Bayesian Dynamic Factor Models for High-dimensional Matrix-valued Time Series

Wei Zhang

Purdue University

August 2024

Abstract

High-dimensional matrix-valued time series are of significant interest in economics and finance, with prominent examples including cross region macroeconomic panels and firms' financial data panels. We introduce a class of Bayesian matrix dynamic factor models that utilize matrix structures to identify more interpretable factor patterns and factor impacts. Our model accommodates time-varying volatility, adjusts for outliers, and allows cross-sectional correlations in the idiosyncratic components. To determine the dimension of the factor matrix, we employ an importance-sampling estimator based on the cross-entropy method to estimate marginal likelihoods. Through a series of Monte Carlo experiments, we show the properties of the factor estimators and the performance of the marginal likelihood estimator in correctly identifying the true dimensions of the factor matrices. Applying our model to a macroeconomic dataset and a financial dataset, we demonstrate its ability in unveiling interesting features within matrix-valued time series.

Keywords: Matrix-valued time series, dynamic factor models, approximate factor models, time-varying volatility, Bayesian model comparison

1 Introduction

Recently, matrix-valued time series models have gained significant attention due to their ability to capture complex multidimensional relationships among data series. These models are particularly promising for empirical research in macroeconomics and finance, where the availability of such multi-dimensional data has been increasing. A prominent example is macroeconomic indicators collected across multiple countries. These datasets can be modeled as matrices, with rows representing countries and columns representing indicators. A commonly used approach to modeling such data is to stack the matrix into a long vector and to apply standard multivariate methods. However, this approach overlooks the inherent structure of the data, where variables within the same row (country) or column (indicator often exhibit stronger relationships.

Research on statistical methods for matrix-valued time series is still evolving, and matrix factor models are becoming more prominent due to their ability to reduce dimensions, particularly in high-dimensional contexts. Wang et al. (2019) introduce a factor model for such matrix-valued time series, where both factors and factor loadings are unknown matrices, and the idiosyncratic component is assumed to be white noise. Subsequent work has built on this framework. For example, Chen et al. (2020) incorporate prior knowledge using linear constraints, Liu and Chen (2019) develop a threshold version, and Chen et al. (2024) extend the model to include time-varying loadings. However, these studies all focus on static factors and do not account for serial and cross-sectional correlations in idiosyncratic components, both of which are essential for understanding the evolving nature of economic relationships and risks.

This paper contributes to two strands of the literature. First, we extend the growing research on factor models for matrix-valued time series by incorporating an autoregressive (AR) process for factor evolution. The application of AR processes in factor models is first introduced by Sargent et al. (1977) and has since shown to be valuable for macroeconomic modeling (see e.g., Stock and Watson, 2012; Bai and Wang, 2015; Poncela et al., 2021), because of their ability to capture the persistency in macroeconomic data. Additionally, this extension allows for forecasting which makes our model a more practical tool for empirical macroeconomics and financial economics.

The other strand of the literature we contribute to is approximate factor models that

could be dated back to Chamberlain and Rothschild (1983). Unlike exact factor models that assume a diagonal covariance matrix for the idiosyncratic components, approximate factor models allow weak serial or cross-sectional correlations. We build a framework for matrix dynamic factor model that allows both serial and cross-sectional correlations. Particularly, for the time dimension, we allow a common stochastic volatility, fat-tailed errors and COVID-19 outliers. This is motivated by the increasing recognition of the need for time-varying volatility in modeling many macroeconomic datasets (see, e.g., Cross and Poon, 2016; Marcellino et al., 2016; Kastner et al., 2017; Thorsrud, 2020; Li and Scharth, 2022; Chan, 2023). In addition, the unexpected extreme movements in many macroeconomic variables at the onset of the COVID-19 pandemic have underlined the need to allow for fat-tailed errors and potential outliers (see, e.g., Lenza and Primiceri, 2022; Carriero et al., 2024a,b).

For the cross-section dimension, we allow cross-row and cross-column correlations in idiosyncratic components. In our macroeconomic panel example, this implies that we allow individual risks to be correlated across countries or indicators. We achieve this by employing a Kronecker structure in the idiosyncratic components where the covariance matrix of the vectorized error is a Kronecker product of column-wise and row-wise covariance matrices. We impose inverse-Wishart priors on the two covariance matrices, with prior means set as diagonal matrices. This data-driven approach offers a more flexible framework compared to exact factor models. In addition, compared to a full covariance matrix, this Kronecker structure greatly reduces the number of parameters and improves the efficiency of our Bayesian estimation.

In order to determine the factor matrix dimensions, we adopt a Bayesian approach and estimate marginal likelihoods. Given the challenges of computing marginal likelihoods in high-dimensional settings, we employ an importance-sampling technique proposed by Chan and Eisenstat (2015), which is based on a cross-entropy approach. This technique offers two main benefits: the importance sampling generates independent draws instead of correlated Markov chain Monte Carlo (MCMC) draws, and it provides an efficient estimator by using an importance-sampling density that is "closest" to the posterior in terms of Kullback-Leibler divergence. As a result, the marginal likelihood estimator is efficient with a low variance, and the simulation typically requires only a few thousand draws or fewer.

Our paper is closely related to two recent works. The first is Yu et al. (2024), who introduce a matrix autoregressive process for factor matrices under the framework of matrix factor model proposed by Wang et al. (2019). While they introduce a matrix factor model with white noise series that allow contemporary correlations in vectorized errors, we focus on building a framework that allows time-varying volatility, cross-row and cross-column correlations in idiosyncratic components. Methodologically, we adopt a full Bayesian approach with identification restrictions for clearer interpretation, whereas Yu et al. (2024) use a two-step procedure to estimate the column spaces aimed at forecasting. The second paper is Yuan et al. (2023), who propose a dynamic factor model using a two-way matrix factor framework distinct from Wang et al. (2019).

Through a series of Monte Carlo experiments, we demonstrate that the proposed estimator works well in practice. In particular, we show that our factor estimates are close to their true values. In addition, our results suggest that the larger the sample size is, the more accurate factor estimates are. This is in line with statistical theories for static matrix factor model analyzed in Chen and Fan (2023). Moreover, we find that the marginal likelihood estimator can correctly identify the true dimensions of factor matrices under a variety specifications of sample sizes.

We illustrate the empirical benefits of the MDFM using two datasets. The first dataset includes 10 quarterly macroeconomic indicators for 19 countries, covering 115 quarters from 1995.Q1 to 2023.Q3. The second dataset is a 10×10 Fama-French monthly panel spanning from January 1990 to June 2024 (414 observations). In both applications, we determine the dimension of factor matrices using the marginal likelihood estimator. Several key findings emerge from our analysis. First, the estimated factor loadings reveal clear patterns that can be used to group both rows (countries or sizes) or columns (indicators or book equity to market equity ratios). Second, the factors exhibit significant dynamics, as shown by the autoregressive estimates for the factor evolution processes. Third, we observe strong evidence of time-varying volatility in both applications, while cross-sectional correlations are found in the first application. Overall, these results demonstrate that our flexible modeling framework is empirically valuable for capturing complex dependencies in real-world datasets.

