Bayesian Dynamic Factor Models for High-Dimensional Matrix-Valued Time Series

Wei Zhang

Department of Economics, Purdue University

The 94th SEA Conference, November 23, 2024

Matrix-valued time series are of great interests in economics

Figure 1: Growth Rates of Seven Macroeconomic Series for G7 Countries

Matrix-valued time series are of great interests in economics

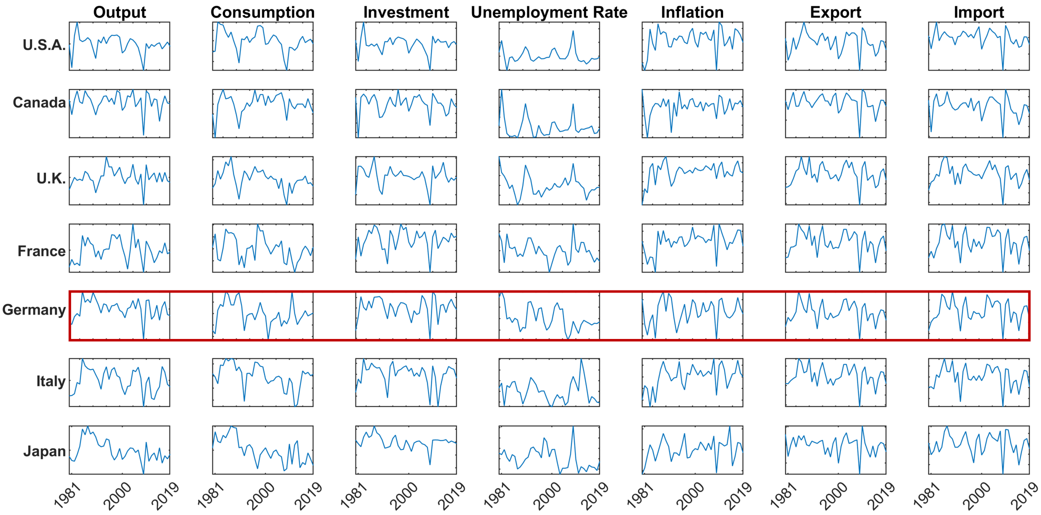

Figure 2: Growth Rates of Seven Macroeconomic Series for G7 Countries

A standard dynamic factor model

Assume we observe k indicators for Germany, denoted as $\mathbf{y}_t$. Consider the following dynamic factor model:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{M}\mathbf{f}_t + \varepsilon_t, \\ \mathbf{f}_t &= \mathbf{H}_\rho \mathbf{f}_{t-1} + \nu_t. \end{aligned} \quad (1)$$

- $\mathbf{f}_t$: a $p \times 1$ vector of factors.
- $\mathbf{M}$: a $k \times p$ loading matrix. p is the number of factors.
- $\mathbf{H}_\rho$: a k-dimensional diagonal matrix consisting of autoregressive coefficients.

Matrix-valued time series are of great interests in economics

Figure 3: Growth Rates of Seven Macroeconomic Series for G7 Countries

Standard multivariate time-series models

For an macroeconomic panel, in traditional multivariate time-series analysis, we usually stack such matrices into vectors:

$$\mathbf{y}_t = \Big[\underbrace{y^{US}_{output,t}, \dots, y^{US}_{import,t}}_{US}, \underbrace{y^{Canada}_{output,t}, \dots, y^{Canada}_{import,t}}_{Canada}, \dots, \underbrace{y^{Japan}_{output,t}, \dots, y^{Japan}_{import,t}}_{Japan}\Big]'. \quad (2)$$

Standard multivariate time-series models

For an macroeconomic panel, in traditional multivariate time-series analysis, we usually stack such matrices into vectors:

$$\mathbf{y}_t = \left[\underbrace{y^{US}_{output,t}, \ldots, y^{US}_{import,t}}_{US}, \underbrace{y^{Canada}_{output,t}, \ldots, y^{Canada}_{import,t}}_{Canada}, \ldots, \underbrace{y^{Japan}_{output,t}, \ldots, y^{Japan}_{import,t}}_{Japan} \right]' . \quad (2)$$

A popular dynamic factor model for international macroeconomic panels:

$$y_{i,t} = b^{global}_i f^{global}_t + b^{region}_i f^{region}_{r,t} + b^{country}_i f^{country}_{c,t} + \varepsilon_{i,t}, \quad i = 1, \ldots, n \times k. \quad (3)$$

- Used in Kose et al. (2003, 2008), Crucini et al. (2011), Miranda-Agrippino et al. (2015), Ha et al. (2023)

Limitations of standard dynamic factor models

- *Cross-sectional dependencies*: Relationships between variables across entities (e.g., countries, indicators) are important but difficult to capture.
  - Can geography tell the whole story?
- *Dimensionality*: As the number of variables increases, standard models face estimation challenges.
  - If we have 15 countries (5 regions) and 20 indicators, the dimension of the loading space in (3) would be: $15 \times 20 + 15 + 15 = 330$

Development of factor models for matrix-valued series is still in its initial stage

Idiosyncratic Components	Common Factors	
	Static	Dynamic
White noises	Wang et al. (2019) Matrix factor model	Yu et al. (2024) Matrix autoregressions for factors
	Liu and Chen (2019) A threshold variant	
	Chen et al. (2020) A constrained version	
	Chen et al. (2022) Tensor factor model	
Cross-sectional correlations	Chen et al. (2024) Time-varying loadings	
Correlations across time		
Time-varying volatility Outliers		

Contributions: Tailoring the model for macroeconomics studies

By tailoring the model for macroeconomic studies, we make the following two contributions:

1. Incorporation of dynamic factors
   - *Persistency* in macroeconomic data
   - Forecasting
2. Accommodating *time-varying volatility*, *cross-sectional correlation* and *outlier adjustments* in idiosyncratic components
   - Time-varying volatilities in macroeconomic data
   - Flexible for correlation in individual risks
   - Adjusting for outliers instead of removing them

Contributions: Tailoring the model for macroeconomics studies

<table>
<tr><th rowspan="2">Idiosyncratic Components</th><th colspan="2">Common Factors</th></tr>
<tr><th>Static</th><th>Dynamic</th></tr>
<tr><td rowspan="4">White noises</td><td>Wang et al. (2019)
Matrix factor model</td><td rowspan="4">Yu et al. (2024)
Matrix autoregressions for factors</td></tr>
<tr><td>Liu and Chen (2019)
A threshold variant</td></tr>
<tr><td>Chen et al. (2020)
A constrained version</td></tr>
<tr><td>Chen et al. (2022)
Tensor factor model</td></tr>
<tr><td>Cross-sectional correlations</td><td rowspan="2">Chen et al. (2024)
Time-varying loadings</td><td rowspan="4">Our model</td></tr>
<tr><td>Correlations across time</td></tr>
<tr><td>Time-varying volatility</td><td rowspan="2"></td></tr>
<tr><td>Outliers</td></tr>
</table>

The Model

Dynamic factor Models for matrix-valued time series (MDFM)

