Tensor Principal Component Analysis with Applications in Finance

Andrii Babii   Eric Ghysels   Junsu Pan

UNC-Chapel Hill

March 18, 2024

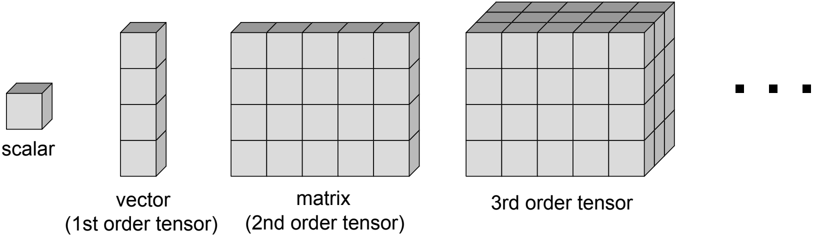
Presentation based on:

- Babii, A. and Ghysels, E. and Pan, Junsu (2023) *Tensor Principal Component Analysis*, `arXivpreprintarXiv:2212.12981`
- Babii, A. and Ghysels, E. and Pan, Junsu (2024) *Missing Financial Data: Filling the Tensor Blanks*, Work in progress

Overview

What is a Tensor?

A tensor is a multi-dimensional panel dataset — an array generalizing vectors and matrices to higher dimensions

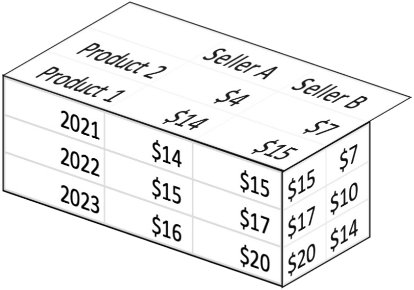


Long Table vs. Wide Table vs. Tensor

	Price		
Year	Seller A	Seller B	Product
2021	$14	$15	1
2022	$15	$17	1
2023	$16	$20	1
2021	$4	$7	2
2022	$5	$10	2
2023	$6	$14	2

	Price			
	Product 1		Product 2	
Year	Seller A	Seller B	Seller A	Seller B
2021	$14	$15	$4	$7
2022	$15	$17	$5	$10
2023	$16	$20	$6	$14

3-dimensional tensor $\rightarrow$



Examples of tensors in finance

- Decile sorted portfolios: (a) time, (b) characteristics and (c) decile. Usually deciles of characteristics sorted portfolios are collapsed into high-low spreads, yielding a panel
- Firm characteristics: (a) time, (b) firm and (c) characteristic. Typically one does panel analysis for each time period separately, see Bryzgalova, Lerner, Lettau, and Pelger (2022)
- International asset pricing has cross-section within countries, time and country dimensions. Typically one dimension is omitted
- Related literature: Lettau (2022) who looks at mutual funds

Big Picture

- The tensor structure is often ignored to apply standard methods.

Big Picture

- The tensor structure is often ignored to apply standard methods.
- This paper: apply insights from the PCA literature to tensors.

Big Picture

- The tensor structure is often ignored to apply standard methods.
- This paper: apply insights from the PCA literature to tensors.
- Our contribution:

Big Picture

- The tensor structure is often ignored to apply standard methods.
- This paper: apply insights from the PCA literature to tensors.
- Our contribution:
  - PCA-type estimators for d-dimensional panel data;

Big Picture

- The tensor structure is often ignored to apply standard methods.
- This paper: apply insights from the PCA literature to tensors.
- Our contribution:
  - PCA-type estimators for d-dimensional panel data;
  - statistical test for the number of latent factors;

Big Picture

- The tensor structure is often ignored to apply standard methods.
- This paper: apply insights from the PCA literature to tensors.
- Our contribution:
    - PCA-type estimators for d-dimensional panel data;
    - statistical test for the number of latent factors;
    - easy to use: closed-form estimators, avoid non-convex optimization, sequential computation of factors.

Table of Contents

Factor Model for a 2D Panel

- A classical 2D factor model

$$y_{it} = \sum_{r=1}^{R} \lambda_{ir} f_{tr} + u_{it},$$

  where
  - f_{tr} is a latent factor driving co-movement;
  - λ_{ir} is a latent exposure;
  - $i = 1, \dots, N$ is cross-section and $t = 1, \dots, T$ is time.

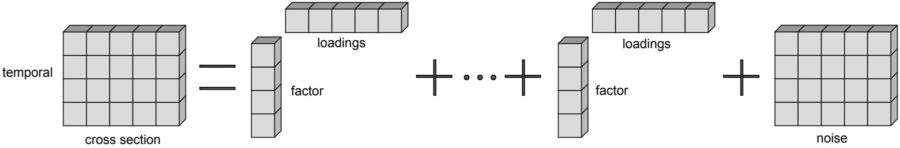
- Example: asset pricing
  - y_{it} is the excess returns of the i^{th} asset at time t;
  - explaining co-movement of asset returns with a small number of factors.

Factor Model for a 2D Panel

- Matrix notation:

$$\mathbf{Y} = \sum_{r=1}^{R} \lambda_r \otimes f_r + \mathbf{U},$$

where $\mathbf{Y} \in \mathbb{R}^{N \times T}$ and $\lambda_r \otimes f_r = \lambda_r f_r^\top$ is a tensor (outer) product



- **Objective:** identify and estimate the loadings $\lambda_r \in \mathbb{R}^N$ and the factors $f_r \in \mathbb{R}^T$.

Factor Models

Factor models are ubiquitous in economics/finance. A few examples are:

- Asset pricing: Ross (1976), Chamberlain and Rothschild (1983).
- Business cycle analysis: Sargent and Sims (1977).
- Consumer theory: Lewbel (1991).
- Forecasting with 'big data': Stock and Watson (2002).
- Unobserved heterogeneity and interactive fixed effects: Bai (2009), Moon and Weidner (2015), Cunha, Heckman, and Schennach (2010).

A Tensor Factor Model: 3-dimensional Case

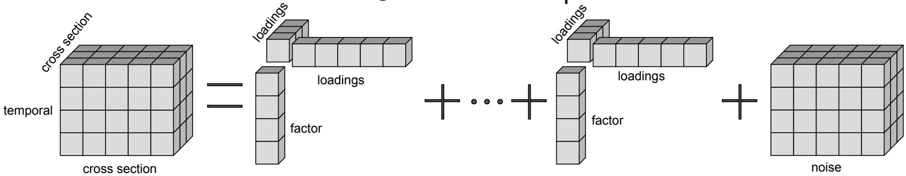
- Tensor factor model for 3-dimensional panel data

$$y_{ijt} = \sum_{r=1}^{R} \lambda_{ir}\mu_{jr}f_{tr} + u_{ijt},$$

where $\mu_{j,r}$ are loadings for the third dimension $j = 1, \ldots, J$. In tensor form

$$\mathbf{Y} = \sum_{r=1}^{R} \lambda_r \otimes \mu_r \otimes f_r + \mathbf{U},$$

where $\mathbf{Y} \in \mathbb{R}^{N \times J \times T}$ and $\otimes$ is the tensor product



- **Objective:** identify and estimate loadings $\lambda_r \in \mathbb{R}^N$, $\mu_r \in \mathbb{R}^J$, and factors $f_r \in \mathbb{R}^T$.

A Tensor Factor Model: d-dimensional Case

- This can be generalized to d dimensions

$$y_{i_1 i_2 \cdots i_d} = \sum_{r=1}^{R} v_{1,r}^{i_1} v_{2,r}^{i_2} \cdots v_{d,r}^{i_d} + u_{i_1 i_2 \cdots i_d},$$

  where $i_j = 1, \ldots, N_j$ and $j = 1, \ldots, d$.
- Collecting the data in a d-dimensional tensors $\mathbf{Y} \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d}$,

$$\mathbf{Y} = \sum_{r=1}^{R} \bigotimes_{j=1}^{d} v_{j,r} + \mathbf{U},$$

  where $\bigotimes_{j=1}^{d} v_{j,r} = v_{1,r} \otimes v_{2,r} \otimes \cdots \otimes v_{d,r}$.
- This is called the Canonical Polyadic Decomposition (CP-decomposition).

A Tensor Factor Model: d-dimensional Case

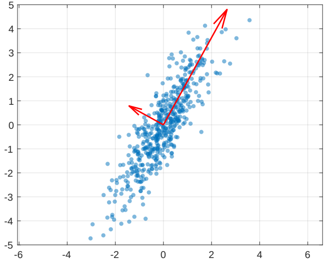
- For identification purposes, we normalize the loadings/factors

$$\mathbf{Y} = \sum_{r=1}^{R} \sigma_r \bigotimes_{j=1}^{d} m_{j,r} + \mathbf{U}, \qquad \mathbb{E}\mathbf{U} = 0,$$

  where $\sigma_r = \prod_{j=1}^{d} \|v_{j,r}\|$ is a scale and $m_{j,r} = v_{j,r}/\|v_{j,r}\|$ is a unit-norm loading/factor.
- σ_r interpreted as a signal strength.
- **Objective:** identify and estimate the unit-norm loadings/factors $m_{j,r} \in \mathbb{R}^{N_j}$ and scale components $(\sigma_r)_{r=1}^{R}$.
- Tucker (1958) introduced a more general decomposition which does not restrict the same number of factors along each tensor dimension.

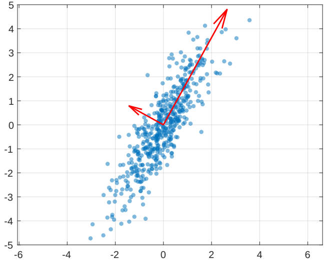
Table of Contents

Principal Component Analysis



- PCA constructs linearly weighted combination of variables
  - explaining most variance;
  - are mutually orthogonal.

Principal Component Analysis



- PCA constructs linearly weighted combination of variables
  - explaining most variance;
  - are mutually orthogonal.

- Solution: Eigendecomposition of $\mathbf{Y}^\top \mathbf{Y} = \hat{\Gamma}\hat{D}^2\hat{\Gamma}^\top$, where
  - $\hat{\Gamma} \rightarrow$ eigenvectors
  - $\hat{D}^2 \rightarrow$ eigenvalues.

2D Factor Model via PCA

- 2D factor model

$$\mathbf{Y} = \sum_{r=1}^{R} \lambda_r \otimes f_r + \mathbf{U} = \Lambda F^\top + \mathbf{U}$$

where matrices Λ, F collect the vectors λ_r, f_r.

