Rank and Factor Loadings Estimation in Time Series Tensor Factor Model by Pre-averaging

Clifford Lam (joint work with Weilin Chen)

LSE Department of Statistics Research Showcase, June 14-15, 2022

Outline of the Talk

- Tensor factor models - model and examples
- Basic tensor manipulations
- Pre-averaging and results
- Projection - Re-estimation
- Projection - Rank determination for core rank tensor
- Simulation studies
- Summary and Future Research

Tensor Time Series

- Tensor (multi-dimensional array) time series examples:
  - Genomics - Multiple gene-gene interaction network of correlations from DNA microarray.
  - Neuroimaging analysis - Tensor response (e.g. MRI 3-dimensional array) and vector predictors. Decomposition of regression coefficient tensor.
  - Economics - import-export volume time series of products among different countries.
  - Finance - 10 by 10 Fama-French return time series (e.g. 100 portfolios formed on 10 sizes and 10 Book-to-Market ratios/Operating profitability).

Can we find *simplifying* structures? *Factors* driving the dynamics of a particular category of variables?

Tensor Factor Models

- For a panel time series $\mathbf{x}_t \in \mathbb{R}^p$ (*order-1 tensor*), a multi-factor model is
$$\mathbf{x}_t = C_t + \epsilon_t = \mathbf{A}\mathbf{f}_t + \epsilon_t, \quad t = 1, \ldots, T.$$

- If $\mathbf{x}_t \in \mathbb{R}^{d_1 \times d_2}$, an *order-2 tensor*, then the *Tucker decomposition* of the common component C_t is
$$C_t = \mathbf{A}_1 \mathbf{f}_t \mathbf{A}_2^{\mathrm{T}}.$$

- Two factor loading matrices, $\mathbf{A}_1 \in \mathbb{R}^{d_1 \times r_1}$, $\mathbf{A}_2 \in \mathbb{R}^{d_2 \times r_2}$.
- The factor series is $\mathbf{f}_t \in \mathbb{R}^{r_1 \times r_2}$. $\mathbf{A}_1$ is relevant to the dynamics of the row variables, as $\mathbf{A}_2$ does for the columns.

Tensor Factor Models

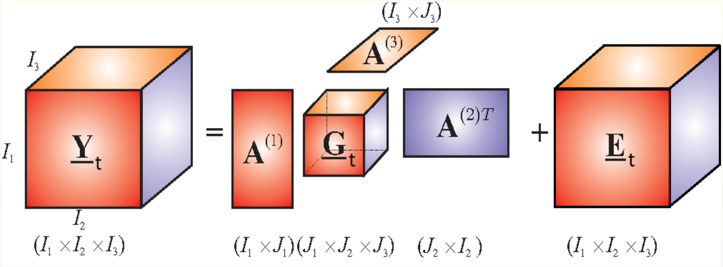


Figure: Tensor factor model for order-3 tensor time series. [A. Phan and A. Cichocki (2011)]

Notations and Basic Manipulations

- For a general *order-K tensor* $\mathcal{X}_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, write $\mathcal{X}_t = \mathcal{C}_t + \mathcal{E}_t$. The *Tucker decomposition* of $\mathcal{C}_t$ is

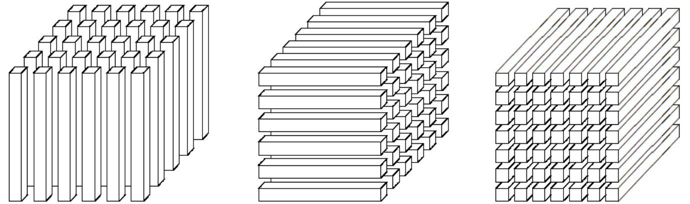
$$\mathcal{C}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \cdots \times_K \mathbf{A}_K.$$

- $\mathcal{F}_t$ is also called the *core tensor*.

- The notation $\times_k$ represents the k-mode product of a tensor $\mathcal{F} \in \mathbb{R}^{r_1 \times \cdots r_K}$ with a matrix $\mathbf{A} \in \mathbb{R}^{d \times r_k}$: $\mathcal{F} \times_k \mathbf{A} \in \mathbb{R}^{r_1 \times \cdots r_{k-1} \times d \times r_{k+1} \times \cdots \times r_K}$, where

$$(\mathcal{F} \times_k \mathbf{A})_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_K} = \sum_{i_k=1}^{r_k} f_{i_1 i_2 \cdots i_K} a_{j i_k}.$$

- Mode-k fibres of a tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is defined by fixing all indices but the k-th.