The rest of this paper is organized as follows. In Section 2 we specify the proposed dynamic matrix factor model, with detailed discussion of motivation, identification, priors,

Bayesian estimation and a useful extension. Section 3 introduces an importance-sampling marginal likelihood estimator for the purpose of determining the dimension of factor matrix. Monte Carlo studies are presented in Section 4 to demonstrate the accuracy of the factor estimates as well as the performance of the marginal likelihood estimator. In Section 5, we illustrate the usefulness of our model employing two empirical applications. Lastly, Section 6 concludes.

2 The Dynamic Factor Model for Matrix-valued Time Series

Building on the framework established by Wang et al. (2019), we introduce a dynamic factor model for matrix-valued time series. We follow the spirit of the approximate dynamic factor model proposed by Chamberlain and Rothschild (1983) and allow cross-row and cross-column correlations. We then incorporate time-varying volatility and outlier adjustments, discuss identification restrictions, and outline the Bayesian priors and estimation process. Lastly, we extend the model to include row-specific and column-specific stochastic volatility.

2.1 The Model

Consider observing an $n \times k$ data matrix $\mathbf{Y}_t$ at time t. To better illustrate, assume $\mathbf{Y}_t$ is a macroeconomic data matrix drawn from multiple nations at time t, where the rows correspond to n countries and the columns correspond to k variables. Consider the following dynamic factor model:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{F}_t\mathbf{B}' + \mathbf{E}_t, \quad \text{vec}(\mathbf{E}_t) \sim \mathcal{N}(\mathbf{0}, \omega_t\mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r), \tag{1}$$

$$\text{vec}(\mathbf{F}_t) = \mathbf{H}_{\boldsymbol{\rho}_1}\text{vec}(\mathbf{F}_{t-1}) + \ldots + \mathbf{H}_{\boldsymbol{\rho}_q}\text{vec}(\mathbf{F}_{t-q}) + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_t), \tag{2}$$

where $\mathbf{A}$ is a $n \times p_1$ matrix of factor loadings, $\mathbf{B}$ is a $k \times p_2$ matrix of factor loadings, $\mathbf{F}_t$ is a $p_1 \times p_2$ latent matrix-valued time series of common factors, $\mathbf{E}_t$ is a $n \times k$ idiosyncratic component, $\text{vec}(\,\cdot\,)$ is a vectorizing function, $\mathbf{H}_{\boldsymbol{\rho}_l}$ is a diagonal matrix of autoregressive coefficients $(\rho_{1,l}, \ldots, \rho_{p_1p_2,l})'$, $l = 1, \ldots, q$, and $\mathbf{\Lambda}_t$ is a covariance matrix for the error in

factor evolution process.

In model (1)-(2), we assume that each matrix series in the data set, $\mathbf{Y}_t$, can be expressed as the sum of two orthogonal components: the common components, $\mathbf{A}\mathbf{F}_t\mathbf{B}'$, and the idiosyncratic components, $\mathbf{E}_t$. The common components capture the part of the series that comove with the economy, while the idiosyncratic components represent the individual risks. The dimension of $\mathbf{F}_t$, i.e., (p_1, p_2), is typically much smaller than the dimension of the data matrix $\mathbf{Y}_t$, i.e., (n, k). This corresponds to the assumption in factor model analysis that these high-dimensional macro series can be explained by few shocks; (See, e.g., Sargent et al., 1977; Giannone et al., 2004; Bok et al., 2018; Alessi and Kerssenfischer, 2019).

The bilinear form in the common components $\mathbf{A}\mathbf{F}_t\mathbf{B}$ is crucial for capturing the interrelationships within the rows and columns of the data matrix. Particularly, in the macroeconomic panel, $\mathbf{A}$ could capture country-specific sensitivities to the latent factors, while $\mathbf{B}$ could capture the different responses of various economic indicators to these factors. Specifically, the i-th row of the data matrix can be expressed as follows:

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

where $\mathbf{Y}_{i,.,t}$ denotes the i-th row of the data matrix, $\mathbf{A}_{i,.}$ represents the i-th row of the loading matrix $\mathbf{A}$, and $\mathbf{E}_{i,.,t}$ is the i-th row of the matrix $\mathbf{E}_t$.

It is evident that the i-th row of the data matrix is a linear combination of the rows of $\mathbf{F}_t\mathbf{B}'$, with the elements in $\mathbf{A}_{i,.}$ serving as the linear coefficients. Likewise, the j-th column of the data matrix represents the linear combination of the columns of $\mathbf{A}\mathbf{F}_t$, with the elements in the j-th column of $\mathbf{B}'$ as the linear coefficients. Hence in the context of multinational macroeconomic data, each column of $\mathbf{F}_t$ captures the comovement for each variable at time t, while each row of $\mathbf{F}_t$ can be interpreted as the latent factor that influences each corresponding country. $\mathbf{A}$ illustrates the pattern of the impact of common factors on each country (row), whereas $\mathbf{B}$ illustrates that impact on each indicator (column).

2.1.1 Important Features of MDFM

Compared to the matrix factor model introduced by Wang et al. (2019), our model has two key features. One is the incorporation of factor dynamics. The other is that we allow cross-sectional correlations and time-varying volatility in the idiosyncratic components.

The motivation to incorporate factor dynamics is straightforward. In real world data, many economic indicators have a tendency of remain above or below their long-term trends for extended periods after experiencing shocks. For example, if GDP growth slows due to a recession, persistence would describe how long it takes for the economy to recover and return to its pre-recession growth rate. An autogressive process such as (2) is useful in capturing this persistent comovement among these indicators and thus have been proven to be useful for economic modeling and forecasting.

An alternative specification for the evolution process for the factor matrix is a matrix autoregressive process (MAR).[1] Yu et al. (2024) introduce a model with an one-lag MAR for the factor matrix. Given the different focus of this paper, we leave the MAR specification for the factor evolution process as a topic for future research.

The other important feature of model (1)-(2) is the Kronecker structure of the covariance of the vectorized error, vec($\mathbf{E}_t$). This Kronecker structure offers flexibility and straightforward interpretation. Firstly, it allows cross-sectional correlation in the idiosyncratic components. For this reason, model (1) is not an exact factor model but an approximate factor model. Approximate factor models are proposed by Chamberlain and Rothschild (1983) for asset pricing. Later, approximate factor models have also been proven to be useful in macroeconomics (see, e.g., Forni et al., 2001; Stock and Watson, 2005).

Additionally, the Kronecker structure in the idiosyncratic components separates the contemporaneous correlations by columns and rows. In particular, for any row, the conditional covariance is $\text{Cov}(\mathbf{Y}'_{i,.,t}|\mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{r,i,i} \mathbf{\Sigma}_c$, for $i = 1, \ldots, n$. Similarly, for any column, the conditional covariance is $\text{Cov}(\mathbf{Y}_{.,j,t}|\mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{c,j,j} \mathbf{\Sigma}_r$, $j = 1, \ldots, k$. In summary, $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$ represent the row-wise and column-wise covariances, respectively, which are not explained by the common components.

[1] For a detailed discussion about matrix autoregression, see e.g., Hoff (2015), Chen et al. (2021) and Chan and Qi (2024).

