} - P_{\hat{F}, \hat{V}_\varepsilon} (\hat{V}_y - \hat{V}_\varepsilon) P'_{\hat{F}, \hat{V}_\varepsilon} = \hat{V}_y - \hat{V}_\varepsilon - M_{\hat{F}, \hat{V}_\varepsilon} (\hat{V}_y - \hat{V}_\varepsilon) M'_{\hat{F}, \hat{V}_\varepsilon}$, where the first equality is because the three terms on the RHS are all equal to $\hat{F}\hat{F}'$ by (FA2). The

T	Size (%) 6	12	24	Global Power (%) 6	12	24	Local Power (%) 6	12	24	$\hat{k}_{LR}$ 6	12	24
$n = 500$	6.0	5.2	6.7	100	100	100	80	100	100	2.0	2.0	2.1
	(2.8)	(0.3)	(0.4)	(0.1)	(0.0)	(0.0)	(20.5)	(0.0)	(0.0)	(0.1)	(0.1)	(0.2)
$n = 1000$	5.6	4.9	5.5	100	100	100	81	100	100	2.0	2.0	2.0
	(2.3)	(0.3)	(0.3)	(0.0)	(0.0)	(0.0)	(21.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)
$n = 5000$	5.3	5.0	4.9	100	100	100	85	100	100	2.0	2.0	2.0
	(0.9)	(0.3)	(0.3)	(0.0)	(0.0)	(0.0)	(20.4)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)

Table 1: For each sample size combination (n, T), we provide the average size and power in % for the statistic $LR(2)$ (first three panels), and the average of the estimated number $\hat{k}_{LR}$ of factors obtained by sequential testing (last panel). Nominal size is 5% for the first three panels, and $\alpha_n = 10/n$ for the last panel. Global power refers to the global alternative $\bar{\kappa} = 0$, and local power refers to the local alternative $\bar{\kappa} = 0.5$. In parentheses, we report the standard deviations for size, power, and $\hat{k}_{LR}$ across 100 different draws of the factor path.

conclusion follows from (FA1) and $\hat{V}_\varepsilon$ being diagonal.

Proof of Lemma 2: We have $\mathscr{Z}_n = \frac{1}{\sqrt{n}}(W\Sigma W' - Tr(\Sigma)I_T)$. Hence, $(\mathscr{Z}_n)_{tt} = \frac{1}{\sqrt{n}}\sum_{i,j}(w_{i,t}w_{j,t} - 1_{\{i=j\}})\sigma_{ij} = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n}\zeta_{m,n}^{tt}$, with $\zeta_{m,n}^{tt} = \sum_{i\in I_m}[w_{i,t}^2 - 1]\sigma_{ii} + 2\sum_{\substack{i,j\in I_m\\ i<j}} w_{i,t}w_{j,t}\sigma_{ij}$, together with $(\mathscr{Z}_n)_{ts} = \frac{1}{\sqrt{n}}\sum_{i,j} w_{i,t}w_{j,s}\sigma_{ij} = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n}\zeta_{m,n}^{ts}, t \neq s$, with $\zeta_{m,n}^{ts} = \sum_{i\in I_m} w_{i,t}w_{i,s}\sigma_{ii} + \sum_{\substack{i,j\in I_m\\ i<j}} w_{i,t}w_{j,s}\sigma_{ij} + \sum_{\substack{i,j\in I_m\\ i>j}} w_{i,t}w_{j,s}\sigma_{ij}, t\neq s$, so that $vech(\mathscr{Z}_n) = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n} vech(\zeta_{m,n})$, where $\zeta_{m,n}$ is the $T\times T$ matrix having element $\zeta_{m,n}^{ts}$ in position (t,s). Hence, $vech(\mathscr{Z}_n)$ is the row sum of a triangular array $\{vech(\zeta_{m,n})\}_{1\leq m\leq n}$ of independent centered random vectors. Let $\Omega_{m,n} := V[vech(\zeta_{m,n})]$. Using Assumption 2 (a), we compute (i) $E[(\zeta_{m,n}^{tt})^2] = \sum_{i\in I_m}(E[w_{i,t}^4] - 1)\sigma_{ii}^2 + 2\sum_{\substack{i,j\in I_m\\ i\neq j}}\sigma_{ij}^2$; (ii) $E[(\zeta_{m,n}^{ts})^2] = \sum_{i\in I_m} E[w_{i,t}^2 w_{i,s}^2]\sigma_{ii}^2 + \sum_{\substack{i,j\in I_m\\ i\neq j}}\sigma_{ij}^2, t\neq s$; (iii) $E[\zeta_{m,n}^{tt}\zeta_{m,n}^{ss}] = \sum_{i\in I_m} E[w_{i,t}^2 w_{i,s}^2 - 1]\sigma_{ii}^2, t\neq s$; (iv) $E[\zeta_{m,n}^{tt}\zeta_{m,n}^{rp}] = \sum_{i\in I_m} E[w_{i,t}^2 w_{i,r}w_{ip}]\sigma_{ii}^2, r\neq p$; (v) $E[\zeta_{m,n}^{ts}\zeta_{m,n}^{rp}] = \sum_{i\in I_m} E[w_{i,t}w_{i,s}w_{i,r}w_{i,p}]\sigma_{ii}^2, t\neq s, r\neq p$. It follows that $V[vech(\mathscr{Z}_n)] = \frac{1}{n}\sum_{m=1}^{J_n}\Omega_{m,n} = D_n + \kappa_n I_{\frac{T(T+1)}{2}} = \Omega_n$. The eigenvalues of D_n are bounded away from 0 under Assumption A.6 (b), because for any unit vector $\xi\in\mathbb{R}^{T(T+1)/2}$, we have $\xi' D_n\xi \geq \frac{1}{n}\sum_{i=1}^n 1_{i\in\bar{S}}\sigma_{ii}^2\xi' V[vech(w_i w_i')]\xi \geq \underline{c}\frac{1}{n}\sum_{i=1}^n 1_{i\in\bar{S}}\sigma_{ii}^2 \geq \underline{c}\left(1 - \frac{1}{n}\sum_{i=1}^n(1 - 1_{i\in\bar{S}})\sigma_{ii}\right)^2 \geq \underline{c}\left(1 - \bar{C}\frac{1}{n}\sum_{i=1}^n(1 - 1_{i\in\bar{S}})\right)^2 \geq \frac{\underline{c}}{4}$, for all n. We use the multivariate Lyapunov condition $\|\Omega_n^{-1/2}\|^4\frac{1}{n^2}\sum_{m=1}^{J_n} E[\|vech(\zeta_{m,n})\|^4] \to 0$ to invoke a CLT. Since $\|A^{-1/2}\|^4 \leq \frac{k^2}{\delta_k^2(A)}$ and $\|x\|^4 \leq k\sum_{j=1}^k x_j^4$, for any $k\times k$ positive semi-definite matrix A and $k\times 1$ vector x, it suffices to check that $\frac{1}{n^2}\sum_{m=1}^{J_n} E[(\zeta_{m,n}^{ts})^4] \to 0$, for all t,s. Besides, we can show that there exists a constant $M>0$, such that $E[(\zeta_{m,n}^{ts})^4] \leq M b_{m,n}^{2(1+\delta)}$, for all m,n,t,s. We get $\frac{1}{n^2}\sum_{m=1}^{J_n} E[(\zeta_{m,n}^{ts})^4] \leq M\frac{1}{n^2}\sum_{m=1}^{J_n} b_{m,n}^{2(1+\delta)} = Mn^{2\delta}\sum_{m=1}^{J_n} B_{m,n}^{2(1+\delta)} = o(1)$, under Assumption 2 (d). Then, $\Omega_n^{-1/2} vech(\mathscr{Z}_n) \Rightarrow N(0, I_{\frac{T(T+1)}{2}})$ by the multivariate Lyapunov CLT. Under Assumptions A.6 (a)-(c), $\Omega_n \to \Omega$ follows from the Slutsky theorem, and Ω is positive definite.

Proof of Lemma 3: We have $c_j(\lambda) = \frac{1}{j!}E[Q^j] = \frac{1}{j!}\frac{d^j\Psi(0)}{du^j}$ where $\Psi(u) := E[\exp(uQ)] = \exp[\psi(u)]$ is the Moment Generating Function (MGF) of $Q = \frac{1}{2}\sum_{j=1}^{df}(\sqrt{\nu}_j X_j + \sqrt{1-\nu_j}\lambda_j)^2$ with $X_j \sim i.i.d.N(0,1)$. By the independence of variables X_j, we get $\Psi(u) = \prod_{j=1}^{df} E[\exp(\frac{u}{2}(\sqrt{\nu}_j X_j + \sqrt{1-\nu_j}\lambda_j)^2]$ where $E[\exp(\frac{u}{2}(\sqrt{\nu}_j X_j + \sqrt{1-\nu_j}\lambda_j)^2] = (1-\nu_j u)^{-1/2} e^{\frac{1}{2}\frac{(1-\nu_j)u}{1-\nu_j u}\lambda_j^2}$, for $u < 1/\nu_j$. Thus we get the log MGF $\psi(u) = \frac{1}{2}\sum_{j=1}^{df}\left[-\log(1-\nu_j u) + \frac{(1-\nu_j)u}{1-\nu_j u}\lambda_j^2\right]$, for $u < 1/\nu_{df}$. Its lth order derivative evaluated at $u = 0$ is

$$\psi^{(l)}(0) = \frac{(l-1)!}{2} \sum_{j=1}^{df} \nu_j^{l-1} \left[\nu_j + l(1-\nu_j)\lambda_j^2\right], \quad l \geq 0. \tag{D.1}$$

By using the Faa di Bruno formula for the derivatives of a composite function, we have $\frac{d^l}{du^l} e^{\psi(u)} = e^{\psi(u)} B_l(\psi'(u), \psi''(u), ..., \psi^{(l)}(u))$, where B_l is the lth complete exponential Bell's polynomial (Bell (1934)). Hence, $\Psi^{(l)}(0) = B_l(\psi'(0), \psi''(0), ..., \psi^{(l)}(0))$. The complete Bell's polynomials satisfy the recurrence relation $B_{l+1}(x_1, x_2, ..., x_{l+1}) = \sum_{i=0}^{l} \binom{l}{i} B_{l-i}(x_1, ..., x_{l-i}) x_{i+1}$. Thus, $\Psi^{(l+1)}(0) = \sum_{i=0}^{l} \binom{l}{i} \Psi^{(l-i)}(0) \psi^{(i+1)}(0)$. After standardization with the factorial term, and using equation (D.1), the conclusion follows.

Proof of Lemma 4: The proof is in four steps. (i) We first show that (c_i) is increasing, i.e., $G_i^c := c_{i+1} - c_i \geq 0$ for all i. For this purpose, from the recursive relation defining c_{i+1} we have:

$$\begin{aligned}
c_{i+1} &= \frac{1}{i}\left(a_1(c_{i-1} + G_{i-1}^c) + a_2(c_{i-2} + G_{i-2}^c) + \cdots + a_{i-1}(c_1 + G_1^c) + a_i\right) \\
&= \frac{1}{i}\left((a_1 - 1)G_{i-1}^c + (a_2 - 1)G_{i-2}^c + \cdots + (a_{i-1} - 1)G_1^c + (a_i - 1)\right) \\
&\quad + \frac{1}{i}\left(G_{i-1}^c + G_{i-2}^c + \cdots + G_1^c + 1\right) + \frac{1}{i}\left(a_1 c_{i-1} + a_2 c_{i-2} + \cdots + a_{i-1}\right).
\end{aligned}$$

The second term in the RHS is equal to $\frac{1}{i} c_i$. Using $a_1 c_{i-1} + a_2 c_{i-2} + \cdots + a_{i-1} = (i-1)c_i$, the third term in the RHS is equal to $\frac{i-1}{i} c_i$. Thus, by bringing these two terms in the LHS, we get $G_i^c = \frac{1}{i}\left((a_1 - 1)G_{i-1}^c + (a_2 - 1)G_{i-2}^c + \cdots + (a_{i-1} - 1)G_1^c + (a_i - 1)\right)$, for all $i \geq 2$, with $G_1^c = a_1 - 1$. Since $a_i \geq 1$ for all i, we get $G_i^c \geq 0$ for all $i \geq 1$ by an induction argument .

(ii) We now strengthen the result in step (i) and show that $H_i^c := c_{i+1} - c_i \frac{\zeta + i - 1}{i} \geq 0$ for all i, with $\zeta = \max\{\frac{df-1}{2}, 1\}$. Similarly as in step (i), we have

$$\begin{aligned}
c_{i+1} &= \frac{1}{i}\left((a_1 - \zeta)G_{i-1}^c + (a_2 - \zeta)G_{i-2}^c + \cdots + (a_{i-1} - \zeta)G_1^c + (a_i - \zeta)\right) \\
&\quad + \frac{\zeta}{i}\left(G_{i-1}^c + G_{i-2}^c + \cdots + G_1^c + 1\right) + \frac{1}{i}\left(a_1 c_{i-1} + a_2 c_{i-2} + \cdots + a_{i-1}\right),
\end{aligned}$$

where the second term in the RHS equals $\frac{\zeta}{i} c_i$, and the third term equals $\frac{i-1}{i} c_i$. Thus, we get $H_i^c = \frac{1}{i}\left((a_1 - \zeta)G_{i-1}^c + (a_2 - \zeta)G_{i-2}^c + \cdots + (a_{i-1} - \zeta)G_1^c + (a_i - \zeta)\right)$, for all i. By step (i), we have $G_i^c \geq 0$ for $i \geq 1$. Using the separation condition $a_i \geq \zeta$ for all i, we get $H_i^c \geq 0$ for all i.

(iii) We show that $H_i^g := g_{i+1} - g_i \frac{\zeta+i-1}{i} \leq 0$ for all $i \geq 1$. For $df = 2$ this statement follows with $\zeta = 1$ since $g_{i+1} = \frac{1}{2i}(g_i + g_{i-1} + ... + 1) = \frac{2i-1}{2i} g_i$ and hence (g_i) is decreasing. Let us now consider the case $df \geq 3$ with $\zeta = \frac{df-1}{2}$. As above we have $H_i^g = \frac{1}{i}\sum_{l=1}^{i}(b_l - \zeta)G_{i-l}^g$, where $G_i^g := g_{i+1} - g_i$. We plug in $b_l - \zeta = \frac{1}{2}\sum_{j=2}^{df-1}(\rho_j^l - 1) = \frac{1}{2}\sum_{j=2}^{df-1}(\rho_j - 1)(1 + \rho_j + ... + \rho_j^{l-1}) = \frac{1}{2}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{k=1}^{l}\rho_j^{k-1}$. Thus, we get:

$$
\begin{aligned}
H_i^g &= \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{l=1}^{i}\sum_{k=1}^{l}\rho_j^{k-1}G_{i-l}^g = \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{k=1}^{i}\rho_j^{k-1}\sum_{l=k}^{i}G_{i-l}^g \\
&= \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{k=1}^{i}\rho_j^{k-1}g_{i-k+1} = \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\left(g_i + \rho_j g_{i-1} + ... + \rho_j^{i-1}\right) \leq 0.
\end{aligned}
$$

(iv) The inequalities established in steps (ii) and (iii) imply $\frac{c_{i+1}}{c_i} \geq \frac{\zeta+i-1}{i}$ and $\frac{g_{i+1}}{g_i} \leq \frac{\zeta+i-1}{i}$ for all i. Then, we get $\frac{c_{i+1}}{c_i} \geq \frac{g_{i+1}}{g_i}$, that is equivalent to $\frac{c_{i+1}}{g_{i+1}} \geq \frac{c_i}{g_i}$, for all i, because the sequences c_i and g_i are strictly positive. The conclusion follows.