(a) Mode-1 (column) fibers: $\mathbf{x}_{:jk}$ (b) Mode-2 (row) fibers: $\mathbf{x}_{i:k}$ (c) Mode-3 (tube) fibers: $\mathbf{x}_{ij:}$

Figure: Fibres of order-3 tensors. (Figure from Kolda and Bader (2009))

- k-mode product $\mathcal{F} \times_k \mathbf{A}$ is to sort all mode-k fibres of $\mathcal{F}$ in columns, pre-multiply them with $\mathbf{A}$, then put them back into their corresponding places.

Notations and Basic Manipulations

- Mode-k flattening/unfolding/matricization of $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is to put all mode-k fibres as columns into a matrix $\mathsf{mat}_k(\mathcal{X})$ of size $d_k \times d_{\text{-}k}$, with $d_{\text{-}k} := \prod_{j \neq k} d_j$.

- If $\mathcal{C}_t = \mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \cdots \times_K \mathbf{A}_K$, then

$$\begin{aligned} \mathsf{mat}_k(\mathcal{C}_t) &= \mathbf{A}_k \mathsf{mat}_k(\mathcal{F}_t)(\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^{\mathrm{T}} \\ &=: \mathbf{A}_k \mathsf{mat}_k(\mathcal{F}_t) \mathbf{A}_{\text{-}k}^{\mathrm{T}}. \end{aligned}$$

★ We want to estimate $\mathbf{A}_1, \ldots, \mathbf{A}_K$, and determine the ranks of the *core tensor* $r_1, \ldots, r_K$ from data $\mathcal{X}_t = \mathcal{C}_t + \mathcal{E}_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, $t = 1, \ldots, T$.

Statistical and Econometric Factor Models

Two different types of assumptions for time series factor models:

- 'Statistical Factor Model' (Lam, Yao and Bathia (2011))
  - Common factors accommodate all dynamics. White noise, but allowing strong cross-correlations.
- 'Econometrics Factor Model' (Bai and Ng (2002))
  - Common factors have impact on most of the series. The noise has weak serial dependence and weak cross-correlations.
- Recent developments are based on statistical factor model assumptions.

★ No current literature on tensor factor models under econometric assumptions.

Pre-averaging idea

- Multiply Mode-k unfolded data by a vector $\mathbf{u}$:

$$\begin{aligned}\mathbf{y}_t &:= \mathsf{mat}_k(\mathcal{X}_t - \bar{\mathcal{X}})\mathbf{u} \\ &= \mathbf{A}_k \mathsf{mat}_k(\mathcal{F}_t - \bar{\mathcal{F}})\mathbf{A}_{\text{-}k}^{\mathrm{T}}\mathbf{u} + \mathsf{mat}_k(\mathcal{E}_t - \bar{\mathcal{E}})\mathbf{u}.\end{aligned}$$

- If $\mathbf{A}_k = \mathbf{U}_k \mathbf{G}_k \mathbf{V}_k^{\mathrm{T}}$, $\mathbf{A}_{\text{-}k}^{\mathrm{T}} = \mathbf{V}_{\text{-}k}\mathbf{G}_{\text{-}k}\mathbf{U}_{\text{-}k}^{\mathrm{T}} \Rightarrow \mathbf{u} = \mathbf{U}_{\text{-}k,(1)}$ inflates signal most, but unknown.

- Set $\mathbf{u} = \mathbf{1}_S$ for some set $S \subseteq [d_{\text{-}k}]$. $\mathbf{U}_{\text{-}k}^{\mathrm{T}}\mathbf{1}_S$ can be small $\Rightarrow$ Try random S.

- Estimate $\mathbf{A}_k$ (or part of it) by finding the first $z_k (\leq r_k)$ eigenvectors of the covariance matrix of $\mathbf{y}_t$.

- Set $z_k = 1$ for estimating the best direction $\mathbf{U}_{k,(1)}$, which usually is the most accurately estimated.