Consider the following dynamic factor model

$$\begin{aligned}
\mathbf{Y}_t &= \mathbf{A}\mathbf{F}_t\mathbf{B}' + \mathbf{E}_t, \\
\text{vec}(\mathbf{F}_t) &= \mathbf{H}_{\rho_1}\text{vec}(\mathbf{F}_{t-1}) + \ldots + \mathbf{H}_{\rho_q}\text{vec}(\mathbf{F}_{t-q}) + \mathbf{u}_t
\end{aligned} \tag{4}$$

- $\mathbf{Y}_t$: an $n \times k$ matrix of observed data at time t
- $\mathbf{A}$: an $n \times p_1$ matrix of factor loadings
- $\mathbf{B}$: a $k \times p_2$ matrix of factor loadings
- $\mathbf{F}_t$: a $p_1 \times p_2$ factor matrix; vec($\mathbf{F}_t$): vectorized factors of $\mathbf{F}_t$
- $\mathbf{E}_t$: a $n \times k$ idiosyncratic component
- $\mathbf{H}_{\rho_l}$: a diagonal matrix of autoregressive coefficients $(\rho_{1,l}, \ldots, \rho_{p_1 p_2,l})'$

Interpretations

For the i-th row of the observed matrix $\mathbf{Y}_t$:

$$\mathbf{Y}_{i,\cdot,t} = \mathbf{A}_{i,\cdot}\mathbf{F}_t\mathbf{B}' + \mathbf{E}_{i,\cdot,t}, \quad i = 1, \ldots, n. \tag{5}$$

Similarly, for the j-th column:

$$\mathbf{Y}_{\cdot,j,t} = \mathbf{A}\mathbf{F}_t\mathbf{B}'_{\cdot,j} + \mathbf{E}_{\cdot,j,t}, \quad j = 1, \ldots, k. \tag{6}$$

A Kronecker structure in the covariance

$$\text{vec}(\mathbf{E}_t) \sim \mathcal{MN}(\mathbf{0}_{nk}, \omega_t \mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r), \tag{7}$$

- $\mathbf{\Sigma}_r$: a covariance matrix with dimension $n \times n$
- $\mathbf{\Sigma}_c$: a covariance matrix with dimension $k \times k$.

For any row, the conditional covariance is

$$\text{Cov}(\mathbf{Y}'_{i,\cdot,t} | \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{r,i,i} \mathbf{\Sigma}_c, \quad i = 1, \ldots, n,$$

For any column, the conditional covariance is

$$\text{Cov}(\mathbf{Y}_{\cdot,j,t} | \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{c,j,j} \mathbf{\Sigma}_r, \quad j = 1, \ldots, k,$$

$\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$ represent the row-wise and column-wise covariances, respectively, that are not be explained by the common components.

Extension: time-varying volatlity

Specification 1: Time-varying volatility (Carriero et al., 2016)
Let $\omega_t = \exp(h_t)$, where h_t is a latent variable following an AR(1) process:

$$h_t = \phi h_{t-1} + u_t^h, \quad u_t^h \sim \mathcal{N}(0, \sigma_h^2), \tag{8}$$

Specification 2. The explicit outlier component (Stock and Watson, 2016)
Let $\omega_t = o_t^2$, where o_t^2 follows a mixture distribution that distinguishes between regular observations $o_t = 1$ and outliers with $o_t \geqslant 2$. The probability that outliers occur is p, which is assumed to have a beta prior.
Specification 3. Fat-tailed innovations (Jacquier et al., 2004)
Let $\omega_t = q_t^2$, where q_t^2 follows an inverse-gamma distribution: $q_t^2 \sim \mathcal{IG}(l/2, l/2)$.
Then the marginal distribution of the vectorized error has a multivariate t distribution with zero mean, scale matrix $\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$, and degree of freedom l

A more flexible specification

Identification

Identification problems

Problem 1:

Model (4) can be written as

$$\mathbf{Y}_t = \mathbf{A}\mathbf{C}^{-1}\mathbf{C}\mathbf{F}_t\mathbf{D}'(\mathbf{D}')^{-1}\mathbf{B}' + \mathbf{E}_t, \tag{9}$$

where **C** and **D** are $p_1 \times p_1$ and $p_2 \times p_2$ invertible matrices.

Problem 2:

Covariance matrix can only be identified up to scale:

$$m\boldsymbol{\Sigma}_c \otimes m^{-1}\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r, \forall m \in \mathbb{R} \setminus \{0\} \tag{10}$$

Identification conditions

1. Factor and idiosyncratic component are uncorrelated
2. $\text{Cov}(\mathbf{u}_t)$ is a positive-definite diagonal matrix
3. **A** and **B** are lower triangular matrices with ones on their diagonals
4. The $(1,1)$ element of $\boldsymbol{\Sigma}_c$ is normalized to be 1.

The next two conditions can be used as a substitute for assumptions 2 and 3 above:

2.* $\text{Cov}(\mathbf{u}_t)$ is an identity matrix

3.* One of the matrices of factor loadings, **A** or **B**, are lower-triangular matrices with ones on the diagonal, while the other one is a lower-triangular matrix with strictly positive diagonal elements.

Prior and Posterior

Determining the Dimensions of the Factor Matrix

We estimate marginal likelihoods to determine the dimensions of the factor matrix

- We use a cross-entropy (CE) method to estimate marginal likelihood.
- The importance sampling estimator can be obtained from

$$\widehat{p}_{IS}(\mathbf{y}) = \frac{1}{N}\sum_{n=1}^{N}\frac{p(\mathbf{y}|\boldsymbol{\theta}_n)p(\boldsymbol{\theta}_n)}{g(\boldsymbol{\theta}_n)}, \tag{11}$$

  where $g(\boldsymbol{\theta}_n)$ is importance densities evaluated at importance draws $\boldsymbol{\theta}_n$.
- CE method is used to find the importance densities within a certain parametric family (Chan and Eisenstat, 2015)

Multinational Macroeconomic Panel

Data

- The dataset includes 10 quarterly indicators of 19 countries from 1995.Q1 to 2023.Q3 for 115 quarters. Display data
- The countries include developed countries from North America, Europe, Asia and Oceania.
- The indicators include real GDP, price indices, labor unit cost, unemployment rates, international trade as well as household consumption.
- Each time series is adjusted for stationarity and standardized by demeaning and dividing by their standard deviations.

The log marginal likelihood estimates suggest an 1×2 factor matrix.