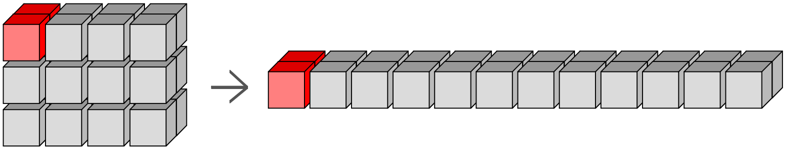
- Identifying loadings $\Lambda \in \mathbb{R}^{N\times R}$ via PCA assuming that $F^\top F = I_R$:

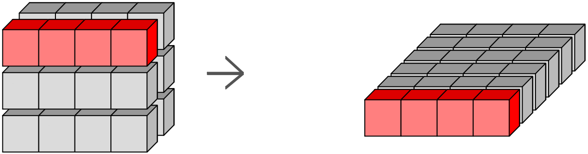
$$\mathbf{Y}\mathbf{Y}^\top = \Lambda\Lambda^\top + \text{noise} = \hat{\Gamma}\hat{D}^2\hat{\Gamma}^\top, \quad (1)$$

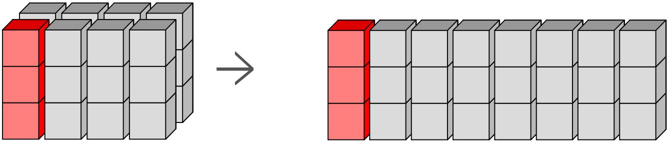
where
  - $\hat{\Gamma} \to$ eigenvectors of $\mathbf{Y}\mathbf{Y}^\top$
  - $\hat{D}^2 \to$ eigenvalues of $\mathbf{Y}\mathbf{Y}^\top$

- Estimate $\hat{\Lambda}$: the first R eigenvectors of $\mathbf{Y}\mathbf{Y}^\top$.
- Estimate $\hat{F}$: the first R eigenvectors of $\mathbf{Y}^\top\mathbf{Y}$, or regress Y on $\hat{\Lambda}$.

TPCA - Tensor Matricization







Tensor Matricization

Let $\mathbf{Y}$ be a $3 \times 4 \times 2$ dimensional tensor of the following two frontal slices:

$$\mathbf{Y}_1 = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix} \quad \mathbf{Y}_2 = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}.$$

Then the mode-1, 2 and 3 matricization of $\mathbf{Y}$ are respectively:

$$\mathbf{Y}_{(1)} = \begin{bmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{bmatrix}_{3\times 8},$$

$$\mathbf{Y}_{(2)} = \begin{bmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 18 \\ 7 & 8 & 9 & 19 & 20 & 21 \\ 10 & 11 & 12 & 22 & 23 & 24 \end{bmatrix}_{4\times 6},$$

$$\mathbf{Y}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 & \cdots & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & \cdots & 21 & 22 & 23 & 24 \end{bmatrix}_{2\times 12}.$$

Some intuition

- Tensor factor model for 3-dimensional panel data

$$y_{ijt} = \sum_{r=1}^{R} \lambda_{ir}\mu_{jr}f_{tr} + u_{ijt},$$

where $i = 1, \ldots, N$, $j = 1, \ldots, J$, and $t = 1, \ldots, T$.

- Collecting the data in a three-dimensional tensors $\mathbf{Y} \in \mathbb{R}^{N \times J \times T}$,

$$y_{ijt} = \sum_{r=1}^{R} [\mu_{jr}f_{tr}]\lambda_{ir} + u_{ijt},$$

$$y_{ijt} = \sum_{r=1}^{R} [\lambda_{ir}\mu_{jr}]f_{tr} + u_{ijt},$$

$$y_{ijt} = \sum_{r=1}^{R} [\lambda_{ir}f_{tr}]\mu_{jr} + u_{ijt},$$

Some notation/definitions

- **Khatri-Rao Product:** The KR product between two matrices $A \in \mathbb{R}^{I \times K}$ and $B \in \mathbb{R}^{J \times K}$, $A \odot B \in \mathbb{R}^{(IJ) \times K}$ corresponds to column-wise Kronecker Product:

$$A \odot B := \begin{bmatrix} a_1 \otimes_K b_1 & a_2 \otimes_K b_2 & \cdots & a_K \otimes_K b_K \end{bmatrix}$$

- **Hadamard Product:** The Hadamard product between two same-sized matrices $A \in \mathbb{R}^{I \times J}$ and $B \in \mathbb{R}^{I \times J}$, $A \circ B \in \mathbb{R}^{I \times J}$ corresponds to the element-wise matrix product:

$$A \circ B := \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2J}b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \cdots & a_{IJ}b_{IJ} \end{bmatrix}$$

Tensor Decomposition via PCA

- Recall that the tensor factor model for $\mathbf{Y} \in \mathbb{R}^{N \times J \times T}$ is

$$\mathbf{Y} = \sum_{r=1}^{R} \sigma_r \lambda_r \otimes \mu_r \otimes f_r + \mathbf{U},$$

  where the unit-norm λ_r, μ_r, and f_r can be collected in matrices Λ, M, F.

- Matricizing the tensor

$$\mathbf{Y}_{(1)} = \Lambda D (F \odot M)^\top + \mathbf{U}_{(1)},$$

  where $\mathbf{Y}_{(1)}$ and $\mathbf{U}_{(1)}$ are $N \times JT$ matrices and $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_R)$.

- Likewise, we could reshape the 3-way factor model as

$$\mathbf{Y}_{(2)} = MD(F \odot \Lambda)^\top + \mathbf{U}_{(2)} \qquad \text{or} \qquad \mathbf{Y}_{(3)} = FD(M \odot \Lambda)^\top + \mathbf{U}_{(3)},$$

  where $\mathbf{Y}_{(2)}, \mathbf{U}_{(2)}$ are $J \times NT$ matrices and $\mathbf{Y}_{(3)}, \mathbf{U}_{(3)}$ are $T \times JN$ matrices.

TPCA Algorithm

For a 3 dimensional tensor:

- Unfold the tensor $\mathbf{Y} \in \mathbb{R}^{N \times J \times T} \quad \rightarrow \quad \mathbf{Y}_{(1)}, \mathbf{Y}_{(2)}, \mathbf{Y}_{(3)}$
- Estimate loadings and factors:
  - $\hat{\Lambda} \rightarrow$ first R eigenvectors of $\mathbf{Y}_{(1)} \mathbf{Y}_{(1)}^\top$;
  - $\hat{M} \rightarrow$ first R eigenvectors of $\mathbf{Y}_{(2)} \mathbf{Y}_{(2)}^\top$;
  - $\hat{F} \rightarrow$ first R eigenvectors of $\mathbf{Y}_{(3)} \mathbf{Y}_{(3)}^\top$.
- Estimate the scale components:
  - $\hat{\sigma} \rightarrow$ first R eigenvalues of $\mathbf{Y}_{(1)} \mathbf{Y}_{(1)}^\top$.

Alternatively:

- Estimate factors $\hat{F}$ by regressing $\mathbf{Y}_{(3)}$ on the product of $\hat{\Lambda}$ and $\hat{M}$.
- Estimate scale $\hat{\sigma}_r \rightarrow \|\hat{f}_r\|$.

Appendix: Derivation

Table of Contents

Asymptotics of PCA tensor decomposition

Theorem

Suppose that $M_j^\top M_j = I_R$ *the idiosyncratic errors* $\mathbf{U}$ *are i.i.d. with* $\mathbb{E}\mathbf{U} = 0$ *and finite fourth moment. Then for all*

$$\|\hat{m}_{j,r} - m_{j,r}\| = O_P\left(\frac{\sqrt{N_j}\operatorname{tr}(D) + \left(N_j \vee \prod_{k\neq j} N_k\right)}{\delta_r}\right), \quad \forall r \leq R, j \leq d.$$

where δ_r *is the eigengap, namely:*

$$\delta_r = \min_{k\neq r} |\sigma_k^2 - \sigma_r^2| \wedge \sigma_r^2$$

which measures the strength of r^{th} *factor. Note that if we ignore the eigengap for the smallest eigenvalue (distance to zero) we have* $\delta_r = \min_{k\neq r} |\sigma_k^2 - \sigma_r^2|$

Rates of Consistency

- Strong tensor factor model: For any r^{th} factor, the signal strength σ_r increases proportional to $\sqrt{\left(\prod_{j=1}^{d} N_j\right) \cdot d_r}$ for some $d_1 > d_2 > \cdots > d_R > 0$ as $N_1, \ldots, N_d \to \infty$.

- Tensor dimensions improve the convergence rate

$$\|\hat{m}_{j,r} - m_{j,r}\| = O_P\left(\sqrt{\frac{1}{\prod_{k \neq j} N_k}}\right),$$

for the j^{th} dimension.

Appendix: Asymptotics

Asymptotic distribution: factors/loadings

- Suppose additionally that $\nu \in \mathbb{R}^{N_j}$ is such that

$$\lim_{N_j \to \infty} \sqrt{N_j} \langle m_{j,k}, \nu \rangle > 0.$$

- Examples:
  - i^{th} element of factors/loadings vector $m_{j,r} \in \mathbb{R}^{N_j}$: $\nu \in \mathbb{R}^{N_j}$ is the all zeros vector except for i^{th} coordinate equal to 1;
  - average factors/loadings: $\nu = N_j^{-1}(1, 1, \ldots, 1)$.
- Under some conditions on tensor dimensions, e.g. $N_1 \sim N_2 \sim N_3$ in the 3D case

$$\prod_{k \neq j} \sqrt{N_k} \langle \hat{m}_{j,k} - m_{j,k}, \nu \rangle \xrightarrow{d} N \left(0, \sigma^2 \sum_{k \neq r} \omega_{j,k}^2(\nu) \frac{d_r + d_k}{(d_r - d_k)^2} \right).$$

Testing the Number of Factors

Consider the following hypotheses

$$\mathrm{H}_0: \le k \text{ factors} \quad \text{vs.} \quad \mathrm{H}_1: \text{the number of factors is } > k\text{, but } \le K.$$

Eigenvalue ratio statistics inspired by Onatski (2009)

$$S_j = \max_{k<r\le K} \frac{\hat{\sigma}^2_{r,j} - \hat{\sigma}^2_{r+1,j}}{\hat{\sigma}^2_{r+1,j} - \hat{\sigma}^2_{r+2,j}}, \qquad 1 \le j \le d.$$

Theorem

Then under H_0, $S_j \xrightarrow{d} Z$, where Z can be approximated using type-1 Tracy-Widom distribution. Under H_1, we have $S_j \uparrow \infty$ for every $j \le d$.