LSE

Assumptions on the Errors

(E1) (Decomposition of error) *Assume that*

$$\mathrm{mat}_k(\mathcal{E}_t) = (\boldsymbol{\xi}_{t,1}^{(k)}, \ldots, \boldsymbol{\xi}_{t,d_{\text{-}k}}^{(k)}), \quad \textit{where}$$
$$\boldsymbol{\xi}_{t,\ell}^{(k)} := \boldsymbol{\Psi}_{\ell}^{(k)} \mathbf{e}_t^{(k)} + (\boldsymbol{\Sigma}_{\epsilon,\ell}^{(k)})^{1/2} \boldsymbol{\epsilon}_{t,\ell}^{(k)},$$

with $E(\mathbf{e}_t^{(k)}) = \mathbf{0}$, $E(\boldsymbol{\xi}_{t,\ell}^{(k)}) = \mathbf{0}$, $\mathbf{e}_t^{(k)} \in \mathbb{R}^{r_e}$ *independent of* $\boldsymbol{\epsilon}_{s,\ell}^{(k)}$, $\boldsymbol{\epsilon}_{t,\ell}^{(k)}$ *independent of* $\boldsymbol{\epsilon}_{t,m}^{(k)}$ *for* $\ell \neq m$, $\mathrm{var}(\mathbf{e}_t^{(k)}) = \mathbf{I}_{r_e}$ *and* $\mathrm{var}(\boldsymbol{\epsilon}_{t,\ell}^{(k)}) = \mathbf{I}_{d_k}$ *for each* $s, t \in [T]$, $\ell, m \in [d_{\text{-}k}]$, $k \in [K]$. *Also, each* $\boldsymbol{\Sigma}_{\epsilon,\ell}^{(k)}$ *has non-vanishing diagonals with tr*$(\boldsymbol{\Sigma}_{\epsilon,\ell}^{(k)}) = O(d_k)$. *Moreover, denote* $\boldsymbol{\Psi}^{(k)} := \sum_{\ell=1}^{d_{\text{-}k}} \boldsymbol{\Psi}_l^{(k)}$ *and* $\boldsymbol{\Sigma}_{\epsilon}^{(k)} := \sum_{\ell=1}^{d_{\text{-}k}} \boldsymbol{\Sigma}_{\epsilon,\ell}^{(k)}$. *Then we assume* $\|\boldsymbol{\Psi}^{(k)} \boldsymbol{\Psi}^{(k)\mathrm{T}}\| = O(d_{\text{-}k})$ *and* $\|\boldsymbol{\Sigma}_{\epsilon}^{(k)}\| = O(d_{\text{-}k})$.

With (E1), each mode-k fibre of $\mathcal{E}_t$ is a sum of two independent parts. The first part $\boldsymbol{\Psi}_{\ell}^{(k)} \mathbf{e}_t^{(k)}$ is similar to a common component in a factor model, but it is too weak to be detected. This promotes (weak) cross-correlations among the fibres.

Assumptions on the Errors

(E2) (Time series) *The elements in* $\mathbf{e}_t^{(k)} = (e_{t,j}^{(k)})$ *and* $\boldsymbol{\epsilon}_{t,\ell}^{(k)} = (\epsilon_{t,\ell,j}^{(k)})$ *are following weakly stationary general linear processes, such that with* $\ell \in [d_{\text{-}k}]$, $t \in [T]$ *and* $k \in [K]$,

$$e_{t,j}^{(k)} = \sum_{q \geq 0} a_{e,q} z_{e,t-q,j}^{(k)}, \quad j \in [r_e],$$

$$\epsilon_{t,\ell,j}^{(k)} = \sum_{q \geq 0} a_{\epsilon,q} z_{\epsilon,t-q,\ell,j}^{(k)}, \quad j \in [d_k],$$

where the coefficients $a_{e,q}$ *and* $a_{\epsilon,q}$ *are such that* $\sum_{q \geq 0} a_{e,q}^2 = \sum_{q \geq 0} a_{\epsilon,q}^2 = 1$ *and* $\sum_{q \geq 0} |a_{e,q}| \leq C$, $\sum_{q \geq 0} |a_{\epsilon,q}| \leq C$ *for some constant C. For each* $k \in [K]$, *the series of random variables* $\{z_{e,t,j}^{(k)}\}$ *and* $\{z_{\epsilon,t,\ell,j}^{(k)}\}$ *are independent of each other, with i.i.d. elements having mean 0 and variance 1.*

With (E2), the error variables are serially correlated in general. Together with (E1), (weak) serial and cross-sectional dependence within and among fibres are allowed for the errors.