E Additional theory

E.1 Pseudo likelihood and PML estimator

The FA estimator is the PML estimator based on the Gaussian likelihood function obtained from the pseudo model $y_i = \mu + F\beta_i + \varepsilon_i$ with $\beta_i \sim N(0, I_k)$ and $\varepsilon_i \sim N(0, V_\varepsilon)$ mutually independent and i.i.d. across $i = 1, ..., n$. Then, $y_i \sim N(\mu, \Sigma(\theta))$ under this pseudo model, where $\Sigma(\theta) := FF' + V_\varepsilon$ and $\theta := (vec(F)', diag(V_\varepsilon)')' \in \mathbb{R}^r$ with $r = (k+1)T$. It yields the pseudo log-likelihood function $\hat{L}(\theta, \mu) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2n}\sum_{i=1}^{n}(y_i - \mu)'\Sigma(\theta)^{-1}(y_i - \mu) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Tr\left(\hat{V}_y\Sigma(\theta)^{-1}\right) - \frac{1}{2}(\bar{y} - \mu)'\Sigma(\theta)^{-1}(\bar{y} - \mu)$, up to constants, where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\hat{V}_y = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})'$. We concentrate out parameter μ to get its estimator $\hat{\mu} = \bar{y}$. Then, estimator $\hat{\theta} = (vec(\hat{F})', diag(\hat{V}_\varepsilon)')'$ is defined by the maximization of

$$
\hat{L}(\theta) := -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Tr\left(\hat{V}_y\Sigma(\theta)^{-1}\right), \tag{E.1}
$$

subject to the normalization restriction that $F'V_\varepsilon^{-1}F$ is a diagonal matrix, with diagonal elements ranked in decreasing order.[43]

E.2 Global identification and consistency

The population criterion $L_0(\theta)$ is defined in Appendix A, with $V_y = V_y^0 = \Sigma(\theta_0) = F_0F_0' + V_\varepsilon^0$.

Lemma 5 *The following conditions are equivalent: a) the true value θ_0 is the unique maximizer of $L_0(\theta)$ for $\theta \in \Theta$; b) $\Sigma(\theta) = \Sigma(\theta_0)$, $\theta \in \Theta \Rightarrow \theta = \theta_0$, up to sign changes in the columns of F. They yield the global identification in the FA model.*

In Lemma 5, condition a) is the standard identification condition for a M-estimator with population criterion $L_0(\theta)$. Condition (b) is the global identification condition based on the variance matrix as in Anderson and Rubin (1956). Condition (b) corresponds to our Assumption A.4.

Let us now establish the consistency of the FA estimators in our setting. Write $\hat{V}_y = \frac{1}{n}\sum_{i=1}^n(\varepsilon_i - \bar{\varepsilon})(\varepsilon_i - \bar{\varepsilon})' + F[\frac{1}{n}\sum_{i=1}^n(\beta_i - \bar{\beta})(\beta_i - \bar{\beta})']F' + F[\frac{1}{n}\sum_{i=1}^n(\beta_i - \bar{\beta})(\varepsilon_i - \bar{\varepsilon})'] + [\frac{1}{n}\sum_{i=1}^n(\varepsilon_i - \bar{\varepsilon})(\beta_i - \bar{\beta})']F'$, where $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^n \varepsilon_i$ and $\bar{\beta} = \frac{1}{n}\sum_{i=1}^n \beta_i$. Under the normalization in Assumption A.1 we have:

$$\hat{V}_y = \frac{1}{n}\varepsilon\varepsilon' - \bar{\varepsilon}\bar{\varepsilon}' + FF' + F\left(\frac{1}{n}\varepsilon\beta\right)' + \left(\frac{1}{n}\varepsilon\beta\right)F'. \tag{E.2}$$

Lemma 6 *Under Assumptions 1, 2, and A.2, A.3, as $n \to \infty$, we have: (a) $\bar{\varepsilon} = o_p(\frac{1}{n^{1/4}})$, (b) $\frac{1}{n}\varepsilon\varepsilon' \xrightarrow{p} V_\varepsilon^0$, and (c) $\frac{1}{n}\varepsilon\beta \xrightarrow{p} 0$.*

From Equation (E.2) and Lemma 6, we have $\hat{V}_y \xrightarrow{p} V_y^0$. Thus, $\hat{L}(\theta)$ converges in probability to $L_0(\theta)$ as $n \to \infty$, uniformly over Θ compact. From standard results on M-estimators, we get consistency of $\hat{\theta}$. Moreover, from $\bar{y} = \mu + \bar{\varepsilon}$, we get the consistency of $\hat{\mu}$.

[43]If the risk-free rate vector is considered observable, we can rewrite the model as $\tilde{y}_i = F\tilde{\beta}_i + \varepsilon_i = \mu + F\beta_i + \varepsilon_i$, where $\tilde{y}_i = y_i - r_f$ is the vector of excess returns and $\mu = F\mu_{\tilde{\beta}}$. It corresponds to a constrained model with parameters θ and $\mu_{\tilde{\beta}}$. The maximization of the corresponding Gaussian pseudo likelihood function leads to a constrained FA estimator, that we do not consider in this paper since it does not match a standard FA formulation.

Proposition 5 *Under Assumptions 1, 2, and A.2-A.4, the FA estimators $\hat{F}$, $\hat{V}_\varepsilon$ and $\hat{\mu}$ are consistent as $n \to \infty$ and T is fixed.*

Anderson and Rubin (1956) establish consistency in Theorem 12.1 (see beginning of the proof, page 145) within a Gaussian ML framework. Anderson and Amemiya (1988) provide a version of this result in their Theorem 1 for generic distribution of the data, dispensing for compacity of the parameter set but using a more restrictive identification condition.

E.3 Asymptotic expansions of estimators $\hat{V}_\varepsilon$ and $\hat{F}$

The FA estimators $\hat{V}_\varepsilon$ and $\hat{F}$ are consistent M-estimators under nonlinear constraints, and admit expansions at first order for fixed T and $n \to \infty$, namely $\hat{V}_\varepsilon = \tilde{V}_\varepsilon + \frac{1}{\sqrt{n}}\Psi_\varepsilon + o_p(\frac{1}{\sqrt{n}})$ and $\hat{F}_j = F_j + \frac{1}{\sqrt{n}}\Psi_{F_j} + o_p(\frac{1}{\sqrt{n}})$ (see Appendix E.5.1). The next proposition (new to the literature) characterizes the diagonal random matrix Ψ_ε and the random vectors Ψ_{F_j} by using conditions (FA1) and (FA2) in Section 2 (see proof at the end of the section).

Proposition 6 *Under Assumptions 1, 2, and A.1-A.4, A.6, we have (a) for $j = 1, ..., k$*

$$\Psi_{F_j} = R_j(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \Lambda_j \Psi_\varepsilon V_\varepsilon^{-1}F_j, \tag{E.3}$$

where $R_j := \frac{1}{2\gamma_j}P_{F_j,V_\varepsilon} + \frac{1}{\gamma_j}M_{F,V_\varepsilon} + \sum_{\ell=1,\ell\neq j}^{k} \frac{1}{\gamma_j-\gamma_\ell}P_{F_\ell,V_\varepsilon}$ and $\Lambda_j := -\sum_{\ell=1,\ell\neq j}^{k} \frac{\gamma_l}{\gamma_j-\gamma_\ell}P_{F_\ell,V_\varepsilon}$ and $P_{F_j,V_\varepsilon} = F_j(F_j'V_\varepsilon^{-1}F_j)^{-1}F_j'V_\varepsilon^{-1} = \frac{1}{\gamma_j}F_jF_j'V_\varepsilon^{-1}$ is the GLS orthogonal projection onto F_j. Further, (b) the diagonal matrix Ψ_ε is such that:

$$diag\left(M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M_{F,V_\varepsilon}'\right) = 0. \tag{E.4}$$

Equation (E.3) yields the asymptotic expansion of the eigenvectors by accounting for estimation errors of matrix $\hat{V}_y\hat{V}_\varepsilon^{-1}$ (first term) and of the normalization constraint (second term). To interpret Equation (E.4), we can observe that the matrix $M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M_{F,V_\varepsilon}'$ yields the first-order term in the asymptotic expansion of $\sqrt{n}\hat{S}$ (up to the left- and right-multiplication by diagonal matrix $V_\varepsilon^{-1/2}$). Thus, Equation (E.4) is implied by the property that the diagonal terms of matrix $\hat{S}$ are equal to zero as stated in Lemma 1 (c).

Let us now give the explicit expression of Ψ_ε. By using $M_{F,V_\varepsilon}\Psi_y M'_{F,V_\varepsilon} = M_{F,V_\varepsilon}Z_n M'_{F,V_\varepsilon}$, we can rewrite Equation (E.4) as $diag\left(M_{F,V_\varepsilon}(Z_n - \Psi_\varepsilon)M'_{F,V_\varepsilon}\right) = 0$. Now, since Ψ_ε is diagonal, we have $diag\left(M_{F,V_\varepsilon}\Psi_\varepsilon M'_{F,V_\varepsilon}\right) = M^{\odot 2}_{F,V_\varepsilon} diag(\Psi_\varepsilon)$, where $M^{\odot 2}_{F,V_\varepsilon} = M_{F,V_\varepsilon} \odot M_{F,V_\varepsilon}$. Thus, we get:

$$M^{\odot 2}_{F,V_\varepsilon} diag(\Psi_\varepsilon) = diag(M_{F,V_\varepsilon}Z_n M'_{F,V_\varepsilon}). \tag{E.5}$$

To have a unique solution for vector $diag(\Psi_\varepsilon)$, we need the non-singularity of the $T \times T$ matrix $M^{\odot 2}_{F,V_\varepsilon}$. It is the local identification condition in the FA model stated in Assumption A.5. Let us write $G = [g_1 : \cdots : g_{T-k}]$. Then, we have $M_{F,V_\varepsilon} = GG'V_\varepsilon^{-1} = \sum_{j=1}^{T-k} g_j(V_\varepsilon^{-1}g_j)'$, and so we get the Hadamard product $M^{\odot 2}_{F,V_\varepsilon} = \sum_{i,j=1}^{T-k}[g_i(V_\varepsilon^{-1}g_i)'] \odot [g_j(V_\varepsilon^{-1}g_j)'] = \left[\sum_{i,j=1}^{T-k}(g_i \odot g_j)(g_i \odot g_j)'\right]$ $V_\varepsilon^{-2} = 2\left(\boldsymbol{X}'\boldsymbol{X}\right)V_\varepsilon^{-2}$.[44] Hence, we can state the local identification condition in Assumption A.5 as a full-rank condition for matrix $\boldsymbol{X}$, analogously as in linear regression (Lemma 7). In Lemma 7 in Appendix E.4 i), we also show equivalence with invertibility of the bordered Hessian, i.e., the Hessian of the Lagrangian function in a constrained M-estimation.

Under Assumption A.5, we get from Equation (E.5):

$$\Psi_\varepsilon = \mathcal{T}_{F,V_\varepsilon}(Z_n), \tag{E.6}$$

where $\mathcal{T}_{F,V_\varepsilon}(V) := diag\left([M^{\odot 2}_{F,V_\varepsilon}]^{-1} diag(M_{F,V_\varepsilon} V M'_{F,V_\varepsilon})\right)$, for any matrix V. Mapping $\mathcal{T}_{F,V_\varepsilon}(\cdot)$ is linear and such that $\mathcal{T}_{F,V_\varepsilon}(V) = V$, for a diagonal matrix V. We have $diag(M_{F,V_\varepsilon}Z_n M'_{F,V_\varepsilon}) = diag\left(GZ_n^*G'\right) = 2\boldsymbol{X}'vech\left(Z_n^*\right)$,[45] and so

$$diag(\Psi_\varepsilon) = V_\varepsilon^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'vech\left(Z_n^*\right). \tag{E.7}$$

[44] Let us recall the following property of the Hadamard product: $(ab') \odot (cd') = (a \odot c)(b \odot d)'$ for conformable vectors a, b, c, d. The last equalitiy because $\boldsymbol{X}' = \left[\frac{1}{\sqrt{2}}g_1 \odot g_1 \; : \; \cdots \; : \; \frac{1}{\sqrt{2}}g_{T-k} \odot g_{T-k} \; : \; \{g_i \odot g_j\}_{i<j}\right]$ (see beginning of the proof of Proposition 2 (a)) .

[45] We have $diag(GAG') = 2\boldsymbol{X}'vech(A)$ for any $T \times T$ symmetric matrix A; see beginning of the proof of Proposition 1 (a).

Anderson and Rubin (1956), Theorem 12.1, show that the FA estimator is asymptotically normal if $\sqrt{n}(\hat{V}_y - V_y)$ is asymptotically normal. They use a linearization of the first-order conditions similar as the one of Proposition 6. Their Equation (12.16) corresponds to our Equation (E.4). However, they only provide an implicit characterization of the Ψ_{F_j} and not an explicit expression for Ψ_ε and Ψ_{F_j} in terms of asymptotically Gaussian random matrices like Z_n as we do. These key developments pave the way to establishing the asymptotic distributions of estimators $\hat{F}$ and $\hat{V}_\varepsilon$ in general settings, that we cover in Appendix E.5.

