SUPPLEMENTARY MATERIALS FOR CODING

Latent Factor Analysis in Short Panels

Alain-Philippe Fortin, Patrick Gagliardini, and Olivier Scaillet

This file provides implementation details for the codes developed for the empirics and simulation experiments. Section 1 provides details for implementation of the LR test for the number of latent factors. Section 2 provides details for implementation of canonical correlation estimators to investigate the stability of the factor structure within windows. Section 3 gives details for implementation of the RS and KP tests for spanning. Section 4 gives a table under equally-spaced grids of values for ν_j and λ_j for the numerical checks of conditions (12). Section 5 reports Monte Carlo results to assess size and power of the LR test in finite samples. Section 6 collects the analytical characterization of the coefficients $c_k(\lambda_1, ..., \lambda_{df})$ and the maximum value of k as a function of T.

1 Implementation of LR test

We get the FA estimates via the zigzag routine (Magnus and Neudecker (2007), p. 407). We compute the p-value for the LR test by simulating a large number of draws (10,000) from the variate $\sum_{j=1}^{df} \hat{\mu}_j \chi_j^2(1)$, where $\hat{\mu}_j$ are the nonzero eigenvalues of $\hat{\Omega}_{\bar{Z}^*}$ (see Proposition 4). To regularize the estimate $\hat{\Omega}_{\bar{Z}^*}$, we use the parametric structure from (D.14), and get the $T-1$ estimated parameters by least squares, as detailed in OA Section D.4.3 i).

2 Implementation of canonical correlation estimators

We divide each window of 20 months into two overlapping subperiods of 16 months (overlap of 12 months). Let Y_1 and Y_2 denote the resulting two panels of returns, with the last 12 months of panel 1 corresponding to the first 12 months of panel 2. Moreover, let k_c denote the number of common factors between the two panels, defined as the number of factors having the same loadings in both panels. We first estimate the total number of latent factors k_1 and k_2 in each panel by sequential testing as in Section 6.1. Here, k_1 and k_2 include both common and specific factors. Then, we obtain the estimates $\hat{F}_l, \hat{\beta}_l, \hat{V}_{l,\varepsilon}$, $l = 1, 2$, by performing FA separately on each panel. We define the estimators of canonical correlations among the betas as the square roots of the eigenvalues of $\hat{R} = (\frac{\hat{\beta}_1'\hat{\beta}_2}{n} - \hat{B}_{1,2})(\frac{\hat{\beta}_2'\hat{\beta}_1}{n} - \hat{B}_{2,1})$, where $\hat{B}_{1,2} = (\hat{F}_1'\hat{V}_{1,\varepsilon}^{-1}\hat{F}_1)^{-1}\hat{F}_1' \begin{bmatrix} 0 & 0 \\ \hat{V}_{c,\varepsilon}^{-1} & 0 \end{bmatrix} \hat{F}_2(\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}$, $\hat{B}_{2,1} = \hat{B}_{1,2}'$, and $\hat{V}_{c,\varepsilon}$ is a consistent estimator of the idiosyncratic variances in the overlapping subperiod of 12 months, i.e., an average of the estimates from panel 1 and panel 2; see also Andreou et al. (2019, AGGR) in a large n, T setting. The term $\hat{B}_{1,2}$ corrects for the bias induced by the error-in-variable (EIV) problem coming from beta estimates with fixed T, and by the overlap of the two subperiods.