Log marginal likelihood estimates

The 19 countries are categorized into 3 groups

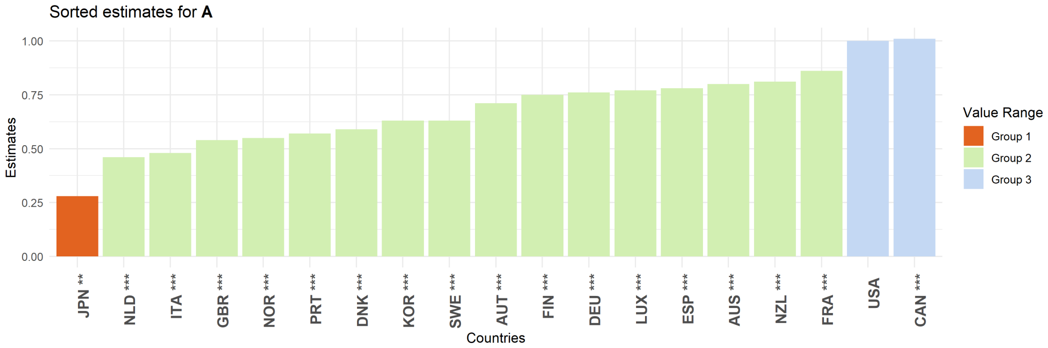

Figure 4: Bar plots of sorted estimates for loading matrix **A**. The 19 countries are categorized into 3 distinct groups based on the posterior probabilities that the differences between neighboring values are greater than 0. The stars on the country labels show the significance level of the corresponding estimates. There is no significance level on USA because we fix the corresponding element in **A** to be 1.

The 10 indicators are categorized into 4 groups

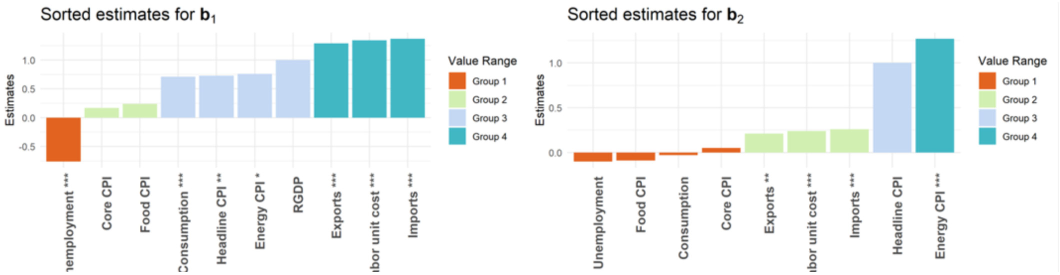

Figure 5: Bar plots of sorted estimates for **B**. The 10 indicators are categorized into 4 distinct groups according to the first column (left) or the second column (right) of **B**. The stars on the variable labels show the significance level of the corresponding estimates.

Interpretations of factor estimates

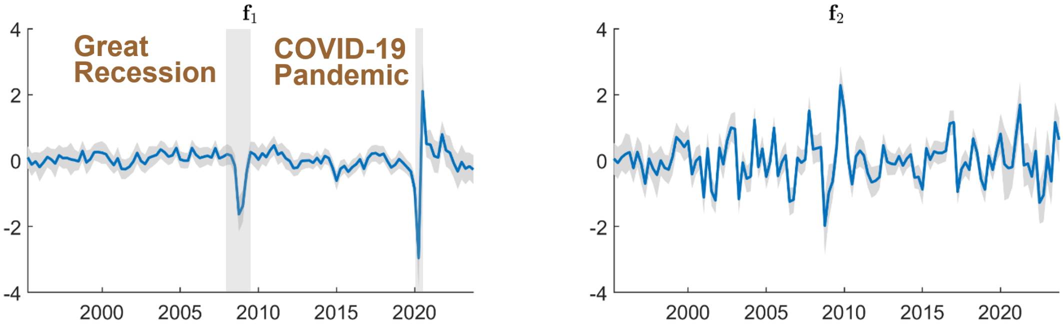

Figure 6: Plots of estimates for the factor matrix and 90% credible intervals. The first column of the matrix (left) impacts all indicators, likely representing a international business cycle. The second factor influences all the indicators except for real GDP, starting with headline CPI., likely capturing price dynamics.

The second factor co-move with oil prices

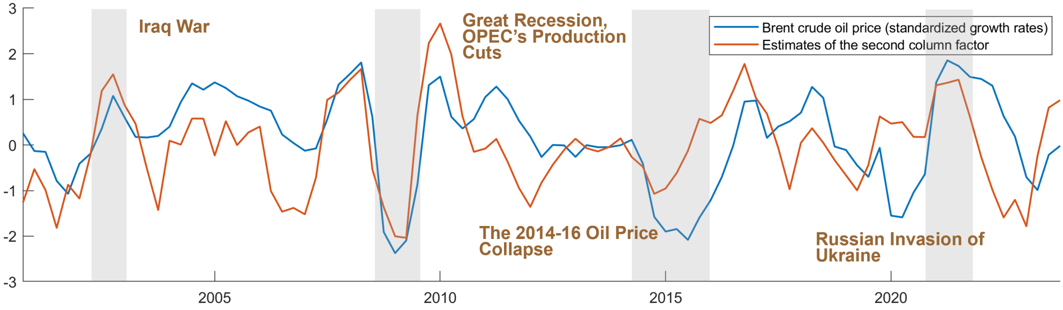

Figure 7: Yearly moving average of standardized growth rates of Brent crude oil price and the second column factor estimates. The comovement between the two series is evident.

Significant cross-indicator correlations in idiosyncratic components

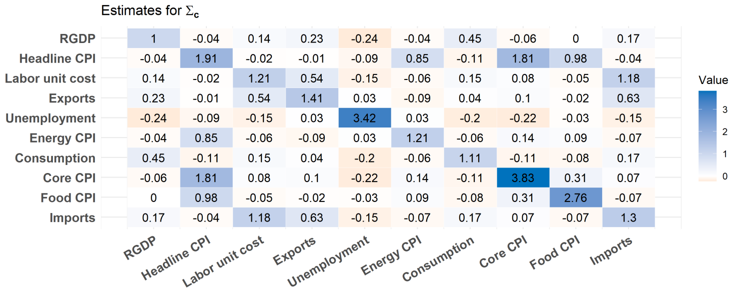

Figure 8: Heatmap of estimates for $\mathbf{\Sigma}_c$. Headline CPI is positively correlated with its disaggregated components. Unemployment is negatively correlated with real GDP, labor unit cost, etc. Labor unit cost is positively correlated to exports and imports.