Moreover, the Kronecker structure facilitates computation. Particularly, with the Kronecker structure, when combined with a natural conjugate prior on the loading matrices ($\mathbf{A}$ and $\mathbf{B}$) and the row and column covariance matrices ($\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$), the conditional posteriors of $(\mathbf{A}, \mathbf{\Sigma}_r)$ and $(\mathbf{B}, \mathbf{\Sigma}_c)$ follow normal-inverse-Wishart distributions, allowing for efficient sampling. The details will be elaborated in Section 2.4.

The specification of the idiosyncratic components is flexible also because we can incorporate time-varying volatility and outlier adjustments. Particularly, the latent variable ω_t can accommodate various time-varying volatility models, as illustrated in Section 2.1.2. For a homoskedastic setting, we can assume $\omega_t = 1$ for $t = 1, \ldots, T$.

Model (1)-(2) can be extended into its tensor version, with the factor framework for tensor time series proposed by Chen et al. (2022). A tensor autoregression can be employed for modeling the factor evolution process in substitute of (2) as well.[2]

2.1.2 Time-varying Volatility and Outlier Adjustment

It is very important to allow for time-varying volatility in modeling macroeconomic data in empirical macroeconomics or finance.[3] The conditionally Gaussian framework in (2) can accommodate a variety of stochastic volatility processes. We extend our model by accommodating three popular specifications in the literature: the common stochastic volatility model of Carriero et al. (2016), the explicit outlier component of Stock and Watson (2016), and the t-distributed innovations of Jacquier et al. (2004).

Specification 1. Common stochastic volatility
An important example is the common stochastic volatility model introduced in Carriero et al. (2015). In particular, let $\omega_t = \mathrm{e}^{h_t}$, and assume that the log-volatility h_t follows a stationary AR(1) process with 0 mean:

$$h_t = \phi h_{t-1} + u_t^h, \quad u_t^h \sim \mathcal{N}(0, \sigma_h^2), \tag{3}$$

[2]See Li and Xiao (2021) for the introduction of tensor autoregressive models.

[3]See, for example, the discussions on the importance of incorporating time-varying volatility in vector autoregressions (VARs) by Cross and Poon (2016), Chan and Eisenstat (2018), and Chan (2023), and in factor models by Aguilar and West (2000), Chib et al. (2006), Kastner et al. (2017), and Li and Scharth (2022).

for $t = 2, ..., T$, where it is assumed $|\phi| < 1$ and the initial state h_1 is assumed to have a Gaussian prior: $h_1 \sim \mathcal{N}(0, \sigma_h^2/(1-\phi^2))$. In our example of multinational macroeconomic dataset, the log-volatility h_t can be interpreted as the level of global macroeconomic uncertainty that cannot be explained by the comovement captured by the factor matrix.

Specification 2. The explicit outlier component
Another widely-adopted specification after the COVID-19 pandemic is the explicit outlier component proposed in Stock and Watson (2016). In specific, the outlier indicators enter the model in a scale factor, denoted $\omega_t = o_t^2$. o_t 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 $p_{\mathbf{o}}$, which is assumed to have a beta prior.

Specification 3. Fat-tailed innovations
This specification characterizes the infrequent occurrences of outliers by incorporating a latent variable $\omega_t = q_t^2$, 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 $\mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r$, and degree of freedom l. t-distributions have fatter tails than normal distribution, and thus may provide better fit for data with infrequent occurrences of outliers.

2.1.3 Relations to Vectorized Dynamic Factor Models

A natural competitor to model (1)-(2) is a standard dynamic factor model (DFM) defined as following:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{M}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \\ \mathbf{f}_t &= \mathbf{H}_\rho \mathbf{f}_{t-1} + \boldsymbol{\nu}_t. \end{aligned} \tag{4}$$

where $\mathbf{M}$ is a $nk \times p$ loading matrix, while $\mathbf{f}_t$ is a $p \times 1$ vector of factors. p is the number of factors and we assume $p = p_1 \times p_2$ for better comparison to MDFM. The matrix $\mathbf{H}_\rho$ is a k-dimensional diagonal matrix consisting of autoregressive coefficients for the evolution process of these factors.

In fact, model (1) is a restrained version of (4):

$$\text{vec}(\mathbf{y}_t) = (\mathbf{B} \otimes \mathbf{A})\text{vec}(\mathbf{F}_t) + \text{vec}(\mathbf{E}_t). \tag{5}$$

A key advantage of MDFM in contrast to DFM in (4) is that MDFM has a smaller parameter space. Specifically, in (1), we need to estimate $np_1 + kp_2$ loading parameters, while we need to estimate $nk \times p_1 p_2$ loading coefficients in (4). Therefore, MDFM significantly reduces the dimension of the problem and thus reduces the computational time significantly.

Another competitor to model (1) is the multilevel factor models. A prominent example in macroeconomics is the model proposed by Kose et al. (2003), shown in (6). This model is designed to capture the multilevel business cycles, with global factors representing the international business cycle, regional factors and country-specific factors show their corresponding regional and national business cycles. While this model is useful to monitor the business cycles geographically, it does not allow the data to reveal the underlying factor structure and their interactions between the two different cross-sections endogenously. Moreover, similar to (4), (6) has a larger paramter space than (1).[4]

$$y_{i,t} = b_i^{global} f_t^{global} + b_i^{region} f_{r,t}^{region} + b_i^{country} f_{c,t}^{country} + \varepsilon_{i,t}, \quad i = 1, \ldots, n \times k. \tag{6}$$

While the proposed MDFM provides a flexible framework for matrix-valued time series, it is vital to understand its limitations. Firstly, as mentioned before, model (1) is a special case of a standard DFM with a Kronecker structure in the factor loadings, as shown in (5). This means that if the factor structure implicitly defined in (1) does not exist, then it would be proper to use the standard DFM. To address this problem, He et al. (2023) developed a family of randomised tests to check whether such factor structure exists. Theoretically, information critera or Bayesian model comparison methods can be developed for this purpose. We leave this topic for future research.

Another limitation in our framework is that we assume the elements in factor matrix are i.i.d., while they might be correlated in real world. A more general framework with a q-lag matrix autoregressive process for factor evolution equation would be ideal. However, considering correlation among factors, idiosyncratic components as well as time-varying volatility all together would tremendously increase the complexity of the model. The

[4]We need to estimate $3nk$ loading coefficients in (6). This difference is large even when we have a medium-size matrix. For instance, if we have a panel for the euro area that includes 10 countries, each with 20 series. Assuming reasonably that $p_1 = 3$ and $p_2 = 5$, then we end up needing to estimate 130 loading parameters in (1), compared to 400 in (6).

proposed framework provides a proper balance between flexibility and complexity.

2.2 Identification

Similar to standard DFMs, MDFMs defined in (1)-(2) cannot be identified without further restrictions. For this reason, research in this field focuses on estimating the column space of the factor loadings, as which is uniquely identifiable. This approach proves beneficial when the objective is to group countries (rows) or variables (columns) based on the pattern of the column space of the loading matrices and to make forecasts using the estimates of common components. However, this strategy may pose challenges for the interpretation of the factors.