Using the FA expansions from Section 3.1, we can show that $\hat{R} = R + O_p(\frac{1}{\sqrt{n}})$, where R is a symmetric matrix whose k_c largest eigenvalues are equal to 1, while the remaining eigenvalues are strictly less than 1. It allows for consistent estimation of k_c (see also AGGR). Indeed, let $g(n)$ be a function satisfying $g(n) \to 0$ and $\sqrt{n}g(n) \to \infty$, and let $\xi(j) = \hat{\rho}_j - 1 + g(n)$ for $j = 1, \ldots, \min(\hat{k}_1, \hat{k}_2)$, where $\hat{\rho}_j$ denotes the jth largest canonical correlation estimate. Finally, let $\hat{k}_c = \max\{0 \leq k \leq \underline{k} : \xi(k) > 0\}$, where we define $\xi(0) := g(n)$. Then, we can show that $\hat{k}_c = k_c$ with probability approaching 1 as $n \to \infty$. To implement this estimator of the number of common factors, we use $g(n) = (\log n)^2/\sqrt{n}$, scaled by the average of estimated canonical correlations. Unreported Monte Carlo experiments show good finite-sample performance of that choice. We define the fraction of common factors as $\hat{k}_c/\min(\hat{k}_1, \hat{k}_2)$. We leave the formal derivation of the asymptotic properties of such a procedure in a large-n fixed-T setting to future research.

3 Implementation of spanning tests

In this section, we give estimator $\hat{\Omega}_\Psi$ and prove its consistency. To start let us provide the explicit expression of Ω_Ψ. We use $\eta_{i,t} f_t = (f_t' \otimes I_{k^O})(\tilde{\beta}_i \otimes [z_{i,t} - E(z_{i,t})])$. By the CLT, $vec(\Psi'_{F^0,n}) \Rightarrow N(0, \Omega_\Psi^*)$, where $\Omega_\Psi^* = (\Omega_{\Psi,t,s}^*)$ has $k^O \times k^O$ blocks $\Omega_{\Psi,t,s}^* = (f_t' \otimes I_{k^O}) Q_{\eta\eta,ts}(f_s \otimes I_{k^O}) + 1_{t,s} V_{\varepsilon,tt} Q_{zz,tt}$ with $Q_{\eta\eta,t,s} = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{J_n}\sum_{i,j\in I_m} E[vec(\eta_{i,t})vec(\eta_{j,s})']$ and $Q_{zz,tt} := \lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{J_n}\sum_{i,j\in I_m}\sigma_{i,j}E[z_{i,t}z_{j,t}']$. Then, $\Omega_\Psi = \mathcal{K}_{k^O,T}\Omega_\Psi^*\mathcal{K}_{k^O,T}'$ and we use the commutation matrix $\mathcal{K}_{k^O,T}$ because Ω_Ψ is the asymptotic variance of the vec of $\Psi_{F^O,n}$ instead of $\Psi'_{F^O,n}$. To ease estimation of $Q_{\eta\eta,t,s}$ we assume that the $z_{i,t}$ are i.i.d. across i within a block, weakly stationary across t, $t = 1, ..., T$, with $E[z_{i,t}] = \mu_{z,m}$ and $Cov(z_{i,t}, z_{i,s}) = V_{z,m}(t-s)$ for $i \in I_m$, and mutually independent across blocks. Then, we have $Q_{\eta\eta,ts} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n(\tilde{\beta}_i\tilde{\beta}_i') \otimes Cov(z_{i,t}, z_{i,s}) = \lim_{n\to\infty}\sum_m B_{m,n} Q_{\tilde{\beta},m} \otimes V_{z,m}(t-s) = Q_{\tilde{\beta}} \otimes \bar{V}_z(t-s)$ with $Q_{\tilde{\beta},m} := \frac{1}{b_{m,n}}\sum_{i\in I_m}(\tilde{\beta}_i\tilde{\beta}_i')$, $Q_{\tilde{\beta}} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \tilde{\beta}_i\tilde{\beta}_i' = I_k + \mu_{\tilde{\beta}}\mu_{\tilde{\beta}}'$, $\mu_{\tilde{\beta}} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\tilde{\beta}_i$, $\bar{V}_z(t-s) = \lim_{n\to\infty}\sum_m B_{m,n}V_{z,m}(t-s)$, and we assume that $Q_{\tilde{\beta},m}$ and $V_{z,m}(h)$ are asymptotically uncorrelated across blocks. Then, the estimator is $\hat{\Omega}_\Psi = \mathcal{K}_{k^O,T}\hat{\Omega}_\Psi^*\mathcal{K}_{k^O,T}'$, where matrix $\hat{\Omega}_\Psi^* = (\hat{\Omega}_{\Psi,t,s}^*)$ has $k^O \times k^O$ blocks given by $\hat{\Omega}_{\Psi,t,s}^* = (\hat{f}_t' \otimes I_{k^O})\hat{Q}_{\eta\eta,ts}(\hat{f}_s \otimes I_{k^O}) + 1_{t,s}\hat{V}_{\varepsilon,tt}\hat{Q}_{zz,tt}$. Here, $\hat{Q}_{zz,tt} = \frac{1}{n(T-k)}\sum_m\sum_{i,j\in I_m}(z_{i,t}z_{j,t}')(\hat{\varepsilon}_j'\hat{V}_\varepsilon^{-1}\hat{\varepsilon}_i)$ where the GLS residuals are $\hat{\varepsilon}_i = M_{\hat{F},\hat{V}_\varepsilon}(y_i - r_f)$, and $\hat{Q}_{\eta\eta,ts} = \hat{Q}_{\tilde{\beta}} \otimes \hat{\bar{V}}_z(t-s)$, where $\hat{\bar{V}}_z(h) = \sum_m B_{m,n}\hat{V}_{z,m}(h)$ with $\hat{V}_{z,m}(h) = \frac{1}{(b_{m,n}-1)(T-h)}\sum_{i\in I_m}\sum_{t=h+1}^T(z_{i,t} - \bar{z}_{m,t})(z_{i,t-h} - \bar{z}_{m,t-h})'$ and $\bar{z}_{m,t} = \frac{1}{b_{m,n}}\sum_{i\in I_m} z_{i,t}$, and $\hat{Q}_{\tilde{\beta}} = I_k + \hat{\mu}_{\tilde{\beta}}(\hat{\mu}_{\tilde{\beta}})'$ with $\hat{\mu}_{\tilde{\beta}} = \frac{1}{n}\sum_{i=1}^n\hat{\tilde{\beta}}_i$ being the average of the estimated loadings $\hat{\tilde{\beta}}_i = \left(\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F}\right)^{-1}\hat{F}'\hat{V}_\varepsilon^{-1}(y_i - r_f)$.