Assumptions on the Factors

Similar to (E2), the factors in $\mathcal{F}_t$ are assumed to follow general linear processes.

(F1) *Let* $\mathbf{f}_{t,\ell}^{(k)} = (f_{t,\ell,j}^{(k)})$ *be the* ℓ*-th column vector in* $\text{mat}_k(\mathcal{F}_t)$, $\ell \in [r_{\text{-}k}]$, *where* $r_{\text{-}k} := \prod_{\ell \neq k} r_\ell$. *We assume that* $\text{var}(\mathbf{f}_{t,\ell}^{(k)}) = \mathbf{I}_{r_k}$ *(the identity matrix with size* r_k*), and* $\text{cov}(\mathbf{f}_{t,\ell_1}^{(k)}, \mathbf{f}_{t,\ell_2}^{(k)}) = \mathbf{0}$ *for* $\ell_1 \neq \ell_2$.
Then we can write

$$f_{t,\ell,j}^{(k)} = \sum_{q \geq 0} a_{f,q} z_{f,t-q,\ell,j}^{(k)}, \quad j \in [r_k],$$

where we have $\sum_{q \geq 0} a_{f,q}^2 = 1$ *and* $\sum_{q \geq 0} |a_{f,q}| \leq C$ *for some constant C.* *For each* $k \in [K]$, *the series of random variables* $\{z_{f,t,\ell,j}^{(k)}\}$ *has i.i.d. elements having zero mean and variance 1.*

LSE

Assumptions on the model parameters

(L1) (Factor Strength) *We assume that, for $k \in [K]$, $\mathbf{A}_k$ is of full rank, $r_k = o(T^{1/3})$, and as $d_k \to \infty$,*

$$\mathbf{D}_k^{-1/2} \mathbf{A}_k^{\mathrm{T}} \mathbf{A}_k \mathbf{D}_k^{-1/2} \to \Sigma_{\mathbf{A},k},$$

where $\mathbf{D}_k = \text{diag}(\mathbf{A}_k^{\mathrm{T}} \mathbf{A}_k)$ is a diagonal matrix, and $\Sigma_{\mathbf{A},k}$ is positive definite with all eigenvalues bounded away from 0 and infinity. Let $(\mathbf{D}_k)_j$ be the j-th diagonal element of $\mathbf{D}_k$, then we assume $(\mathbf{D}_k)_j \asymp d_k^{\alpha_{k,j}}$ for $j \in [r_k]$, and $0 < \alpha_{k,r_k} \leq \cdots \leq \alpha_{k,2} \leq \alpha_{k,1} \leq 1$.

(L1) states that the factors can have **different strengths**. It **generalizes the assumption of Bai and Ng (2021) to tensor time series with mixed strengths of factors**.

Can show also the j-th singular values in $\mathbf{G}_k$ is of order $d_k^{\alpha_{k,j}}$.

Assumptions on the model parameters

(L2) (Signal Cancellation of maximum eigenvalue ratio sample) *For* $k \in [K]$, *and for the* m*-th sample (of fibres) out of* M_0, *define*
$s_{k,max} := \max_{|\mathcal{S}_{k,m}|=n_k, m\in[M_0]} \left[\sum_{j=1}^{r_k} \left(\sum_{i\in\mathcal{S}_{k,m}} (\mathbf{A}_k)_{ij}\right)^2\right]$ *and*
$s_{\text{-}k,max} := \prod_{l\in[K]\setminus\{k\}} s_{l,max}$. *Then we assume*
$\frac{d_{\text{-}k}}{s_{\text{-}k,max}}\left(1+\frac{d_k}{T}\right) = o\left(d_k^{\alpha_{k,z_k}}\right)$, *for some* $z_k \leq r_k$.

- $\mathcal{S}_{k,m} \subseteq [d_{\text{-}k}]$ is set by the user through choosing $n_\ell \asymp d_\ell$ for each $\ell \in [K]$.
- Consider many $\mathcal{S}_{k,m}$ and choose those that have large $s_{k,max}$, and hence large $s_{\text{-}k,max}$.