Significant cross-country correlations in idiosyncratic components

Estimates for Σ_r

	USA	GBR	AUS	DEU	JPN	FRA	NOR	CAN	NLD	AUT	DNK	ESP	FIN	SWE	ITA	KOR	LUX	NZL	PRT
USA	0.17	0.06	0.06	0.06	0.05	0.05	0.04	0.08	0.08	0.06	0.05	0.06	0.06	0.07	0.06	0.08	0.03	0.06	0.06
GBR	0.06	0.23	0.06	0.09	0.06	0.09	0.07	0.05	0.11	0.1	0.09	0.09	0.09	0.1	0.1	0.05	0.09	0.07	0.1
AUS	0.06	0.06	0.37	0.06	0.08	0.06	0.07	0.05	0.07	0.06	0.06	0.06	0.08	0.09	0.07	0.08	0.05	0.12	0.07
DEU	0.06	0.09	0.06	0.25	0.06	0.12	0.07	0.04	0.15	0.15	0.12	0.1	0.12	0.13	0.12	0.08	0.09	0.07	0.11
JPN	0.05	0.06	0.08	0.06	0.88	0.06	0.04	0.03	0.07	0.06	0.04	0.05	0.05	0.06	0.04	0.09	0.04	0.07	0.06
FRA	0.05	0.09	0.06	0.12	0.06	0.2	0.08	0.05	0.12	0.12	0.1	0.11	0.1	0.11	0.12	0.07	0.09	0.08	0.11
NOR	0.04	0.07	0.07	0.07	0.04	0.08	0.44	0.06	0.1	0.08	0.09	0.06	0.07	0.1	0.06	0.05	0.04	0.06	0.06
CAN	0.08	0.05	0.05	0.04	0.03	0.05	0.06	0.18	0.07	0.04	0.05	0.04	0.05	0.06	0.04	0.05	0.03	0.06	0.05
NLD	0.08	0.11	0.07	0.15	0.07	0.12	0.1	0.07	0.3	0.15	0.13	0.13	0.12	0.13	0.14	0.09	0.1	0.08	0.12
AUT	0.06	0.1	0.06	0.15	0.06	0.12	0.08	0.04	0.15	0.25	0.13	0.11	0.14	0.13	0.12	0.07	0.1	0.09	0.11
DNK	0.05	0.09	0.06	0.12	0.04	0.1	0.09	0.05	0.13	0.13	0.33	0.1	0.1	0.12	0.11	0.06	0.07	0.08	0.11
ESP	0.06	0.09	0.06	0.1	0.05	0.11	0.06	0.04	0.13	0.11	0.1	0.21	0.1	0.1	0.12	0.06	0.1	0.08	0.12
FIN	0.06	0.09	0.08	0.12	0.05	0.1	0.07	0.05	0.12	0.14	0.1	0.1	0.35	0.14	0.11	0.09	0.08	0.09	0.11
SWE	0.07	0.1	0.09	0.13	0.06	0.11	0.1	0.06	0.13	0.13	0.12	0.1	0.14	0.29	0.12	0.08	0.08	0.08	0.1
ITA	0.06	0.1	0.07	0.12	0.04	0.12	0.06	0.04	0.14	0.12	0.11	0.12	0.11	0.12	0.23	0.08	0.09	0.08	0.11
KOR	0.08	0.05	0.08	0.08	0.09	0.07	0.05	0.05	0.09	0.07	0.06	0.06	0.09	0.08	0.08	0.37	0.05	0.1	0.06
LUX	0.03	0.09	0.05	0.09	0.04	0.09	0.04	0.03	0.1	0.1	0.07	0.1	0.08	0.08	0.09	0.05	0.37	0.08	0.08
NZL	0.06	0.07	0.12	0.07	0.07	0.08	0.06	0.06	0.08	0.09	0.08	0.08	0.09	0.08	0.08	0.1	0.08	0.35	0.08
PRT	0.06	0.1	0.07	0.11	0.06	0.11	0.06	0.05	0.12	0.11	0.11	0.12	0.11	0.1	0.11	0.06	0.08	0.08	0.24

Value: 0.8, 0.6, 0.4, 0.2

Figure 9: Heatmap of estimates for $\mathbf{\Sigma}_r$. Idiosyncratic risks for countries in European Union are correlated. UK is weakly correlated to EU

Significant stochastic volatility

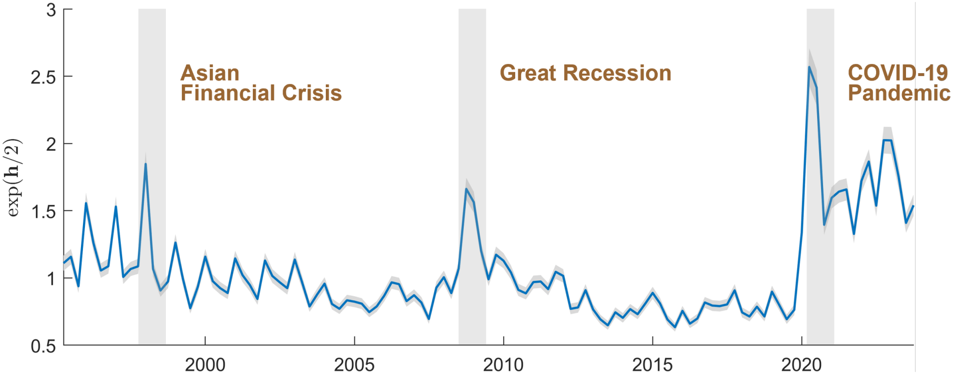

Figure 10: Estimates for stochastic volatility.

⏩ Fama-French 10 × 10 Panel

Conclusion

We have:

- Developed a new dynamic factor model designed for matrix-valued time series
- Proposed an effective method to estimate this model and an approach to determine the dimension of the factor matrix.
- Evaluated the performance of our estimators in practice using Monte Carlo experiments.
- Illustrated the usefulness using empirical applications.

Future direction

- Model Development:
  1. Extend the model for data with higher-dimensional structure, such as a tensor dynamic factor model.
  2. Develop a sparse matrix factor model to focus on identifying and estimating only the most relevant factors.
- Macroeconomic applications
  1. Monetary shock transmission mechanism
  2. Technology-spillover effects
  3. Trade networks

. . .

Thank you!

Check out the paper:

Bibliography I

Carriero, A., T. E. Clark, and M. Marcellino (2016): "Common drifting volatility in large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390.

Chan, J. C. and E. Eisenstat (2015): "Marginal likelihood estimation with the cross-entropy method," *Econometric Reviews*, 34, 256–285.

Chan, J. C. and I. Jeliazkov (2009): "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101–120.

Chan, J. C. and Y. Qi (2024): "Large Bayesian Matrix Autoregressions," *Available at SSRN 4855762*.

Chib, S. and E. Greenberg (1994): "Bayes inference in regression models with ARMA (p, q) errors," *Journal of Econometrics*, 64, 183–206.

Cong, Y., B. Chen, and M. Zhou (2017): "Fast simulation of hyperplane-truncated multivariate normal distributions," .

Bibliography II

Crucini, M. J., M. A. Kose, and C. Otrok (2011): "What are the driving forces of international business cycles?" *Review of Economic Dynamics*, 14, 156–175.

Ha, J., M. A. Kose, and F. Ohnsorge (2023): "One-stop source: A global database of inflation," *Journal of International Money and Finance*, 137, 102896.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004): "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122, 185–212.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003): "International business cycles: World, region, and country-specific factors," *American Economic Review*, 93, 1216–1239.

——— (2008): "Understanding the evolution of world business cycles," *Journal of international Economics*, 75, 110–130.

Miranda-Agrippino, S., H. Rey, et al. (2015): *World asset markets and the global financial cycle*, vol. 21722, National Bureau of Economic Research Cambridge, MA.

Bibliography III

Nobile, A. (2000): "Comment: Bayesian multinomial probit models with a normalization constraint," *Journal of Econometrics*, 99, 335–345.

Stock, J. H. and M. W. Watson (2016): "Core inflation and trend inflation," *Review of Economics and Statistics*, 98, 770–784.