Let us now show consistency of $\hat{\Omega}_\Psi$. Let us first consider $\hat{Q}_{zz,tt}$. Let $\hat{\Psi} := \frac{1}{n}\sum_m\sum_{i,j\in I_m}(z_{i,t}z_{j,t}') \otimes (\hat{\varepsilon}_i\hat{\varepsilon}_j')$ and use that $\hat{\varepsilon}_i = M_{\hat{F},\hat{V}_\varepsilon}F\tilde{\beta}_i + M_{\hat{F},\hat{V}_\varepsilon}\varepsilon_i$. Then, $\hat{\Psi} = \hat{\Psi}_1 + \hat{\Psi}_2 + (\hat{\Psi}_2)' + \hat{\Psi}_3$, where $\hat{\Psi}_1 = \frac{1}{n}\sum_m\sum_{i,j\in I_m}(z_{i,t}z_{j,t}') \otimes (M_{\hat{F},\hat{V}_\varepsilon}F\tilde{\beta}_i\tilde{\beta}_j'F'M_{\hat{F},\hat{V}_\varepsilon})$, $\hat{\Psi}_2 = \frac{1}{n}\sum_m\sum_{i,j\in I_m}(z_{i,t}z_{j,t}') \otimes (M_{\hat{F},\hat{V}_\varepsilon}F\tilde{\beta}_i\varepsilon_j'M_{\hat{F},\hat{V}_\varepsilon}')$, $\hat{\Psi}_3 = \frac{1}{n}\sum_m\sum_{i,j\in I_m}(z_{i,t}z_{j,t}') \otimes (M_{\hat{F},\hat{V}_\varepsilon}\varepsilon_i\varepsilon_j'M_{\hat{F},\hat{V}_\varepsilon}')$. Using $M_{\hat{F},\hat{V}_\varepsilon}F = O_p(\frac{1}{\sqrt{n}})$ and boundedness conditions on the $\tilde{\beta}_i$ and $z_{i,t}$, we get $\hat{\Psi}_1 = O_p\left(\frac{1}{n^2}\sum_m b_{m,n}^2\right) = O_p\left(\sum_m B_{m,n}^2\right) = o_p(1)$. Further, from $\sum_{i\in I_m} z_{i,t} \otimes \varepsilon_i = O_p\left(\sqrt{b_{m,n}}\right)$ uniformly in m, we get $\hat{\Psi}_2 = O_p\left(\frac{1}{n^{3/2}}\sum_m b_{m,n}^{3/2}\right) = O_p\left(\sum_m B_{m,n}^{3/2}\right) = o_p(1)$, using $\sum_m B_{m,n}^{3/2} = o(1)$ as in the proof of Proposition 4 (c). Finally,