Proof of Proposition 6: From (E.2) and Lemma 6 we have $\hat{V}_y = \tilde{V}_y + \frac{1}{\sqrt{n}}\Psi_y + o_p(\frac{1}{\sqrt{n}})$, where $\tilde{V}_y = FF' + \tilde{V}_\varepsilon$ and $\Psi_y = \frac{1}{\sqrt{n}}(\varepsilon\beta F' + F\beta'\varepsilon') + \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right)$. Let us substitute this expansion for $\hat{V}_y$ into (FA2) and rearrange to obtain $\hat{F}\hat{\Gamma} - FF'\hat{V}_\varepsilon^{-1}\hat{F} = \frac{1}{\sqrt{n}}\Psi_y\hat{V}_\varepsilon^{-1}\hat{F} + (\tilde{V}_\varepsilon\hat{V}_\varepsilon^{-1} - I_T)\hat{F} + o_p(\frac{1}{\sqrt{n}})$, where $\hat{\Gamma} = \hat{F}'\hat{V}_\varepsilon^{-1}\hat{F} = diag(\hat{\gamma}_1, \ldots, \hat{\gamma}_k)$. From $\hat{V}_\varepsilon = \tilde{V}_\varepsilon + \frac{1}{\sqrt{n}}\Psi_\varepsilon + o_p(\frac{1}{\sqrt{n}})$, we have $\tilde{V}_\varepsilon\hat{V}_\varepsilon^{-1} - I_T = -\frac{1}{\sqrt{n}}\Psi_\varepsilon\hat{V}_\varepsilon^{-1} + o_p(\frac{1}{\sqrt{n}})$. Substituting into the above equation and right multiplying both sides by $(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}$ gives $\hat{F}\hat{\mathcal{D}} - F = \frac{1}{\sqrt{n}}(\Psi_y - \Psi_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{F}(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1} + o_p(\frac{1}{\sqrt{n}})$, where $\hat{\mathcal{D}} := \hat{\Gamma}(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}$. By the root-$n$ convergence of the FA estimates (see Section E.5.1), we get

$$\hat{F}\hat{\mathcal{D}} - F = \frac{1}{\sqrt{n}}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}}), \tag{E.8}$$

and $\hat{\mathcal{D}} = I_k + O_p(\frac{1}{\sqrt{n}})$, where $\Gamma = diag(\gamma_1, ..., \gamma_k)$. We can push the expansion by plugging into (E.8) the expansion of $\hat{\mathcal{D}}$. We have $F'\hat{V}_\varepsilon^{-1}\hat{F} = [I_k - (\hat{F} - F)'\hat{V}_\varepsilon^{-1}\hat{F}\hat{\Gamma}^{-1}]\hat{\Gamma}$, so that $\hat{\mathcal{D}} = [I_k - (\hat{F} - F)'\hat{V}_\varepsilon^{-1}\hat{F}\hat{\Gamma}^{-1}]^{-1} = I_k + (\hat{F} - F)'V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}})$. By plugging into (E.8), we get:

$$\hat{F} - F + F[(\hat{F} - F)'V_\varepsilon^{-1}F\Gamma^{-1}] = \frac{1}{\sqrt{n}}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}}). \tag{E.9}$$

By multiplying both sides with M_{F,V_ε}, we get $M_{F,V_\varepsilon}(\hat{F} - F) = \frac{1}{\sqrt{n}}M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}})$. Then, $\hat{F} - F = \frac{1}{\sqrt{n}}M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + \frac{1}{\sqrt{n}}FA + o_p(\frac{1}{\sqrt{n}})$, where A is a random $k \times k$ matrix to be determined next. By plugging into (E.9), we get $F(A + A') = P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}})$. By multiplying both sides by $\frac{1}{2}\Gamma^{-1}F'V_\varepsilon^{-1}$ and using $F'V_\varepsilon^{-1}P_{F,V_\varepsilon} = F'V_\varepsilon^{-1}$, we get the symmetric part of matrix A, i.e., $\frac{1}{2}(A + A') = \frac{1}{2}\Gamma^{-1}F'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1}$

(we include higher-order terms in the remainder $o_p(\frac{1}{\sqrt{n}})$). Thus, $\hat{F} - F = \frac{1}{\sqrt{n}}\Psi_F + o_p(\frac{1}{\sqrt{n}})$, where

$$\Psi_F = M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + \frac{1}{2}P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + F\tilde{A}, \qquad \text{(E.10)}$$

and $\tilde{A} = \frac{1}{2}(A - A')$ is an antisymmetric $k \times k$ random matrix. To find the antisymmetric matrix $\tilde{A} = (\tilde{a}_{\ell,j})$, we use that $\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F}$ is diagonal. Plugging the expansions of the FA estimates, for the term at order $1/\sqrt{n}$ we get that the out-of-diagonal elements of matrix $\Psi_F'V_\varepsilon^{-1}F + F'V_\varepsilon^{-1}\Psi_F - F'V_\varepsilon^{-1}\Psi_\varepsilon V_\varepsilon^{-1}F = \frac{1}{2}\Gamma^{-1}F'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F + \frac{1}{2}F'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + \Gamma\tilde{A} - \tilde{A}\Gamma - F'V_\varepsilon^{-1}\Psi_\varepsilon V_\varepsilon^{-1}F$ are nil. Setting the (ℓ, j) element of this matrix equal to 0, we get $\tilde{a}_{\ell,j} = -\tilde{a}_{j,\ell} = \frac{1}{\gamma_j - \gamma_\ell}\left[\frac{1}{2}(\frac{1}{\gamma_j} + \frac{1}{\gamma_\ell})F_\ell'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j - F_\ell'V_\varepsilon^{-1}\Psi_\varepsilon V_\varepsilon^{-1}F_j\right]$, for $j \neq \ell$. Then, from Equation (E.10), the jth column of Ψ_F is $\Psi_{F_j} = \frac{1}{\gamma_j}M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \frac{1}{2\gamma_j}P_{F_j,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \sum_{\ell=1:\ell\neq j}^k \frac{1}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j - \sum_{\ell=1:\ell\neq j}^k \frac{\gamma_\ell}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}\Psi_\varepsilon V_\varepsilon^{-1}F_j$, where we use $P_{F,V_\varepsilon} = \sum_{\ell=1}^k P_{F_\ell,V_\varepsilon}$. Part (a) follows.

Let us now prove part (b). The asymptotic expansion of condition (FA1) yields:

$$diag(\Psi_y) = diag\left(\sum_{j=1}^k (F_j\Psi_{F_j}' + \Psi_{F_j}F_j') + \Psi_\varepsilon\right). \qquad \text{(E.11)}$$

From part (a) and the definition of P_{F_j,V_ε} we have $\sum_{j=1}^k \Psi_{F_j}F_j' = \frac{1}{2}\sum_{j=1}^k P_{F_j,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F_j,V_\varepsilon}' + M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F,V_\varepsilon}' + \sum_{\ell\neq j}\frac{\gamma_j}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F_j,V_\varepsilon}' - \sum_{\ell\neq j}^k \frac{\gamma_\ell\gamma_j}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}\Psi_\varepsilon P_{F_j,V_\varepsilon}' =: N_1 + N_2 + N_3 + N_4$, where $P_{F,V_\varepsilon} = \sum_{j=1}^k P_{F_j,V_\varepsilon} = I_T - M_{F,V_\varepsilon}$ and $\sum_{\ell\neq j}$ denotes the double sum over $j, \ell = 1, ..., k$ such that $\ell \neq j$. Matrix N_1 is symmetric and it contributes $2N_1$ to the RHS of (E.11). Instead, matrix N_4 is antisymmetric (it can be seen by interchanging indices j and ℓ in the summation) and it does not contribute to the RHS of (E.11). For matrix N_3 we have $N_3 + N_3' = \sum_{\ell\neq j}\frac{\gamma_j}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F_j,V_\varepsilon}' + \sum_{\ell\neq j}\frac{\gamma_\ell}{\gamma_\ell - \gamma_j}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F_j,V_\varepsilon}' = \sum_{\ell\neq j}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F_j,V_\varepsilon}' = \sum_{\ell,j}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F_j,V_\varepsilon}' - \sum_j P_{F_j,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F_j,V_\varepsilon}' = P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F,V_\varepsilon}' - 2N_1$, where we have interchanged ℓ and j in the first equality when writing N_3'. Thus, we get:

$$\begin{aligned}\sum_{j=1}^k (F_j\Psi_{F_j}' + \Psi_{F_j}F_j') &= M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F,V_\varepsilon}' + P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M_{F,V_\varepsilon}' + P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P_{F,V_\varepsilon}' \\ &= (\Psi_y - \Psi_\varepsilon) - M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M_{F,V_\varepsilon}'. \qquad \text{(E.12)}\end{aligned}$$

Then, Equation (E.11) with (E.12) yields Equation (E.4).

E.4 Local analysis of the first-order conditions of FA estimators

Consider the criterion $L(\theta) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Tr\left(V_y\Sigma(\theta)\right)$, where V_y is a p.d. matrix in a neighbourhood of V_y^0. In our Assumptions, θ_0 is an interior point of Θ. Let $\theta^* = (vec(F^*)', diag(V_\varepsilon^*)')'$ denote the maximizer of $L(\theta)$ subject to $\theta \in \Theta$. According to Anderson (2003), the first-order conditions (FOC) for the maximization of $L(\theta)$ are: (a) $diag(V_y) = diag(F^*(F^*)' + V_\varepsilon^*)$ and (b) F^* is the matrix of eigenvectors of $V_y(V_\varepsilon^*)^{-1}$ associated to the k largest eigenvalues $1+\gamma_j^*$ for $j = 1, ..., k$, normalized such that $(F^*)'(V_\varepsilon^*)^{-1}F^* = diag(\gamma_1^*, ..., \gamma_k^*)$.

i) Local identification

Let $V_y = V_y^0$. The true values F_0 and V_ε^0 solve the FOC. Let $F = F_0 + \epsilon\Psi_F^\epsilon$ and $V_\varepsilon = V_\varepsilon^0 + \epsilon\Psi_{V_\varepsilon}^\epsilon$, where ϵ is a small scalar and $\Psi_F^\epsilon, \Psi_{V_\varepsilon}^\epsilon$ are deterministic conformable matrices, be in a neighbourhood of F_0 and V_ε^0 and solve the FOC up to terms $O(\epsilon^2)$. The model is locally identified if, and only if, it implies $\Psi_{V_\varepsilon}^\epsilon = 0$ and $\Psi_F^\epsilon = 0$.

Lemma 7 *Under Assumption 1, the following four conditions are equivalent: (a) Matrix $M_{F_0,V_\varepsilon^0}^{\odot 2}$ is non-singular, (b) Matrix $\boldsymbol{X}$ is full-rank, (c) Matrix $\Phi^{\odot 2}$ is non-singular, where $\Phi := V_\varepsilon^0 - F_0(F_0'(V_\varepsilon^0)^{-1}F_0)^{-1}F_0'$, (d) Matrix $B_0'J_0B_0$ is non-singular, where $J_0 := -\frac{\partial^2 L_0(\theta_0)}{\partial\theta\partial\theta'}$ and B_0 is any full-rank $r \times (r - \frac{1}{2}k(k-1))$ matrix such that $\frac{\partial g(\theta_0)}{\partial\theta'}B_0 = 0$, for $g(\theta) = \{[F'V_\varepsilon^{-1}F]_{i,j}\}_{i<j}$ the $\frac{1}{2}k(k-1)$ dimensional vector of the constraints. They yield the local identification of our model.*

In Lemma 7, condition (a) corresponds to Assumption A.5 and is equivalent to condition (b) that $\boldsymbol{X}$ is full-rank. Condition (c) is used in Theorem 5.9 of Anderson and Rubin (1956) to show local identification. Condition (d) involves the second-order partial derivatives of the population criterion function. While the Hessian matrix J_0 itself is singular because of the rotational invariance of the model to latent factors, the second-order partial derivatives matrix along parameter directions, which are in the tangent plan to the contraint set, is non-singular. Condition (d) is equivalent to invertibility of the bordered Hessian.

ii) Local misspecification

Now, let $V_y = V_y^0 + \epsilon\Psi_y^\epsilon$ be in a neighbourhood of V_y^0. Let $F^* = F_0 + \epsilon\Psi_F^\epsilon + O(\epsilon^2)$ and $V_\varepsilon^* = V_\varepsilon^0 + \epsilon\Psi_{V_\varepsilon}^\epsilon + O(\epsilon^2)$ be the solutions of the FOC. Consider $V_y - \Sigma^*$, where $\Sigma^* = F^*(F^*)' + V_\varepsilon^*$, i.e., the difference between variance V_y and its k-factor approximation with population FA. We want to find the first-order development of $V_y - \Sigma^*$ for small ϵ. From the FOC, we have that the diagonal of such symmetric matrix is null, but not necessarily the out-of-diagonal elements.

From the arguments in the proof of Proposition 6, Equations (E.11) and (E.12), we get:

$$\Psi_F^\epsilon F_0' + F_0(\Psi_F^\epsilon)' = \Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon - M_{F_0,V_\varepsilon^0}(\Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon)M_{F_0,V_\varepsilon^0}', \tag{E.13}$$

$$diag(M_{F_0,V_\varepsilon^0}(\Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon)M_{F_0,V_\varepsilon^0}') = 0. \tag{E.14}$$

As in Section E.3, Equation (E.14) yields:

$$diag(\Psi_{V_\varepsilon}^\epsilon) = (V_\varepsilon^0)^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}' vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right). \tag{E.15}$$

Now, using Equation (E.13), we get $V_y - \Sigma^* = \epsilon\left(\Psi_y^\epsilon - F_0(\Psi_F^\epsilon)' - \Psi_F^\epsilon F_0' - \Psi_{V_\varepsilon}^\epsilon\right) + O(\epsilon^2)$ $= \epsilon M_{F_0,V_\varepsilon^0}(\Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon)M_{F_0,V_\varepsilon^0}' + O(\epsilon^2) = \epsilon G_0\Delta^* G_0' + O(\epsilon^2)$, where $\Delta^* := G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0 - G_0'(V_\varepsilon^0)^{-1}\Psi_{V_\varepsilon}^\epsilon(V_\varepsilon^0)^{-1}G_0$. Using that $vech(G_0' diag(a) G_0) = \boldsymbol{X}a$, and Equation (E.15), the vectorized form of matrix Δ^* is: $vech(\Delta^*) = vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right) - \boldsymbol{X}(V_\varepsilon^0)^{-2} diag(\Psi_{V_\varepsilon}^\epsilon) = M_{\boldsymbol{X}} vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right)$. Thus, we have shown that, at first order in ϵ, the difference between $V_y = V_y^0 + \epsilon\Psi_y^\epsilon$ and the FA k-factor approximation Σ^* is $\epsilon G_0\Delta^* G_0'$, with $vech(\Delta^*) = M_{\boldsymbol{X}} vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right)$. It shows that the small perturbation $\epsilon\Psi_y^\epsilon$ around V_y^0 keeps the DGP within the k-factor specification (at first order) if, and only if, we have that vector $vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right)$ is spanned by the columns of $\boldsymbol{X}$.

Consider $\Psi_y^\epsilon = H\xi\xi' H'$, where $H := [F_0 : G_0]$ and vector $\xi = (\xi_F', \xi_G')'$ are partitioned in k and $T-k$ dimensional components, which corresponds to a local alternative with $(k+1)$th factor

$H\xi$ and small loading ϵ in the perturbation $\epsilon\Psi_y^\epsilon$. Then, we have $G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0 = \xi_G\xi_G'$ since $F_0'(V_\varepsilon^0)^{-1}G_0 = 0$ and $G_0'(V_\varepsilon^0)^{-1}G_0 = I_{T-k}$. Thus, $vech(\Delta^*) = M_{\boldsymbol{X}} vech\,(\xi_G\xi_G')$. Hence, it is only the component of $vech\,(\xi_G\xi_G')$ that is orthogonal to the range of $\boldsymbol{X}$, which generates a local deviation from a k-factor specification through the multiplication by the projection matrix $M_{\boldsymbol{X}}$. It clarifies the role of the projector in the local power. On the contrary, the component spanned by the columns of $\boldsymbol{X}$ can be "absorbed" in the k-factor specification by a redefinition of the factor F and the variance V_ε through F^* and V_ε^*.