A more flexible specification for the time-varying volatility

Consider the following more flexible specification for the time-varying volatility:

$$\text{vec}(\mathbf{E}_t) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}_t\right), \tag{12}$$

- Independent stochastic volatility series: $\mathbf{D}_t = \text{diag}(e^{h_{1,1,t}}, e^{h_{2,1,t}}, \ldots, e^{h_{n,k,t}})$
- Autoregressive processes:

$$h_{i,j,t} = \phi_{i,j} h_{i,j,t-1} + u_{i,j,t}, \quad u_{i,j,t} \sim \mathcal{N}(0, \sigma^2_{h,i,j}), \quad t = 2, \ldots, T \tag{13}$$

- The initial states are assumed to follow Gaussian priors.
- Tradeoff: Flexibility vs Complexity

Back

The posterior sampler

Step 1. Sampling from $(\mathbf{A}', \mathbf{\Sigma}_r | \mathbf{Y}, \mathbf{B}, \mathbf{F}, \mathbf{\Sigma}_c)$
We sample $(\mathbf{A}', \mathbf{\Sigma}_r)$ conditional on the latent factors and other parameters from a normal-inverse-Wishart distribution:

$$(\mathbf{A}', \mathbf{\Sigma}_r | \cdot) \sim \mathcal{NIW}(\widehat{\mathbf{A}}', \mathbf{K}_{\mathbf{A}'}^{-1}, \widehat{\nu}_r, \widehat{\mathbf{S}}_r),$$

where

$$\mathbf{K}_{\mathbf{A}'} = \mathbf{V}_{\mathbf{A}'}^{-1} + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t \mathbf{B}' \mathbf{\Sigma}_c^{-1} \mathbf{B} \mathbf{F}_t', \quad \widehat{\mathbf{A}}' = \mathbf{K}_{\mathbf{A}'}^{-1} \left(\mathbf{V}_{\mathbf{A}'}^{-1} \mathbf{A}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t \mathbf{B}' \mathbf{\Sigma}_c^{-1} \mathbf{Y}_t' \right)$$

$$\widehat{\nu}_r = \nu_r + Tk, \quad \widehat{\mathbf{S}}_r = \mathbf{S}_r + \mathbf{A}_0 \mathbf{V}_{\mathbf{A}'}^{-1} \mathbf{A}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{Y}_t \mathbf{\Sigma}_c^{-1} \mathbf{Y}_t' - \widehat{\mathbf{A}} \mathbf{K}_{\mathbf{A}'} \widehat{\mathbf{A}}'.$$

Back

Priors

- A Natural conjugate prior for the loadings:

$$\boldsymbol{\Sigma}_r \sim \mathcal{IW}(\nu_r, \mathbf{S}_r), \quad (\text{vec}(\mathbf{A}')|\boldsymbol{\Sigma}_r) \sim \mathcal{N}(\text{vec}(\mathbf{A}_0'), \boldsymbol{\Sigma}_r \otimes \mathbf{V}_{\mathbf{A}'}),$$
$$\boldsymbol{\Sigma}_c \sim \mathcal{IW}(\nu_c, \mathbf{S}_c), \quad (\text{vec}(\mathbf{B}')|\boldsymbol{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\mathbf{B}_0'), \boldsymbol{\Sigma}_c \otimes \mathbf{V}_{\mathbf{B}'}). \tag{14}$$

- A truncated normal prior for the autoregressive coefficients:

$$\rho_{j,k} \sim \mathcal{TN}(\rho_{j,k0}, V_{\rho_{j,k}}), \quad j = 1, ..., p_1, \quad k = 1, ..., p_2.$$

- An inverse-gamma prior for $\lambda_{j,k}^2$: $\mathcal{IG}(\nu_{\lambda_{j,k}}, S_{\lambda_{j,k}})$
- The first q values of $\mathbf{F}_t$ are treated unknown:

$$f_{j,k,l} \sim \mathcal{N}\left(0, \frac{\lambda_{j,k}^2}{1 - \sum_{m=1}^{q} \rho_{j,k,m}^2}\right), \quad l = 1, \ldots, q. \tag{15}$$

Back

The posterior sampler

1. Sample from $(\mathbf{A}', \boldsymbol{\Sigma}_r | \cdot)$ from a normal-inverse-Wishart distribution: $(\mathbf{A}', \boldsymbol{\Sigma}_r | \cdot) \sim \mathcal{NIW}(\widehat{\mathbf{A}}', \mathbf{K}_{\mathbf{A}'}^{-1}, \widehat{\nu}_r, \widehat{\mathbf{S}}_r)$.
   - Apply Algorithm 2 in Cong et al. (2017) for the restrictions on the lower-triangular structure of $\mathbf{A}$
2. Sample from $(\mathbf{B}', \boldsymbol{\Sigma}_c | \cdot)$ from a normal-inverse-Wishart distribution: $(\mathbf{B}, \boldsymbol{\Sigma}_c | \cdot) \sim \mathcal{NIW}(\widehat{\mathbf{B}}', \mathbf{K}_{\mathbf{B}'}^{-1}, \widehat{\nu}_c, \widehat{\mathbf{S}}_c)$.
   - Apply Algorithm 2 in Cong et al. (2017) for the restrictions on the lower-triangular structure of $\mathbf{B}$.
   - Apply the algorithm proposed by Nobile (2000) for the restriction on $\boldsymbol{\Sigma}_c$.
3. Sample from $(\text{vec}(\mathbf{F}_t) | \cdot)$, $t = 1, \ldots, T$ from a normal distribution.
4. Sample from $(\lambda_{j,k}^2 | \cdot)$, $\quad j = 1, ..., p_1, k = 1, ..., p_2$ from an inverse gamma distribution.
5. Sample from $(\rho_{j,k} | \cdot)$, $\quad j = 1, ..., p_1, k = 1, ..., p_2$ using an Metropolis-Hastings algorithm following Chib and Greenberg (1994) and Chan and Jeliazkov (2009).

Details

The posterior sampler

Apply Algorithm 2 in Cong et al. (2017) or Algorithm 1 in Chan and Qi (2024) to efficiently sample $(\text{vec}(\mathbf{A}')|\cdot) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}'), \mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})$ such that $\mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}') = \mathbf{a}_0$. In particular, one can first sample $\text{vec}(\mathbf{A}'_u)$ from the unconstrained conditional posterior distribution in Step 1, and then return

$$\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}'_u) + (\mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})\mathbf{M}'_{\mathbf{A}'}\left(\mathbf{M}_{\mathbf{A}'}(\mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})\mathbf{M}'_{\mathbf{A}'}\right)^{-1}(\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}'_u)),$$

which can be realized by the following four steps:

1. Compute $\mathbf{C} = \mathbf{C}_{\mathbf{\Sigma}_r^{-1}} \otimes \mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$, where $\mathbf{C}_{\mathbf{\Sigma}_r^{-1}}$ is the lower Cholesky factor of $\mathbf{\Sigma}_r^{-1}$, and $\mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$ is the lower Cholesky factor of $\mathbf{K}_{\mathbf{A}'}$;
2. Solve $\mathbf{C}\mathbf{C}'\mathbf{U} = \mathbf{M}'_{\mathbf{A}'}$ for $\mathbf{U}$;
3. Solve $\mathbf{M}_{\mathbf{A}'}\mathbf{U}\mathbf{V} = \mathbf{U}'$ for $\mathbf{V}$;
4. Return $\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}'_u) + \mathbf{V}'(\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}'_u))$.