$\hat{\Psi}_3 = (I_{k^O} \otimes M_{\hat{F},\hat{V}_\varepsilon}) \left(\frac{1}{n}\sum_m \sum_{i,j\in I_m}(z_{i,t}z'_{j,t}) \otimes(\varepsilon_i\varepsilon'_j)\right)(I_{k^O} \otimes M'_{\hat{F},\hat{V}_\varepsilon}) = (I_{k^O} \otimes M_{F,V_\varepsilon})(Q_{zz,tt} \otimes V_\varepsilon)(I_{k^O} \otimes M'_{F,V_\varepsilon}) + o_p(1) = Q_{zz,tt} \otimes (M_{F,V_\varepsilon}V_\varepsilon) + o_p(1)$. By post-multiplication with $I_{k^O} \otimes \hat{V}_\varepsilon^{-1}$, taking the trace in the second term of the Kronecker product and dividing by $T-k$, we deduce that $\hat{Q}_{zz,tt} = \frac{1}{n(T-k)}\sum_m \sum_{i,j\in I_m}(z_{i,t}z'_{j,t})(\hat{\varepsilon}'_j\hat{V}_\varepsilon^{-1}\hat{\varepsilon}_i)$ is a consistent estimator for $Q_{zz,tt}$. Further, we have $\hat{V}_{z,m}(h) = \frac{1}{T-h}\sum_{t=h+1}^T \hat{V}_{z,m,t}(h)$ with $\hat{V}_{z,m,t}(h) = \frac{1}{b_{m,n}-1}\sum_{i\in I_m}(z_{i,t}-\bar{z}_{m,t})(z_{i,t-h}-\bar{z}_{m,t-h})'$. We use that $\hat{\hat{V}}_{z,m,t}(h)$ is an unbiased estimator of $V_{z,m}(h)$ for any t. Then, $E[\hat{V}_{z,m}(h)] = V_{z,m}(h)$ and $E[\hat{\hat{V}}_z(h)] = V_z(h) + o(1)$. Moreover, $\|V(vec[\hat{V}_{z,m,t}(h)])\| \leq C/b_{m,n}$ uniformly in t, and $\|V(vec[\hat{V}_{z,m}(h)])\| \leq C/b_{m,n}$ because T is finite. Then, $\|V(vec[\hat{\hat{V}}_z(h)])\| \leq C\sum_m B_{m,n}^2/b_{m,n} = O(1/n) = o(1)$. Consistency of estimator $\hat{\hat{V}}_z(h)$ in mean-square error sense, and hence in probability, follows. Finally, $\hat{\mu}_{\tilde{\beta}}$ is a consistent estimator of $\mu_{\tilde{\beta}}$ because $\bar{\varepsilon} = o_p(1)$.