E.5 Feasible asymptotic normality of the FA estimators

E.5.1 Asymptotic expansions

We first establish the asymptotic expansion of $\hat{\theta}$ along the lines of pseudo maximum likelihood estimators (White (1982)). The sample criterion is $\hat{L}(\theta)$ given in Equation (E.1), where $\theta = (vec(F)', diag(V_\varepsilon)')'$ is subject to the nonlinear vector constraint $g(\theta) := \{[F'V_\varepsilon^{-1}F]_{i,j}\}_{i<j} = 0$, i.e., matrix $F'V_\varepsilon^{-1}F$ is diagonal. By standard methods for constrained M-estimators, we consider the FOC of the Lagrangian function: $\frac{\partial \hat{L}(\hat{\theta})}{\partial\theta} - \frac{\partial g(\hat{\theta})'}{\partial\theta}\hat{\lambda}_L = 0$ and $g(\hat{\theta}) = 0$, where $\hat{\lambda}_L$ is the $\frac{1}{2}k(k-1)$ dimensional vector of estimated Lagrange multipliers. Define vector $\tilde{\theta} := \left(vec(F_0)', diag(\tilde{V}_\varepsilon)'\right)'$, which also satisfies the constraint $g(\tilde{\theta}) = 0$ by the in-sample factor normalization. We apply the mean value theorem to the FOC around $\tilde{\theta}$ and get:

$$\hat{J}(\bar{\theta})\sqrt{n}(\hat{\theta}-\tilde{\theta}) + A(\hat{\theta})\sqrt{n}\hat{\lambda}_L = \sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial\theta}, \quad \text{(E.16)}$$

$$A(\bar{\theta})'\sqrt{n}(\hat{\theta}-\tilde{\theta}) = 0, \quad \text{(E.17)}$$

where $\hat{J}(\theta) := -\frac{\partial^2 \hat{L}(\theta)}{\partial\theta\partial\theta'}$ is the $r \times r$ Hessian matrix, $A(\theta) := \frac{\partial g(\theta)'}{\partial\theta}$ is the $r \times \frac{1}{2}k(k-1)$ dimensional gradient matrix of the constraint function, and $\bar{\theta}$ is a mean value vector between $\hat{\theta}$ and $\tilde{\theta}$ componentwise. Matrix $A(\theta)$ is full rank for θ in a neighbourhood of θ_0. For any θ define the $r \times (r - \frac{1}{2}k(k-1))$ matrix $B(\theta)$ with orthonormal columns that span the orthogonal complement of the range of $A(\theta)$. Matrix function $B(\theta)$ is continuous in θ in a neighbourhood of θ_0.[46] Then, by multiplying Equation (E.16) times $B(\hat{\theta})'$ to get rid of the Lagrange multiplier vector, using the identity $I_r = A(\theta)(A(\theta)'A(\theta))^{-1}A(\theta)' + B(\theta)B(\theta)'$ for $\theta = \bar{\theta}$ and Equation (E.17), we get $[B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})]B(\bar{\theta})'\sqrt{n}(\hat{\theta} - \tilde{\theta}) = B(\hat{\theta})'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta}$. By the uniform convergence of $\hat{J}(\theta)$ to $J(\theta) := -\frac{\partial^2 L_0(\theta)}{\partial \theta \partial \theta'}$, and the consistency of the FA estimator $\hat{\theta}$ (Section E.2), matrix $B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})$ converges to $B_0'J_0B_0$, where $J_0 := J(\theta_0)$ and $B_0 := B(\theta_0)$. Matrix $B_0'J_0B_0$ is invertible under the local identification Assumption A.5 (see Lemma 7 condition d)). Then, $B(\bar{\theta})'\sqrt{n}(\hat{\theta} - \tilde{\theta}) = [B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})]^{-1}B(\hat{\theta})'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta}$ w.p.a. 1. By using again $I_r = A(\bar{\theta})(A(\bar{\theta})'A(\bar{\theta}))^{-1}A(\bar{\theta})' + B(\bar{\theta})B(\bar{\theta})'$ and Equation (E.17), we get $\sqrt{n}(\hat{\theta} - \tilde{\theta}) = B(\bar{\theta})[B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})]^{-1}B(\hat{\theta})'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta}$. The distributional results established below imply $\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} = O_p(1)$. Thus, we get $\sqrt{n}$-consistency:

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}) = B_0(B_0'J_0B_0)^{-1}B_0'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} + o_p(1). \tag{E.18}$$

Let us now find the score $\frac{\partial \hat{L}(\theta)}{\partial \theta}$. We have $\frac{\partial \hat{L}(\theta)}{\partial \theta} = \left(\frac{\partial vec(\Sigma(\theta))}{\partial \theta'}\right)' vec\left(\frac{\partial \hat{L}(\theta)}{\partial \Sigma}\right)$, where $vec\left(\frac{\partial \hat{L}(\theta)}{\partial \Sigma}\right) = \frac{1}{2}\left(\Sigma(\theta)^{-1} \otimes \Sigma(\theta)^{-1}\right) vec\left(\hat{V}_y - \Sigma(\theta)\right)$. Moreover, by using $vec(\Sigma(\theta)) = \sum_{j=1}^k F_j \otimes F_j + [e_1 \otimes e_1 \,:\, \cdots \,:\, e_T \otimes e_T] diag(V_\varepsilon)$, where e_t is the t-th column of I_T, we get: $\frac{\partial vec(\Sigma(\theta))}{\partial \theta'} = [(I_T \otimes F_1) + (F_1 \otimes I_T) \,:\, \cdots \, e_1 \otimes e_1 \,:\, \cdots \,:\, e_T \otimes e_T]$. Thus, we get: $\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} = \frac{1}{2}\left(\frac{\partial vec(\Sigma(\tilde{\theta}))}{\partial \theta'}\right)' \left(\tilde{V}_y^{-1} \otimes \tilde{V}_y^{-1}\right) \sqrt{n} vec\left(\hat{V}_y - \tilde{V}_y\right)$. From Equation (E.2) and Lemma 6 we have $\hat{V}_y = \tilde{V}_y + \frac{1}{\sqrt{n}}(Z_n + W_nF' + FW_n') + o_p(\frac{1}{\sqrt{n}})$, where $W_n := \frac{1}{\sqrt{n}}\varepsilon\beta$. Thus, $\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} = \frac{1}{2}\left(\frac{\partial vec(\Sigma(\theta_0))}{\partial \theta'}\right)' \left(V_y^{-1} \otimes V_y^{-1}\right) vec\left(W_nF' + FW_n' + Z_n\right) + o_p(1)$ and, from Equation (E.18), we get:

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}) = \frac{1}{2}B_0\left(B_0'J_0B_0\right)^{-1}B_0'\left(\frac{\partial vec(\Sigma(\theta_0))}{\partial \theta'}\right)' \left(V_y^{-1} \otimes V_y^{-1}\right) vec\left(W_nF' + FW_n' + Z_n\right) + o_p(1). \tag{E.19}$$

[46] Matrix $B(\theta)$ is uniquely defined up to rotation and sign changes in their columns. We can pick a unique representer such that matrix $B(\theta)$ is locally continuous, e.g., by taking $B(\theta) = \tilde{B}(\theta)[\tilde{B}(\theta)'\tilde{B}(\theta)]^{-1/2}$, where matrix $\tilde{B}(\theta)$ consists of the first $r - \frac{1}{2}k(k-1)$ columns of $I_r - A(\theta)[A(\theta)'A(\theta)]^{-1}A(\theta)'$, if those columns are linearly independent.

E.5.2 Asymptotic normality

In this subsection, we establish the asymptotic normality of estimators $\hat{F}$ and $\hat{V}_\varepsilon$. From Lemma 2, as $n \to \infty$ and T is fixed, we have the Gaussian distributional limit $Z_n \Rightarrow Z$ with $vech(Z) \sim N(0, \Omega_Z)$, where the asymptotic variance Ω_Z is related to the asymptotic variance Ω of $\mathscr{Z}$ such that $Cov(Z_{ts}, Z_{rp}) = \sqrt{V_{\varepsilon,tt}V_{\varepsilon,ss}V_{\varepsilon,rr}V_{\varepsilon,pp}}Cov(\mathscr{Z}_{ts}, \mathscr{Z}_{rp})$. Moreover, $Z_n^* \Rightarrow Z^* = G'V_\varepsilon^{-1}ZV_\varepsilon^{-1}G$ and $\bar{Z}_n := Z_n - \mathcal{T}_{F,V_\varepsilon}(Z_n) \Rightarrow \bar{Z}$, where $\bar{Z} = Z - \mathcal{T}_{F,V_\varepsilon}(Z) = Z - V_\varepsilon^2 diag\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z^*)\right)$ (see (E.7)). The distributional limit of W_n is given next.

Lemma 8 *Under Assumptions 1, 2 and A.2, A.3, A.8, as $n \to \infty$, (a) we have $W_n \Rightarrow \bar{W}$, where $vec(\bar{W}) \sim N(0, \Omega_W)$ with $\Omega_W = Q_\beta \otimes V_\varepsilon$, and (b) if additionally $E[w_{i,t}w_{i,r}w_{i,s}] = 0$, for all t, r, s and i, then Z and $\bar{W}$ are independent.*

We get the following proposition from Lemmas 2 and 8 (see proof at the end of the section).

Proposition 7 *Under Assumptions 1-2 and A.1-A.6, A.8, as $n \to \infty$ and T is fixed, for $j = 1, .., k$:*

$$\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) \Rightarrow V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z^*), \tag{E.20}$$

$$\sqrt{n}(\hat{F}_j - F_j) \Rightarrow R_j(\bar{W}F' + F\bar{W}' + \bar{Z})V_\varepsilon^{-1}F_j + \Lambda_j\{[V_\varepsilon(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z^*)] \odot F_j\}, \tag{E.21}$$

$$\sqrt{n}(\hat{F}_j\hat{\mathcal{D}} - F_j) \Rightarrow \frac{1}{\gamma_j}(\bar{W}F' + F\bar{W}' + \bar{Z})V_\varepsilon^{-1}F_j, \tag{E.22}$$

where deterministic matrices R_j and Λ_j are defined in Proposition 6, and $\hat{\mathcal{D}} := \hat{\Gamma}(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}$ and $\hat{\Gamma} := diag(\hat{\gamma}_1, ..., \hat{\gamma}_k)$.

The joint asymptotic Gaussian distribution of the FA estimators involves the Gaussian matrices Z^*, $\bar{Z}$ and $\bar{W}$, the former two being symmetric. The asymptotic distribution of $\hat{V}_\varepsilon$ involves re-centering around $\tilde{V}_\varepsilon = \frac{1}{n}\sum_{i=1}^n E[\varepsilon_i\varepsilon_i']$, i.e., the finite-sample average cross-moments of errors, and not V_ε. For the asymptotic distribution of any functional that depends on F up to one-to-one transformations of its columns, we can use the Gaussian law of (E.22) involving $\bar{W}$ and $\bar{Z}$ only.

The asymptotic expansions (E.20)-(E.21) characterize explicitly the matrices $C_1(\theta)$ and $C_2(\theta)$ that appear in Theorem 2 in Anderson and Amemiya (1988). Their derivation is based on an asymptotic normality argument treating $\hat{\theta}$ as a M-estimator, see Section C.2. However, neither the asymptotic variance nor a feasible CLT are given in Anderson and Amemiya (1988). We cannot use their results for our empirics.

To further compare our Proposition 7 with Theorem 2 in Anderson and Amemiya (1988), let $\bar{Z} = Z - \mathcal{T}_{F,V_\varepsilon}(Z) = \check{Z} - \mathcal{T}_{F,V_\varepsilon}(\check{Z})$, where $\check{Z} := Z - diag(Z)$ is the symmetric matrix of the off-diagonal elements of Z with zeros on the diagonal.[47] Hence, the zero-mean Gaussian matrix $\bar{Z}$ only involves the off-diagonal elements of Z. Moreover, since $V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(\Delta_n^*) = V_\varepsilon^2 diag(V_\varepsilon^{-1}\Delta_n V_\varepsilon^{-1}) = diag(\Delta_n)$ for a diagonal matrix Δ_n and $\Delta_n^* := G'V_\varepsilon^{-1}\Delta_n V_\varepsilon^{-1}G$, we can write the asymptotic expansion of $\hat{V}_\varepsilon$ as $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) = V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(\check{Z}_n^*) + diag(Z_n) + o_p(1)$, where $\check{Z}_n^* = G'V_\varepsilon^{-1}\check{Z}_n V_\varepsilon^{-1}G$ and $\check{Z}_n := Z_n - diag(Z_n)$. Thus, we get: $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) \Rightarrow V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(\check{Z}^*) + diag(Z)$, where $\check{Z}^* = G'V_\varepsilon^{-1}\check{Z}V_\varepsilon^{-1}G$. Hence, the asymptotic distribution of the FA estimators depends on the diagonal elements of Z via term $diag(Z)$ in the asymptotic distribution of $\hat{V}_\varepsilon$. In Theorem 2 in Anderson and Amemiya (1988), this term does not appear because in their results the asymptotic distribution of $\hat{V}_\varepsilon$ is centered around $diag(\frac{1}{n}\varepsilon\varepsilon')$ instead of $\tilde{V}_\varepsilon$. Our recentering around $\tilde{V}_\varepsilon$ avoids a random bias term.

Finally, by applying the CLT to (E.19), the asymptotic distribution of vector $\hat{\theta}$ is:

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}) \Rightarrow \frac{1}{2}B_0\left(B_0'J_0B_0\right)^{-1}B_0'\left(\frac{\partial vec(\Sigma(\theta_0))}{\partial\theta'}\right)'\left(V_y^{-1}\otimes V_y^{-1}\right)vec\left(\bar{W}F' + F\bar{W}' + Z\right). \quad \text{(E.23)}$$

The Gaussian asymptotic distribution in (E.23) matches those in (E.20) and (E.21) written for the components, and its asymptotic variance yields the 'sandwich formula". The result in (E.23) is analogue to Theorem 2 in Anderson and Amemiya (1988), for different factor normalization and recentering of the variance estimator.

Proof of Proposition 7: From (E.7), we have the asymptotic expansion: $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) = diag(\Psi_\varepsilon) + o_p(1) = V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z_n^*) + o_p(1)$. Moreover, from Proposition 6 (a) and using

[47] Here, $diag(Z)$ is the diagonal matrix with the same diagonal elements as Z.