Back

The posterior sampler

Step 2. Sampling from $(\mathbf{B}', \boldsymbol{\Sigma}_c | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \boldsymbol{\Sigma}_r)$
Similar to step 1, $(\mathbf{B}, \boldsymbol{\Sigma}_c)$ are drawn from a normal-inverse-Wishart distribution:

$$(\mathbf{B}, \boldsymbol{\Sigma}_c | \cdot\,) \sim \mathcal{NIW}(\widehat{\mathbf{B}}', \mathbf{K}_{\mathbf{B}'}^{-1}, \widehat{\nu}_c, \widehat{\mathbf{S}}_c),$$

where

$$\mathbf{K}_{\mathbf{B}'} = \mathbf{V}_{\mathbf{B}'}^{-1} + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t' \mathbf{A}' \boldsymbol{\Sigma}_r^{-1} \mathbf{A} \mathbf{F}_t, \quad \widehat{\mathbf{B}}' = \mathbf{K}_{\mathbf{B}'}^{-1} \left(\mathbf{V}_{\mathbf{B}'}^{-1} \mathbf{B}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t' \mathbf{A}' \boldsymbol{\Sigma}_r^{-1} \mathbf{Y}_t \right)$$

$$\widehat{\nu}_c = \nu_c + Tn, \quad \widehat{\mathbf{S}}_c = \mathbf{S}_c + \mathbf{B}_0 \mathbf{V}_{\mathbf{B}'}^{-1} \mathbf{B}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{Y}_t' \boldsymbol{\Sigma}_r^{-1} \mathbf{Y}_t - \widehat{\mathbf{B}} \mathbf{K}_{\mathbf{B}'} \widehat{\mathbf{B}}'.$$

Back

The posterior sampler

We sample $(\mathbf{B}', \mathbf{\Sigma}_c | \cdot)$ in two steps. First, we sample $\mathbf{\Sigma}_c$ marginally from $(\mathbf{\Sigma}_c | \cdot) \sim \mathcal{IW}(\widehat{\mathbf{S}}_c, \widehat{\nu}_c)$ with the restriction that $\sigma_{c,1,1} = 1$. We use the algorithm by Nobile (2000) for this step, outlined below:

1. Exchange row/column 1 and n in the matrix $\widehat{\mathbf{S}}_c$. Denote this matrix as $\widehat{\mathbf{S}}_c^{Trans}$.
2. Construct a lower triangular matrix such that
   - δ_{ii} equal to the square root of $\chi^2_{\widehat{\nu}_c+1-i}$ for $i = 1, \ldots, n-1$;
   - $\delta_{nn} = (l_{nn})^{-1}$, where l_{nn} is the (n, n)-th element in the Cholesky decomposition of $(\widehat{\mathbf{S}}_c^{Trans})^{-1}$, denoted as $\mathbf{L}$
   - δ_{ij} equal to $\mathcal{N}(0, 1)$ random variates, $i > j$.
3. Set $\mathbf{\Sigma}_c = (\mathbf{L}^{-1})'(^{-1})'^{-1}\mathbf{L}^{-1}$.
4. Exchange the row/column 1 and n of $\mathbf{\Sigma}_c$ back.

Then we simulate from a normal distribution for $\mathbf{B}$:

$$(\text{vec}(\mathbf{B}') | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r \mathbf{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{B}}'), \mathbf{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}'}^{-1}),$$

which can be done using the algorithm depicted in step 1.

The posterior sampler

Step 3. Sampling from $(\text{vec}(\mathbf{F}_t)|\mathbf{Y}_t, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \omega^2, \rho)$, $t = 1, \ldots, T$

We sample the factors by t. Specifically, conditional on parameters, $\text{vec}(\mathbf{F}_t)$ from a normal distribution:

$$(\text{vec}(\mathbf{F}_t)| \cdot) \sim \mathcal{N}(\widehat{\mathbf{f}}_t, \mathbf{K}_{\mathbf{f}_t}^{-1}),$$

where

$$\mathbf{K}_{\mathbf{f}_t} = \omega_t^{-1}\mathbf{B}'\boldsymbol{\Sigma}_c^{-1}\mathbf{B} \otimes \mathbf{A}'\boldsymbol{\Sigma}_r^{-1}\mathbf{A} + \boldsymbol{\Lambda}_t^{-1}$$

$$\widehat{\mathbf{f}}_t = \mathbf{K}_{\mathbf{f}_t}^{-1}\left[\omega_t^{-1}(\mathbf{B}'\boldsymbol{\Sigma}_c^{-1} \otimes \mathbf{A}'\boldsymbol{\Sigma}_r^{-1})\text{vec}(\mathbf{Y}_t)\right] \quad \text{for } t = 1, \ldots, q,$$

$$\widehat{\mathbf{f}}_t = \mathbf{K}_{\mathbf{f}_t}^{-1}\left[\omega_t^{-1}(\mathbf{B}'\boldsymbol{\Sigma}_c^{-1} \otimes \mathbf{A}'\boldsymbol{\Sigma}_r^{-1})\text{vec}(\mathbf{Y}_t) + \boldsymbol{\Lambda}_t^{-1}\sum_{m=1}^{q}\mathbf{H}_{\boldsymbol{\rho}_m}\mathbf{f}_{t-m}\right] \quad \text{for } t = q+1, \ldots, T,$$

where for $t = 1, \ldots, q$, $\boldsymbol{\Lambda}_t = \text{diag}(\boldsymbol{\lambda}^2/(1 - \sum_{m=1}^{q}\boldsymbol{\rho}_m^2))$, and for $t = 2, \ldots, T$, $\boldsymbol{\Lambda}_t = \text{diag}(\boldsymbol{\lambda}^2)$. $\boldsymbol{\rho}_m = (\rho_{1,1,m}, \ldots, \rho_{p_1,p_2,m})'$, $\boldsymbol{\lambda} = (\lambda_{1,1}, \ldots, \lambda_{p_1,p_2})'$.
$\mathbf{H}_{\boldsymbol{\rho}_m} = \text{diag}(\rho_{1,1,m}, \rho_{2,1,m}, \ldots, \rho_{p_1,p_2,m})$.

The posterior sampler

Step 4. Sampling from $(\lambda^2_{j,k}|\mathbf{f}_{j,k},\boldsymbol{\rho}_{j,k}), \quad j=1,...,p_1, k=1,...,p_2$

It is clear that $(\lambda^2_{j,k}|\mathbf{f}_{j,k},\boldsymbol{\rho}_{j,k}) \sim \mathcal{IG}(\widehat{\nu}_{\lambda_{j,k}}, \widehat{S}_{\lambda_{j,k}})$, where $\widehat{\nu}_{\lambda_{j,k}} = \nu_{\lambda_{j,k}} + \frac{T}{2}$, and $\widehat{S}_{\lambda_{j,k}} = S_{\lambda_{j,k}} + \frac{1}{2}\left[\sum_{t=1}^{q} f^2_{j,k,t}(1-\sum_m \rho^2_{j,k,m}) + \sum_{t=q+1}^{T}(f_{j,k,t} - \rho_{j,k,1} f_{j,k,t-1} - ... - \rho_{j,k,q} f_{j,k,q})^2\right]$.

