Identification through Sparsity in Factor Models: The ℓ_1-rotation criterion

Simon Freyaldenhoven
Federal Reserve Bank of Philadelphia

Factor Models

$$\underset{(T\times n)}{X} = \underset{(T\times r)}{F}\underset{(r\times n)}{\Lambda^{*\prime}} + \underset{(T\times n)}{e}$$

Learn this structure $\Leftrightarrow$ Estimate Λ^* and F

Rotational Indeterminacy

$$\underset{(T\times n)}{X} = \underset{(T\times r)}{F}\underset{(r\times n)}{\Lambda^{*\prime}} + \underset{(T\times n)}{e}$$

Fix rotation of estimates Λ^0, F^0, such that:

1. $\frac{\Lambda^{0\prime}\Lambda^0}{n} = \mathrm{I}$
2. $\frac{F^{0\prime}F^0}{T} = D$, where D denotes a diagonal matrix

$\Rightarrow$ Estimates will be rotations of true loadings and factors.

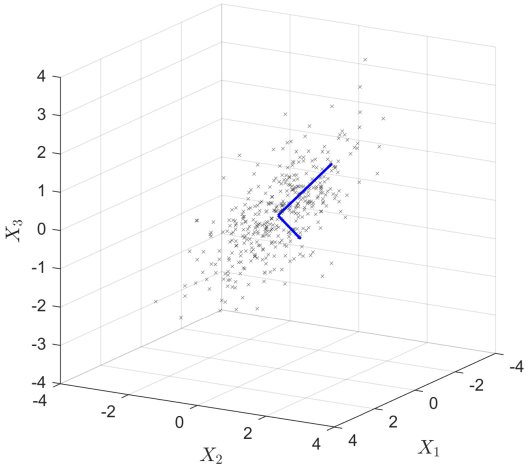
Rotational Indeterminacy - A simple example

For a given t,

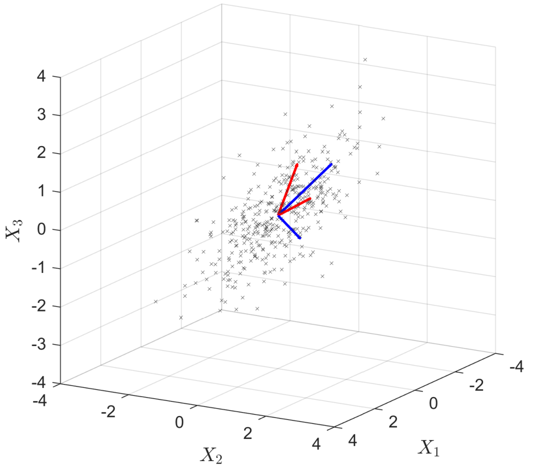
- 3 observed outcomes: X_1, X_2, X_3.
- 2 Factors F_1, F_2, with $F_k \sim N(0, I_2)$
- X follows simple factor structure with $i.i.d.$ $e_i \sim N(0, 1)$, and

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}.$$

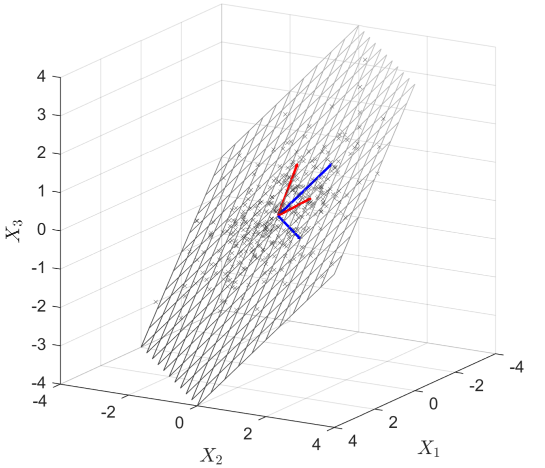
Rotational Indeterminacy - A simple example



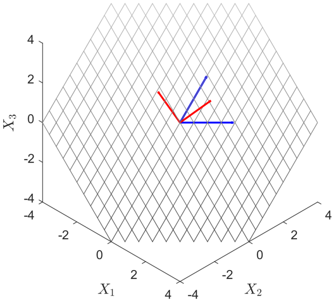
Rotational Indeterminacy - A simple example