In DFMs, a commonly imposed set of restrictions is that the factor loading matrix is a lower triangular matrix with ones on the diagonal, accompanied by the assumption that the idiosyncratic error $\text{vec}(\mathbf{E}_t)$ is independent of the latent factors $\text{vec}(\mathbf{F}_t)$. We follow that spirit and propose a set of sufficient identification assumptions for MDFMs. The proofs are in the Appendix A.

Assumption 1 Factors and idiosyncratic errors are independent of each other.

Assumption 2 Factor series are independent of each other. $\mathbf{H}_{\boldsymbol{\rho}_l}$ is diagonal matrix, for $l = 1, \ldots, q$. $\text{Cov}(\mathbf{u}_t) = \mathbf{I}_{p_1 p_2}$.

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

Proposition 1 *Consider the matrix dynamic factor model in* (1) *and* (2)*. Under Assumptions 1-3, the dynamic factors* $\mathbf{F}_t$ *and the loading matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are uniquely identified.*

Note that a variation is that we restrict the diagonal elements of both $\mathbf{A}$ and $\mathbf{B}$ to be ones, while allowing $\text{Cov}(\mathbf{u}_t)$ to be a positive diagonal matrix rather than an identity matrix.

Assumption 4 Factor series are independent of each other. $\mathbf{H}_{\boldsymbol{\rho}_l}$ is diagonal matrix, for $l = 1, \ldots, q$. $\mathrm{Cov}(\mathbf{u}_t)$ is a positive definite diagonal matrix.

Assumption 5 Factor loading matrices, $\mathbf{A}$ and $\mathbf{B}$ are lower-triangular matrices with ones on their diagonal.

Proposition 2 *Consider the matrix dynamic factor model in* (1) *and* (2)*. Under Assumptions 1, 4, and 5, the dynamic factors* $\mathbf{F}_t$ *and the loading matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are uniquely identified.*

Different from the standard DFM, with the structure in the covariance matrix for the vectorized error, shown in (2), the covariances $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$ can only be identified up to scale. That is, there exist a constant $m \in \mathbb{R} \setminus \{0\}$, such that $\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r = \widetilde{\boldsymbol{\Sigma}}_c \otimes \widetilde{\boldsymbol{\Sigma}}_r$, where $\widetilde{\boldsymbol{\Sigma}}_c = m\boldsymbol{\Sigma}_c$ and $\widetilde{\boldsymbol{\Sigma}}_r = m^{-1}\boldsymbol{\Sigma}_r$. To fix the scale, we normalize the $(1,1)$ element of $\boldsymbol{\Sigma}_c$ to be 1.

In Section 2.4, we discuss how to efficiently draw from the posterior given these identification restrictions.

2.3 Priors

We use a natural conjugate prior for the transpose of factor loadings: $\mathbf{A}'$ and $\mathbf{B}'$. In addition, we use inverse-Wishart prior for $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$:

$$\begin{aligned} \boldsymbol{\Sigma}_r &\sim \mathcal{IW}(\nu_r, \mathbf{S}_r), \quad (\mathrm{vec}(\mathbf{A}')|\boldsymbol{\Sigma}_r) \sim \mathcal{N}(\mathrm{vec}(\mathbf{A}_0'), \boldsymbol{\Sigma}_r \otimes \mathbf{V}_{\mathbf{A}'}), \\ \boldsymbol{\Sigma}_c &\sim \mathcal{IW}(\nu_c, \mathbf{S}_c), \quad (\mathrm{vec}(\mathbf{B}')|\boldsymbol{\Sigma}_c) \sim \mathcal{N}(\mathrm{vec}(\mathbf{B}_0'), \boldsymbol{\Sigma}_c \otimes \mathbf{V}_{\mathbf{B}'}). \end{aligned} \tag{7}$$

In practice, we can adopt diagonal matrices for hyperparameter matrices $\mathbf{S}_r$ and $\mathbf{S}_c$, based on the belief that there is no cross-sectional correlation in idiosyncratic components, and let the data reveal whether such correlation exists.

Then we are able to obtain the following joint density function of $(\mathbf{A}', \boldsymbol{\Sigma}_r)$:

$$p(\mathbf{A}', \boldsymbol{\Sigma}_r) \propto |\mathbf{V}_{\mathbf{A}'}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_r|^{-\frac{\nu_r+n+p_1+1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_r^{-1}\mathbf{S}_r)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_r^{-1}(\mathbf{A}'-\mathbf{A}_0')'\mathbf{V}_{\mathbf{A}'}^{-1}(\mathbf{A}'-\mathbf{A}_0')\right)}. \tag{8}$$

Similarly, we obtain the following joint density function of $(\mathbf{B}', \mathbf{\Sigma}_c)$:

$$p(\mathbf{B}', \mathbf{\Sigma}_c) \propto |\mathbf{V}_{\mathbf{B}'}|^{-\frac{k}{2}} |\mathbf{\Sigma}_c|^{-\frac{\nu_c+k+p_2+1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\mathbf{\Sigma}_c^{-1}\mathbf{S}_c)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}_c^{-1}(\mathbf{B}'-\mathbf{B}_0')'\mathbf{V}_{\mathbf{B}'}^{-1}(\mathbf{B}'-\mathbf{B}_0')\right)}. \tag{9}$$

The autoregressive coefficient $\rho_{j,k}$ is assumed to have a truncated normal prior on the interval $(-1, 1)$:

$$\rho_{j,k,l} \sim \mathcal{TN}(\rho_{j,k,l,0}, V_{\rho_{j,k,l}}), \quad j = 1, ..., p_1, \quad k = 1, ..., p_2, \quad l = 1, \ldots, q.$$

The prior variance $\lambda_{j,k}^2$ is assumed to have a inverse-gamma prior: $\mathcal{IG}(\nu_{\lambda_{j,k}}, S_{\lambda_{j,k}})$. We also treat the first q values of $\mathbf{F}_t$ as unknown, and use the following prior

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

2.4 Bayesian Estimation

From the model described in (1)-(2), we obtain the following likelihood function:

$$\begin{aligned} p(\mathbf{Y}|\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r, \boldsymbol{\omega}) =& (2\pi)^{-\frac{Tnk}{2}} |\mathbf{\Sigma}_c|^{-\frac{Tn}{2}} |\mathbf{\Sigma}_r|^{-\frac{Tk}{2}} \prod_{t=1}^{T} \omega_t^{-\frac{nk}{2}} \times \\ & \prod_{t=1}^{T} \mathrm{e}^{-\frac{1}{2\omega_t}\mathrm{tr}\left(\mathbf{\Sigma}_c^{-1}(\mathbf{Y}_t-\mathbf{A}\mathbf{F}_t\mathbf{B}')'\mathbf{\Sigma}_r^{-1}(\mathbf{Y}_t-\mathbf{A}\mathbf{F}_t\mathbf{B}')\right)}. \end{aligned} \tag{10}$$

With the natural conjugate prior for the matrices of factor loadings, we can make full use of the Kronecker structure in idiosyncratic components and achieve efficient computation. To sample these loadings with identification restrictions outlined in Section 2.2, we adopt the approaches proposed by Cong et al. (2004). To sample the covariance matrix $\mathbf{\Sigma}_c$ with the first element fixed at 1, we adopt the algorithm proposed by Nobile (2000).