★ If $\widetilde{\mathbf{\Sigma}}_y$ is the covariance matrix corresponding to $\mathcal{S}_{k,m}$, then

$$\frac{\lambda_1(\widetilde{\mathbf{\Sigma}}_y)}{\lambda_j(\widetilde{\mathbf{\Sigma}}_y)} \asymp \frac{d_k^{\alpha_{k,1}}}{\frac{d_{\text{-}k}}{s_{\text{-}k,m}}\left(1+\frac{d_k}{T}\right)}, \quad r_k+1 \leq j \leq \lfloor c\min(T,d_k)\rfloor - r_k,$$

for some constant $c > 0$. Hence we choose $\mathcal{S}_{k,m}$ by taking those leading to largest eigenvalue ratios.

LSE

Pre-Averaging Estimator

- With different $\mathcal{S}_{k,m}$ for different samples (only retain those with large $s_{\text{-}k,max}$), we can construct different covariance matrices for each such $\mathcal{S}_{k,m}$.
- The pre-averaging estimator is the z_k eigenvectors corresponding to the largest z_k eigenvalues of the sum of all the covariance matrices above.

Theorem

Under Assumptions (E1), (E2), (F1), (L1), (L2), (R1), (R2), with $n_l \asymp d_l$ for $l \neq k$, let $c_{k,max} := \min\left\{1 + \frac{d_k}{T}, \frac{r_k d_k}{T}\right\} \frac{d_{\text{-}k}}{s_{\text{-}k,max}} + d_k^{\alpha_{k,1}} \left(1 + \frac{d_k^2}{T^2}\right) \frac{d_{\text{-}k,}^2}{s_{\text{-}k,max}^2}$, then

$$\|\hat{\mathbf{U}}_{k,pre,(z_k)} - \mathbf{U}_{k,(z_k)}\|^2 = O_p\left(d_k^{-2\alpha_{k,z_k}}\left[d_k^{2\alpha_{k,1}}\frac{r_k}{T} + c_{k,max}\right]\right).$$

★ $\hat{\mathbf{U}}_{k,pre,(1)}$ serves as a good projection direction, i.e., take $\mathbf{c} = \hat{\mathbf{U}}_{k,pre,(1)}$.

Projection: Re-estimation through iterations

- The new projected data: $\check{\mathbf{q}}_k^{(0)} := \hat{\mathbf{U}}_{k,pre,(1)}$.

$$\mathbf{y}_{t,i}^{(k)} := \mathsf{mat}_k(\mathcal{X}_t - \bar{\mathcal{X}})\check{\mathbf{q}}_k^{(i-1)} \Rightarrow \widetilde{\mathbf{\Sigma}}_{y,i}^{(k)} := T^{-1}\sum_{t=1}^{T}\mathbf{y}_{t,i}^{(k)}\mathbf{y}_{t,i}^{(k)\mathrm{T}}.$$

- At the i-th step: For each $k \in [K]$, estimate $\mathbf{A}_k$ by $\check{\mathbf{q}}_k^{(i)}$, the eigenvector corresponding to the largest eigenvalue of $\widetilde{\mathbf{\Sigma}}_{y,i}^{(k)}$. Repeat for several times (usually results in convergence).

Theorem

Under all previous assumptions, at the mth step of iteration, and

$$r = O(r_e), \quad d_k = O\left(\prod_{j=1}^{K} d_j^{\alpha_{j,1}}\right) = (r_e + \sqrt{T})\sum_{j=1}^{d_{\text{-}k}} \|\Psi_j^{(k)}\|^2,$$

$$\max_{j\in[d_{\text{-}k}]} \|\mathbf{\Sigma}_{\epsilon,j}^{(k)}\| = O\left(\prod_{j=1}^{K} d_j^{\alpha_{j,1}}\sqrt{\frac{r}{T}}\right), \; K\left(r + \max_{j\in[d_{\text{-}k}]}\|\mathbf{\Sigma}_{\epsilon,j}^{(k)}\|\right)\prod_{j=1}^{K} d_j^{1-\alpha_{j,1}} = o(T),$$

we have for each $k \in [K]$, $\|\check{\mathbf{q}}_k^{(m)} - \mathbf{U}_{k,(1)}\| = O_P(\sqrt{r/T})$.