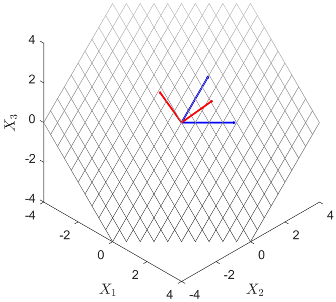
Rotational Indeterminacy - A simple example



Rotational Indeterminacy - A simple example



Rotational Indeterminacy - A simple example



$$\Lambda^* = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\Lambda^0 = \begin{bmatrix} 0.77 & -0.61 \\ 1.60 & -0.03 \\ 0.86 & 0.59 \end{bmatrix}$$

Rotational Indeterminacy - in practice

Stock and Watson [2002]:

> *"Because the factors are identified only up to a $k \times k$ matrix, detailed discussion of the individual factors is unwarranted."*

This Paper

Main insight

Suppose loadings are "sparse" (there are local factors).

Then, individual loading vectors are identified.

Local Factors

Natural concept in many economic settings:

- Industry specific factors
- Country specific factors
- Character traits manifest in some but not all observational outcomes

These will be identified

The Main Idea

The Idea

1. Estimate the space spanned by the loading vectors
2. Find rotation that minimizes l_0-norm of loadings

$\Rightarrow$ If true factor loadings are sparse, this will be the argmin.

The Idea

1. Estimate the space spanned by the loading vectors
2. Find rotation that minimizes l_0-norm of loadings

In general, infeasible in practise

The Idea - feasible

1. Estimate the space spanned by the loading vectors
2. Find rotation that minimizes l_1-norm of loadings

$\Rightarrow$ If true factor loadings are sparse, this will be the argmin.

An example with two factors

Exemplary DGP

$$\underset{(224\times 207)}{X} = \underset{(224\times 2)}{F}\underset{(2\times 207)}{\Lambda^{*T}} + \underset{(224\times 207)}{e} \tag{1}$$

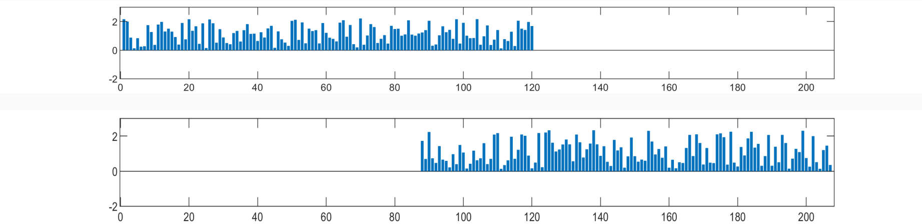
- $F_t \overset{i.i.d.}{\sim} N(0, \Sigma_F)$, with

$$\Sigma_F = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}.$$

- Either $\lambda^*_{ik} \overset{i.i.d.}{\sim} U(0.1, 2.9)$, or $\lambda^*_{ik} = 0$, such that

$$\Lambda^* = \begin{bmatrix} \lambda^*_{1:120,1} & 0 \\ 0 & \lambda^*_{(n+1)-120:n,2} \end{bmatrix}.$$

True loading matrix Λ^*



Estimation of Λ^*

Under standard regularity conditions, obtain estimates $\lambda^0_{\bullet 1}, \lambda^0_{\bullet 2}$, such that

$$\begin{aligned} \lambda^0_{\bullet 1} &= H_{11}\lambda^*_{\bullet 1} + H_{12}\lambda^*_{\bullet 2} + o_p(1) \\ \lambda^0_{\bullet 2} &= H_{21}\lambda^*_{\bullet 1} + H_{22}\lambda^*_{\bullet 2} + o_p(1), \end{aligned} \tag{2}$$

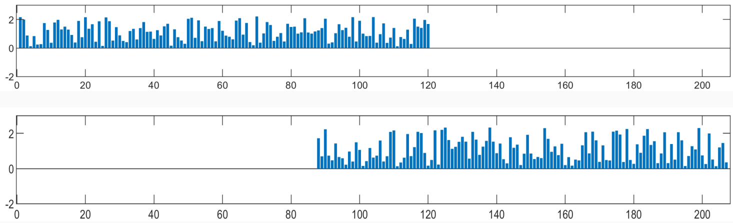
where H is an unknown non-singular rotation matrix (e.g. Bai 2003).