We conclude the section with a computational note. We can dispense of using the commutation matrix for obtaining the values of the test statistics and critical values. Indeed we have $\mathcal{K}'_{k^O,T}(\hat{V}_{k^O-r} \otimes \hat{U}_{T-r}) = \mathcal{K}_{T,k^O}(\hat{V}_{k^O-r} \otimes \hat{U}_{T-r}) = (\hat{U}_{T-r} \otimes \hat{V}_{k^O-r})\mathcal{K}_{T-r,k^O-r}$. Then, we get $\hat{\Omega}_{\mathcal{S}} = \mathcal{K}'_{T-r,k^O-r}\hat{\Omega}^*_{\mathcal{S}}\mathcal{K}_{T-r,k^O-r}$ with $\hat{\Omega}^*_{\mathcal{S}} = (\hat{U}_{T-r} \otimes \hat{V}_{k^O-r})'\hat{\Omega}^*_{\Psi}(\hat{U}_{T-r} \otimes \hat{V}_{k^O-r})$. Matrices $\hat{\Omega}_{\mathcal{S}}$ and $\hat{\Omega}^*_{\mathcal{S}}$ have the same eigenvalues, and $\mathscr{S}_{KP} = nvec(\hat{\mathcal{S}}'_{22})'(\hat{\Omega}^*_{\mathcal{S}})^{-1}vec(\hat{\mathcal{S}}'_{22})$. This symmetry holds because we can construct the KP and RS statistics from the SVD of $(\hat{F}^O)'$.

4 Numerical checks of conditions (12) with equally-spaced grids

To further investigate the validity of Inequalities (12), we conduct a series of experiments using equally-spaced grids of values for ν_j and λ_j. Because the distribution of a weighted sum of non-central chi-square variates is invariant under permutations of the pairs (ν_j, λ_j), $j = 1, ..., df$, w.l.o.g. we can assume that they are ranked in increasing order of the ν_j parameters, i.e., we rank the eigenvalues of the variance-covariance matrix $\Omega_{\bar{Z}^*}$ in increasing order. Thus, we check Inequalities (12) only for the grid points with $0 \leq \nu_2 \leq \nu_3 \leq \cdots \leq \nu_{df}$. In each dimension ν_j and λ_j we use the same number N of discretization points, and we adapt N to df in order to keep the total number of grid points G smaller than 10^8, see Table 1.

Let us focus first on the cases with $\bar{\nu} = 0.7$. When $\underline{\lambda} = 0.5$, we observe no violations of Inequalities (12) for all considered df. It confirms the findings in OA Section E that violations are confined to small values of the non-centrality parameters. When $\underline{\lambda} = 0.1$, we observe some violations for values of df up to 7. For instance, with $df = 4$, we find 483 violations out of the $2.2 \cdot 10^6$ grid points, which correspond to frequency $0.220‰$ in Table 1. They all involve at least one non-centrality parameter at the lowest grid value $\lambda_j = 0.1$, especially in combination with a large value ν_j for the same j. Such combinations of low non-centrality parameters for the large eigenvalues correspond to alternatives that are close to the null hypothesis. Moreover, the violations of Inequalities (12) typically occur for the coefficients with large m, say 12 or larger, and the negative values are very small - of order 10^{-9} or smaller. These findings are similar for the other values of $df \leq 7$.

Overall, Table 1 of SMC corroborates the findings in Table 1 of the paper obtained with Monte Carlo draws, namely we do not observe violations of Inequalities (12) for values of $\underline{\lambda}$ sufficiently large, and $\bar{\nu}$ sufficiently small. The numerically larger frequencies of violations reported in Table 1 of SMC are explained by the grid yielding by construction higher weights for λ_j very close to the lower boundary $\underline{\lambda}$ compared to uniform random draws.

5 Monte Carlo results

This section explores the finite sample properties of the test statistic $LR(k)$. We first introduce the Data Generating Process (DGP) that we use in our Monte Carlo analysis, and then present the results for the size and power of the LR statistic.

5.1 Data Generating Process

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 (2022)) via the ARCH $h_t = 0.6 + 0.5h_{t-1}z_{t-1}^2$, with $z_t \sim IIN(0,1)$. 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 $c_i = 1 - \alpha_i$ to ensure the normalization $V[\varepsilon_{i,t}/h_t^{1/2}] = \sigma_{ii} = \frac{c_i}{1-\alpha_i} = 1$. 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 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 from (D.14). We get the $T - 1$ estimated parameters by least squares, as detailed in OA Section D.4.3 i). The p-values are computed over $5,000$ draws.