$\Psi_y - \Psi_\varepsilon = W_n F' + F W_n' + \bar{Z}_n$, we have: $\sqrt{n}(\hat{F}_j - F_j) = R_j(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \Lambda_j \Psi_\varepsilon V_\varepsilon^{-1} F_j + o_p(1) = R_j(W_n F' + F W_n' + \bar{Z}_n)V_\varepsilon^{-1}F_j + \Lambda_j[diag(\Psi_\varepsilon) \odot (V_\varepsilon^{-1}F_j)] + o_p(1) = R_j(W_n F' + F W_n' + \bar{Z}_n)V_\varepsilon^{-1}F_j + \Lambda_j\{[V_\varepsilon(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' vech(Z_n^*)] \odot F_j\} + o_p(1)$. Lemmas 2 and 8 yield (E.20)-(E.21), together with (E.22) from (E.8) since $\Psi_y - \Psi_\varepsilon \Rightarrow \bar{W}F' + F\bar{W}' + \bar{Z}$.

E.5.3 Feasible CLT for the FA estimators

i) Feasible CLT for Z_n via a parametric estimator of the asymptotic variance

We first show that, under strengthening of Assumption 2, we get a parametric structure for the variance $V[vech(Z)] = \Omega_Z(V_\varepsilon, \vartheta)$ with a vector of unknown parameters ϑ of dimension $T+1$.

Assumption 3 *The standardized errors processes $w_{i,t}$ in Assumption 2 are (a) stationary martingale difference sequences (mds), and (b) $E[w_{i,t}^2 w_{i,r} w_{i,s}] = 0$, for $t > r > s$.*

Assumption 3 holds e.g. for conditionally homoskedastic mds, and for ARCH processes (see below). Let $\mathscr{Z} := V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}$. Then, using Lemma 2, under Assumptions 2 and 3, we have $V[\mathscr{Z}_{t,t}] = \psi(0) + 2\kappa$, $V[\mathscr{Z}_{t,s}] = \psi(t-s) + q + \kappa$ and $Cov(\mathscr{Z}_{t,t}, \mathscr{Z}_{s,s}) = \psi(t-s)$, where $\psi(t-s) := \lim_{n\to\infty} \frac{1}{n}\sum_i Cov(w_{i,t}^2, w_{i,s}^2)\sigma_{ii}^2$. Quantity $\psi(t-s)$ depends on the difference $t-s$ only, by stationarity. The other covariance terms between elements of $\mathscr{Z}$ vanish. Then, we have $\Omega = [\psi(0) - 2q]D(0) + \sum_{h=1}^{T-1}\psi(h)D(h) + (q+\kappa)I_{T(T+1)/2}$, where $D(0) = \sum_{t=1}^{T} vech(E_{t,t})vech(E_{t,t})'$ and $D(h) = \tilde{D}(h) + \bar{D}(h)$ with $\tilde{D}(h) = \sum_{t=1}^{T-h}[vech(E_{t,t})vech(E_{t+h,t+h})' + vech(E_{t+h,t+h})vech(E_{t,t})']$ and $\bar{D}(h) = \sum_{t=1}^{T-h} vech(E_{t,t+h} + E_{t+h,t})vech(E_{t,t+h} + E_{t+h,t})'$ for $h = 1, ..., T-1$, and where $E_{t,s}$ denote the $T \times T$ matrix with entry 1 in position (t,s) and 0 elsewhere. Hence, with $Z = V_\varepsilon^{1/2}\mathscr{Z}V_\varepsilon^{1/2}$, we get a parametrization $\Omega_Z(V_\varepsilon, \vartheta)$ for $V[vech(Z)]$ with $\vartheta = (q+\kappa, \psi(0) - 2q, \psi(1), ..., \psi(T-1))'$.

Then, we obtain a parametric structure for $M_{\boldsymbol{X}}\Omega_{Z^*}M_{\boldsymbol{X}} = M_{\boldsymbol{X}}\boldsymbol{R}'\Omega_Z\boldsymbol{R}M_{\boldsymbol{X}}$.

Lemma 9 *Under Assumptions 1-3 and A.1-A.6, we have:*

$$M_{\boldsymbol{X}}\Omega_{Z^*}M_{\boldsymbol{X}} = \sum_{h=1}^{T-1}[\psi(h) + q + \kappa]M_{\boldsymbol{X}}\boldsymbol{R}'\bar{D}(h)\boldsymbol{R}M_{\boldsymbol{X}}. \tag{E.24}$$

Hence, the parametric structure $M_{\boldsymbol{X}}\Omega_{Z^*}M_{\boldsymbol{X}}(V_\varepsilon, G, \tilde{\vartheta})$ depends linearly on vector $\tilde{\vartheta}$ that stacks the $T-1$ parameters $\psi(h)+q+\kappa$, for $h=1,...,T-1$. It does not involve parameter $\psi(0)$, i.e., the quartic moment of errors, because the asymptotic expansion of the LR statistic does not involve the diagonal terms of Z. Moreover, the unknown parameters appear through the linear combinations $\psi(h)+q+\kappa$ that are the scaled variances of the out-of-diagonal elements of Z. We can estimate the unknown parameters in $\tilde{\vartheta}$ by least squares applied on (E.24), using the nonparametric estimator $M_{\hat{\boldsymbol{X}}}\hat{\Omega}_{Z^*}M_{\hat{\boldsymbol{X}}}$ defined in Proposition 1, after half-vectorization and replacing V_ε and G by their FA estimates. It yields a consistent estimator of $M_{\boldsymbol{X}}\Omega_{Z^*}M_{\boldsymbol{X}}$ incorporating the restrictions implied by Assumption 3.

To get a feasible CLT for the FA estimates, we need to estimate the additional parameters $\psi(0)-2q$ and $q+\kappa$. We consider the matrix $\hat{\Omega}_{Z^*}$ from Proposition 1, that involves fourth-order moments of residuals.

Lemma 10 *Under Assumptions 1-3 and A.1-A.6, and* $\sqrt{n}\sum_{m=1}^{J_n} B_{m,n}^2 = o(1)$*, up to pre- and post-multiplication by an orthogonal matrix and its transpose, we have* $\hat{\Omega}_{Z^*} = \boldsymbol{R}'\tilde{\Xi}_n\boldsymbol{R} + o_p(1)$*, where* $\tilde{\Xi}_n = [\psi_n(0)-2q_n]D(0) + \sum_{h=1}^{T-1}\psi_n(h)D(h) + (q_n+\kappa_n)I_{T(T+1)/2} + (q_n+\xi_n)vech(I_T)vech(I_T)'$ *and* $\xi_n := \dfrac{1}{n}\sum_{m=1}^{J_n}\sum_{i\neq j\in I_m}\sigma_{ii}\sigma_{jj}$.

With blocks of equal size, the condition $\sqrt{n}\sum_{m=1}^{J_n} B_{m,n}^2 = o(1)$ holds if $J_n = n^{\bar{\alpha}}$ and $\bar{\alpha} > 1/2$. Now, we have the relation $3D(0) + \sum_{h=1}^{T-1} D(h) - vech(I_T)vech(I_T)' = I_{T(T+1)/2}$, which implies $3\boldsymbol{R}'D(0)\boldsymbol{R} + \sum_{h=1}^{T-1}\boldsymbol{R}'D(h)\boldsymbol{R} - vech(I_{T-k})vech(I_{T-k})' = I_p$. Hence, matrix

$$\begin{aligned}\boldsymbol{R}'\tilde{\Xi}_n\boldsymbol{R} &= [\psi_n(0)+q_n+3\kappa_n]\boldsymbol{R}'D(0)\boldsymbol{R} + \sum_{h=1}^{T-1}[\psi_n(h)+q_n+\kappa_n]\boldsymbol{R}'D(h)\boldsymbol{R} \\ &\quad +(\xi_n-\kappa_n)vech(I_{T-k})vech(I_{T-k})' \end{aligned} \tag{E.25}$$

depends on $T+1$ linear combinations of the elements of $\vartheta_n = (q_n+\kappa_n, \psi_n(0)-2q_n, \psi_n(1), ..., \psi_n(T-1))'$ and $\xi_n-\kappa_n$. Thus, the linear system (E.25) is rank-deficient to identify ϑ_n. Moreover, in Assumption A.3 (b), κ_n is defined as a double sum over squared covariances scaled by n, and is assumed to converge to a constant κ. Such a convergence is difficult to assume for ξ_n since ξ_n is a double sum over products of two variances scaled by n.

We apply half-vectorization on (E.25), replace the LHS by its consistent estimate $\hat{\hat{\Xi}}$, and plug-in the FA estimates in the RHS. From Lemma 10, least squares estimation on such a linear regression yields consistent estimates of linear combinations $\psi(0) + q + 3\kappa$ and $\psi(h) + q + \kappa$ for $h = 1, \ldots, T-1$. Consistency of those parameters applies independently of $\xi_n - \kappa_n$ converging as $n \to \infty$, or not.[48] In order to identify the components of ϑ, we need an additional condition. We use the assumption $\psi(T-1) = 0$. That condition is implied by serial uncorrelation in the squared standardized errors after lag $T-1$, that is empirically relevant in our application with monthly returns data. Then, parameter $q + \kappa$ is estimated by $\widehat{\psi_n(T-1) + q_n} + \kappa_n$, and by difference we get the estimators of $\psi(0) - 2q$ and $\psi(h)$, for $h = 1, ..., T-2$.

Let us now discuss the case of ARCH errors. Suppose the $w_{i,t}$ follow independent ARCH(1) processes with Gaussian innovations that are independent across assets, i.e., $w_{i,t} = h_{i,t}^{1/2} z_{i,t},\ z_{i,t} \sim IIN(0,1),\ h_{i,t} = c_i + \alpha_i w_{i,t-1}^2$ with $c_i = 1 - \alpha_i$. Then $E[w_{i,t}] = 0$, $E[w_{i,t}^2] = 1$, $\eta_i := V[w_{i,t}^2] = \frac{2}{1-3\alpha_i^2}$, $Cov(w_{i,t}^2, w_{i,t-h}^2) = \eta_i \alpha_i^h$. Moreover, $E[w_{i,t} w_{i,r} w_{i,s} w_{i,p}] = 0$ if one index among t, r, s, p is different from all the others. Indeed, without loss of generality, suppose t is different from s, p, r. By the law of iterated expectation: $E[\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,p}\varepsilon_{i,r}] = E[E[\varepsilon_{i,t} | \{z_{i,\tau}^2\}_{\tau=-\infty}^{\infty}, \{z_{i,\tau}\}_{\tau \neq t}]\varepsilon_{i,s}\varepsilon_{i,p}\varepsilon_{i,r}] = E[h_{i,t}^{1/2} E[z_{i,t} | z_{i,t}^2]\varepsilon_{i,s}\varepsilon_{i,p}\varepsilon_{i,r}] = 0$. Then, Assumption 3 holds. The explicit formula of Ω involves $\psi(h) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \frac{2\alpha_i^h}{1-3\alpha_i^2} \sigma_{ii}^2$, for $h = 0, 1, ..., T-1$. Hence, setting $\psi(T-1) = 0$ is a mild assumption for identification purpose since α_i^{T-1} is small. If $\alpha_i = 0$ for all i, i.e., no ARCH effects, we have $\psi(0) = 2q$ and $\psi(h) = 0$ for $h > 0$, so that $\Omega = (q+\kappa) I_{\frac{T(T+1)}{2}}$.

ii) Feasible CLT for W_n

Let us now establish a feasible CLT for W_n. In order to estimate matrix Q_β in the asymptotic

[48]To see this, write the half-vectorization of the RHS of (E.25) as $\chi\eta_n$, where χ is the $\frac{p(p+1)}{2} \times (T+1)$ matrix of regressors and η_n the $(T+1) \times 1$ vector of unknown parameters. Then, $vech(\hat{\Omega}_{Z^*}) = \hat{\chi}\eta_n + o_p(1)$, by Lemma 10, the consistency of the FA estimates, and the last column of χ not depending on unknown parameters. Thus, $\hat{\eta}_n := (\hat{\chi}'\hat{\chi})^{-1}\hat{\chi}' vech(\hat{\Omega}_{Z^*}) = \eta_n + o_p(1)$. In particular, we also have $\widehat{\xi_n - \kappa_n} = \xi_n - \kappa_n + o_p(1)$.

variance Ω_W in Lemma 8, we use the estimated betas and residuals, and combine them with a temporal sample splitting approach to cope with the EIV problem caused by the fixed T setting. Specifically, let us split the time spell into two consecutive sub-intervals with T_1 and T_2 observations, with $T_1 + T_2 = T$ and such that $T_1 > k$ and $T_2 \geq k$. The factor model in the two sub-intervals reads $y_{1,i} = \mu_1 + F_1\beta_i + \varepsilon_{1,i}$ and $y_{2,i} = \mu_2 + F_2\beta_i + \varepsilon_{2,i}$, and let $V_{1,\varepsilon}$ and $V_{2,\varepsilon}$ denote the corresponding diagonal matrices of error average unconditional variances.[49] The conditions $T_1 > k$ and $T_2 \geq k$ are needed because we estimate residuals and betas in the first and the second sub-intervals, namely $\hat{\varepsilon}_{1,i} = M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}(y_{1,i} - \bar{y}_1)$ and $\hat{\beta}_i = (\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}(y_{2,i} - \bar{y}_2)$. Here, $\hat{F}_j$ and $\hat{V}_{j,\varepsilon}$ for $j = 1, 2$ are deduced from the FA estimates in the full period of T observations. Define $\hat{\Psi}_\beta = \frac{1}{n}\sum_m \sum_{i,j\in I_m}(\hat{\beta}_i\hat{\beta}_j') \otimes (\hat{\varepsilon}_{1,i}\hat{\varepsilon}_{1,j}')$. By using $\hat{\varepsilon}_{1,i} = (M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}F_1)\beta_i + M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}(\varepsilon_{1,i} - \bar{\varepsilon}_1)$, $M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}F_1 = O_p(\frac{1}{\sqrt{n}})$ and $\frac{1}{n^2}\sum_m b_{m,n}^2 = \sum_m B_{m,n}^2 = o(1)$, we get $\hat{\Psi}_\beta = (I_k \otimes M_{\hat{F}_1,\hat{V}_{1,\varepsilon}})\left(\frac{1}{n}\sum_m \sum_{i,j\in I_m}(\hat{\beta}_i\hat{\beta}_j') \otimes [(\varepsilon_{1,i} - \bar{\varepsilon}_1)(\varepsilon_{1,j} - \bar{\varepsilon}_1)']\right)(I_k \otimes M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}') + o_p(1)$. Now, we use $\hat{\beta}_i = \left[(\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}F_2\right]\beta_i + (\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}(\varepsilon_{2,i} - \bar{\varepsilon}_2)$, and $\bar{\varepsilon}_1 = o_p(n^{-1/4})$, $\bar{\varepsilon}_2 = o_p(n^{-1/4})$ from Lemma 6 (a), as well as the the mds condition in Assumption 3. We get $\hat{\Psi}_\beta = \hat{\Psi}_{\beta,1} + \hat{\Psi}_{\beta,2} + o_p(1)$, where $\hat{\Psi}_{\beta,1} = (I_k \otimes M_{F_1,V_{1,\varepsilon}})\left(\frac{1}{n}\sum_m \sum_{i,j\in I_m}(\beta_i\beta_j') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}')\right)(I_k \otimes M_{F_1,V_{1,\varepsilon}}')$ and $\hat{\Psi}_{\beta,2} = \left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)\left(\frac{1}{n}\sum_m \sum_{i,j\in I_m}(\varepsilon_{2,i}\varepsilon_{2,j}') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}')\right)\left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)'$. We use $\frac{1}{n}\sum_m \sum_{i,j\in I_m}(\beta_i\beta_j') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}') = Q_\beta \otimes V_{1,\varepsilon} + o_p(1)$, and $\frac{1}{n}\sum_m \sum_{i,j\in I_m}(\varepsilon_{2,i}\varepsilon_{2,j}') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}') = \Omega_{21} + o_p(1)$, where Ω_{21} is the sub-block of matrix Ω_Z that is the asymptotic variance of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_{2,i} \otimes \varepsilon_{1,i} \Rightarrow N(0, \Omega_{21})$. Then, $\hat{\Psi}_\beta = Q_\beta \otimes (M_{F_1,V_{1,\varepsilon}}V_{1,\varepsilon}) + \left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)\Omega_{21}\left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)' + o_p(1)$. Thus, we get a consistent estimator of $Q_\beta \otimes (V_{1,\varepsilon}^{-1/2}M_{F_1,V_{1,\varepsilon}}V_{1,\varepsilon}^{1/2})$ by subtracting to $\hat{\Psi}_\beta$ a consistent estimator of the second term on the RHS (bias term),[50] and then by pre- and post-multiplying times $(I_k \otimes \hat{V}_{1,\varepsilon}^{-1/2})$.