Simulated Distribution of S_j

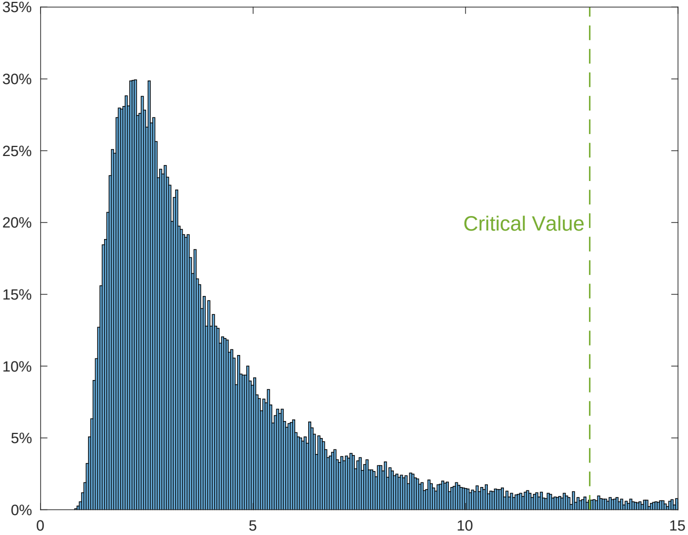


Table of Contents

Model Complexity

- Consider the following model for $\mathbf{Y} = \{y_{i,j,t}\} \in \mathbb{R}^{N\times J\times T}$:

$$y_{i,j,t} = \sum_{r=1}^{R} \sigma_r \lambda_{i,r} \mu_{j,r} f_{t,r} + u_{i,j,t}, \qquad \mathbb{E}(u_{i,j,t}) = 0,$$

- If pooling the tensor into a matrix and then applying PCA to $\mathbf{Y}_{(3)}\mathbf{Y}_{(3)}^\top$, where $\mathbf{Y}_{(3)} \in \mathbb{R}^{(NJ)\times T}$

$$y_{i,j,t} = \sum_{r=1}^{R} \sigma_r \beta_{i,j,r} f_{t,r} + u_{i,j,t}, \qquad \mathbb{E}(u_{i,j,t}) = 0.$$

  where $\beta_{i,j,r} = \lambda_{i,r}\mu_{j,r}$.
- Number of parameters:
  - Tensor $\rightarrow R \times (N + J + T)$;
  - Pooling $\rightarrow R \times (NJ + T)$.

Model Complexity: # of Parameters/Sample Size

Model	Number of factors 1	2	3	4	5
	Panel A: $T = 100, N = 30, J = 20$				
Tensor	0.25%	0.50%	0.75%	1.00%	1.25%
Pooling	1.17%	2.33%	3.50%	4.67%	5.83%
	Panel B: $T = 50, N = 50, J = 50$				
Tensor	0.12%	0.24%	0.36%	0.48%	0.60%
Pooling	2.04%	4.08%	6.12%	8.16%	10.2%
	Panel C: $T = 50, N = 100, J = 100$				
Tensor	0.05%	0.10%	0.15%	0.20%	0.25%
Pooling	2.01%	4.02%	6.03%	8.04%	10.05%

Model Fit: Tensor vs. Pooling

$R^2 := 1 - \text{RSS}/\text{TSS}$, where $\text{RSS} = \sum_{i,j,t} \hat{u}_{i,j,t}^2$ and $\text{TSS} = \sum_{i,j,t} y_{i,j,t}^2$.

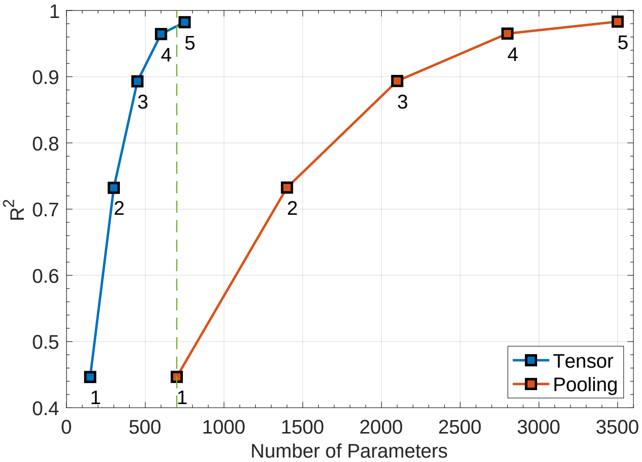


Figure: $100 \times 30 \times 20$ - 5 Factors

▸ Appendix: DGP

Tensor Decomposition via ALS Algorithm

- The low-rank tensor approximation is a non-convex optimization problem

$$\min_{\hat{\mathbf{Y}}} \|\mathbf{Y} - \hat{\mathbf{Y}}\| \quad \text{where} \quad \hat{\mathbf{Y}} = \sum_{r=1}^{R} \lambda_r \otimes \mu_r \otimes f_r$$

- **Alternating Least Squares (ALS) algorithm,** Kolda and Bader (2009), is the workhorse of tensor decomposition in numerical analysis.

Tensor Decomposition via ALS Algorithm

- For the 3-way tensor, the ALS algorithm solves sequentially the following steps repeatedly until convergence (where $[\cdot]^+$ is the Penrose inverse)

$$\hat{\Lambda} = \underset{\Lambda}{\text{argmin}} \|Y_{(1)} - \Lambda(F \odot M)^T\| = Y_{(1)}[(F \odot M)^T]^+$$
$$\hat{M} = \underset{M}{\text{argmin}} \|Y_{(2)} - M(F \odot \Lambda)^T\| = Y_{(2)}[(F \odot \Lambda)^T]^+$$
$$\hat{F} = \underset{F}{\text{argmin}} \|Y_{(3)} - F(M \odot \Lambda)^T\| = Y_{(3)}[(M \odot \Lambda)^T]^+$$

Tensor Decomposition via ALS Algorithm

- For the 3-way tensor, the ALS algorithm solves sequentially the following steps repeatedly until convergence (where $[\cdot]^+$ is the Penrose inverse)

$$\hat{\Lambda} = \underset{\Lambda}{\text{argmin}} \|Y_{(1)} - \Lambda(F \odot M)^T\| = Y_{(1)}[(F \odot M)^T]^+$$
$$\hat{M} = \underset{M}{\text{argmin}} \|Y_{(2)} - M(F \odot \Lambda)^T\| = Y_{(2)}[(F \odot \Lambda)^T]^+$$
$$\hat{F} = \underset{F}{\text{argmin}} \|Y_{(3)} - F(M \odot \Lambda)^T\| = Y_{(3)}[(M \odot \Lambda)^T]^+$$

- Requires initial starting values, and convergence is not guaranteed.

Tensor Decomposition via ALS Algorithm

- For the 3-way tensor, the ALS algorithm solves sequentially the following steps repeatedly until convergence (where $[\cdot]^+$ is the Penrose inverse)

$$\hat{\Lambda} = \underset{\Lambda}{\text{argmin}} \|Y_{(1)} - \Lambda(F \odot M)^T\| = Y_{(1)}[(F \odot M)^T]^+$$
$$\hat{M} = \underset{M}{\text{argmin}} \|Y_{(2)} - M(F \odot \Lambda)^T\| = Y_{(2)}[(F \odot \Lambda)^T]^+$$
$$\hat{F} = \underset{F}{\text{argmin}} \|Y_{(3)} - F(M \odot \Lambda)^T\| = Y_{(3)}[(M \odot \Lambda)^T]^+$$

- Requires initial starting values, and convergence is not guaranteed.

- Requires prior knowledge of rank R, whereas TPCA allows sequential computation of factors/loadings.

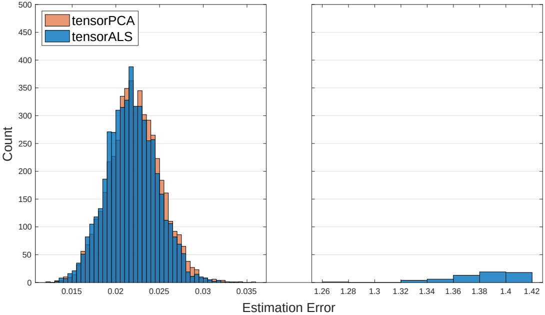
MC Experiment: TPCA vs. ALS

- Same DGP,
  - allow the true number of factors $R \in \{1, 2, 3, 4\}$,
  - always fit a one-factor model without the knowledge of the true R.

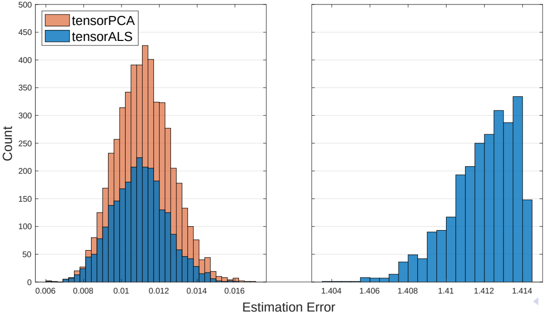
- Performance evaluation: the norm of estimation errors

$$\mathbb{L}_\lambda = \|\hat{\lambda}_r - \lambda_r\|,$$
$$\mathbb{L}_\mu = \|\hat{\mu}_r - \mu_r\|,$$
$$\mathbb{L}_f = \|\hat{f}_r - f_r\|.$$