Back

The posterior sampler

Step 5. Sampling from $(\boldsymbol{\rho}_{j,k}|\mathbf{f}_{j,k},\lambda^2_{j,k})$, $\quad j=1,...,p_1,\, k=1,...,p_2$
Note that $\boldsymbol{\rho}_{j,k}$ is a $q\times 1$ vector: $\boldsymbol{\rho}_{j,k}=(\rho_{j,k,1},\ldots,\rho_{j,k,q})'$. We rewrite (**??**) as follows:

$$\widetilde{\mathbf{f}}_{j,k}=\widetilde{\mathbf{F}}_{j,k}\boldsymbol{\rho}_{j,k}+\mathbf{u}_{j,k},\quad \mathbf{u}_{j,k}\sim\mathcal{N}(\mathbf{0},\lambda_{j,k}\mathbf{I}_{T-q}), \tag{16}$$

where $\widetilde{\mathbf{f}}_{j,k}=(f_{j,k,q+1},\ldots,f_{j,k,T})'$, and

$$\widetilde{\mathbf{F}}_{j,k}=\begin{bmatrix} f_{j,k,1} & f_{j,k,2} & \cdots & f_{j,k,q} \\ f_{j,k,2} & f_{j,k,3} & \cdots & f_{j,k,q+1} \\ \vdots & \cdots & \cdots & \vdots \\ f_{j,k,T-q} & f_{j,k,T-q+1} & \cdots & f_{j,k,T} \end{bmatrix}_{(T-q)\times q}$$

Back

The posterior sampler

Following Chib and Greenberg (1994) and Chan and Jeliazkov (2009), we design an Metropolis-Hastings algorithm with proposal $\boldsymbol{\rho}^*_{j,k} \sim \mathcal{N}(\widehat{\boldsymbol{\rho}}_{j,k}, \mathbf{K}^{-1}_{\boldsymbol{\rho}_{j,k}})$, where $\mathbf{K}_{\boldsymbol{\rho}_{j,k}} = \mathbf{V}^{-1}_{\boldsymbol{\rho}_{j,k}} + \widetilde{\mathbf{F}}'_{j,k}\widetilde{\mathbf{F}}_{j,k}/\lambda^2_{j,k}$, $\widehat{\boldsymbol{\rho}}_{j,k} = \mathbf{K}^{-1}_{\boldsymbol{\rho}_{j,k}}(\mathbf{V}^{-1}_{\boldsymbol{\rho}_{j,k}}\boldsymbol{\rho}_{j,k,0} + \widetilde{\mathbf{F}}'_{j,k}\widetilde{\mathbf{f}}_{j,k}/\lambda^2_{j,k})$. The proposed value $\boldsymbol{\rho}^*_{j,k}$ is accepted with probablity

$$\alpha_{MH}(\boldsymbol{\rho}_{j,k}, \boldsymbol{\rho}^*_{j,k}) = \min\left\{1, \frac{f_{\mathcal{N}}(\mathbf{f}_{j,k,1:q}|\mathbf{0}, \lambda^2_{j,k}/(1-\sum_m \rho^{*2}_{j,k,m})\mathbf{I}_q)}{f_{\mathcal{N}}(\mathbf{f}_{j,k,1:q}|\mathbf{0}, \lambda^2_{j,k}/(1-\sum_m \rho^{2}_{j,k,m})\mathbf{I}_q)}\right\}.$$

Back

Simulation Results

Data generating process

The parameters are drawn as follows

- Free parameters in $\mathbf{A}$ and $\mathbf{B}$ are sampled from $\mathcal{U}(0,1)$
- $\rho_{j,k} \sim \mathcal{U}(0.8, 0.9)$
- $\mathbf{\Sigma}_c$ to $0.3\mathbf{I}_k$, $\mathbf{\Sigma}_r$ to $0.5\mathbf{I}_n$
- $\lambda^2_{j,k} = 1$ for $j = 1, \ldots, p_1$, $k = 1, \ldots, k$

Sample size

- $(n, k) \in \{(10, 10), (20, 15), (30, 20)\}$
- $T \in \{200, 500, 1000\}$
- The factor matrices are preset to dimensions $(p_1, p_2) = (3, 2)$ or $(p_1, p_2) = (5, 5)$