[49] We can take the two sub-intervals as the halves of the time span. If this choice does not meet conditions $T_1 > k$ and $T_2 \geq k$ in a subperiod, we take the second sub-interval such that $T_2 = k$, and add to the first sub-interval a sufficient number of dates from the preceeding subperiod in order to get $T_1 = k + 1$.

[50] Sample splitting makes the estimation of the bias easier, but we can avoid such a splitting at the expense of a more complicated debiasing procedure.

To get a consistent estimator of Q_β, we apply a linear transformation that amounts to computing the trace of the second term of a Kronecker product, and divide by $Tr(V_{1,\varepsilon}^{-1/2} M_{F_1,V_{1,\varepsilon}} V_{1,\varepsilon}^{1/2}) = T_1 - k$. Thus: $\hat{Q}_\beta = \frac{1}{n(T_1-k)} \sum_m \sum_{i,j \in I_m} (\hat{\beta}_i \hat{\beta}_j')(\hat{\varepsilon}_{1,j}' \hat{V}_{1,\varepsilon}^{-1} \hat{\varepsilon}_{1,i}) - \frac{1}{T_1-k} \sum_{j=1}^{T_1} (I_k \otimes e_j') \left\{ \left([(\hat{F}_2' \hat{V}_{2,\varepsilon}^{-1} \hat{F}_2)^{-1} \hat{F}_2' \hat{V}_{2,\varepsilon}^{-1}] \otimes [\hat{V}_{1,\varepsilon}^{-1/2} M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}] \right) \hat{\Omega}_{21} \left([\hat{V}_{2,\varepsilon}^{-1} \hat{F}_2 (\hat{F}_2' \hat{V}_{2,\varepsilon}^{-1} \hat{F}_2)^{-1}] \otimes [M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}' \hat{V}_{1,\varepsilon}^{-1/2}] \right) \right\} (I_k \otimes e_j)$, where the e_j are T_1-dimensional unit vectors, and $\hat{\Omega}_{21}$ is obtained from Subsection E.5.3 i). If estimate $\hat{Q}_\beta$ is not positive definite, we regularize it by deleting the negative eigenvalues.

iii) Joint feasible CLT

To get a feasible CLT for the FA estimators from (E.20)-(E.21), we need the joint distribution of the Gaussian matrix variates Z and W. Under the condition of Lemma 8 (b), the estimates of the asymptotic variances of $vech(Z)$ and $vec(W)$ are enough, since these vectors are independent. Otherwise, to estimate the covariance $Cov(vech(Z), vec(W))$, we need to extend the approaches of the previous subsections.

E.5.4 Special cases

In this subsection, we particularize the asymptotic distributions of the FA estimators for three special cases along the lines of Section 4, plus a fourth special case that allows us to further discuss the link with Anderson and Amemiya (1988).

i) Gaussian errors

When the errors admit a Gaussian distribution $\varepsilon_i \overset{ind}{\sim} N(0, \sigma_{ii} V_\varepsilon)$ with diagonal V_ε, matrix $\frac{1}{\sqrt{q}} V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}$ is in the GOE for dimension T, i.e., $\frac{1}{\sqrt{q}} vech(V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}) \sim N(0, I_{T(T+1)/2})$, where $q = \lim_{n\to\infty} \frac{1}{n} \sum_i \sigma_{ii}^2$. Moreover, $vec(W) \sim N(0, Q_\beta \otimes V_\varepsilon)$, where $Q_\beta = \lim_{n\to\infty} \frac{1}{n} \sum_i \sigma_{ii} \beta_i \beta_i'$, mutually independent of Z because of the symmetry of the Gaussian distribution.

ii) Quasi GOE errors

As an extension of the previous case, here let us suppose that the errors meet Assumption 2, the Conditions (a) and (b) in Proposition 2 plus additionally (c) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n V(\varepsilon_{i,t}^2) = \eta V_{\varepsilon,tt}^2$, for a constant $\eta > 0$, and (d) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n E[\varepsilon_{i,t}^2 \varepsilon_{i,r} \varepsilon_{i,p}] = 0$ for $r \neq p$. This setting allows e.g. for conditionally homoskedastic mds processes in the errors, but excludes ARCH effects. Then, the arguments in Lemma 2 imply $vech(V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}) \sim N(0, \Omega)$ with $\Omega = \begin{pmatrix} (\eta/2+\kappa)I_T & 0 \\ 0 & (q+\kappa)I_{\frac{1}{2}T(T-1)} \end{pmatrix}$. The distribution of $V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}$ is similar to (scaled) GOE holding in the Gaussian case up to the variances of diagonal and of out-of-diagonal elements being different when $\eta \neq 2q$. Hence, contrasting with test statistics, the asymptotic distributions of FA estimates differ in cases i) and ii) beyond scaling factors. It is because the asymptotic distributions of FA estimates involve diagonal elements of Z as well.

iii) Spherical errors

Let us consider the case $\varepsilon_i \overset{ind}{\sim} (0, \sigma_{ii} V_\varepsilon)$ where $V_\varepsilon = \bar{\sigma}^2 I_T$, with independent components across time and the normalization $\lim_{n\to\infty} \frac{1}{n}\sum_i \sigma_{ii} = 1$. By repeating the arguments of Section E.3 for the constrained FA estimators (see Section 5.3), we get $Tr(M_F(\Psi_y - \Psi_\varepsilon)M_F) = 0$ instead of equation (E.4). It yields the asymptotic expansions $\sqrt{n}(\hat{\sigma}^2 - \tilde{\sigma}^2) = \frac{1}{T-k} Tr(M_F Z_n) + o_p(1) = \frac{\bar{\sigma}^2}{T-k} Tr(Z_n^*) + o_p(1)$, and $\sqrt{n}(\hat{F}_j - F_j) = \frac{1}{\bar{\sigma}^2} R_j(\Psi_y - \Psi_\varepsilon)F_j - \frac{1}{\bar{\sigma}^2}\Lambda_j \Psi_\varepsilon F_j + o_p(1) = \frac{1}{\bar{\sigma}^2} R_j(W_n F' + F W_n' + \bar{Z}_n)F_j + o_p(1)$, where we use $\Psi_y - \Psi_\varepsilon = W_n F' + F W_n' + \bar{Z}_n$, $\Psi_\varepsilon = \frac{1}{T-k} Tr(M_F Z_n) I_T$ and $\Lambda_j F_j = 0$, and $\bar{Z}_n = Z_n - \frac{1}{T-k} Tr(M_F Z_n) I_T$. Moreover, by sphericity, we have $R_j = \frac{1}{2\gamma_j} P_{F_j} + \frac{1}{\gamma_j} M_F + \sum_{\ell=1,\ell\neq j}^k \frac{1}{\gamma_j - \gamma_\ell} P_{F_\ell}$. Thus, we get $\sqrt{n}(\hat{\sigma}^2 - \tilde{\sigma}^2) \Rightarrow \frac{\bar{\sigma}^2}{T-k} Tr(Z^*)$ and $\sqrt{n}(\hat{F}_j - F_j) \Rightarrow \frac{1}{\bar{\sigma}^2} R_j(WF' + FW' + \bar{Z})F_j$.[51] The Gaussian matrix Z is such that $Z_{tt} \sim N(0, \eta)$ and $Z_{t,s} \sim N(0, q)$ for $t \neq s$, mutually independent, where $\eta = \lim_{n\to\infty} \frac{1}{n}\sum_i V[\varepsilon_{i,t}^2]$, and $vec(W) \sim N(0, Q_\beta \otimes I_T)$. Variables Z and W are independent if $E[\varepsilon_{i,t}^3] = 0$. FGS, Section 4.3.1, explain how we can estimate q and η by solving a system of two linear equations based on estimated moments of $\hat{\varepsilon}_{i,t}$.

iv) Cross-sectionally homoskedastic errors and link with Anderson and Amemiya (1988)

Let us now make the link with the distributional results in Anderson and Amemiya (1988). In our setting, the analogous conditions as those in their Corollary 2 would be: (a) random effects

[51] The asymptotic distribution of estimator $\hat{\sigma}^2$ coincides with that derived in FGS with perturbation theory methods. The asymptotic distribution of the factor estimates slightly differs from that given in FGS, Section 5.1, because of the different factor normalization adopted by FA compared to PCA even under sphericity.

for the loadings that are i.i.d. with $E[\beta_i] = 0$, $V[\beta_i] = I_k$, (b) error terms are i.i.d. $\varepsilon_i \sim (0, V_\varepsilon)$ with $V_\varepsilon = diag(V_{\varepsilon,11}, ..., V_{\varepsilon,TT})$ such that $E[\varepsilon_{i,t}\varepsilon_{i,r}\varepsilon_{i,s}\varepsilon_{i,p}] = V_{\varepsilon,tt}V_{\varepsilon,ss}$, for $t = r > s = p$, and $= 0$, otherwise, and (c) β_i and ε_i are mutually independent. Thus, $\sigma_{ii} = 1$ for all i, i.e., errors are cross-sectionally homoskedastic. Under the aforementioned Conditions (a)-(c), the Gaussian distributional limits Z and W are such that $V[Z_{tt}] = \eta_t V_{\varepsilon,tt}^2$, for $\eta_t := V[\varepsilon_{i,t}^2]/V_{\varepsilon,tt}^2$, $V[Z_{ts}] = V_{\varepsilon,tt}V_{\varepsilon,ss}$, for $t \neq s$, all covariances among different elements of Z vanish, and $V[vec(W)] = I_k \otimes V_\varepsilon$. Equations (E.20)-(E.21) yield the asymptotic distributions of the FA estimates. In particular, they do not depend on the distribution of the β_i. Moreover, the distribution of the out-of-diagonal elements of Z does not depend on the distribution of the errors, while, for the diagonal term, we have $\eta_t = 2$ for Gaussian errors. As remarked in Section E.5.2, if the asymptotic distribution of estimator $\hat{V}_\varepsilon$ is centered around the realized matrix $\frac{1}{n}\sum_i \varepsilon_i\varepsilon_i'$ instead of its expected value, that distribution involves the out-of-diagonal elements of Z, and the elements of W. Hence, in that case, the asymptotic distribution of the FA estimates is the same independent of the errors being Gaussian or not, and depends on F and V_ε only, as found in Anderson and Amemiya (1988).

E.6 Orthogonal transformations and maximal invariant statistic

In this subsection, we consider the transformation $\mathcal{O}$ that maps matrix $\hat{G}$ into $\hat{G}O$, where O is an orthogonal matrix in $\mathbb{R}^{(T-k)\times(T-k)}$, and the transformation $\mathcal{O}_D$ that maps matrix $\hat{\boldsymbol{D}}$ into $\hat{\boldsymbol{D}}O_D$, where O_D is an orthogonal matrix in $\mathbb{R}^{df\times df}$. These transformations are induced from the freedom in chosing the orthonormal bases spanning the orthogonal complements of $\hat{F}$ and $\hat{\boldsymbol{X}}$. We show that they imply a group of orthogonal transformations on the vector $\hat{W} = \sqrt{n}\hat{\boldsymbol{D}}' vech(\hat{S}^*)$, with $\hat{S}^* = \hat{G}'\hat{V}_\varepsilon^{-1}(\hat{V}_y - \hat{V}_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}$, and establish the maximal invariant.

Under the transformation $\mathcal{O}$, matrix $\hat{S}^*$ is mapped into $O^{-1}\hat{S}^*O$. This transformation is mirrored by a linear mapping at the level of the half-vectorized form $vech(\hat{S}^*)$. In fact, this mapping is norm-preserving, since $\|vech(S)\|^2 = \frac{1}{2}\|S\|^2$ and $\|O^{-1}SO\| = \|S\|$ for any conformable symmetric matrix S and orthogonal matrix O. This mapping is characterized in the next lemma.

Lemma 11 *For any symmetric matrix S and orthogonal matrix O in $\mathbb{R}^{m\times m}$, we have $vech(O^{-1}SO) = \mathscr{R}(O)vech(S)$, where $\mathscr{R}(O) = \frac{1}{2}A_m'(O' \otimes O')A_m$ is an orthogonal matrix, and A_m is the duplication matrix defined in Appendix B. Transformations $\mathscr{R}(O)$ with orthogonal O have the structure of a group: (a) $\mathscr{R}(I_m) = I_{\frac{1}{2}m(m+1)}$, (b) $\mathscr{R}(O_1)\mathscr{R}(O_2) = \mathscr{R}(O_2O_1)$, and (c) $[\mathscr{R}(O)]^{-1} = \mathscr{R}(O^{-1})$.*

With this lemma, we can give the transformation rules under $\mathcal{O}$ for a set of relevant statistics in the next proposition. We denote generically with $\widetilde{\cdot}$ a quantity computed with $\hat{G}O$ instead of $\hat{G}$.