In population, $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$ are linear combinations of the true loading vectors $\lambda^*_{\bullet 1}$ and $\lambda^*_{\bullet 2}$.

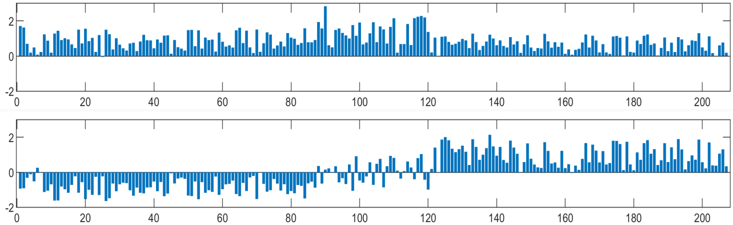
Observation 1

Linear combinations of sparse loading vectors are generally dense

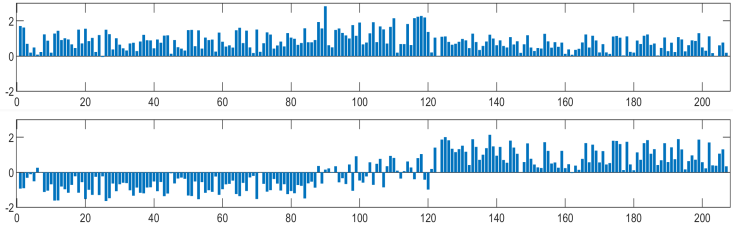
Let $\lambda^0_{\bullet 1} = H_{11}\lambda^*_{\bullet 1} + H_{12}\lambda^*_{\bullet 2}$ with $H_{11}, H_{12} \neq 0$. Then, generally $\lambda^0_{i1} \neq 0$ for $i = 1, \ldots, n$.



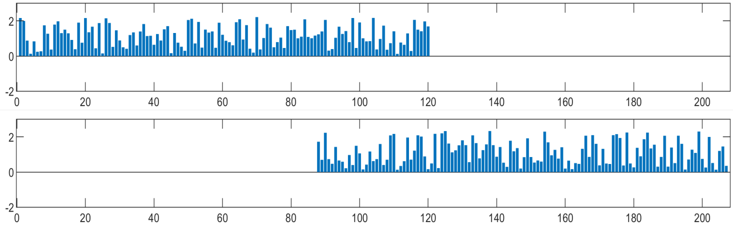
PCA estimate Λ^0



PCA estimate Λ^0



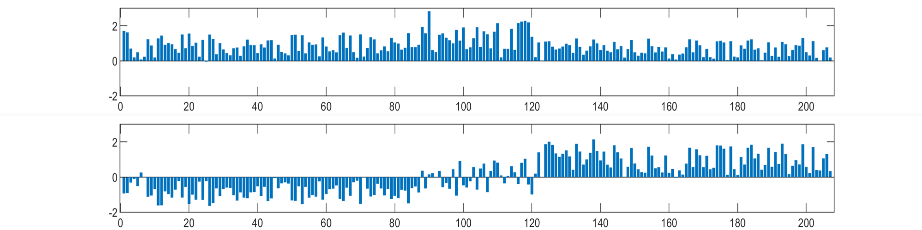
Compare to Λ^*:



Observation 2

There exists a linear combination of the estimated loading vectors that is sparse

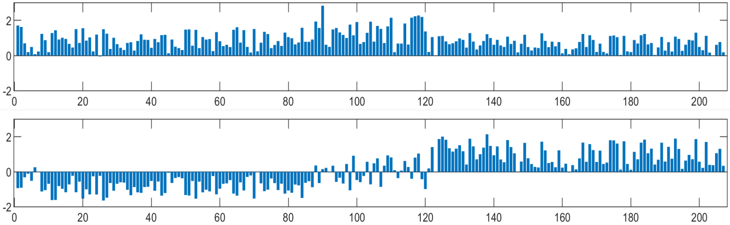
There must exist weights w_1 and w_2, such that $\lambda^*_{\bullet 1} = w_1 \lambda^0_{\bullet 1} + w_2 \lambda^0_{\bullet 2}$. But then, if $\lambda^*_{\bullet 1}$ is sparse, there must exist a linear combination of $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$ that is sparse.