Projection: Re-estimation through iterations

Theorem

If r_k is known, we can obtain r_k eigenvectors of $\widetilde{\mathbf{\Sigma}}_{y,m+1}^{(k)}$ as an estimator of the factor loading space of $\mathbf{A}_k$. Then there exists $\check{\mathbf{U}}_k \in \mathbb{R}^{d_k \times r_k}$ with $\check{\mathbf{U}}_k^{\mathrm{T}} \check{\mathbf{U}}_k = \mathbf{I}_{r_k}$ such that the r_k eigenvectors obtained above is $\check{\mathbf{U}}_k$ multiplied with some orthogonal matrix, with

$$
\|\check{\mathbf{U}}_k - \mathbf{U}_k\| = O_P\Bigg\{ g_s^{-1/2} d_k^{\alpha_{k,1}-\alpha_{k,r_k}} \Bigg[\sqrt{\frac{r}{T}} \Bigg(\sqrt{d_k} + K\sqrt{\frac{rd}{T}} + \sqrt{r_e S_\psi^{(k)}} \Bigg) + g_s^{-1/2} \Bigg(\max_{j \in [d_{\text{-}k}]} \|\mathbf{\Sigma}_{\epsilon,j}^{(k)}\| \Bigg[1 + \frac{K^2 rd}{T^2} \Bigg] + S_\psi^{(k)} \Bigg) \Bigg] \Bigg\},
$$

where $g_s := \prod_{j=1}^K d_j^{\alpha_{j,1}}$, $S_\psi^{(k)} := \sum_{j=1}^{d_{\text{-}k}} \|\mathbf{\Psi}_j^{(k)}\|^2$.

Consider $d_1 \asymp \cdots \asymp d_K \asymp T$, K and r_k being constants, all factors for $\mathbf{A}_k$ are pervasive.

- $\|\check{\mathbf{U}}_k - \mathbf{U}_k\| = O_P(T^{-3/4})$ if $r_e = O(d_k^{1/2})$, $\|\mathbf{\Psi}_j^{(k)}\| = O(1)$ and $\|\mathbf{\Sigma}_{\epsilon,j}^{(k)}\| = O(d_{\text{-}k})$. Improves to $O_P(T^{-1})$ if $r_e \asymp d_k$ but $S_\psi^{(k)} = O(d_{\text{-}k}/T)$.

Projection: Core tensor rank estimation

Define the correlation matrix for each $k \in [K]$:

$$\widetilde{\mathbf{R}}_{y,m+1}^{(k)} := \mathsf{diag}^{-1/2}(\widetilde{\boldsymbol{\Sigma}}_{y,m+1}^{(k)})\widetilde{\boldsymbol{\Sigma}}_{y,m+1}^{(k)}\mathsf{diag}^{-1/2}(\widetilde{\boldsymbol{\Sigma}}_{y,m+1}^{(k)}).$$

- Our estimator for r_k for each $k \in [K]$ is then defined to be

$$\hat{r}_k := \max\{j : \lambda_j(\widetilde{\mathbf{R}}_{y,m+1}^{(k)}) > 1 + \eta_T,\ j \in [d_k]\},$$

where $\eta_T \to 0$ as $T \to \infty$.

Theory: Core Tensor Rank Estimation

Theorem

Let all previous assumptions hold, and

$$d_k^{-\alpha_{k,r_k}}\big(\max_{j\in[d_{\text{-}k}]}\|\mathbf{\Sigma}_{\epsilon,j}^{(k)}\|+S_\psi^{(k)}\big)=o\Bigg(\prod_{j=1;j\neq k}^{K}d_j^{\alpha_{k,1}}\Bigg).$$

Then as $T,d_k\to\infty$, we have for each $k\in[K]$,

$$\lambda_j(\widetilde{\mathbf{R}}_{y,m+1}^{(k)})=\begin{cases}\asymp d_k^{\alpha_{k,j}}(1+O_P\{a_T(\alpha_{k,1}-\alpha_{k,r_k})\}), & j\in[r_k];\\ \leq 1+O_P\{a_T(0)\}, & j\in[d_k]/[r_k],\end{cases}$$

where for $0\leq\delta\leq 1/2$,

$$a_T(\delta):=d_k^{\delta}\sqrt{\frac{r}{T}}+Krd_k^{\alpha_{k,1}/2}\prod_{j=1}^{K}d_j^{(1-\alpha_{j,1})/2}\frac{1}{T}=o(1).$$

Hence $\hat{r}_k$ is a consistent estimator for r_k if we choose $\eta_T=Ca_T(0)$ for some constant $C>0$.