Distribution of Errors: TPCA vs. ALS

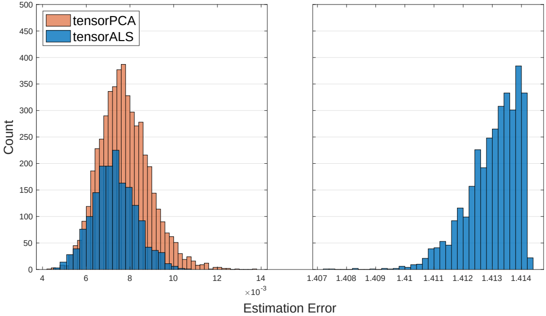


(a) 1 Factor

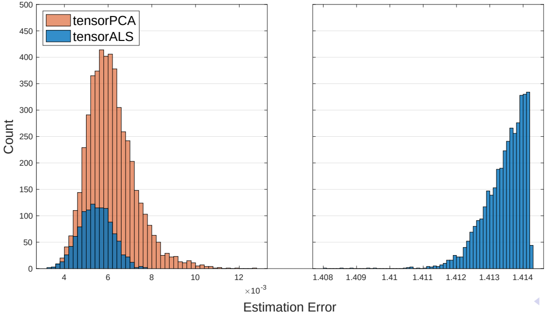


(b) 2 Factors

Distribution of Errors: TPCA vs. ALS



(a) 3 Factors



(b) 4 Factors

MC Experiment: Eigenvalue Ratio Test

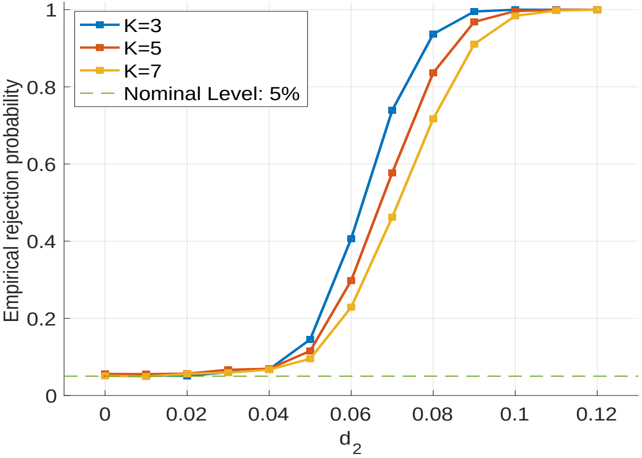
- We generate a 2-factor model, and test

$$\mathrm{H}_0: \leq 1 \text{ factors} \quad \text{vs.} \quad \mathrm{H}_1: \text{the number of factors is} > 1 \text{, but} \leq K.$$

- The signal strength
  - $\sigma_r = d_r \times \sqrt{NJT}$ with $d_1 = 2$,
  - gradually increase d_2 to study the power properties of the test.
  - $d_2 = 0$ implies a 1-factor model.

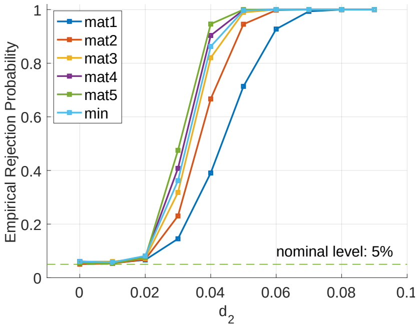
Power Curves of Eigenvalue Ratio Test

$\mathrm{H}_0: \leq 1$ factors vs. $\mathrm{H}_1:$ the number of factors is > 1, but $\leq K$.



Power Curves of Eigenvalue Ratio Test

- 5 dimensional tensor - Gaussian Errors



- Student's t-distributed idiosyncratic errors also work with the empirical power climbs slightly slower to 1.

Table of Contents

High-minus-low Risk Premia

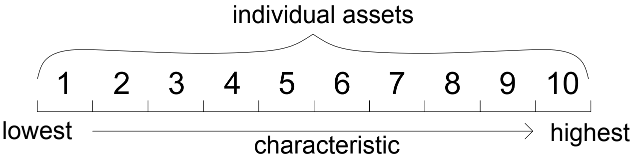
- Risk Premium: excess return that an investor expects to receive for taking on additional risk compared to a risk-free investment
- Examples: market, firm size, book-to-market ratio, operating profitability, investment strategies, *etc.*
- Market risk is the primary risk of assets

$$\text{Market Risk Premium} = \text{Expected Return} - \text{Risk Free Rate}$$

where Expected Return is return of S&P 500 or value-weighted average of all US common shares, and Risk Free Rate is treasury bill rate.

High-minus-low Risk Premia

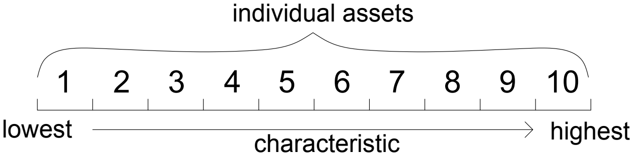
- Constructing characteristic risk premium: The difference between the returns of portfolios with the highest and lowest characteristic.



- This can be done at any point of time t and for any characteristic i, making it a 3-dimensional tensor.

High-minus-low Risk Premia

- Constructing characteristic risk premium: The difference between the returns of portfolios with the highest and lowest characteristic.



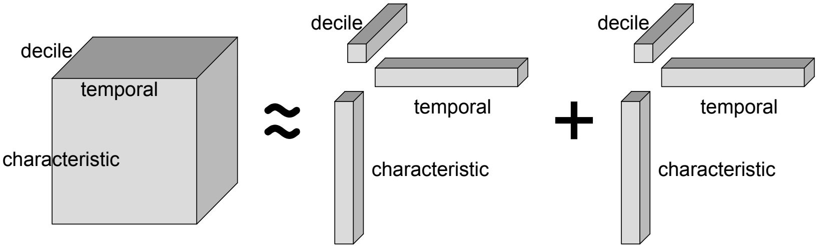
- This can be done at any point of time t and for any characteristic i, making it a 3-dimensional tensor.
- Question: Are characteristic risk premia biased by the market risk?

High-minus-low risk premia

- Consider a 3D factor model:

$$y_{i,j,t} = \sum_{r=1}^{R} \sigma_r \lambda_{i,r} \mu_{j,r} f_{t,r} + u_{i,j,t}, \qquad \mathbb{E}(u_{i,j,t}) = 0,$$

where $y_{i,j,t}$ is the excess return of the j^{th} quantile of the i^{th} characteristic at time t.



High-minus-low risk premia

- Consider a 3D factor model:

$$y_{i,j,t} = \sum_{r=1}^{R} \sigma_r \lambda_{i,r} \mu_{j,r} f_{t,r} + u_{i,j,t}, \qquad \mathbb{E}(u_{i,j,t}) = 0,$$

where $y_{i,j,t}$ is the excess return of the j^{th} quantile of the i^{th} characteristic at time t.

- High-minus-low risk premia:

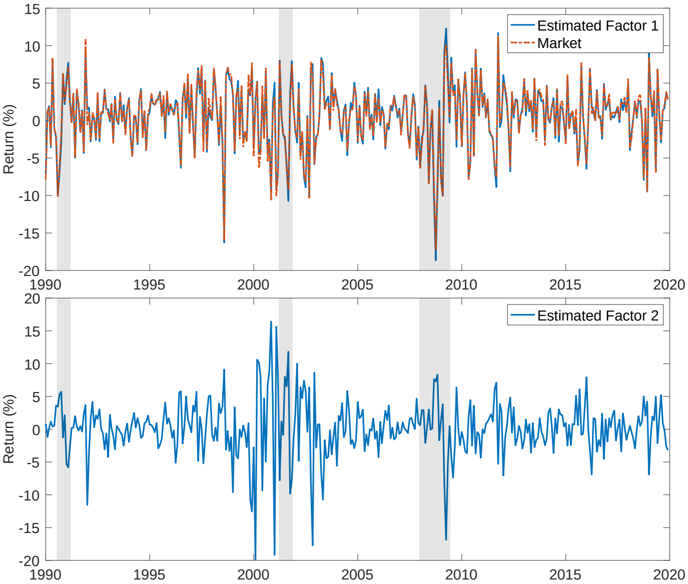
$$y_{it}^{10-1} = \sigma_1 \lambda_{i,1}(\mu_{10,1} - \mu_{1,1}) f_{t,1} + \sigma_2 \lambda_{i,2}(\mu_{10,2} - \mu_{1,2}) f_{t,2} + u_{it}^{10-1},$$

where $y_{it}^{10-1} = y_{i,10,t} - y_{i,1,t}$ and $u_{it}^{10-1} = u_{i,10,t} - u_{i,1,t}$.

Datasource

- Data source: "Open Source Cross-Sectional Asset Pricing" database created by Chen and Zimmermann (2023).
- Monthly portfolio returns, sorted into 10 deciles based on firm level characteristics, from Jan. 1990 to Dec. 2020
- The number of characteristics is 133.
- The 3D tensor we consider is
  - size $N \times J \times T$, with $N = 133$, $J = 10$, $T = 360$,
  - the total number of observations $NJT = 478{,}800$.
- Compare the estimates of the 3D factor model from both TPCA and ALS.

Estimated Factors



▸ Appendix: First Factor is Market

Estimated Loadings Specific to Deciles

Decile	1	2	3	4	5	6	7	8	9	10	10 - 1
						Tensor PCA					
$\hat{\mu}_1$	0.3779	0.3423	0.3170	0.3021	0.2924	0.2846	0.2866	0.2961	0.3097	0.3406	-0.0373
$\hat{\mu}_2$	0.5259	0.3719	0.2289	0.1225	0.0216	-0.0689	-0.1660	-0.2769	-0.3823	-0.5119	-1.0378
						ALS					
$\hat{\mu}_1$	0.4417	0.3925	0.3481	0.3186	0.2927	0.2700	0.2567	0.2517	0.2541	0.2752	-0.1665
$\hat{\mu}_2$	-0.3585	-0.3308	-0.2346	-0.1433	-0.0369	0.0752	0.2003	0.3261	0.4483	0.5762	0.9347

- Exposures of deciles to the first factor - Market:
  - largest on the two extremes and smaller in the middle
  - TPCA is more symmetric than ALS

Estimated Loadings Specific to Deciles

Decile	1	2	3	4	5	6	7	8	9	10	10 - 1
						Tensor PCA					
$\hat{\mu}_1$	0.3779	0.3423	0.3170	0.3021	0.2924	0.2846	0.2866	0.2961	0.3097	0.3406	-0.0373
$\hat{\mu}_2$	0.5259	0.3719	0.2289	0.1225	0.0216	-0.0689	-0.1660	-0.2769	-0.3823	-0.5119	-1.0378
						ALS					
$\hat{\mu}_1$	0.4417	0.3925	0.3481	0.3186	0.2927	0.2700	0.2567	0.2517	0.2541	0.2752	-0.1665
$\hat{\mu}_2$	-0.3585	-0.3308	-0.2346	-0.1433	-0.0369	0.0752	0.2003	0.3261	0.4483	0.5762	0.9347

- Exposures of deciles to the second factor:
  - both TPCA and ALS are monotone
  - sign indeterminant

Message on High-Minus-Low Portfolios

- High-minus-low portfolios (risk premia):

$$y_{it}^{10-1} = \sigma_1 \lambda_{i,1}(\mu_{10,1} - \mu_{1,1}) f_{t,1} + \sigma_2 \lambda_{i,2}(\mu_{10,2} - \mu_{1,2}) f_{t,2} + u_{it}^{10-1}$$

- For TPCA, symmetry of μ_1 implies the first term on the right-hand side cancels out. Not the case with ALS.
- Constructing risk premia using high-minus-low portfolios:
  - eliminates the risks from the market;
  - delivers the highest risk premium associated with one characteristic, *i.e.,* 10-1 is better than 9-2 or 8-3.