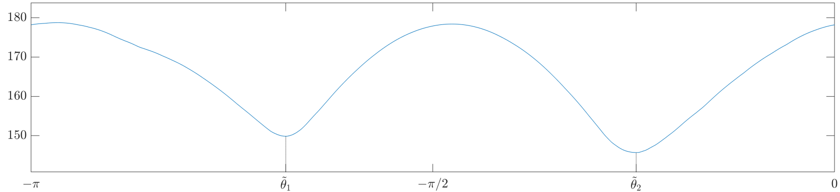
Finding the sparse rotation

Our proposal

Find rotation that minimizes ℓ_1-norm across rotations of Λ^0.

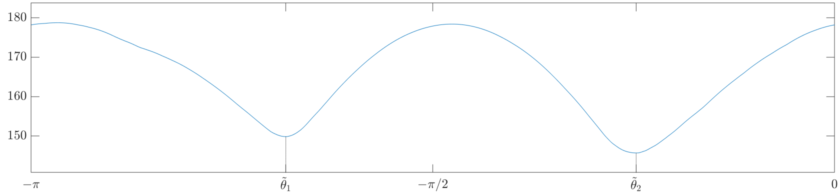


ℓ_1-norm of loadings across all rotations



$\|\lambda_{\bullet k}\|_1 = \|sin(\theta)\lambda^0_{\bullet 1} + cos(\theta)\lambda^0_{\bullet 2}\|_1$ as a function of θ.

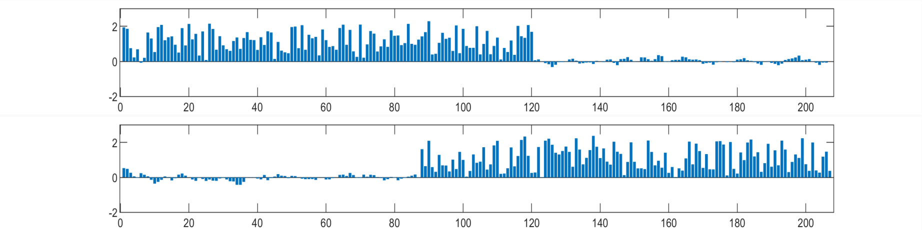
ℓ_1-norm of loadings across all rotations



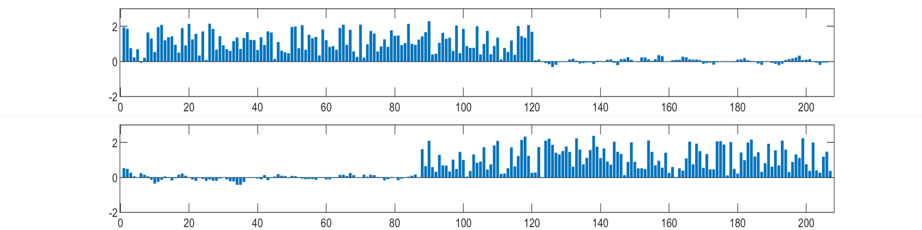
$$\|\lambda_{\bullet k}\|_1 = \|sin(\theta)\lambda^0_{\bullet 1} + cos(\theta)\lambda^0_{\bullet 2}\|_1$$ as a function of θ.

Proposed estimate $\Rightarrow$
$$\tilde{\lambda}_{\bullet 1} = sin(\tilde{\theta}_1)\lambda^0_{\bullet 1} + cos(\tilde{\theta}_1)\lambda^0_{\bullet 2}$$
$$\tilde{\lambda}_{\bullet 2} = sin(\tilde{\theta}_2)\lambda^0_{\bullet 1} + cos(\tilde{\theta}_2)\lambda^0_{\bullet 2}$$

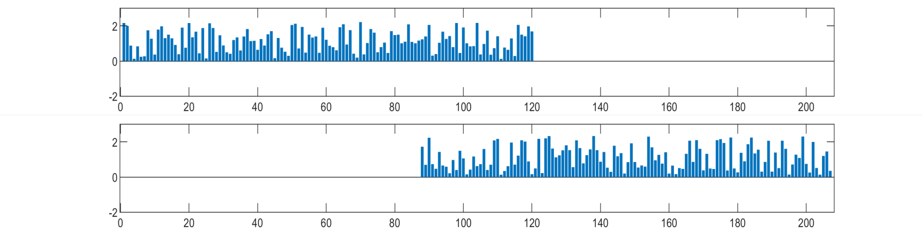
Rotated estimate $\tilde{\Lambda}$



Rotated estimate $\tilde{\Lambda}$



Compare to Λ^*:



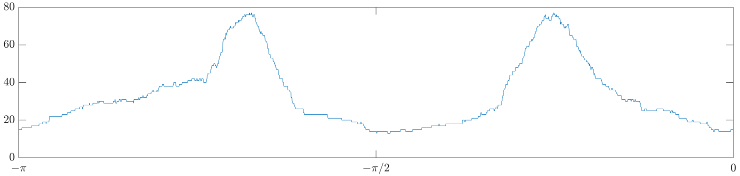
Second contribution: Testing for the presence of local factors

Intuition:

1. If no local factors present: No sparse rotation exists
2. If local factors present: Sparse rotation exists

Number of small loadings in $\lambda_{\bullet k} = sin(\theta)\lambda^0_{\bullet 1} + cos(\theta)\lambda^0_{\bullet 2}$

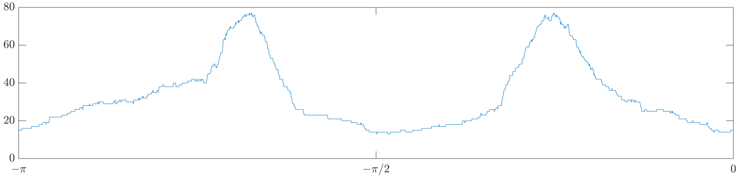
Example DGP:



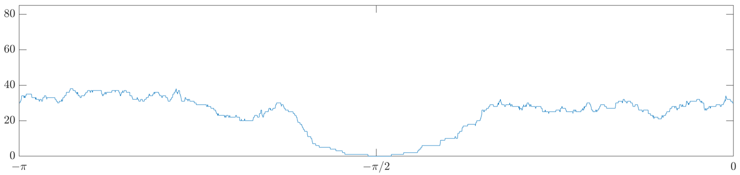
- "small" : $|\lambda_{ik}| < 1/log(n)$

Number of small loadings in $\lambda_{\bullet k} = sin(\theta)\lambda^0_{\bullet 1} + cos(\theta)\lambda^0_{\bullet 2}$

Example DGP:



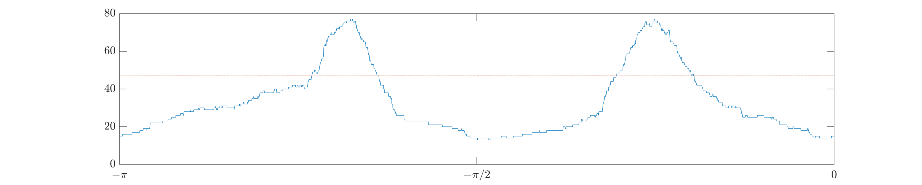
“Dense" DGP:



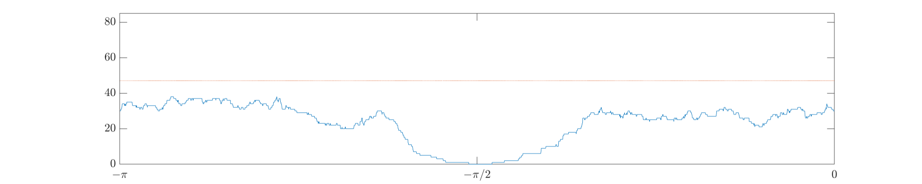
- “small” : $|\lambda_{ik}| < 1/log(n)$

Testing for the presence of local factors

Example DGP:



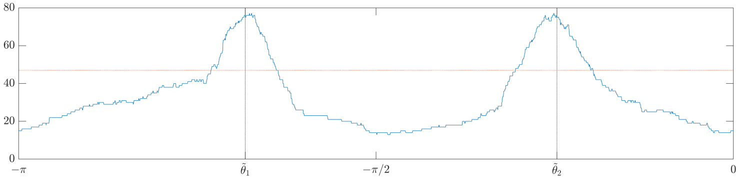
“Dense" DGP:



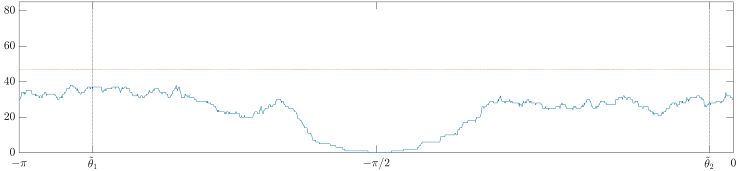
- Horizontal dashed red line represents critical value.