Tuning parameter selection: Bootstrapping fibres

- If $\alpha_{j,1} = 1$ for each $j \in [K]$, and $d_k \asymp T$, then $a_T(0) \asymp KrT^{-1/2}$. It means that our search for η_T can be in the form $CT^{-1/2}$.

- Bootstrapping mode-k fibres: choose the $d_{\text{-}k}$ fibres randomly with replacement, and project them accordingly. That is, for $b = 1, \ldots, B$,

$$\mathbf{y}_{t,m+1,b}^{(k)} := \mathsf{mat}_k(\mathcal{X}_t - \bar{\mathcal{X}})\mathbf{W}_b\mathbf{W}_b^{\mathrm{T}}\check{\mathbf{q}}_{\text{-}k}^{(m)},$$

- The i-th column of $\mathbf{W}_b$ is $\mathbf{0}$ except at the j-th position (j randomly chosen from $[d_{\text{-}k}]$), it is replaced by an i.i.d. Bernoulli r.v.

Tuning parameter selection: Bootstrapping fibres

- For a constant C, we calculate

$$\hat{r}_k^{(b)}(C) := \max\{j : \lambda_j(\widetilde{\mathbf{R}}_{y,b}^{(k)}) > 1 + CT^{-1/2},\ j \in [d_k]\}.$$

- We propose to choose C with

$$\hat{C} := \min_{C>0} \widehat{\mathsf{Var}}(\{\hat{r}_k^{(b)}(C)\}_{b\in[B]}),$$

since $\lambda_j(\widetilde{\mathbf{R}}_{y,b}^{(k)}), b \in [B]$ are less stable for $j > r_k$ as compared to when $j \in [r_k]$.

- Finally, our estimator for r_k is

$$\check{r}_k := \text{Mode of } \{\hat{r}_k^{(b)}(\hat{C})\}_{b\in[B]}.$$

Simulation Settings

- Generate $\mathbf{A}_k = \mathbf{B}_k \mathbf{R}_k$, where the elements in $\mathbf{B}_k \in \mathbb{R}^{d_k \times r_k}$ are i.i.d. $U(u_1, u_2)$, and $\mathbf{R}_k \in \mathbb{R}^{r_k}$ is diagonal with the jth element being $d_k^{-\zeta_{k,j}}$, $0 \le \zeta_{k,j} \le 0.5$.
- Elements in $\mathcal{F}_t$, $\mathbf{e}_t^{(k)}$ and $\boldsymbol{\epsilon}_{t,\ell}^{(k)}$ are independent AR(5). $\boldsymbol{\Psi}^{(1)}$ with i.i.d. standard normal entries, but has an independent probability of 0.7 being set exactly to 0.
- $K = 2$, $d_1 = d_2 = 40$, $T = 100$ and $r_1 = r_2 = 2$.

(Ia) All strong factors $\zeta_{k,j} = 0$ for all k, j. $u_1 = -2$, $u_2 = 2$ so that columns of $\mathbf{A}_k$ sum to normal magnitude (small s_k).

(IIa) One strong factor with $\zeta_{k,1} = 0$ and $\zeta_{k,2} = 0.2$ for all k. $u_1 = -2$, $u_2 = 2$.

(IIIa) Two weak factors with $\zeta_{k,1} = 0.1$ and $\zeta_{k,2} = 0.2$ for all k. $u_1 = -2$, $u_2 = 2$.

- Setting (Ib)(IIb)(IIIb) are the same as (Ia)(IIa)(IIIa), except that $u_1 = 0$, $u_2 = 2$ so that column sums of $\mathbf{A}_k$ have large magnitude (large s_k).

Estimation of Factor Loading Spaces

Competitors to compare are TOPUP/TIPUP from Chen et al (2021), iTOPUP/iTIPUP from Han et al (2022), and HOSVD/HOOI.