Specifically, posterior draws can be obtained by sequentially sampling from: (1) $p(\mathbf{A}', \mathbf{\Sigma}_r|\mathbf{Y}, \mathbf{B}, \mathbf{F}, \mathbf{\Sigma}_c)$; (2) $p(\mathbf{B}', \mathbf{\Sigma}_c|\mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r)$; (3) $p(\mathrm{vec}(\mathbf{F}_t)|\mathbf{Y}_t, \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c, \boldsymbol{\omega}^2, \boldsymbol{\rho})$, $t = 1, \ldots, T$; (4) $p(\lambda_{j,k}^2|\mathbf{f}_{j,k}, \rho_{j,k})$, $j = 1, ..., p_1, k = 1, ..., p_2$; (5) $p(\boldsymbol{\rho}_{j,k}|\mathbf{f}_{j,k}, \lambda_{j,k}^2)$, $j = 1, ..., p_1, k = 1, ..., p_2$; (6) $p(\omega_t|\mathbf{A}, \mathbf{F}_t, \mathbf{B}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r)$, $t = 1, \ldots, T$. Sampling the stochastic

volatility and outlier adjustments specified in Section 2.1.2 (Step 6) is typically easy as they amount to fitting a univariate time-series model. Therefore, we put this part in Appendix B. In the following, we provide details on implementing Step 1-5.

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

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

With the constraints for identification, we cannot directly sample from the above normal-inverse-Wishart distribution. Here we outline the sampling scheme for $\mathbf{A}'$ with the structure constraints. To that end, we first represent the restrictions as a system of linear restrictions. For example, for $\mathbf{A}'$, we represent the restrictions that $\mathbf{A}$ is a lower triangular matrix with ones on the diagonal using $\mathbf{M}_{\mathbf{A}'} \text{vec}(\mathbf{A}') = \mathbf{a}_0$. Assuming $n > p_1$, $\mathbf{M}_{\mathbf{A}'}$ is a $p_1(p_1+1)/2 \times np_1$ selection matrix, and $\mathbf{a}_0$ is a $p_1(p_1+1)/2 \times 1$ vector consisting of ones and zeros. Then we apply Algorithm 2 in Cong et al. (2004) 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 $\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}'_u) + \mathbf{V}'(\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}'_u))$.

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

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

where

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

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

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 $\mathbf{\Delta}$ 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 $\boldsymbol{L}$
   - δ_{ij} equal to $\mathcal{N}(0,1)$ random variates, $i > j$.
(3) Set $\mathbf{\Sigma}_c = (\boldsymbol{L}^{-1})'(\mathbf{\Delta}^{-1})'\mathbf{\Delta}^{-1}\boldsymbol{L}^{-1}$.
(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.

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

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

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

Step 4. Sampling from $(\lambda_{j,k}^2|\mathbf{f}_{j,k}, \boldsymbol{\rho}_{j,k}), \quad j = 1, ..., p_1, k = 1, ..., p_2$
It is clear that $(\lambda_{j,k}^2|\mathbf{f}_{j,k}, \boldsymbol{\rho}_{j,k}) \sim \mathcal{IG}(\widehat{\nu}_{\lambda_{j,k}}, \widehat{S}_{\lambda_{j,k}})$, where $\widehat{\nu}_{\lambda_{j,k}} = \nu_{\lambda_{j,k}} + \frac{T}{2}$, and $\widehat{S}_{\lambda_{j,k}} = S_{\lambda_{j,k}} + \frac{1}{2}\left[\sum_{t=1}^{q} f_{j,k,t}^2(1 - \sum_m \rho_{j,k,m}^2) + \sum_{t=q+1}^{T}(f_{j,k,t} - \rho_{j,k,1}f_{j,k,t-1} - ... - \rho_{j,k,q}f_{j,k,q})^2\right]$.

Step 5. Sampling from $(\boldsymbol{\rho}_{j,k}|\mathbf{f}_{j,k}, \lambda_{j,k}^2), \quad j = 1, ..., p_1, k = 1, ..., p_2$
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 (2) as follows:

$$\widetilde{\mathbf{f}}_{j,k} = \widetilde{\mathbf{F}}_{j,k}\boldsymbol{\rho}_{j,k} + \mathbf{u}_{j,k}, \quad \mathbf{u}_{j,k} \sim \mathcal{N}(\mathbf{0}, \lambda_{j,k}\mathbf{I}_{T-q}), \tag{11}$$

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

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

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

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

2.5 Extension: Row-wise and Column-wise Stochastic Volatility

A more flexible way to model time-varying volatility is to incorporate multiple stochastic volatility processes, as first proposed by Cogley and Sargent (2005) in a vector autoregression setting. To that end, we assume the idiosyncratic component has a different covariance matrix as follows:

$$\text{vec}(\mathbf{E}_t) \sim \mathcal{N}\left(\mathbf{0}, (\mathbf{D}_{c,t}\mathbf{\Sigma}_c) \otimes (\mathbf{D}_{r,t}\mathbf{\Sigma}_r)\right), \tag{12}$$

where $\mathbf{D}_{c,t}$, $\mathbf{D}_{r,t}$ are diagonal matrices for stochastic volatility, and $\mathbf{\Sigma}_c$ and $\mathbf{\Sigma}_r$ are diagonal covariance matrices. In specific, $\mathbf{D}_{c,t} = \text{diag}(\omega_{c,1,t}, \ldots, \omega_{c,p_2,t})$, $\mathbf{D}_{r,t} = \text{diag}(\omega_{r,1,t}, \ldots, \omega_{r,p_1,t})$, $\mathbf{\Sigma}_c = \text{diag}(\sigma^2_{c,1,1}, \ldots, \sigma^2_{c,p_2,p_2})$ and $\mathbf{\Sigma}_r = \text{diag}(\sigma^2_{r,1,1}, \ldots, \sigma^2_{r,p_1,p_1})$. The log-volatility follows a stationary AR(1) process with 0 mean similar to (3).

Specifically, denote the column log-volatility as $h_{c,j,t} \equiv \log(\omega_{c,j,t})$ for $j = 1, \ldots p_1$ and $h_{r,j,t} \equiv \log(\omega_{r,j,t})$ for $j = 1, \ldots, p_1$, then we assume

$$\begin{aligned} h_{c,j,t} &= \phi_{c,j} h_{c,j-1,t} + u_{c,t}, \quad u_{c,t} \sim \mathcal{N}(0, \sigma^2_{h,c}), \\ h_{r,j,t} &= \phi_{r,j} h_{r,j-1,t} + u_{r,t}, \quad u_{r,t} \sim \mathcal{N}(0, \sigma^2_{h,r}), \end{aligned} \tag{13}$$

for $t = 2, \ldots, T$, where we assume $|\phi_{c,j}| < 1$ and $|\phi_{r,j}| < 1$. The initial states are assumed to follow Gaussian priors.