Proposition 8 *Under Assumptions 1 and A.5, (a) $vech(\widetilde{\hat{S}^*}) = \mathscr{R}(O)vech(\hat{S}^*)$, (b) $\widetilde{\boldsymbol{X}} = \mathscr{R}(O)\boldsymbol{X}$, (c) $I_p - \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}' = \mathscr{R}(O)[I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\mathscr{R}(O)^{-1}$, (d) $\widetilde{\boldsymbol{R}} = \boldsymbol{R}\mathscr{R}(O)^{-1}$, (e) $\widetilde{\boldsymbol{R}}(I_p - \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}') = \boldsymbol{R}(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\mathscr{R}(O)^{-1}$.*

From Proposition 8 (c), under transformation $\mathcal{O}$, matrix $\hat{\boldsymbol{D}}$ is mapped into $\mathscr{R}(O)\hat{\boldsymbol{D}}$. Combining with transformation $\mathcal{O}_D$, we have $\widetilde{\hat{\boldsymbol{D}}} = \mathscr{R}(O)\hat{\boldsymbol{D}}O_D$. Thus, using Proposition 8 (a), under $\mathcal{O}$ and $\mathcal{O}_D$, vector $\hat{W}$ is mapped into $\widetilde{\hat{W}} = \sqrt{n}\widetilde{\hat{\boldsymbol{D}}}'vech(\widetilde{\hat{S}^*}) = O_D'\hat{W}_D$. Thus, statistic $\hat{W}$ is invariant under $\mathcal{O}$, while $\mathcal{O}_D$ operates as the group of orthogonal transformations. The maximal invariant under this group of transformations is the squared norm $\|\hat{W}\|^2 = \hat{W}'\hat{W}$.

Proof of Proposition 8: With $\widetilde{\hat{S}^*} = O^{-1}\hat{S}^*O$, part (a) follows from Lemma 11. Let $\tilde{G} = GO$. Then, for any diagonal matrix Δ, on the one hand, we have $vech(\tilde{G}'\Delta\tilde{G}) = \tilde{\boldsymbol{X}}diag(\Delta)$, and on the other hand, we have $vech(\tilde{G}'\Delta\tilde{G}) = vech(O^{-1}G'\Delta GO) = \mathscr{R}(O)vech(G'\Delta G) = \mathscr{R}(O)\boldsymbol{X}diag(\Delta)$. By equating the two expressions for any diagonal matrix Δ, part (b) follows. Statement (c) is a consequence thereof and $\mathscr{R}(O)$ being orthogonal. Moreover, with $\tilde{Q} = QO$ and using $vech(\tilde{Q}'Z\tilde{Q}) = vech(O^{-1}Q'ZQO) = \mathscr{R}(O)\boldsymbol{R}'vech(Z)$, we deduce part (d). Statement (e) is a consequence of (c) and (d).

E.7 Proofs of Lemmas 5-11

Proof of Lemma 5: The equivalence of conditions (a) and (b) is a consequence of the fact that function $\mathscr{L}(A) = -\frac{1}{2}\log|A| - \frac{1}{2}Tr(V_y^0 A^{-1})$, where A is a p.d. matrix, is uniquely maximized for $A = V_y^0$ (see Magnus and Neudecker (2007), p. 410), and $L_0(\theta) = \mathscr{L}(\Sigma(\theta))$.

Proof of Lemma 6: (a) From Assumption 2, we have $E[\bar{\varepsilon}] = 0$ and $V[\bar{\varepsilon}] = V\left[\frac{1}{n}\sum_{i,k=1}^n s_{i,k}V_\varepsilon^{1/2}w_k\right]$ $= V_\varepsilon^{1/2}\frac{1}{n^2}\sum_{i,j,k,l=1}^n s_{i,k}s_{j,l}E[w_k w_l']V_\varepsilon^{1/2} = (\frac{1}{n^2}\sum_{i,j}^n \sigma_{i,j})V_\varepsilon$ where the $s_{i,k}$ are the elements of $\Sigma^{1/2}$. Now, $\frac{1}{n^2}\sum_{i,j=1}^n \sigma_{i,j} \leq C\frac{1}{n^2}\sum_{m=1}^{J_n} b_{m,n}^{1+\delta} = O(n^{\delta-1}\sum_{m=1}^{J_n} B_{m,n}^{1+\delta}) = O(n^{\delta-1}J_n^{1/2}(\sum_{m=1}^{J_n} B_{m,n}^{2(1+\delta)})^{1/2})$ $= o(n^{-1}J_n^{1/2}) = o(n^{-1/2})$ from the Cauchy-Schwarz inequality and Assumptions 2 (c) and (d). Part (a) follows. To prove part (b), we use $E[\frac{1}{n}\varepsilon\varepsilon'] \to V_\varepsilon^0$ and $V[vech((V_\varepsilon^0)^{-1/2}(\frac{1}{n}\varepsilon\varepsilon')(V_\varepsilon^0)^{-1/2})] = \frac{1}{n}\Omega_n$ from the proof of Lemma 2, and $\frac{1}{n}\Omega_n = o(1)$ by Assumption A.3. Finally, to show part (c), write $\frac{1}{n}\sum_{i=1}^n \varepsilon_i\beta_i' = (V_\varepsilon^0)^{1/2}\frac{1}{n}\sum_{i,j=1}^n s_{i,j}w_j\beta_i'$. Then, $E[\frac{1}{n}\sum_{i=1}^n \varepsilon_i\beta_i'] = 0$ while the variance of $vec(\frac{1}{n}\sum_{i=1}^n \varepsilon_i\beta_i')$ vanishes asymptotically since $V[vec(\frac{1}{n}\sum_{i,j=1}^n s_{i,j}w_j\beta_i')] = \frac{1}{n^2}\sum_{i,j,m,l=1}^n s_{i,j}s_{m,l}$ $(\beta_i\beta_l') \otimes E[w_j w_m'] = \frac{1}{n^2}\sum_{i,l=1}^n \sigma_{i,l}(\beta_i\beta_l') \otimes I_T = o(1)$ under Assumptions 2 and A.2.

Proof of Lemma 7: From the arguments in the proof of Proposition 6 with $\Psi_y = 0$, the solution of the FOC is such that $\Psi_{F,j}^\epsilon = (\Lambda_j^0 - R_j^0)\Psi_{V_\varepsilon}^\epsilon(V_\varepsilon^0)^{-1}F_j$ for $j = 1, ..., k$, and $diag(M_{F_0,V_\varepsilon^0}\Psi_{V_\varepsilon}^\epsilon M_{F_0,V_\varepsilon^0}')$ $= 0$. Since $\Psi_{V_\varepsilon}^\epsilon$ is diagonal, the latter equation yields $M_{F_0,V_\varepsilon^0}^{\odot 2}diag(\Psi_{V_\varepsilon}^\epsilon) = 0$. Under condition (a) of Lemma 7, we get $\Psi_{V_\varepsilon}^\epsilon = 0$, which in turn implies $\Psi_F^\epsilon = 0$. Thus, condition (a) is sufficient for local identification. It is also necessary to get uniqueness of the solution $\Psi_{V_\varepsilon}^\epsilon = 0$. Moreover, conditions (a) and (b) of Lemma 7 are equivalent as shown in Appendix E.3. Further, conditions (a) and (c) are equivalent since $\Phi^{\odot 2} = M_{F_0,V_\varepsilon^0}^{\odot 2}(V_\varepsilon^0)^2$. Finally, let us show that condition (d) of Lemma 7 is both sufficient and necessary for local identification. The FOC for the Lagrangian problem are $\frac{\partial L_0(\theta)}{\partial\theta} - \frac{\partial g(\theta)'}{\partial\theta}\lambda_L = 0$ and $g(\theta) = 0$, where λ_L is the Lagrange multiplier vector. By expanding at first-order around θ_0 and $\lambda_0 = 0$, we get $H_0\begin{pmatrix}\theta - \theta_0 \\ \lambda\end{pmatrix} = 0$, where $H_0 := \begin{pmatrix} J_0 & A_0 \\ A_0' & 0\end{pmatrix}$, with $A_0 = \frac{\partial g(\theta_0)'}{\partial\theta}$, is the bordered Hessian. The parameters are locally identified if, and only if, H_0 is invertible. The latter condition is equivalent to $B_0' J_0 B_0$ being invertible.[52]

Proof of Lemma 8: By Assumption 2, $vec(W_n) = (I_k \otimes V_\varepsilon^{1/2}) \frac{1}{\sqrt{n}} \sum_{m=1}^{J_n} x_{m,n}$ where the $x_{m,n} := \sum_{i,j \in I_m} s_{i,j}(\beta_i \otimes w_j)$ are independent across m. Now, we apply the Liapunov CLT to show $\frac{1}{\sqrt{n}} \sum_{m=1}^{J_n} x_{m,n} \Rightarrow N(0, Q_\beta \otimes I_T)$. We have $E[x_{m,n}] = 0$ and $E[x_{m,n} x_{m,n}'] = \left(\sum_{i,j \in I_m} \sigma_{i,j} \beta_i \beta_j' \right) \otimes I_T$ and, by Assumption A.8, $\Omega_{W,n} := \frac{1}{n} \sum_{m=1}^{J_n} E[x_{m,n} x_{m,n}']$ converges to the positive definite matrix $Q_\beta \otimes I_T$. Let us now check the multivariate Liapunov condition $\|\Omega_{W,n}^{-1/2}\|^4 \frac{1}{n^2} \sum_{m=1}^{J_n} E[\|x_{m,n}\|^4] = o(1)$. Since $\|\Omega_{W,n}^{-1/2}\| = O_p(1)$, it suffices to prove $\frac{1}{n^2} \sum_{m=1}^{J_n} E[(x_{m,n}^{p,t})^4] = o(1)$, for any $p = 1, ..., k$ and $t = 1, ..., T$, where $x_{m,n}^{p,t} := \sum_{i,j \in I_m} s_{i,j} \beta_{i,p} w_{j,t}$. For this purpose, Assumptions A.1 and A.2 yield $E[(x_{m,n}^{p,t})^4] \leq C(\sum_{i,j \in I_m} \sigma_{i,j})^2$. Then, we get $\frac{1}{n^2} \sum_{m=1}^{J_n} E[(x_{m,n}^{p,t})^4] \leq C \frac{1}{n^2} \sum_{m=1}^{J_n} b_{m,n}^{2(1+\delta)} \leq C n^{2\delta} \sum_{m=1}^{J_n} B_{m,n}^{2(1+\delta)} = o(1)$ by Assumptions 2 (c) and (d). Part (a) of Lemma 8 follows. Moreover, $E[vech(\zeta_{m,n}) x_{m,n}'] = 0$ and the proof of Lemma 2 imply part (b).

Proof of Lemma 9: From the proof of Proposition 1 we have $M_{\boldsymbol{X}} \Omega_{Z^*} M_{\boldsymbol{X}} = M_{\boldsymbol{X}} \boldsymbol{R}' \Omega \boldsymbol{R} M_{\boldsymbol{X}}$, where $\Omega = D + \kappa I_{T(T+1)/2} = [\psi(0) - 2q] D(0) + \sum_{h=1}^{T-1} \psi(h) [\tilde{D}(h) + \bar{D}(h)] + (q + \kappa) I_{T(T+1)/2}$. Then, since the columns of $\boldsymbol{R}$ are orthonormal, we get $M_{\boldsymbol{X}} \Omega_{Z^*} M_{\boldsymbol{X}} = [\psi(0) - 2q] M_{\boldsymbol{X}} \boldsymbol{R}' D(0) \boldsymbol{R} M_{\boldsymbol{X}} + \sum_{h=1}^{T-1} \psi(h) M_{\boldsymbol{X}} \boldsymbol{R}' \tilde{D}(h) \boldsymbol{R} M_{\boldsymbol{X}} + \sum_{h=1}^{T-1} \psi(h) M_{\boldsymbol{X}} \boldsymbol{R}' \bar{D}(h) \boldsymbol{R} M_{\boldsymbol{X}} + (q + \kappa) M_{\boldsymbol{X}}$. Now, we show that the the first two terms in this sum are nil. We have $G' E_{t,t} G = Q' V_\varepsilon^{1/2} E_{t,t} V_\varepsilon^{1/2} Q = V_{\varepsilon,tt} Q' E_{t,t} Q$ and thus $vech(G' E_{t,t} G) = V_{\varepsilon,tt} vech(Q' E_{t,t} Q) = V_{\varepsilon,tt} \boldsymbol{R}' vech(E_{t,t})$ (see the proof of Proposition 1). Hence, the kernel of matrix $M_{\boldsymbol{X}}$ is spanned by vectors $\boldsymbol{R}' vech(E_{t,t})$, for $t = 1, ..., T$. We deduce that $M_{\boldsymbol{X}} \boldsymbol{R}' D(0) = 0$ and $M_{\boldsymbol{X}} \boldsymbol{R}' \tilde{D}(h) \boldsymbol{R} M_{\boldsymbol{X}} = 0$. Furthermore, from $I_{T(T+1)/2} = 2 \sum_{t=1}^{T} vech(E_{t,t}) vech(E_{t,t})' + \sum_{t<s} vech(E_{t,s} + E_{s,t}) vech(E_{t,s} + E_{s,t})' = 2D(0) + \sum_{h=1}^{T-1} \bar{D}(h)$, we get $M_{\boldsymbol{X}} = M_{\boldsymbol{X}} \boldsymbol{R}' I_{T(T+1)/2} \boldsymbol{R} M_{\boldsymbol{X}} = \sum_{h=1}^{T-1} M_{\boldsymbol{X}} \boldsymbol{R}' \bar{D}(h) \boldsymbol{R} M_{\boldsymbol{X}}$. The conclusion follows.

Proof of Lemma 10: By the root-n consistency of the FA estimators, $\hat{z}_{m,n}^* = z_{m,n}^* + O_p(\frac{b_{m,n}}{\sqrt{n}})$, uniformly in m, where $z_{m,n}^* = \sum_{i \in I_m} G' V_\varepsilon^{-1} \varepsilon_i \varepsilon_i' V_\varepsilon^{-1} G = \sum_{i \in I_m} Q' e_i e_i' Q$. Under the condition $\frac{1}{n^{3/2}} \sum_{m=1}^{J_n} b_{m,n}^2 = \sqrt{n} \sum_{m=1}^{J_n} B_{m,n}^2 = o(1)$, we have $\hat{\Omega}_{Z^*} = \frac{1}{n} \sum_{m=1}^{J_n} E[vech(z_{m,n}^*) vech(z_{m,n}^*)'] + o_p(1)$, up to pre- and post-multiplication by an orthogonal matrix. Moreover, $vech(z_{m,n}^*) =$

[52] Indeed, we can show $|H_0| = (-1)^{\frac{1}{2}k(k-1)} |A_0' A_0| |B_0' J_0 B_0|$ by using $J_0 A_0 = 0$, where the latter equality follows since the criterion is invariant to rotations of the latent factors.