Highest to Lowest Characteristic Loadings - TPCA

First factor - market

- Firm Age - Momentum
- Idiosyncratic risk
- Price

  ⋮

- Price delay R square
- Frazzini-Pedersen Beta
- Volume Variance

Second factor

- Bid-ask spread
- Idiosyncratic risk
- CAPM beta

  ⋮

- Real estate holdings
- Market leverage
- Earnings Surprise

Summary Statistics of Loadings Specific to Characteristics

	Max	Mean	Min	Std.	> 0
			TPCA		
$\hat{\lambda}_1$	0.1133	0.0863	0.0658	0.0087	100%
$\hat{\lambda}_2$	0.2715	0.0027	-0.2681	0.0870	58.65%
			ALS		
$\hat{\lambda}_1$	0.2740	0.0424	-0.2060	0.0759	81.20%
$\hat{\lambda}_2$	0.2447	0.0748	-0.0254	0.0441	96.24%

- Exposures of characteristics to the first factor - Market:
  - TPCA strictly positive, ALS about 20% negative
  - Std. is significantly larger for ALS than TPCA

Summary Statistics of Loadings Specific to Characteristics

	Max	Mean	Min	Std.	> 0
			TPCA		
$\hat{\lambda}_1$	0.1133	0.0863	0.0658	0.0087	100%
$\hat{\lambda}_2$	0.2715	0.0027	-0.2681	0.0870	58.65%
			ALS		
$\hat{\lambda}_1$	0.2740	0.0424	-0.2060	0.0759	81.20%
$\hat{\lambda}_2$	0.2447	0.0748	-0.0254	0.0441	96.24%

- Exposures of characteristics to the second factor:
  - TPCA about 50% positive
  - ALS more than 96% positive

Table of Contents

Remainder of talk based on:

Babii, A. and Ghysels, E. and Pan, Junsu (2024) *Missing Financial Data: Filling the Tensor Blanks*, Work in progress

Missing Financial Data: Filling the Tensor Blanks

Bryzgalova et al. (2022) models the panel of firm characteristics for each month t as follows:

$$\mathbf{C}_{i,\ell}^{t} = \mathbf{F}_i^t \mathbf{\Lambda}_\ell^{t\top} + \mathbf{U}_{i,\ell}^t \quad \text{with } i = 1, \ldots, N_t \text{ and } \ell = 1, \ldots, L.$$

Using the approach of Xiong and Pelger (2022), the loadings $\mathbf{\Lambda}^t$ are estimated as eigenvectors of the estimated characteristic covariance matrix

$$\hat{\mathbf{\Sigma}}_{\ell,p}^{\mathrm{XS},t} = \frac{1}{Q_{\ell,p}^t} \sum_{i \in Q_{\ell,p}^t} \mathbf{C}_{\ell,i}^t \mathbf{C}_{p,i}^t,$$

where $Q_{\ell,p}^t$ is the set of all stocks that are observed for the two characteristics ℓ and p at time t.

Missing Financial Data: Filling the Tensor Blanks

The cross-sectional factors are estimated by a regression on the estimated loadings $\hat{\mathbf{\Lambda}}^t$:

$$\hat{\mathbf{F}}_i^t = \left(\frac{1}{L} \sum_{\ell=1}^{L} W_{i,\ell}^t \hat{\mathbf{\Lambda}}_\ell^t (\hat{\mathbf{\Lambda}}_\ell^t)^\top \right)^{-1} \left(\frac{1}{L} \sum_{\ell=1}^{L} W_{i,\ell}^t \hat{\mathbf{\Lambda}}_\ell^t (\mathbf{C}_{\ell,i}^t)^\top \right),$$

where $W_{i,\ell}^t = 1$ if characteristic ℓ is observed for stock i at time t and $W_{i,\ell}^t = 0$ otherwise. Due to overfitting, the above equation is replaced by a regularized ridge regression:

$$\hat{\mathbf{F}}_i^{t,\gamma} = \left(\frac{1}{L} \sum_{\ell=1}^{L} W_{i,\ell}^t \hat{\mathbf{\Lambda}}_\ell^t (\hat{\mathbf{\Lambda}}_\ell^t)^\top + \gamma I_K \right)^{-1} \left(\frac{1}{L} \sum_{\ell=1}^{L} W_{i,\ell}^t \hat{\mathbf{\Lambda}}_\ell^t (\mathbf{C}_{\ell,i}^t)^\top \right).$$

Tensor Factor Model for Firm Characteristics

For firm characteristics $\mathbf{C} \in \mathbb{R}^{N \times T \times L}$ with possibly missing entries, consider the tensor factor model

$$\mathbf{C}_{i,t,\ell} = \sum_{r=1}^{R} \sigma_r \mu_{r,i} f_{r,t} \lambda_{r,\ell} + \mathbf{U}_{i,t,\ell},$$

where loadings μ and λ correspond to the firm and characteristic dimensions, respectively. The loadings λ and factors f are estimated as eigenvectors of

$$\hat{\mathbf{\Sigma}}^{(j)}_{\ell,p} = \frac{1}{Q_{\ell,p}} \sum_{i \in Q_{\ell,p}} \mathbf{C}^{(j)}_{\ell,i} \mathbf{C}^{(j)}_{p,i},$$

where $Q_{\ell,p}$ is the set of all columns that are observed for the ℓ^{th} and the p^{th} rows of $\mathbf{C}^{(j)}$, and $\mathbf{C}^{(j)}$ is matricized characteristic tensor along the j^{th} dimension, for $j = 2, 3$.

Tensor Factor Model for Firm Characteristics

The loadings μ are estimated by a cross-sectional regression of the characteristics on the Khatri-Rao product of $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{F}}$, denoted by $\hat{\mathbf{K}} := \hat{\mathbf{\Lambda}} \odot \hat{\mathbf{F}}$, as follows:

$$\hat{\mathbf{M}}_i = \left(\sum_{j=1}^{TL} W_{i,j} \hat{\mathbf{K}}_j (\hat{\mathbf{K}}_j)^\top \right)^{-1} \left(\sum_{j=1}^{TL} W_{i,j} \hat{\mathbf{K}}_j (\mathbf{C}_j^{(1)})^\top \right),$$

where $W_{i,j} = 1$ if the j^{th} column of $\mathbf{C}^{(1)} \in \mathbb{R}^{N \times TL}$ is observed. The missing characteristics are predicted by TPCA as

$$\hat{\mathbf{C}}_{i,t,\ell}^{\text{TPCA}} = \sum_{r=1}^{R} \hat{\sigma}_r \hat{\mu}_{r,i} \hat{f}_{r,t} \hat{\lambda}_{r,\ell}.$$

Backward Information

Similar to Bryzgalova et al. (2023), we also combine backward information by a cross-sectional regression:

$$\mathbf{C}_{i,t,\ell} = \left(\beta^{\ell,t,\text{B-TPCA}}\right)^{\top} \left(\sum_{r=1}^{R} \hat{\sigma}_r \hat{\mu}_{r,i} \hat{f}_{r,t} \hat{\lambda}_{r,\ell} \quad \mathbf{C}_{i,\ell}^{t-1} \quad \hat{\mathbf{e}}_{i,\ell}^{t-1} \right)$$

where the $\hat{\mathbf{e}}_{i,\ell}^{t} = \mathbf{C}_{i,t,\ell} - \hat{\mathbf{C}}_{i,t,\ell}^{\text{TPCA}}$, and estimate

$$\hat{\beta}^{\ell,t,\text{B-TPCA}} = \left(\sum_{i=1}^{N_t} \mathbf{W}_i^{t,\ell} \mathbf{X}_{i,t,\ell}^{\text{B-TPCA}} (\mathbf{X}_{i,t,\ell}^{\text{B-TPCA}})^{\top} \right) \left(\sum_{i=1}^{N_t} \mathbf{W}_i^{t,\ell} \mathbf{X}_{i,t,\ell}^{\text{B-TPCA}} \mathbf{C}_{i,t,\ell} \right)$$

where $\mathbf{W}_i^{t,\ell} = 1$ if $\mathbf{X}_{i,t,\ell}^{\text{B-TPCA}}$ and $C_{i,t,\ell}$ are both observed and 0 otherwise.

Backward-Forward Information

- The B-TPCA model can be extended to include forward information at time $t+1$, where we define the covariates

$$\mathbf{X}_{i,t,\ell}^{\text{BF-TPCA}} = \left(\sum_{r=1}^{R} \hat{\sigma}_r \hat{\mu}_{r,i} \hat{f}_{r,t} \hat{\lambda}_{r,\ell} \quad \mathbf{C}_{i,\ell}^{t-1} \quad \hat{\mathbf{e}}_{i,\ell}^{t-1} \quad \mathbf{C}_{i,\ell}^{t+1} \quad \hat{\mathbf{e}}_{i,\ell}^{t+1} \right),$$

and estimate the model

$$\mathbf{C}_{i,t,\ell} = \left(\beta^{\ell,t,\text{BF-TPCA}}\right)^{\top} \left(\sum_{r=1}^{R} \hat{\sigma}_r \hat{\mu}_{r,i} \hat{f}_{r,t} \hat{\lambda}_{r,\ell} \quad \mathbf{C}_{i,\ell}^{t-1} \quad \hat{\mathbf{e}}_{i,\ell}^{t-1} \quad \mathbf{C}_{i,\ell}^{t+1} \quad \hat{\mathbf{e}}_{i,\ell}^{t+1} \right).$$