Testing for the presence of local factors

Example DGP:



"Dense" DGP:



- $\tilde{\theta}_1$ and $\tilde{\theta}_2$ correspond to minima of the ℓ_1-norm.

Second contribution: Testing for the presence of local factors

1. Find the most sparse rotation in the loading space, $\tilde{\lambda}_{\bullet 1}$
   - Feasible using the ℓ_1-rotation criterion from earlier
2. Count the number of small loadings in $\tilde{\lambda}_{\bullet 1}$
3. Compare it to the number of small loadings that could reasonably be expected under a “dense” loading matrix

Theory

The Main Result

1. Start with orthonormal basis of factor space.
   - Can take any $\sqrt{n}$ consistent estimate.
2. Find rotation that minimizes l_1-norm of loadings
   - Holding l_2-norm constant

$\Rightarrow$ If there are (approximately) local factors, their loading vectors will be an argmin.

The Main Result

Formal defition of a “local factor” in paper is slightly stronger than having a sparse loading vector

Further assume

1. loading vectors are not too close to collinear
2. we have access to a $\sqrt{n}$-consistent initial estimate Λ^0.

The Main Result

$$\min_{R_{\bullet k}} \| \sum_{l=1}^{r} \lambda^0_{\bullet l} R_{lk} \|_1 \qquad \text{such that } R'_{\bullet k} R_{\bullet k} = 1. \tag{3}$$

Theorem 1

Suppose F_k is a local factor and the conditions stated in the paper hold.Then, there exists a local minimum of (3) at $\bar{R}_{\bullet k}$, with $\bar{\lambda}_{\bullet k} = \Lambda^0 \bar{R}_{\bullet k}$, such that

$$\bar{\lambda}_{ik} = \lambda^*_{ik} + O_p(n^{-1/4}) \tag{4}$$

Applications

- **International stock returns**
- Panel of US macroeconomic indicators

The Data

- Daily stock returns across 6 regions
- 687 observations of 272 stocks

Stock index	Number of stocks
Frankfurt	30
London	75
New York	97
Paris	38
Tel Aviv	22

- 8 Factors (Bai and Ng 2002)
- Test suggests local factors are present

Rotated Loading matrix

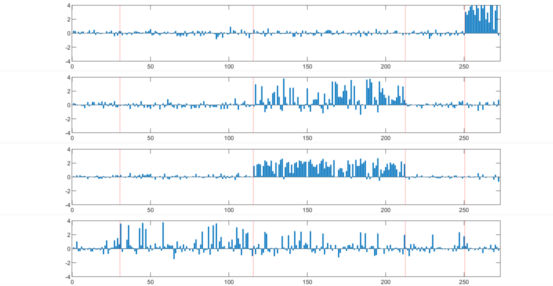


Figure 1: Columns 1-4 of $\tilde{\Lambda}$

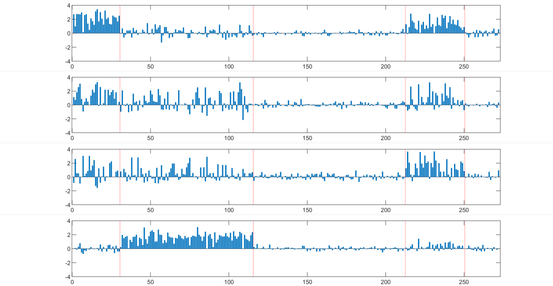


Figure 2: Columns 5-8 of $\tilde{\Lambda}$

Order of geographical regions: Frankfurt, London, New York, Paris, Tel Aviv

Interpretation of individual factors

Factor	Region	Sector
1	Middle East	
2	US	
3	US	
4	Global	Natural Resources (Oil and Mining)
5	Germany, France	
6	Germany, France, UK	
7	Germany, France, UK	
8	UK	

Conclusion

- New method to simplify loading matrix that produces more interpretable estimates (easy to implement!)
- Existing heuristics (e.g. VARIMAX) lack theoretical justification (and perform worse in our simulations)
- We prove that ℓ_1-criterion can identify individual loading vectors of local factors
- Sparsity assumption is testable
- Develop criterion to test for presence of local factors

Thank you!