Compared to the specification in (2), this extension offers greater flexibility in modeling stochastic volatility. This is particularly relevant in practical settings where different time series are likely to exhibit distinct volatility dynamics, making a uniform treatment across series potentially restrictive. However, this flexibility comes at the cost of imposing stricter assumptions of the cross-sectional independence of idiosyncratic errors. In many empirical applications, such assumptions may be too rigid, as correlations between errors are common in both financial and macroeconomic data.[5] It would be valuable for future research to develop a statistical approach for making such decisions within the context of a matrix factor model.

[5]See, e.g., Chamberlain and Rothschild (1983) and Fan et al. (2015) for financial data, and Stock and Watson (2005) and Barigozzi and Luciani (2017) for macroeconomic data.

3 Selecting the Dimension of the Factor Matrix

Determining the numbers of factors is an important and challenging problem. For the standard DFM, several methods have been proposed to select the number of factors, including the information criterion (see, e.g., Bai and Ng, 2002; Hallin and Liška, 2007; Amengual and Watson, 2007), the random matrix theory method (Onatski, 2010), the ratio-based method (see, e.g., Lam and Yao, 2012; Ahn and Horenstein, 2013), and the white noise testing approach (Gao and Tsay, 2022). However, these methods cannot be directly applied to the MDFM.

To date, methods used for matrix factor models include the ratio-based method of Wang et al. (2019), the sequential testing method of He et al. (2023), and the diagonal-path method of Gao and Tsay (2023). Nevertheless, to our knowledge, no method has yet been proposed for a matrix factor model with dynamic factors.

In the Bayesian domain, methods to select the number of factors normally involve the computation of marginal likelihoods under models with different numbers of factors. This approach, however, is criticized for its intensive computational demands, particularly in high-dimensional settings. To address these challenges, several alternative methods have been proposed. For example, Lopes and West (2004) introduced a reversible jump MCMC algorithm to account for uncertainty in the number of factors, while Lee and Song (2002) developed a path sampling approach. Additionally, Bhattacharya and Dunson (2011) and Lee et al. (2022) inferred the number of factors by zeroing out a subset of the loading elements using Bayesian variable selection priors.

In this paper, we adopt the Bayesian approach of estimating marginal likelihoods to select the dimension of the factor matrix. To mitigate computational challenges, we employ an importance-sampling technique proposed by Chan and Eisenstat (2015), which is based on a cross-entropy approach. This technique has two major advantages. First, using the importance-sampling density is convenient as it generates independent draws instead of correlated MCMC draws. Second, it is straightforward to obtain an importance-sampling density that is able to produce an efficient estimator. Specifically, by minimizing the Kullback-Leibler divergence to the posterior distribution, they find that given a parametric density family, the optimal parameters are the maximum likelihood estimators if the posterior samples are treated as observed data. Since the importance-sampling den-

sity is close to the posterior distribution, this marginal likelihood estimator is efficient, exhibiting low variance. Typically, the number of draws needed is a few thousands or less.

To further streamline computation and reduce the estimator's variance, we integrate out the factors from the likelihood function. While a closed-form expression for the integrated likelihood is available, directly computing the inverse of the covariance matrix in high-dimensional datasets is computationally prohibitive. Therefore, we employ the Kalman filter to efficiently integrate out the factors. Further details are provided in Appendix C.

The next step is to find the maximum likelihood estimators for the importance density. In our case, the importance density is denoted as

$$\begin{aligned} f(\boldsymbol{\theta};\mathbf{v}) = f(\mathbf{A},\mathbf{B},\boldsymbol{\Sigma}_r,\boldsymbol{\Sigma}_c,\boldsymbol{\rho},\boldsymbol{\lambda};\mathbf{v}) \\ =f(\mathbf{A};\overline{\mathbf{A}},\overline{\mathbf{D}}_{\mathbf{A}})\cdot f(\mathbf{B};\overline{\mathbf{B}},\overline{\mathbf{D}}_{\mathbf{B}})\cdot f(\boldsymbol{\Sigma}_r;\nu_r,\Psi_r)\cdot \\ f(\boldsymbol{\Sigma}_c;\nu_c,\Psi_c)\cdot f(\boldsymbol{\rho};\overline{\boldsymbol{\rho}},\overline{\mathbf{D}}_{\boldsymbol{\rho}})\cdot f(\boldsymbol{\lambda};\nu_\lambda,S_\lambda). \end{aligned} \tag{14}$$

For the parametric family, we use Gaussian densities for $f(\mathbf{A};\overline{\mathbf{A}},\overline{\mathbf{D}}_{\mathbf{A}})$, and $f(\mathbf{B};\overline{\mathbf{B}},\overline{\mathbf{D}}_{\mathbf{B}})$ where $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ are the means, while $\overline{\mathbf{D}}_{\mathbf{A}}$ and $\overline{\mathbf{D}}_{\mathbf{B}}$ are the covariance matrices. We use inverse Wishart densities for $f(\boldsymbol{\Sigma}_c;\nu_c,\Psi_c)$ and $f(\boldsymbol{\Sigma}_r;\nu_r,\Psi_r)$. The truncated normal density on the interval $(-1,1)$ is used for $f(\boldsymbol{\rho};\overline{\boldsymbol{\rho}},\overline{\mathbf{D}}_{\boldsymbol{\rho}})$, where $\overline{\boldsymbol{\rho}}$ and $\overline{\mathbf{D}}_{\boldsymbol{\rho}}$ are the mean and covariance matrix. Finally, we use inverse-gamma distributions for $f(\boldsymbol{\lambda};\nu_\lambda,S_\lambda)$.

To derive the maximum likelihood estimators for the parameters of the inverse-Wishart distribution, we first estimate the scale matrix. This estimate is then substituted into the likelihood function, and a Newton-type method is applied to estimate the degrees of freedom. Detailed procedures are provided in Appendix C. The maximum likelihood estimators for the parameters of the normal and inverse-gamma distributions are straightforward to compute, so we omit the details here.

4 Monte Carlo Studies

In this section, we demonstrate the performance of the MDFM using Monte Carlo studies. Specifically, we assess the accuracy of the factor estimates by comparing them to their true

values across datasets of varying sizes. Additionally, we evaluate whether the marginal likelihood estimator can accurately identify the true dimensions of the factor matrices.

The data are generated according to (1) and (2) with $q = 1$. The parameters are drawn as follows: the free parameters in $\mathbf{A}$ and $\mathbf{B}$ are sampled from $\mathcal{U}(0, 1)$, $\rho_{j,k} \sim \mathcal{U}(0.8, 0.9)$. We set $\mathbf{\Sigma}_c$ to $0.3\mathbf{I}_k$, $\mathbf{\Sigma}_r$ to $0.5\mathbf{I}_n$, and $\lambda_{j,k}^2$ to 1 for $j = 1, \ldots, p_1$, $k = 1, \ldots, p_2$.

4.1 Performance of Factor Estimators under Different Sample Sizes and Dimensions of Factor Matrices

To assess the accuracy of our factor estimates, we consider sample sizes, we consider sample sizes $(n, k) \in \{(10, 10), (20, 15), (30, 20)\}$ and observation lengths $T \in \{200, 500, 1000\}$. The factor matrices are preset to dimensions $(p_1, p_2) = (3, 2)$ or $(p_1, p_2) = (5, 5)$.