$\boldsymbol{R}'[\sum_{i\in I_m} vech(e_i e_i')] = \frac{1}{2}\boldsymbol{R}'A_T'[\sum_{i\in I_m}(e_i\otimes e_i)]$, and $\sum_{i\in I_m}(e_i\otimes e_i) = \sum_{a,b}\sigma_{a,b}(w_a\otimes w_b)$. Thus, we get $E\left[vech(z_{m,n}^*)vech(z_{m,n}^*)'\right] = \frac{1}{4}\boldsymbol{R}'A_T'\left\{\sum_{a,b,c,d\in I_m}\sigma_{a,b}\sigma_{c,d}E[(w_a\otimes w_b)(w_c\otimes w_d)']\right\}A_T\boldsymbol{R}$. The non-zero contributions to the term in the curly brackets come from the combinations with $a = b = c = d$, $a = b \neq c = d$, $a = c \neq b = d$ and $a = d \neq b = c$, yielding: $\sum_{a,b,c,d}\sigma_{a,b}\sigma_{c,d}E[(w_a\otimes w_b)(w_c\otimes w_d)'] = \sum_a \sigma_{a,a}^2 E[(w_a w_a')\otimes(w_a w_a')] + (\sum_{a\neq c}\sigma_{a,a}\sigma_{c,c})vec(I_T)vec(I_T)' + (\sum_{a\neq b}\sigma_{a,b}^2)(I_{T^2} + K_{T,T}) = \sum_a[\sigma_{a,a}^2 V(w_a\otimes w_a)] + (\sum_a \sigma_{a,a})^2 vec(I_T)vec(I_T)' + (\sum_{a\neq b}\sigma_{a,b}^2)(I_{T^2} + K_{T,T})$. Then, using $w_a\otimes w_a = A_T vech(w_a w_a')$, we get $\frac{1}{4}A_T'\left\{\sum_{a,b,c,d\in I_m}\sigma_{a,b}\sigma_{c,d}E[(w_a\otimes w_b)(w_c\otimes w_d)']\right\}A_T = \sum_a[\sigma_{a,a}^2 V(vech(w_a w_a'))] + (\sum_a\sigma_{a,a})^2 vech(I_T)vech(I_T)' + (\sum_{a\neq b}\sigma_{a,b}^2)I_{\frac{T(T+1)}{2}}$. Then, since $\frac{1}{n}\sum_{i=1}^n \sigma_{i,i}^2 V[vech(w_i w_i')] = D_n$, where matrix D_n is defined in Assumption A.6, we get $\hat{\Omega}_{Z^*} = \boldsymbol{R}'\tilde{\tilde{\Xi}}_n\boldsymbol{R} + o_p(1)$, where $\tilde{\tilde{\Xi}}_n = D_n + (q_n + \xi_n)vech(I_T)vech(I_T)' + \kappa_n I_{\frac{T(T+1)}{2}}$. Moreover, under Assumption 3, and singling out parameter q_n along the diagonal, we have $D_n = [\psi_n(0) - 2q_n]D(0) + \sum_{h=1}^{T-1}\psi_n(h)[\tilde{D}(h) + \bar{D}(h)] + q_n I_{T(T+1)/2}$. The conclusion follows.

Proof of Lemma 11: We use $vec(S) = A_m vech(S)$, where the $m^2 \times \frac{1}{2}m(m+1)$ matrix A_m is such that: (i) $A_m'A_m = 2I_{\frac{1}{2}m(m+1)}$, (ii) $K_{m,m}A_m = A_m$, where $K_{m,m}$ is the commutation matrix for order m, and (iii) $A_m A_m' = I_{m^2} + K_{m,m}$ (see the proof of Proposition 1 and also Theorem 12 in Magnus, Neudecker (2007) Chapter 2.8). Then, $vech(S) = \frac{1}{2}A_m' vec(S)$ by property (i), and $vech(O^{-1}SO) = \frac{1}{2}A_m' vec(O^{-1}SO) = \frac{1}{2}A_m'(O'\otimes O')vec(S) = \frac{1}{2}A_m'(O'\otimes O')A_m vech(S)$, for all symmetric matrix S. It follows $\mathscr{R}(O) = \frac{1}{2}A_m'(O'\otimes O')A_m$. Moreover, by properties (i)-(iii), we have (a) $\mathscr{R}(I_m) = I_{\frac{1}{2}m(m+1)}$, (b) $\mathscr{R}(O_1)\mathscr{R}(O_2) = \frac{1}{4}A_m'(O_1'\otimes O_1')A_m A_m'(O_2'\otimes O_2')A_m = \frac{1}{4}A_m'(O_1'\otimes O_1')(I_{m^2} + K_{m,m})(O_2'\otimes O_2')A_m = \frac{1}{4}A_m'(O_1'O_2'\otimes O_1'O_2')(I_{m^2} + K_{m,m})A_m = \frac{1}{2}A_m'[(O_2O_1)'\otimes(O_2O_1)']A_m = \mathscr{R}(O_2O_1)$, and thus (c) $[\mathscr{R}(O)]^{-1} = \mathscr{R}(O^{-1})$.

F Numerical checks of conditions (6) of Proposition 4

In this section, we check numerically the validity of Inequalities (6) for given df, λ_j, ν_j, and $m = 3, ..., M$, for a large bound M. The idea is to compute the frequency of the LHS of (6) becoming strictly negative over a large number of potential values of λ_j and ν_j, $j = 1, ..., df$, for any given $df > 1$.[53] Table 2 provides those frequencies for $m = 3, ..., 16$ (cumulatively), with λ_j uniformly drawn in $[\underline{\lambda}, \bar{\lambda}]$ for $j = 1, ..., df$, and with $\nu_1 = 0$ [54] and ν_j uniformly drawn in $[0, \bar{\nu}]$, for $j = 2, ..., df$, and different combinations of bounds $\underline{\lambda}$, $\bar{\lambda}$, $\bar{\nu}$, and degrees of freedom $df = 2, ..., 12$. Each frequency is computed from 10^8 draws of λ_j and ν_j, $j = 1, ..., df$. In the SMC, we also report a table of frequencies for large grids of equally-spaced points in $[\underline{\lambda}, \bar{\lambda}]^{df} \times [0, \bar{\nu}]^{df-1}$, which corroborate the findings of this section.

F.1 Calibration of $\bar{\nu}$, $\underline{\lambda}$ and $\bar{\lambda}$

To calibrate the bounds $\bar{\nu}$, $\underline{\lambda}$ and $\bar{\lambda}$ with realistic values, we run the following numerical experiment. For $T = 20$ and $k = 7$, we simulate $10,000$ draws from random $T \times k$ matrix $\tilde{F}$ such that $vec(\tilde{F}) \sim N(0, I_{Tk})$ and set $F = V_\varepsilon^{1/2} U \Gamma^{1/2}$, $U = \tilde{F}(\tilde{F}'\tilde{F})^{-1/2}$, $G = V_\varepsilon^{1/2} Q$, $Q = \tilde{Q}(\tilde{Q}'\tilde{Q})^{-1/2}$, $\tilde{Q}$ are the first $T - k$ columns of $I_T - UU'$, for $V_\varepsilon = diag(V_{\varepsilon,11}, ..., V_{\varepsilon,TT})$, with $V_{\varepsilon,tt} = 1.5$ for $t = 1, ..., 10$, and $V_{\varepsilon,tt} = 0.5$ for $t = 11, ..., 20$, and $\Gamma = Tdiag(4, 3.5, 3, 2.5, 2, 1.5, 1)$, $c_{k+1} = 10T$, and $\xi_{k+1} = e_1$. With these choices, the "signal-to-noise" $\frac{1}{T} F_j' V_\varepsilon^{-1} F_j$ for the seven factors $j = 1, ..., 7$ are $4, 3.5, 3, 2.5, 2, 2.5, 1$, and the "signal-to-noise" for the weak factor is $\frac{1}{T} F_{k+1}' V_\varepsilon^{-1} F_{k+1} = 10n^{-1/2}$. Moreover, the errors follow the ARCH model of Section E.5.3 (i) with ARCH parameters either (a) $\alpha_i = 0.2$ for all i, or (b) $\alpha_i = 0.5$ for all i, and $q = 4$, and $\kappa = 0$ (cross-sectional independence). The choices $\alpha_i = 0.2$, 0.5 both meet the condition $3\alpha_i^2 < 1$ ensuring the existence of fourth-order moments. Moreover, with $q - 1 = 3$, we have a cross-sectional variance of the σ_{ii} that is three times larger than the mean (normalized to 1). For each draw, we compute the $df = 71$ non-zero eigenvalues and associated eigenvectors of $\Omega_{\bar{Z}^*}$, and the values of parameters ν_j and λ_j. In our simulations (a) with $\alpha_i = 0.2$, the draws of $\max_{j=1,...,df} \nu_j$ range between 0.21 and

[53] From Footnote 39, we know that Inequalities (6) are automatically met with $df = 1$. A given value of df may result from several different combinations of T and k, while a given T implies different values of df depending on k. For instance, $df = 2$ applies with $(T, k) = (4, 1)$, $(8, 4)$, and $(13, 8)$, among other combinations. For $T = 20$, the tests for $k = 1, 2, ..., 14$ yield $df = 170, 151, 133, 116, 100, 85, 71, 58, 46, 35, 25, 16, 8, 1$, respectively.

[54] This normalization results from ranking the eigenvalues μ_j, so that μ_1 is the smallest one.

0.30, with 95% quantile equal to 0.28, while the 5% and 95% quantiles of the λ_j are 0.13 and 7.65. Instead, (b) with $\alpha_i = 0.5$, the $\max_{j=1,\ldots,df} \nu_j$ range between 0.70 and 0.79, with 95% quantile equal to 0.77, and the 5% and 95% quantiles of the λ_j are 0.12 and 6.64. To get further insights in the choice of parameters $\bar{\nu}, \underline{\lambda}, \bar{\lambda}$, we also consider the values implied by the FA estimates in our empirical analysis. Here, when testing for the last retained k in a given subperiod, the median across subperiods of $\max_{j=1,\ldots,df} \nu_j$ is 0.76, and smaller than about 0.90 in most subperiods. Similarly, assuming $c_{k+1} = 10T$ and $\xi_{k+1} = e_1$ as above, the median values of the smallest and the largest estimated λ_j are 0.0024 and 5.84. Inspired by these findings, we set $\bar{\lambda} = 7$, and consider $\bar{\nu} = 0.2, 0.7, 0.9, 0.99$, and $\underline{\lambda} = 0, 0.1, 0.5, 1$, to get realistic settings with different degrees of dissimilarity from the case with serially uncorrelated squared errors (increasing with $\bar{\nu}$), and separation of the alternative hypothesis from the null hypothesis (increasing with $\underline{\lambda}$).

F.2 Results with Monte Carlo draws

In Table 2, the entries are nil for $\bar{\nu}$ sufficiently small and $\underline{\lambda}$ sufficiently large, suggesting that the AUMPI property holds for those cases that are closer to the setting with uncorrelated squared errors and sufficiently separated from the null hypothesis. Violations of Inequalities (6) concern $df = 2, 3, 4, 5$.[55] Let us focus on the setting with $\bar{\nu} = 0.7$ and $\underline{\lambda} = 0.1$. We find 3752 violations of Inequalities (6) out of 10^8 simulations, all occurring for $df = 2$, except 65 for $df = 3$. For those draws violating Inequalities (6) for $df = 2$, a closer inspection shows that (a) they feature values ν_2 close to upper bound $\bar{\nu} = 0.7$, and values of λ_2 close to lower bound $\underline{\lambda} = 0.1$, and (b) several of them yield non-monotone density ratios $\frac{f(z;\lambda_1,\lambda_2)}{f(z;0,0)}$, with the non-monotonicity region corresponding to large values of z. As an illustration, let us take the density ratio for $df = 2$ with

[55] A given number of simulated draws become increasingly sparse when considering larger values of df, which makes the exploration of the parameter space more challenging in those cases. However, unreported theoretical considerations show via an asymptotic approximation that the monotone likelihood property holds for $df \to \infty$ since the limiting distribution is then Gaussian. This finding resonates with the absence of violations in Table 2 for the larger values of df.

$\nu_2 = 0.666$, $\lambda_1 = 1.372$, and $\lambda_2 = 0.130$. Here, the eigenvalues of the covariance matrix are $\mu_1 = 1$ (by normalization) and $\mu_2 = (1 - \nu_2)^{-1} = 2.994$, and the non-centrality parameter λ_2 is small. The quantiles of the asymptotic distribution under the null hypothesis for asymptotic size $\alpha = 20\%, 10\%, 5\%, 1\%, 0.1\%$ are $9.3, 12.8, 16.2, 24.5, 36.5$. Non-monotonicity applies for $z \geq 16$. The optimal rejection regions $\{\frac{f(z;\lambda_1,\lambda_2)}{f(z;0,0)} \geq C\}$ correspond to those of the LR test $\{z \geq \tilde{C}\}$, e.g., for asymptotic levels such as $\alpha = 20\%$, but not for $\alpha = 5\%$ or smaller. Indeed, in the latter cases, because of non-monotonicity of the density ratio, the optimal rejection regions are finite intervals in argument z. With $\bar{\nu} = 0.7$, we do not find violations with $\underline{\lambda} = 0.5$ or larger.

	df	2	3	4	5	6	7	8	9	10	11	12
$\bar{\nu} = 0.2$	$\underline{\lambda} = 0$	0.002	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.000	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0	0	0	0	0	0	0	0	0	0	0
$\bar{\nu} = 0.7$	$\underline{\lambda} = 0$	0.051	0.000	0.000	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.037	0.000	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0	0	0	0	0	0	0	0	0	0	0
$\bar{\nu} = 0.9$	$\underline{\lambda} = 0$	0.151	0.004	0.000	0.000	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.134	0.004	0.000	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0.007	0.000	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0	0	0	0	0	0	0	0	0	0	0
$\bar{\nu} = 0.99$	$\underline{\lambda} = 0$	0.426	0.015	0.000	0.000	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.411	0.014	0.000	0.000	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0.218	0.007	0.000	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0.078	0.001	0	0	0	0	0	0	0	0	0

Table 2: Numerical check of Inequalities (6) by Monte Carlo. We display the cumulative frequency of violations in ‰ of Inequalities (6), for $m = 3, ..., 16$, over 10^8 random draws of the parameters $\lambda_j \sim Unif[\underline{\lambda}, \bar{\lambda}]$ and $\nu_j \sim Unif[0, \bar{\nu}]$, for $\bar{\lambda} = 7$, and different combinations of bounds $\underline{\lambda}$, $\bar{\nu}$, and degrees of freedom df. An entry 0.000 corresponds to less than 100 cases out of 10^8 draws.

G Maximum value of k as a function of T

In Table 3, we report the maximal values for the number of latent factors k to have $df \geq 0$, or $df > 0$.

T	1	2	3	4	5	6	7	8	9	10	11	12
$df \geq 0$	0	0	1	1	2	3	3	4	5	6	6	7
$df > 0$	NA	0	0	1	2	2	3	4	5	5	6	7
T	13	14	15	16	17	18	19	20	21	22	23	24
$df \geq 0$	8	9	10	10	11	12	13	14	15	15	16	17
$df > 0$	8	9	9	10	11	12	13	14	14	15	16	17

Table 3: Maximum value of k. We give the maximum admissible value k of latent factors so that the order conditions $df \geq 0$ and $df > 0$ are met, with $df = \frac{1}{2}[(T-k)^2 - T - k]$, for different values of the sample size $T = 1, ..., 24$. Condition $df \geq 0$ is required for FA estimation, and condition $df > 0$ is required for testing the number of latent factors.