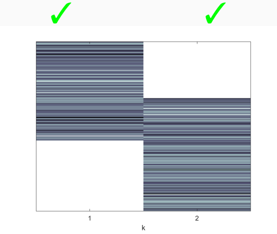
The key assumptions ($r = 2$)

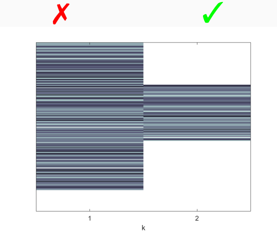
Let $\mathcal{A}_k$ denote the support of $\lambda^*_{\bullet k}$.

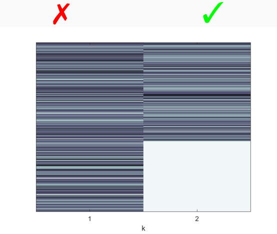
1. Some factors are local, where a factor F_k is local if:
   a) A significant number of entries in $\lambda^*_{\bullet k}$ are equal to zero (e.g. $|\mathcal{A}_k| \leq \alpha n$ for some $\alpha \in [0, 1)$).
   b) No other factor affects only a subset of $\mathcal{A}_k$: $\mathcal{A}_l \not\subset \mathcal{A}_k \forall l \neq k$

2. The loading vectors are not too close to collinearity on their joint support.

When is factor F_k local?







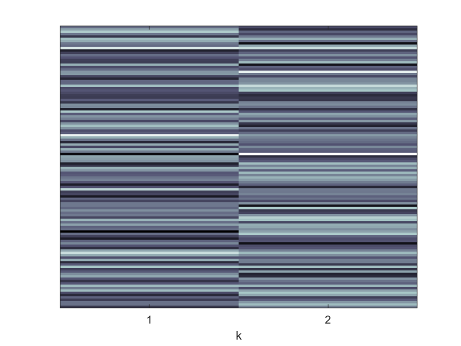
The key assumptions ($r = 2$)

Let $\mathcal{A}_k$ denote the support of $\lambda^*_{\bullet k}$.

1. Some factors are local, where a factor F_k is local if:
   a) A significant number of entries in $\lambda^*_{\bullet k}$ are equal to zero (e.g. $|\mathcal{A}_k| \leq \alpha n$ for some $\alpha \in [0, 1)$).
   b) No other factor affects only a subset of $\mathcal{A}_k$: $\mathcal{A}_l \not\subset \mathcal{A}_k \forall l \neq k$

2. The loading vectors are not too close to collinearity on their joint support.

Loading vectors far from collinear



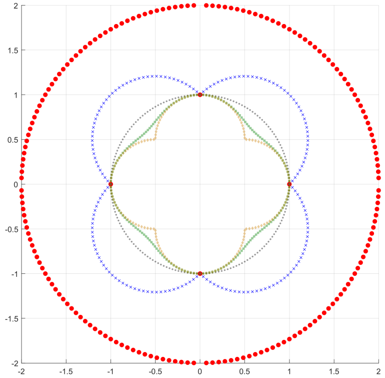
- Ensures no sparse linear combination exists of two dense vectors

Existing Criteria

$$\max_{R:R'R=I} Q(\Lambda^0 R) = Q(\Lambda) = \sum_{k=1}^{r} \left[\sum_{i=1}^{n} \lambda_{ik}^4 - \frac{c}{n} \left(\sum_{i=1}^{n} \lambda_{ik}^2 \right)^2 \right]. \tag{5}$$

Value of c	Criterion
0	Quartimax (Carroll 1953)
1	Varimax (Kaiser 1958)
$r/2$	Equamax (Saunders 1962)

Comparison of criteria



(Pseudo-)Norms across rotations:

- ℓ_0-norm (large, red circles)
- ℓ_1-norm (blue crosses)
- ℓ_2-norm (small, grey circles)
- ℓ_4-norm (green squares)
- ℓ_∞-norm (yellow diamonds).

References

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

John B Carroll. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18:23–38, 1953.

Henry F Kaiser. The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200, 1958.

David R Saunders. Trans-Varimax-some properties of the ratiomax and Equamax criteria for blind orthogonal rotation. *American Psychologist*, 17(6):395–396, 1962.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002.