For models with smaller factor matrix ($p_1 = 3$ and $p_2 = 2$), we use a Gibbs sampling chain of 10,000 iterations after 5,000 burn-in draws. For larger factor matrix orders ($p_1 = 5, p_2 = 5$), we extend the sampling chain to 20,000 iterations after 10,000 burn-in draws.[6] We calculate the posterior mean as the point estimate for the factors and compare these to the true factors. Specifically, we project the true factors onto the estimates to obtain adjusted R^2 values for each factor series in the factor matrix.

Figure 1-2 represent the adjusted R^2 for $(p_1, p_2) = (3, 2)$ and $(p_1, p_2) = (5, 5)$, respectively. Each row corresponds to a different sample size. For example, the first row represents $(n, k) = (10, 10)$, while the second row represents $(20, 15)$. Each column represents a different length of observations with the first column corresponding to $T = 200$. Each color block represents the adjusted R^2 for a specific element in the factor matrix. For instance, the upper-left block of the first subplot in Figure 1 corresponds to the adjusted R^2 from regressing the true value of $f_{1,1,.}$ on the estimates $\widehat{f}_{1,1,.}$, for $(n, k, T) = (10, 10, 200)$.

The color intensity in these figures reflects the magnitude of the adjusted R^2; darker

[6] We found that for $p_1 = 3, p_2 = 2$, convergence is typically achieved within 5,000 burn-in draws, even with initial factor values drawn randomly from a standard normal distribution. However, when the dimension of the factor matrix is large ($p_1 = p_2 = 5$), setting proper initial values is crucial to shorten the Markov chain. Estimates from a standard DFM (1,000 posterior draws after 1,000 burn-in draws) work well as initial values. Geweke statistics are computed to ensure the convergence of Markov chains.

colors indicate higher values. For better visualization, the minimum of the color axis is set to 0.9, as the smallest adjusted R^2 we have obtained is 0.91. More details on the adjusted R^2s are provided in Appendix D.

Overall, our factor estimates closely match the true values. Moreover, a comparison across columns reveals that larger observation lengths T yield better estimates. A comparison across rows shows that larger sample sizes lead to more accurate estimates. Comparing the two figures, it is evident that smaller factor matrix dimensions result in better estimates. These findings suggest that increasing the number of observations and sample size improves the accuracy of factor estimates. This is in line with the theory in standard DFM and static matrix factor model.[7]

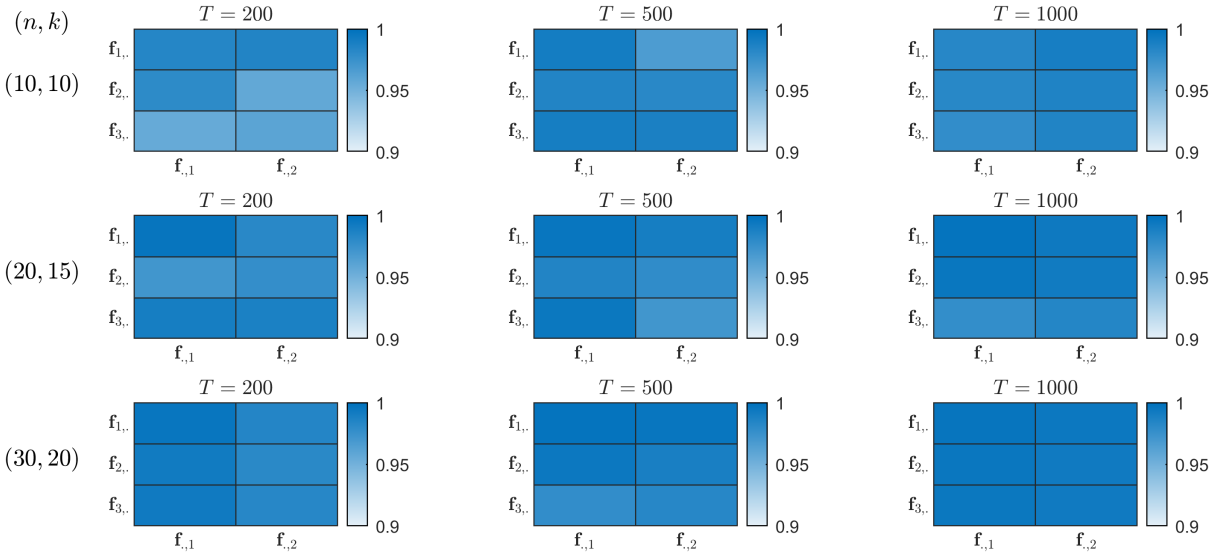

Figure 1: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 3$, $p_2 = 2$

[7]See, e.g., Bai (2003) for inferential theory in vectorized factor model and Chen and Fan (2023) for that in static matrix factor model.

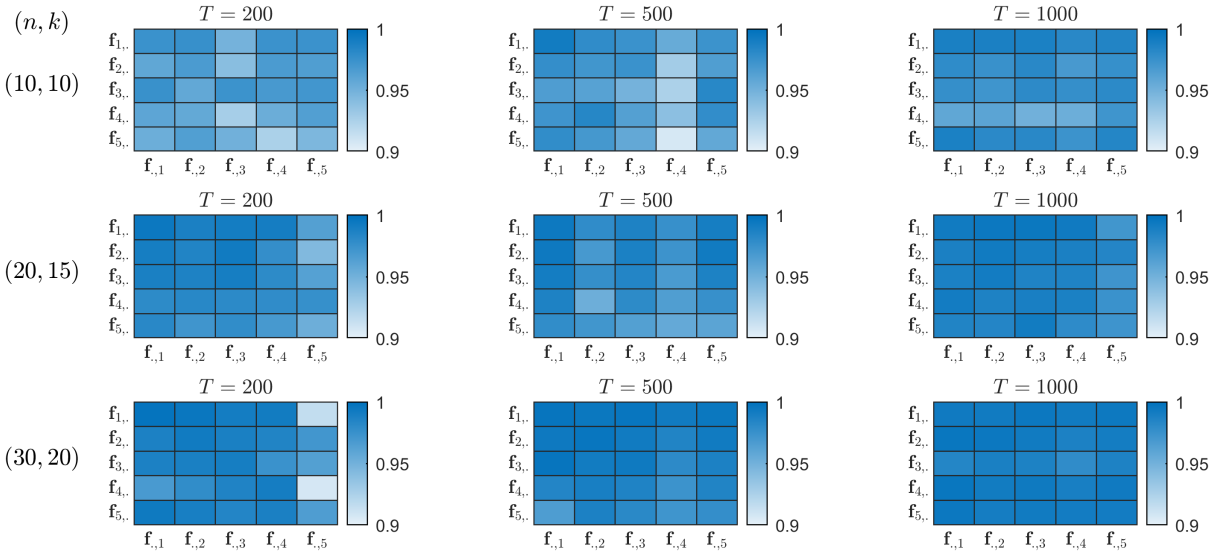

Figure 2: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 5$, $p_2 = 5$

4.2 Performance of the Marginal Likelihood Estimator for Model Selection

To evaluate the performance of the marginal likelihood estimator in correctly identifying the true order of the factor matrix, we estimate log marginal likelihoods for models with a range of orders. Specifically, we use four datasets from Section 4.1:

Dataset 1: $n = 10$, $k = 10$, $T = 500$; true order: $p_1 = 3$, $p_2 = 2$.
Dataset 2: $n = 20$, $k = 15$, $T = 500$; true order: $p_1 = 3$, $p_2 = 2$.
Dataset 3: $n = 10$, $k = 10$, $T = 500$; true order: $p_1 = 5$, $p_2 = 5$.
Dataset 4: $n = 20$, $k = 15$, $T = 500$; true order: $p_1 = 5$, $p_2 = 5$.