Imputation Methods - New tensor-based

- Tensor PCA (TPCA)
  $\hat{\mathbf{C}}_{i,t,\ell}^{\mathrm{TPCA}} = \sum_{r=1}^{R} \hat{\sigma}_r \hat{\mu}_{r,i} \hat{f}_{r,t} \hat{\lambda}_{r,\ell}$
- Backward-TPCA (B-TPCA)
  $\hat{\mathbf{C}}_{i,t,\ell}^{\mathrm{B\text{-}TPCA}} = (\hat{\beta}^{\ell,t,\mathrm{B\text{-}TPCA}})^\top \begin{pmatrix} \hat{\mathbf{C}}_{i,t,\ell}^{\mathrm{TPCA}} & \mathbf{C}_{i,\ell}^{t-1} & \hat{\mathbf{e}}_{i,\ell}^{t-1} \end{pmatrix}$
- Backward-Forward-TPCA (BF-TPCA)
  $\hat{\mathbf{C}}_{i,t,\ell}^{\mathrm{BF\text{-}TPCA}} = (\hat{\beta}^{\ell,t,\mathrm{BF\text{-}TPCA}})^\top \begin{pmatrix} \hat{\mathbf{C}}_{i,t,\ell}^{\mathrm{TPCA}} & \mathbf{C}_{i,\ell}^{t-1} & \hat{\mathbf{e}}_{i,\ell}^{t-1} & \mathbf{C}_{i,\ell}^{t+1} & \hat{\mathbf{e}}_{i,\ell}^{t+1} \end{pmatrix}$

Imputation Methods - Bryzgalova et al. (2022)

- Backward-Forward-XS (BF-XS)
  $\hat{\mathbf{C}}^{\text{BF-XS}}_{i,t,\ell} = (\hat{\beta}^{\ell,t,\text{BF-XS}})^\top \begin{pmatrix} \hat{\mathbf{C}}^{\text{XS}}_{i,t,\ell} & \mathbf{C}^{t-1}_{i,\ell} & \hat{\mathbf{e}}^{t-1}_{i,\ell} & \mathbf{C}^{t+1}_{i,\ell} & \hat{\mathbf{e}}^{t+1}_{i,\ell} \end{pmatrix}$
- Backward-XS (B-XS)
  $\hat{\mathbf{C}}^{\text{B-XS}}_{i,t,\ell} = (\hat{\beta}^{\ell,t,\text{B-XS}})^\top \begin{pmatrix} \hat{\mathbf{C}}^{\text{XS}}_{i,t,\ell} & \mathbf{C}^{t-1}_{i,\ell} & \hat{\mathbf{e}}^{t-1}_{i,\ell} \end{pmatrix}$
- Cross-sectional (XS)
  $\hat{\mathbf{C}}^{\text{XS}}_{i,t,\ell} = \sum_{r=1}^{R} \hat{\sigma}_{r,t}\hat{\mu}_{r,i,t}\hat{\lambda}_{r,\ell,t}$
- Autoregression (AR)
  $\hat{\mathbf{C}}^{\text{AR}}_{i,t,\ell} = (\hat{\beta}^{\ell,t,\text{AR}})^\top \mathbf{C}^{t-1}_{i,\ell}$
- Previous Value (PV)
  $\hat{\mathbf{C}}^{\text{PV}}_{i,t,\ell} = \mathbf{C}^{t-1}_{i,\ell}$
- Cross-sectional median (median)
  $\hat{\mathbf{C}}^{\text{median}}_{i,t,\ell} = 0$

Evaluation Metrics

The accuracy of the imputations are measured by the root-mean-squared errors as

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{N_t}\sum_{i=1}^{N_t}(C_{i,t,\ell} - \hat{C}_{i,t,\ell})^2}.$$

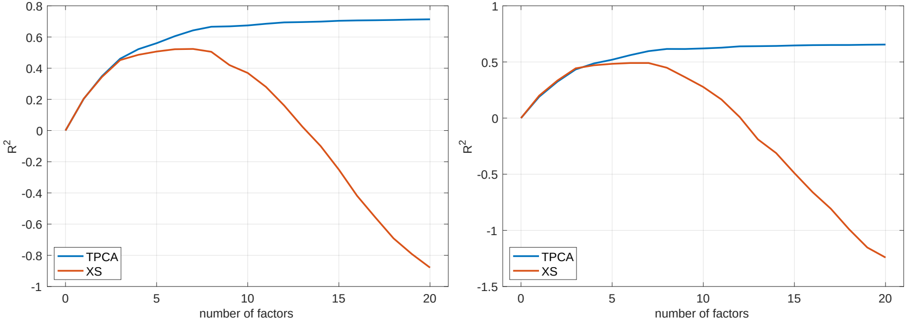
We measure the out-of-sample RMSE by randomly creating missingness following two schemes:

- Missing Completely at Random (MCAR). 10% of the characteristics are masked completely at random.
- Block Missing. We mask 10% of the characteristics in blocks of one year, and 40% of the blocks are at the beginning.

Data

- We use the dataset created by Freyberger, Neuhierl, and Weber (2020). The data cover 35 firm characteristics.
- The dataset is monthly and ranges from January 1966 to December 2020, and there are a total of 13588 firms in the entire sample.
- This number is smaller than 22630 in Bryzgalova et al. (2022), because we only have access to the pre-cleaned dataset that has no missingness in the cross-section of characteristics, i.e., all characteristics exist for all firms at any time period.
- The results calculated based on cross-sectional factor model are also very close to those in Bryzgalova et al. (2022).
- However, this difference does not affect the empirical results because the RMSE are calculated based on simulated (masked) missing characteristics both in this paper and in Bryzgalova et al. (2022).
- Raw characteristics are converted into ranks within range $[-0.5, 0.5]$.

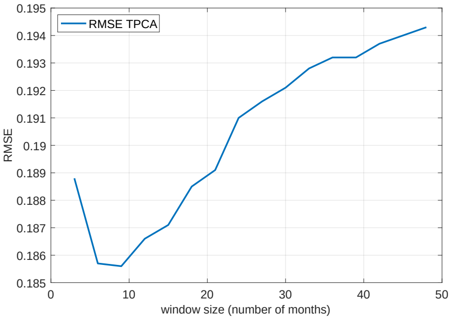
Number of Factors



(a) Missing Completely at Random (b) Block Missing

- Cross-sectional factor model overfits easily.

Window Size for Rolling-Window Estimation



(a) Missing Completely at Random

(b) Block Missing

- The best information comes from within one year.
- The quality of information beyond one year is similar.

Imputation Results

	In-Sample		OOS MCAR		OOS Block	
Method	RMSE	R^2	RMSE	R^2	RMSE	R^2
TPCA (5)	0.1844	0.5747	0.1866	0.5648	0.1984	0.5073
TPCA (10)	0.1591	0.6810	0.1600	0.6802	0.1792	0.5980
XS (5)	0.1784	0.6190	0.2069	0.4875	0.2093	0.4783
XS (10)	0.1244	0.8148	0.1923	0.5574	0.2029	0.5098
B-TPCA (5)	0.1026	0.8674	0.1168	0.8294	0.1934	0.5318
B-TPCA (10)	0.0987	0.8773	0.1098	0.8492	0.1752	0.6157
B-XS (5)	0.0904	0.9021	0.1189	0.8308	0.2038	0.5051
B-XS (10)	0.0670	0.9463	0.1253	0.8120	0.1988	0.5295
Prev. Value	0.1282	0.8031	0.1522	0.7229	0.2814	0.0570
Median	0.2890	0.0000	0.2890	0.0000	0.2898	0.0000
AR(1)	0.1153	0.8407	0.1426	0.7566	0.2810	0.0596

Table of Contents

Conditional Asset Pricing with Imputed Characteristics

- Characteristics data figure prominently in the estimation of many asset pricing models.

- We study the influence of missing characteristics and their imputation on the estimation of conditional asset pricing models. Specifically, we consider the instrumented principal component analysis, or IPCA, proposed by Kelly, Pruitt, and Su (2019). The general IPCA model specification for an excess return $r_{i,t+1}$ is:

$$\begin{aligned} r_{i,t+1} &= \alpha_{i,t} + \beta_{i,t}\mathcal{F}_{t+1} + \varepsilon_{i,t+1} \\ \alpha_{i,t} &= \mathcal{Z}'_{i,t}\Gamma_{\alpha} + \nu_{\alpha,i,t} \\ \beta_{i,t} &= \mathcal{Z}'_{i,t}\Gamma_{\beta} + \nu_{\beta,i,t} \end{aligned}$$

where $\mathcal{F}_t$ are K asset pricing factors.

IPCA estimator

- The model allows for dynamic factor loadings, $\beta_{i,t}$ on a K - vector of latent factors $\mathcal{F}_t$. Loadings depend on L characteristics labeled as instruments (not counting a constant).
- We can rewrite the model with $\alpha_{i,t} = 0 \ \forall \ i$ and t, as follows:

$$r_{t+1} = \mathcal{Z}_t \Gamma_\beta \mathcal{F}_{t+1} + \varepsilon^*_{t+1}$$

where r_{t+1} is an $N \times 1$ vector of individual firm excess returns, $\mathcal{Z}_t$ is a $N \times L + 1$ matrix with individual firm characteristics, and ε^*_{t+1} is the $N \times 1$ vector with entries $\varepsilon^*_{i,t+1} = \varepsilon_{t+1} + \nu_{\alpha,i,t} + \nu_{\beta,i,t}\mathcal{F}_{t+1}$.

IPCA estimator

- The least squares estimator for Γ_β and $\mathcal{F}$ is defined as:

$$\min_{\Gamma_\beta, \mathcal{F}} \sum_{t=1}^{T-1} \left(r_{t+1} - \mathcal{Z}_t \Gamma_\beta \mathcal{F}_{t+1}\right)' \left(r_{t+1} - \mathcal{Z}_t \Gamma_\beta \mathcal{F}_{t+1}\right)$$

with first order conditions:

$$\begin{aligned} \hat{\mathcal{F}}_{t+1} &= (\hat{\Gamma}'_\beta \mathcal{Z}'_t \mathcal{Z}_t \hat{\Gamma}_\beta)^{-1} \hat{\Gamma}'_\beta \mathcal{Z}'_t r_{t+1} \\ \text{vec}(\Gamma'_\beta) &= \left(\sum_{t=1}^{T-1} \mathcal{Z}'_t \mathcal{Z}_t \otimes \hat{\mathcal{F}}_{t+1} \hat{\mathcal{F}}'_{t+1} \right)^{-1} \left(\sum_{t=1}^{T-1} \left[\mathcal{Z}_t \otimes \hat{\mathcal{F}}'_{t+1} \right]' r_{t+1} \right) \end{aligned}$$

This system of first-order conditions has no closed-form solution and must be solved iteratively as an alternating least squares procedure.