Adjusted R^2 from regressing true values on estimates: small factor matrix

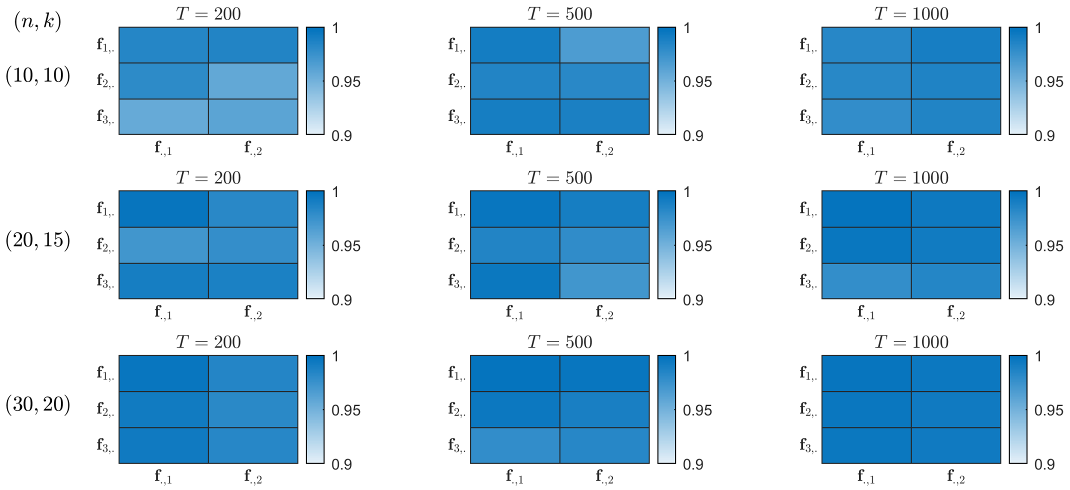

Figure 11: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 3$, $p_2 = 2$

Adjusted R^2 from regressing true values on estimates: large factor matrix

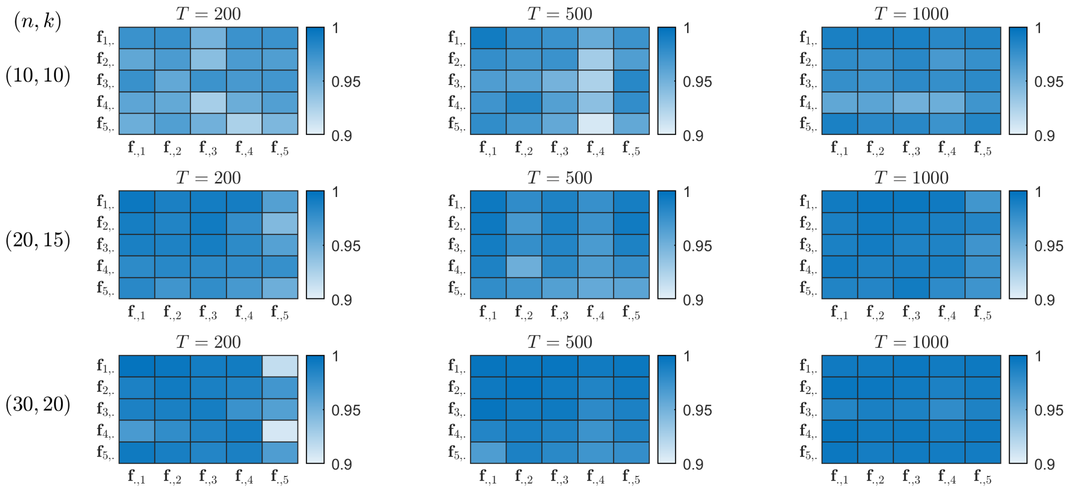

Figure 12: Adjusted R^2 from regressing true factors on estimates: $p_1 = 5, p_2 = 5$

Can marginal likelihoods uncover the true model?

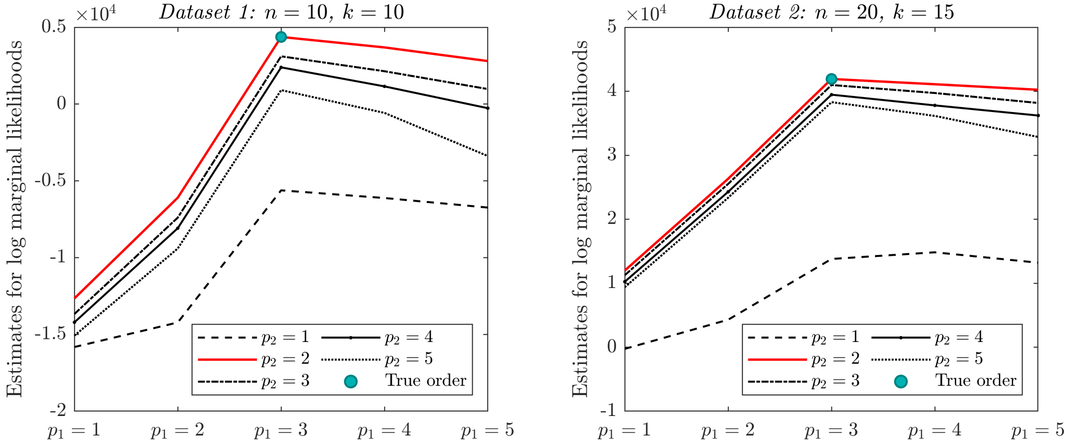

Figure 13: Estimates for log marginal likelihoods when true order of the factor matrix is $p_1 = 3$, $p_2 = 2$

Can marginal likelihoods uncover the true model?

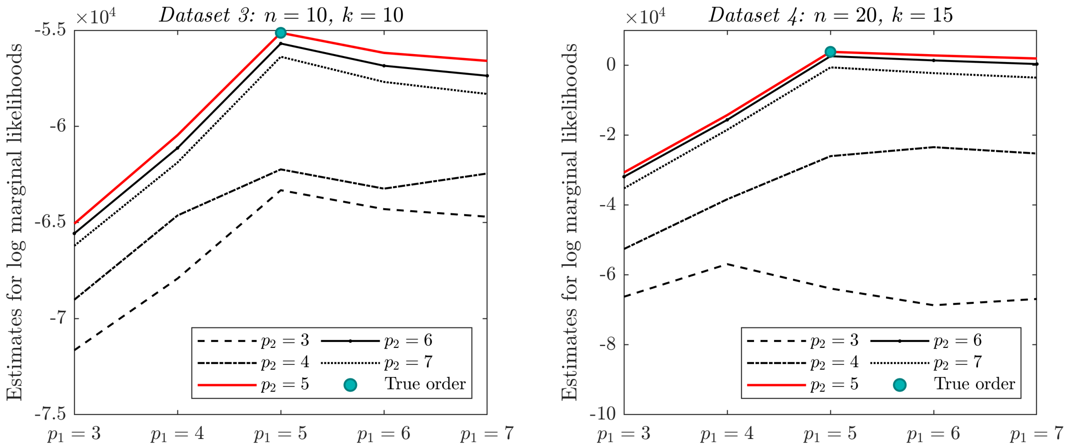

Figure 14: Estimates for log marginal likelihoods when true order of the factor matrix is $p_1 = 5$, $p_2 = 5$

Log marginal likelihood estimates imply an 1×2 factor matrix

Table 1: Log marginal likelihood estimates

	$p_2 = 1$	$p_2 = 2$	$p_2 = 3$
$p_1 = 1$	-17362	**-17325**	-17348
	(0.48)	(0.58)	(0.64)
$p_1 = 2$	-17443	-17439	-17499
	(0.54)	(0.36)	(0.58)
$p_1 = 3$	-17508	-17542	-17598
	(0.55)	(0.68)	(0.82)

Back

Log marginal likelihood estimates suggest an 1×3 factor matrix

Table 2: Estimates for log marginal likelihoods

	$p_2 = 1$	$p_2 = 2$	$p_2 = 3$
$p_1 = 1$	-44630.3 (0.33)	-43928.7 (0.27)	-43638.4 (0.34)
$p_1 = 2$	-43740.5 (0.56)	-43417.9 (0.52)	**-43408.4** (0.38)
$p_1 = 3$	-43867 (0.45)	-43857.2 (0.41)	-43435.4 (0.53)

Back

Correlations among factors suggest a more flexible specification for the factor evolution process

Table 3: Correlation coefficients of the six factor series

	$\mathbf{f}_{1,1}$	$\mathbf{f}_{2,1}$	$\mathbf{f}_{1,2}$	$\mathbf{f}_{2,2}$	$\mathbf{f}_{1,3}$	$\mathbf{f}_{2,3}$
$\mathbf{f}_{1,1}$	1.00	0.14***	0.19***	0.06*	0.16***	-0.07
$\mathbf{f}_{2,1}$	0.14***	1.00	-0.35	0.13***	-0.23	-0.41
$\mathbf{f}_{1,2}$	0.19***	-0.35	1.00	-0.06	**0.47*****	**0.55*****
$\mathbf{f}_{2,2}$	0.06*	0.13***	-0.06	1.00	-0.35	-0.20
$\mathbf{f}_{1,3}$	0.16***	-0.23	**0.47*****	-0.35	1.00	**0.48*****
$\mathbf{f}_{2,3}$	-0.07	-0.41	**0.55*****	-0.20	**0.48*****	1.00