Initial Step	Iterative Step
PRE	PROJ
1.157	0.126
TOPUP	iTOPUP
5.810	1.922
TIPUP	iTIPUP
0.082	1.798
HOSVD	HOOI
0.073	1.783

Table: Average Computational Time (in sec) for Factor Loading Estimators

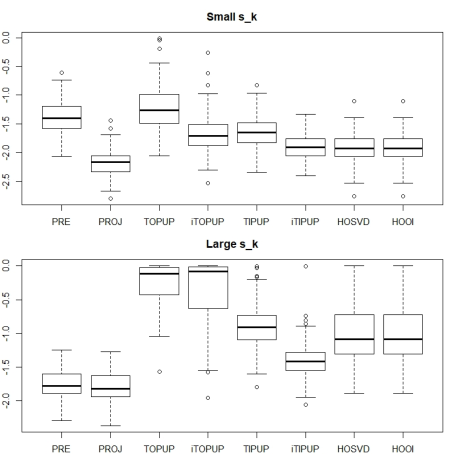


Figure: Box-plot of L_2 Estimation Error (log-scale) of Factor Loading Spaces of $\mathbf{A}_1$ for Setting (II).

Estimation of the Rank of Core Tensor

Setting	Method	$\check{r}_1$	$\check{r}_2$	CorrectProp1	CorrectProp2
Ia	Bootstrap	2.00	2.00	1	1
	iTIP-ER	2.00	2.00	1	1
Ib	Bootstrap	2.00	2.00	1	1
	iTIP-ER	1.79	1.86	0.79	0.86
IIa	Bootstrap	2.00	2.00	1	1
	iTIP-ER	1.89	1.83	0.89	0.83
IIb	Bootstrap	1.95	1.97	0.95	0.97
	iTIP-ER	1.16	1.18	0.16	0.18
IIIa	Bootstrap	1.92	1.99	0.92	0.95
	iTIP-ER	1.92	1.92	0.92	0.92
IIIb	Bootstrap	1.52	1.71	0.52	0.71
	iTIP-ER	1.09	1.09	0.09	0.09

Table: Comparison of the Bootstrapped Rank Estimator with iTIP-ER from Han et al (2022).

Analysis of Matrix-valued Financial Return Data

- Fama-French portfolio returns data on Size and Operating Profitability (OP). 100 returns categorized into 10 different Sizes and 10 different OP levels. Either value-weighted or equal-weighted.
- Monthly data July 1973 to June 2021 ($T = 576$). Market effects (NYSE composite) removed using CAPM.
- Both our bootstrap method and the iTIP-ER gives $\hat{r}_1 = \hat{r}_2 = 2$ for both Size and OP.
- HOOI, iTIPUP and our method all show similar grouping patterns (after varimax rotations).

	Value Weighted	Equal Weight
PROJ	677.3	737.3
iTIPUP	662.1	804.2
HOOI	626.4	683.8

Table: Average Sum of Squared of Residuals.

Summary and Future Research

- With econometric assumptions, a pre-averaging estimator for the factor loading matrices is proved to be consistent with rate of convergence spelt out.
- Re-estimation by iterating the projection step allows for a potentially better rate of convergence.
- Core rank tensor can be estimated by eigenanalyses of correlation matrices from suitably projected data. Bootstrapping tensor fibres help with the search for the optimal tuning parameter.
- Inference of tensor factor models using fibres Bootstrapping?

LSE

References

- Fan, J., Guo, J. and Zheng, S. (2020). Estimating Number of Factors by Adjusted Eigenvalues Thresholding, *JASA*, 0(0):1–10.
- Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. SIAM Review, 51(3):455–500.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.
- Bai, J. and Ng, S. (2021). Approximate Factor Models with Weaker Loadings. Papers 2109.03773, arXiv.org.
- Chen, R., Yang, D., and Zhang, C.-H. (2021). Factor models for high-dimensional tensor time series. *JASA*, 0(0):1–23.
- Han, Y., Zhang, C. H., and Chen, R. (2022). Rank determination in tensor factor model. *Electron. J. Statist.*, 16(1): 1726–1803.
- Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.
- Wang, D., Liu, X., and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1):231–248.