IPCA estimator

- We study the situation where some of the characteristics are missing and imputed with either one of the methods discussed in the previous sections. To that end we define: $\hat{\mathcal{Z}}_t^{\text{BF-TPCA}}$, $\hat{\mathcal{Z}}_t^{\text{B-TPCA}}$, $\hat{\mathcal{Z}}_t^{\text{TPCA}}$, $\hat{\mathcal{Z}}_t^{\text{BF-XS}}$, $\hat{\mathcal{Z}}_t^{\text{B-XS}}$, $\hat{\mathcal{Z}}_t^{\text{XS}}$, $\hat{\mathcal{Z}}_t^{\text{AR}}$ and $\hat{\mathcal{Z}}_t^{\text{median}}$.
- In each case, apart from the constant we use the different imputation methods to complete the $N \times L + 1$ matrix $\mathcal{Z}_t$ for its missing values.

IPCA estimator

- Since imputation implies errors, we expect that the estimates of Γ_β and $\mathcal{F}$ will be biased and our purpose is to study the bias of the various methods. In addition, we also propose an IV estimator to correct the bias, namely we replace the IPCA estimator with:

$$
\begin{aligned}
\hat{\mathcal{B}}_{i,t} &= \left(\hat{\mathcal{F}}_{t-k,t}\hat{\mathcal{F}}'_{t-k,t}\right)^{-1}\hat{\mathcal{F}}'_{t-k,t}r_{i,t-k,t} \qquad i=1,\ldots,N \\
\hat{\mathcal{F}}_{t+1} &= (\hat{\mathcal{B}}'_t\mathcal{Z}_t\hat{\Gamma}_\beta)^{-1}\hat{\mathcal{B}}'_t r_{t+1} \\
\text{vec}(\Gamma'_\beta) &= \left(\sum_{\tau=k+1}^{T-1}(I_K\otimes\mathcal{Z}_\tau)'(I_K\otimes\mathcal{Z}_\tau)\right)^{-1}\left(\sum_{\tau=k+1}^{T-1}(I_K\otimes\mathcal{Z}_\tau)'\text{vec}(\mathcal{B}_\tau)\right)
\end{aligned}
$$

  where $\hat{\mathcal{B}}_{i,t}$ is a rolling sample beta estimator for firm i using returns from $t-k$ to t characterized by the vector $r_{i,t-k,t}$ which is used as instruments in the estimation of $\hat{\mathcal{F}}_{t+1}$ depicted in the second equation.

- The third equation represents the least squares estimator for $\text{vec}(\Gamma'_\beta)$ by regressing $(I_K\otimes\mathcal{Z}_\tau)$ onto $\text{vec}(\mathcal{B}_\tau)$.

Conditional Beta and Factor Model Estimators

Imputation Method	IPCA Estimators		IV Estimators	
Backward-Forward-TPCA	$\hat{\mathcal{F}}_t^{\text{BF-TPCA}}$	$\hat{\Gamma}_\beta^{\text{BF-TPCA}}$	IV-$\hat{\mathcal{F}}_t^{\text{BF-TPCA}}$	IV-$\hat{\Gamma}_\beta^{\text{BF-TPCA}}$
Backward-TPCA	$\hat{\mathcal{F}}_t^{\text{B-TPCA}}$	$\hat{\Gamma}_\beta^{\text{B-TPCA}}$	IV-$\hat{\mathcal{F}}_t^{\text{B-TPCA}}$	IV-$\hat{\Gamma}_\beta^{\text{B-TPCA}}$
Tensor PCA	$\hat{\mathcal{F}}_t^{\text{TPCA}}$	$\hat{\Gamma}_\beta^{\text{TPCA}}$	IV-$\hat{\mathcal{F}}_t^{\text{TPCA}}$	IV-$\hat{\Gamma}_\beta^{\text{TPCA}}$
Backward-Forward-XS	$\hat{\mathcal{F}}_t^{\text{TPCA}}$	$\hat{\Gamma}_\beta^{\text{TPCA}}$	IV-$\hat{\mathcal{F}}_t^{\text{TPCA}}$	IV-$\hat{\Gamma}_\beta^{\text{TPCA}}$
Backward-XS	$\hat{\mathcal{F}}_t^{\text{B-XS}}$	$\hat{\Gamma}_\beta^{\text{B-XS}}$	IV-$\hat{\mathcal{F}}_t^{\text{B-XS}}$	IV-$\hat{\Gamma}_\beta^{\text{B-XS}}$
Cross-sectional	$\hat{\mathcal{F}}_t^{\text{XS}}$	$\hat{\Gamma}_\beta^{\text{XS}}$	IV-$\hat{\mathcal{F}}_t^{\text{XS}}$	IV-$\hat{\Gamma}_\beta^{\text{XS}}$
Autoregression	$\hat{\mathcal{F}}_t^{\text{AR}}$	$\hat{\Gamma}_\beta^{\text{AR}}$	IV-$\hat{\mathcal{F}}_t^{\text{AR}}$	IV-$\hat{\Gamma}_\beta^{\text{AR}}$
Previous Value	$\hat{\mathcal{F}}_t^{\text{PV}}$	$\hat{\Gamma}_\beta^{\text{PV}}$	IV-$\hat{\mathcal{F}}_t^{\text{PV}}$	IV-$\hat{\Gamma}_\beta^{\text{PV}}$
Cross-sectional median	$\hat{\mathcal{F}}_t^{\text{median}}$	$\hat{\Gamma}_\beta^{\text{median}}$	IV-$\hat{\mathcal{F}}_t^{\text{median}}$	IV-$\hat{\Gamma}_\beta^{\text{median}}$

Table of Contents

Conclusions

- Novel approach for analyzing tensor datasets: tensor factor model
- Estimation via the PCA: closed-form expressions, avoids non-convex optimization.
- Asymptotic theory:
  - tensor dimensions can improve the estimation accuracy for factors/loadings.
  - eigenvalue ratio test for selecting the number of factors.
- Monte Carlo experiments:
  - TPCA is superior to ALS.
  - tensor factor model is more effective in reducing dimensions than traditional
- Sorted portfolios application: results with TPCA provide new insights for high-minus-low risk premia calculations.
- Missing characteristics: tensor-based approach provides better results

The End

References

References I

J. Bai. Panel data models with interactive fixed effects. *Econometrica*, 77 (4):1229–1279, 2009.

S. Bryzgalova, S. Lerner, M. Lettau, and M. Pelger. Missing financial data. Review of Financial Studies, (forthcoming), 2022.

S. Bryzgalova, M. Pelger, and J. Zhu. Forest through the trees: Building cross-sections of stock returns. *Journal of Finance* (forthcoming), 2024.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51: 1281–1304, 1983.

A. Y. Chen and T. Zimmermann. Open source cross-sectional asset pricing. *Critical Finance Review*, (forthcoming), 2023.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

References II

J. Freyberger, A. Neuhierl, and M. Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

B. T. Kelly, S. Pruitt, and Y. Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134: 501–524, 2019.

T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

M. Lettau. High dimensional factor models with an application to mutual fund characteristics. Working Paper 29833, National Bureau of Economic Research, 2022.

A. Lewbel. The rank of demand systems: theory and nonparametric estimation. *Econometrica: Journal of the Econometric Society*, pages 711–730, 1991.

H. R. Moon and M. Weidner. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4): 1543–1579, 2015.

References III

A. Onatski. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77:1447–1479, 2009.

S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–60, 1976.

T. J. Sargent and C. A. Sims. Business cycle modeling without pretending to have too much a priori economic theory. In C. A. Sims, editor, *New methods in business cycle research*, pages 145–168, 1977.

J. Stock and M. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179, 2002.

L. R. Tucker. An inter-battery method of factor analysis. *Psychometrika*, 23:111–136, 1958.

R. Xiong and M. Pelger. Large dimensional latent factor modeling with missing observations and applications to causal inference. *Journal of Econometrics* (forthcoming), 2022.

Appendices

Appendix: MC Experiment - Finite Sample Properties

- Recall the convergence rate

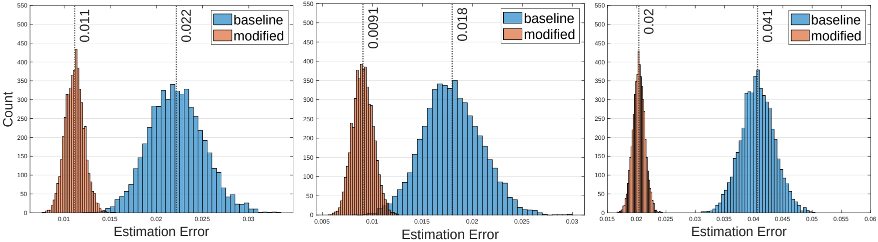
$$\|\hat{m}_{j,r} - m_{j,r}\| = O_P\left(\sqrt{\frac{1}{\prod_{k\neq j} N_k}}\right).$$

- Then the rate would be
  - $\lambda_r \to O_P(1/\sqrt{JT})$,
  - $\mu_r \to O_P(1/\sqrt{NT})$,
  - $f_r \to O_P(1/\sqrt{NJ})$.

Appendix: MC Experiment - Finite Sample Properties

- Generate a strong model with one factor with
  - signal strength $\sigma_1 = \sqrt{NJT}$.
- Baseline model $(T, N, J) = (100, 30, 20)$, and the modified model:
  - double all dimensions, $(T, N, J) =$ (200,60,40),
  - double two dimensions, $(T, N, J) =$ (100,60,40),
  - double only one dimension, $(T, N, J) =$ (100,60,20).
- Performance evaluation: the norm of estimation errors.

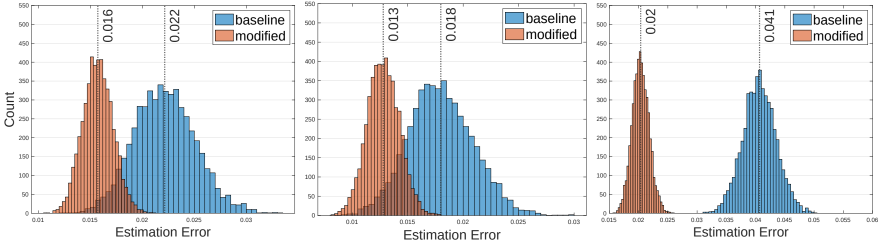
Appendix: Estimation Errors - Increasing Tensor Size



(a) $\hat{\lambda}$ (b) $\hat{\mu}$ (c) $\hat{f}$

- Double all dimensions, all estimation errors are reduced by $1/2$.
- Recall the convergence rate
  - $\lambda_r \to O_P(1/\sqrt{JT})$,
  - $\mu_r \to O_P(1/\sqrt{NT})$,
  - $f_r \to O_P(1/\sqrt{NJ})$.