5.2 Size and power results

We provide the size and power results in % in Table 2. 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 close or equal to 100%, while the local power ranges from 29% to 33% for $T = 6$, 69% to 89% for $T = 12$, and is equal to 100% for $T = 24$. The approximate constancy of local power w.r.t. n, for large n, is coherent with theory implying convergence to asymptotic local power.

6 Additional material

6.1 Analytical characterization of the coefficients $c_k(\lambda_1, ..., \lambda_{df})$

We use $Q(\lambda_1, ..., \lambda_{df}) = \frac{1}{2}(Y_1+Y_2+...+Y_{df})$, where $Y_j = (\sqrt{\gamma_j}X_j+\sqrt{1-\gamma_j}\lambda_j)^2$. By the binomial theorem, we have: $E[Q(\lambda_1, ..., \lambda_{df})^k] = \frac{1}{2^k}E[(Y_1 + Y_2 + ... + Y_{df})^k] = \frac{1}{2^k}\sum_{l=0}^{k} C_{k,l}E[(Y_1 + Y_2 + ... + Y_{df-1})^l]E[Y_{df}^{k-l}] = \sum_{l=0}^{k} C_{k,l}E[Q(\lambda_1, ..., \lambda_{df-1})^l]\frac{1}{2^{k-l}}E[Y_{df}^{k-l}]$, where the $C_{k,l} = \frac{k!}{l!(k-l)!}$ are the combinatorial coefficients. Dividing both sides of the equation by $k!$, we get:

$$c_k(\lambda_1, ..., \lambda_{df}) = \sum_{l=0}^{k} c_l(\lambda_1, ..., \lambda_{df-1})d_{k-l}(\lambda_{df}), \tag{1}$$

where $d_m(\lambda_{df}) := \frac{1}{2^m m!}E[Y_{df}^m]$. We have $E[Y_{df}^m] = E[(\sqrt{\gamma_{df}}X_{df} + \sqrt{1-\gamma_{df}}\lambda_{df})^{2m}]$ $= \sum_{l=0}^{2m} C_{2m,l}\gamma_{df}^{l/2}E(X_{df}^l)(1-\gamma_{df})^{(2m-l)/2}\lambda_{df}^{2m-l} = \sum_{l=0}^{m} C_{2m,2l}\gamma_{df}^l E(X_{df}^{2l})(1-\gamma_{df})^{m-l}(\lambda_{df}^2)^{m-l} = \sum_{l=0}^{m} C_{2m,2l}\gamma_{df}^l \frac{(2l)!}{2^l l!}(1-\gamma_{df})^{m-l}(\lambda_{df}^2)^{m-l}$; so $d_m(\lambda_{df}) = \frac{1}{2^m m!}\sum_{l=0}^{m} C_{2m,2l}\frac{(2l)!}{2^l l!}\gamma_{df}^l(1-\gamma_{df})^{m-l}(\lambda_{df}^2)^{m-l}$.

For computational purposes, we can rewrite (1) in matrix recursive form. Indeed, for a maximum index $\bar{k}$, let us define the vector $c(\lambda_1, ..., \lambda_{df})$ with elements $c_k(\lambda_1, ..., \lambda_{df})$ for $k = 0, 1, ..., \bar{k}$. Further, let us define the lower triangular matrix

$$D(\lambda_{df}) = \begin{pmatrix} 1 & 0 & \dots & & \\ d_1(\lambda_{df}) & 1 & 0 & \dots & \\ d_2(\lambda_{df}) & d_1(\lambda_{df}) & 1 & 0 & \dots \\ \vdots & & \ddots & \ddots & 0 \\ d_{\bar{k}}(\lambda_{df}) & \cdots & & d_1(\lambda_{df}) & 1 \end{pmatrix}.$$

Then, we have the recursion $c(\lambda_1, ..., \lambda_{df}) = D(\lambda_{df})c(\lambda_1, ..., \lambda_{df-1})$. By backward iteration, we get: $c(\lambda_1, ..., \lambda_{df}) = D(\lambda_{df})D(\lambda_{df-1})\cdots D(\lambda_2)c(\lambda_1)$, where $c(\lambda_1)$ is the vector with elements

$\frac{1}{2^k k!}\lambda_1^{2k}$, for $k = 0, 1, ..., \bar{k}$. Such a recursive analytical characterization is particularly useful in the numerical checks of conditions (12) of Proposition 7, when df becomes large (see OA Section E and SMC Section 4).