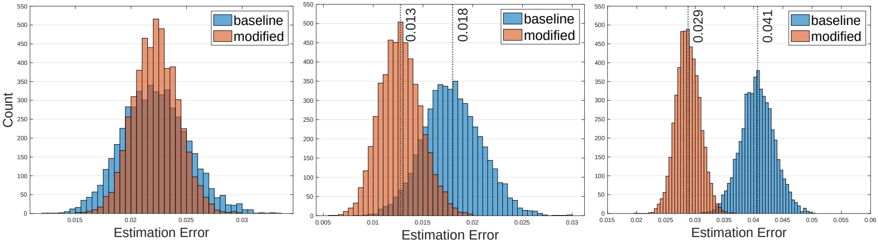
Appendix: Estimation Errors - Increasing Tensor Size



(a) $\hat{\lambda}$ (b) $\hat{\mu}$ (c) $\hat{f}$

- Only double N and J, the estimation errors
  - for both $\hat{\lambda}$ and $\hat{\mu}$ are reduced by $1/\sqrt{2}$,
  - for $\hat{f}$ are reduce by $1/2$.
- Recall the convergence rate
  - $\lambda_r \to O_P(1/\sqrt{JT})$,
  - $\mu_r \to O_P(1/\sqrt{NT})$,
  - $f_r \to O_P(1/\sqrt{NJ})$.

Appendix: Estimation Errors - Increasing Tensor Size



(a) $\hat{\lambda}$ (b) $\hat{\mu}$ (c) $\hat{f}$

- Only double N,
  - no improvement for $\hat{\lambda}$,
  - the estimation errors for $\hat{\mu}$ and $\hat{f}$ are reduced by $1/\sqrt{2}$.
- Recall the convergence rate
  - $\lambda_r \to O_P(1/\sqrt{JT})$,
  - $\mu_r \to O_P(1/\sqrt{NT})$,
  - $f_r \to O_P(1/\sqrt{NJ})$.

Appendix: Tensor Decomposition via PCA

- Recall a 3D tensor factor model

$$\underbrace{\mathbf{Y}}_{N\times J\times T} = \sum_{r=1}^{R} \sigma_r \lambda_r \otimes \mu_r \otimes f_r + \mathbf{U},$$

where $\lambda_r, \mu_r,$ and f_r can be collected in matrices Λ, M, F.

- Matricizing the tensor

$$\underbrace{\mathbf{Y}_{(1)}}_{N\times JT} = \Lambda D (F \odot M)^\top + \underbrace{\mathbf{U}_{(1)}}_{N\times JT},$$

where $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_R)$.

Appendix: Notations

Appendix: Tensor Decomposition via PCA

- Recall a 3D tensor factor model

$$\underbrace{\mathbf{Y}}_{N\times J\times T} = \sum_{r=1}^{R} \sigma_r \lambda_r \otimes \mu_r \otimes f_r + \mathbf{U},$$

where $\lambda_r, \mu_r,$ and f_r can be collected in matrices Λ, M, F.

- Likewise, we could reshape the 3D factor model as

$$\underbrace{\mathbf{Y}_{(2)}}_{J\times NT} = MD(F\odot\Lambda)^\top + \mathbf{U}_{(2)} \qquad \text{or} \qquad \underbrace{\mathbf{Y}_{(3)}}_{T\times JN} = FD(M\odot\Lambda)^\top + \mathbf{U}_{(3)}.$$

Appendix: Derivation

- Recall

$$\underbrace{\mathbf{Y}_{(1)}}_{N \times JT} = \Lambda D (F \odot M)^\top + \underbrace{\mathbf{U}_{(1)}}_{N \times JT},$$

  where $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_R)$.

- If $F^\top F = I_R$, $\Lambda^\top \Lambda = I_R$, and $M^\top M = I_R$, then

$$\begin{aligned} \mathbf{Y}_{(1)} \mathbf{Y}_{(1)}^\top &= \Lambda D (F \odot M)^\top (F \odot M) D \Lambda^\top + \text{noise} \\ &= \Lambda D (F^\top F) \circ (M^\top M) D \Lambda^\top + \text{noise} \\ &= \Lambda D^2 \Lambda^\top + \text{noise} \\ &= \hat{\Gamma} \hat{D}^2 \hat{\Gamma}^\top \end{aligned}$$

  where $\circ$ is element-wise matrix product (Hadamard product).

Appendix: Tensor Decomposition via PCA

- Under orthogonality

$$\mathbf{Y}_{(1)}\mathbf{Y}_{(1)}^{\top} = \Lambda D^2 \Lambda^{\top} + \text{noise}$$
$$= \hat{\Gamma}\hat{D}^2\hat{\Gamma}^{\top}$$

  where
  - $\hat{\Gamma} \rightarrow$ eigenvectors of $\mathbf{Y}_{(1)}\mathbf{Y}_{(1)}^{\top}$,
  - $\hat{D}^2 \rightarrow$ eigenvalues of $\mathbf{Y}_{(1)}\mathbf{Y}_{(1)}^{\top}$.
- PCA estimators:
  - $\hat{\Lambda} \rightarrow$ first R eigenvectors of $\mathbf{Y}_{(1)}\mathbf{Y}_{(1)}^{\top}$;
  - $\hat{M} \rightarrow$ first R eigenvectors of $\mathbf{Y}_{(2)}\mathbf{Y}_{(2)}^{\top}$;
  - $\hat{F} \rightarrow$ first R eigenvectors of $\mathbf{Y}_{(3)}\mathbf{Y}_{(3)}^{\top}$.

Back

Appendix: Tensor Decomposition via PCA

- More generally, the d-dimensional tensor factor model

$$\underbrace{\mathbf{Y}}_{N_1\times\cdots\times N_d} = \sum_{r=1}^{R}\sigma_r \bigotimes_{j=1}^{d} m_{j,r} + \mathbf{U}$$

can be matriced as

$$\underbrace{\mathbf{Y}_{(j)}}_{N_j\times\prod_{k\neq j} N_k} = M_j D\left(\bigodot_{k\neq j} M_k\right)^{\top} + \mathbf{U}_{(j)}, \qquad 1\leq j\leq d,$$

where $\bigodot_{k\neq j} M_k = M_d\odot\cdots\odot M_{j+1}\odot M_{j-1}\odot\cdots\odot M_1$.

- PCA estimators:
  - $\hat{M}_j \rightarrow$ first R eigenvectors of $\mathbf{Y}_{(j)}\mathbf{Y}_{(j)}^{\top}$ for $1,\ldots,J$.

Appendix: Asymptotics of TPCA

Theorem

Suppose that $M_j^\top M_j = I_R$, *the idiosyncratic errors* $\mathbf{U}$ *are i.i.d. with* $\mathbb{E}\mathbf{U} = 0$ *and finite* 4^{th} *moment. Then for all*

$$\|\hat{m}_{j,r} - m_{j,r}\| = O_P\left(\frac{\sqrt{N_j}\operatorname{tr}(D) + \left(N_j \vee \prod_{k\neq j} N_k\right)}{\delta_r}\right), \quad \forall r \leq R, j \leq d.$$

where δ_r *is the eigengap, namely:*

$$\delta_r = \min_{k\neq r} |\sigma_k^2 - \sigma_r^2|$$

measuring the strength of r^{th} *factor.*

Appendix: Asymptotic Distribution: Factors/Loadings

- Let $\nu \in \mathbb{R}^{N_j}$ be such that $\omega_{j,k}(\nu) = \lim_{N_j \to \infty} \sqrt{N_j} \langle m_{j,k}, \nu \rangle > 0$.
- Examples:
  - i^{th} element of factors/loadings vector $m_{j,r} \in \mathbb{R}^{N_j}$: $\nu = (0, 0, \dots, 1, \dots, 0)$;
  - average factors/loadings: $\nu = N_j^{-1}(1, 1, \dots, 1)$.
- If $N_1 \sim N_2 \sim N_3$ in the 3D case

$$\prod_{k \neq j} \sqrt{N_k} \langle \hat{m}_{j,r} - m_{j,r}, \nu \rangle \xrightarrow{d} N \left(0, \sigma^2 \sum_{k \neq r} \omega_{j,k}^2(\nu) \frac{d_r + d_k}{(d_r - d_k)^2} \right).$$

Back

Appendix: MC Experiment DGP

$$y_{i,j,t} = \sum_{r=1}^{R} \sigma_r \lambda_{i,r} \mu_{j,r} f_{t,r} + u_{i,j,t}, \qquad u_{i,j,t} \sim_{\text{i.i.d.}} N(0, s_u^2),$$

$$f_{t,r} = \rho f_{t-1,r} + \varepsilon_{t,r}, \qquad \varepsilon_{t,r} \sim_{\text{i.i.d.}} N(0, s_\varepsilon^2).$$

- Parameters:
  - $\rho = 0.5, s_\varepsilon = 0.1, s_u = 1,$
  - signal strength $\sigma_r = d_r \times \sqrt{NJT}$ with $d_r = R - r + 1,$
- Consider the cases $R = 5$, and $(T, N, J) = (100, 30, 20)$.

Back

Appendix: Notations

- $\odot$ is the Khatri-Rao Product:

$$\underbrace{A}_{M\times K} \odot \underbrace{B}_{N\times K} := \underbrace{\begin{bmatrix} a_1 \otimes_K b_1 & a_2 \otimes_K b_2 & \cdots & a_K \otimes_K b_K \end{bmatrix}}_{(MN)\times K}$$

- $\otimes_K$ is the Kronecker Product:

$$\underbrace{A}_{M\times N} \otimes_K \underbrace{B}_{J\times K} = \underbrace{\begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}}_{MJ\times NK}$$

Back

Appendix: First Factor is Market

- The first factor is market is commonly acknowledged in finance.
- A regression

$$latent\ factor \sim intercept + \beta\ observable\ factor,$$

shows

	Estimate	SE	tStat	pValue
$(intercept)$	0.000	0.000	-1.209	0.228
$observable\ factor$	0.012	0.000	140.39	0.000

R-squared: 0.982, Adjusted R-Squared: 0.982

- $observable\ factor$ is value-weight (excess) returns of all CRSP firms incorporated in the US.

Back