6.2 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 $df > 0$.

7 Additional references

Andreou, E., Gagliardini, P., Ghysels, E. and Rubin, M., 2019. Inference in group factor models with an application to mixed frequency data. Econometrica, 87(4), 1267-1305.

	df	2	3	4	5	6	7	8	9	10	11	12
	N	100	20	10	10	5	5	4	4	3	3	3
$\bar{\nu} = 0.2$	$\underline{\lambda} = 0.01$	1.551	0.466	0.072	0.005	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.047	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
$\bar{\nu} = 0.7$	$\underline{\lambda} = 0.01$	5.734	1.539	0.356	0.034	0.007	0.001	0	0	0	0	0
	$\underline{\lambda} = 0.1$	2.810	1.023	0.220	0.019	0.003	0.000	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
$\bar{\nu} = 0.9$	$\underline{\lambda} = 0.01$	8.855	2.279	0.504	0.052	0.011	0.001	0	0	0	0	0
	$\underline{\lambda} = 0.1$	5.579	1.672	0.375	0.036	0.009	0.001	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0.164	0.064	0.012	0.000	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0	0	0	0	0	0	0	0	0	0	0
$\bar{\nu} = 0.99$	$\underline{\lambda} = 0.01$	12.578	3.641	0.783	0.083	0.018	0.002	0.000	0.000	0	0	0
	$\underline{\lambda} = 0.1$	9.200	2.901	0.627	0.063	0.013	0.002	0.000	0.000	0	0	0
	$\underline{\lambda} = 0.5$	2.517	0.817	0.141	0.010	0.003	0.000	0.000	0.000	0	0	0
	$\underline{\lambda} = 1$	0.823	0.219	0.019	0.000	0	0	0	0	0	0	0

Table 1: Numerical check of Inequalities (12) by grid. We display the frequency of violations in ‰ of Inequalities (12), for $m = 1, ..., 16$, over a grid of equally-spaced points for the parameters $\lambda_j \in [\underline{\lambda}, \bar{\lambda}]$ and $\nu_j \in [0, \bar{\nu}]$, for $\bar{\lambda} = 7$ and different combinations of bounds $\underline{\lambda}$, $\bar{\nu}$ and degrees of freedom df. We use N discretization points in each parameter dimension, as reported in the second line. The number of grid points with $0 \equiv \nu_1 \leq \nu_2 \leq \nu_3 \leq ... \leq \nu_{df}$ is $G = \binom{N+df-2}{df-1} N^{df}$. It ranges between 10^6 and $7.15 \cdot 10^7$ with the chosen values of N.

T	Size (%) 6	12	24	Global Power (%) 6	12	24	Local Power (%) 6	12	24
$n = 500$	5.5	5.1	6.9	100	100	100	29	69	100
	(1.8)	(0.3)	(0.4)	(1.3)	(0.0)	(0.0)	(11.3)	(10.8)	(0.1)
$n = 1000$	5.3	4.9	5.7	100	100	100	33	82	100
	(1.7)	(0.3)	(0.4)	(1.4)	(0.0)	(0.0)	(16.9)	(8.1)	(0.1)
$n = 5000$	5.1	4.8	4.9	99	100	100	30	89	100
	(1.5)	(0.3)	(0.3)	(7.1)	(0.0)	(0.0)	(14.1)	(6.0)	(0.1)

Table 2: For each sample size combination (n, T), we provide the size and power in %. Nominal size for statistic $LR(2)$ is 5%. Power refers to rejection frequencies for statistic $LR(2)$ under global alternative $\bar{\kappa} = 0$ and local alternative $\bar{\kappa} = 0.5$. In parentheses, we report the standard deviations for size and power across 100 different draws of the factor path.

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.