For datasets with a true dimension of factor matrix: $(p_1, p_2) = (3, 2)$, we estimate models with p_1 and p_2 ranging from 1 to 5. For datasets with a true dimension of $(5, 5)$, we estimate models with p_1 and p_2 ranging from 3 to 7.

Figures 3 and 4 presents the estimates for log marginal likelihoods when the true dimension of the factor matrix is $(p_1, p_2) = (3, 2)$ and $(p_1, p_2) = (5, 5)$, respectively.

Two key findings are noteworthy. First, in all the four datasets, the estimates correctly identify the true order; that is, the estimates are the largest when (p_1, p_2) are set to their

true values. Second, the log marginal likelihood estimates exhibit a consistent pattern. Before the true order is reached, the estimates increase monotonically, reflecting an improving model fit. After reaching the true order, the estimates decrease monotonically, indicating that additional factors do not contribute significantly to the model fit and may introduce overfitting. For example, when true order is $(p_1, p_2) = (3, 2)$, the sequence $\log \widehat{p}(\mathbf{y} \,|\, p_1 = 1) < \log \widehat{p}(\mathbf{y} \,|\, p_1 = 2) < \log \widehat{p}(\mathbf{y} \,|\, p_1 = 3)$ and $\log \widehat{p}(\mathbf{y} \,|\, p_1 = 3) > \log \widehat{p}(\mathbf{y} \,|\, p_1 = 4) > \log \widehat{p}(\mathbf{y} \mid p_1 = 5)$ is observed. Similarly, $\log \widehat{p}(\mathbf{y} \mid p_2 = 1) < \log \widehat{p}(\mathbf{y} \mid p_2 = 2)$ and $\log \widehat{p}(\mathbf{y} \mid p_2 = 2) > \log \widehat{p}(\mathbf{y} \mid p_2 = 3) > \log \widehat{p}(\mathbf{y} \mid p_2 = 4) > \log \widehat{p}(\mathbf{y} \mid p_2 = 5)$.

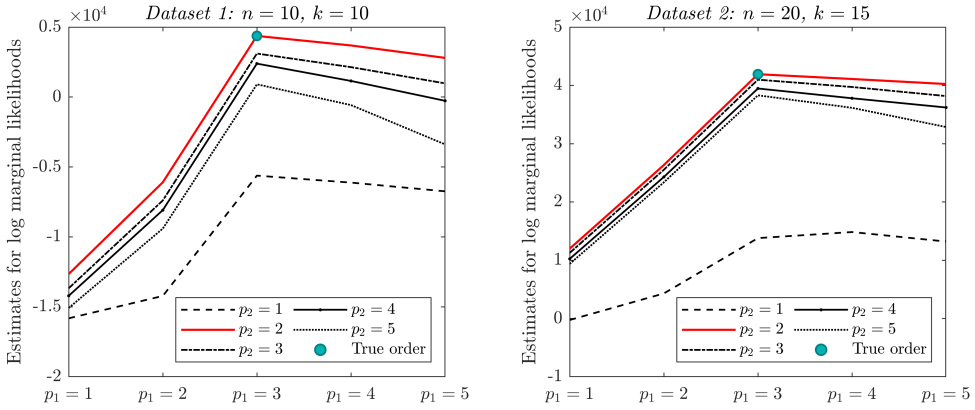

Figure 3: Estimates for log marginal likelihoods when true order of the factor matrix is $p_1 = 3, p_2 = 2$

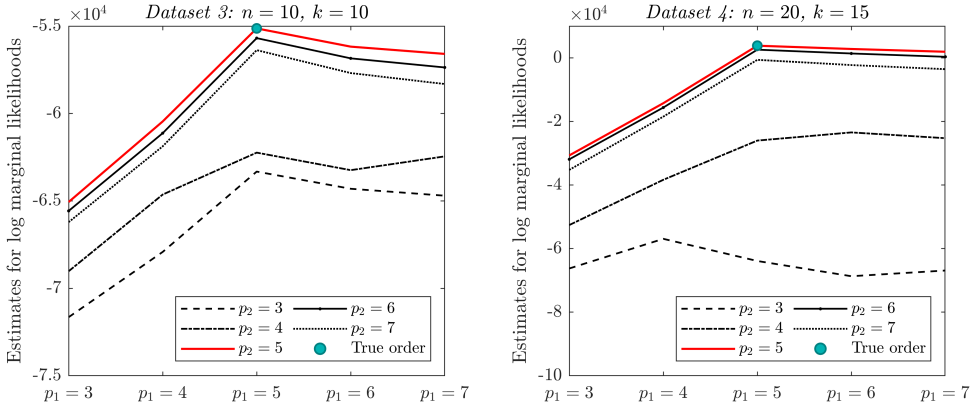

Figure 4: Estimates for log marginal likelihoods when true order of the factor matrix is $p_1 = 5, p_2 = 5$

5 Empirical Application

In this section, we demonstrate the usefulness of matrix dynamic factor model with two applications. In the first application, we use a multinational macroeconomic panel, while in the second application, we use the Fama-French 10×10 panel. The dimensions of factor matrices are determined using the proposed marginal likelihood estimator.

5.1 Multinational Macroeconomic Panel

We apply the matrix dynamic factor model to the macroeconomic panel constructed from OECD database. The dataset comprises 10 quarterly indicators of 19 countries from 1995.Q1 to 2023.Q3 for 115 quarters. The countries include developed economies from North America, Europe, Asia and Oceania. The indicators include real GDP, price indices, labor unit cost, unemployment, international trade and household consumption. Each time series is adjusted for stationarity through first differencing or logarithmic differencing, and standardized by demeaning and dividing by their standard deviations. Detailed descriptions of the dataset and transformation methods are provided in the Appendix E.

Figure 5 shows the transformed time series of macroeconomic indicators of multiple countries. While theoretically one could compute marginal likelihood estimates to assess different orders of countries and variables, this approach is computationally intensive. Instead, we prioritize the economic significance of entities and the relationships among variables. As shown in Figure 5, the first column contains the US variables, due to its status as the largest economy and a global economic bellwether. The UK follows, due to its significant position in the European economy. Australia is placed next, given its critical influence in the Oceania region. Among the indicators, real GDP is prioritized as it serves as the most comprehensive measure of economic activity. Headline CPI follows, given its importance as a key inflation indicator closely tied to monetary policy decisions. Labor unit cost, positioned third, is crucial for insights into productivity and competitiveness.

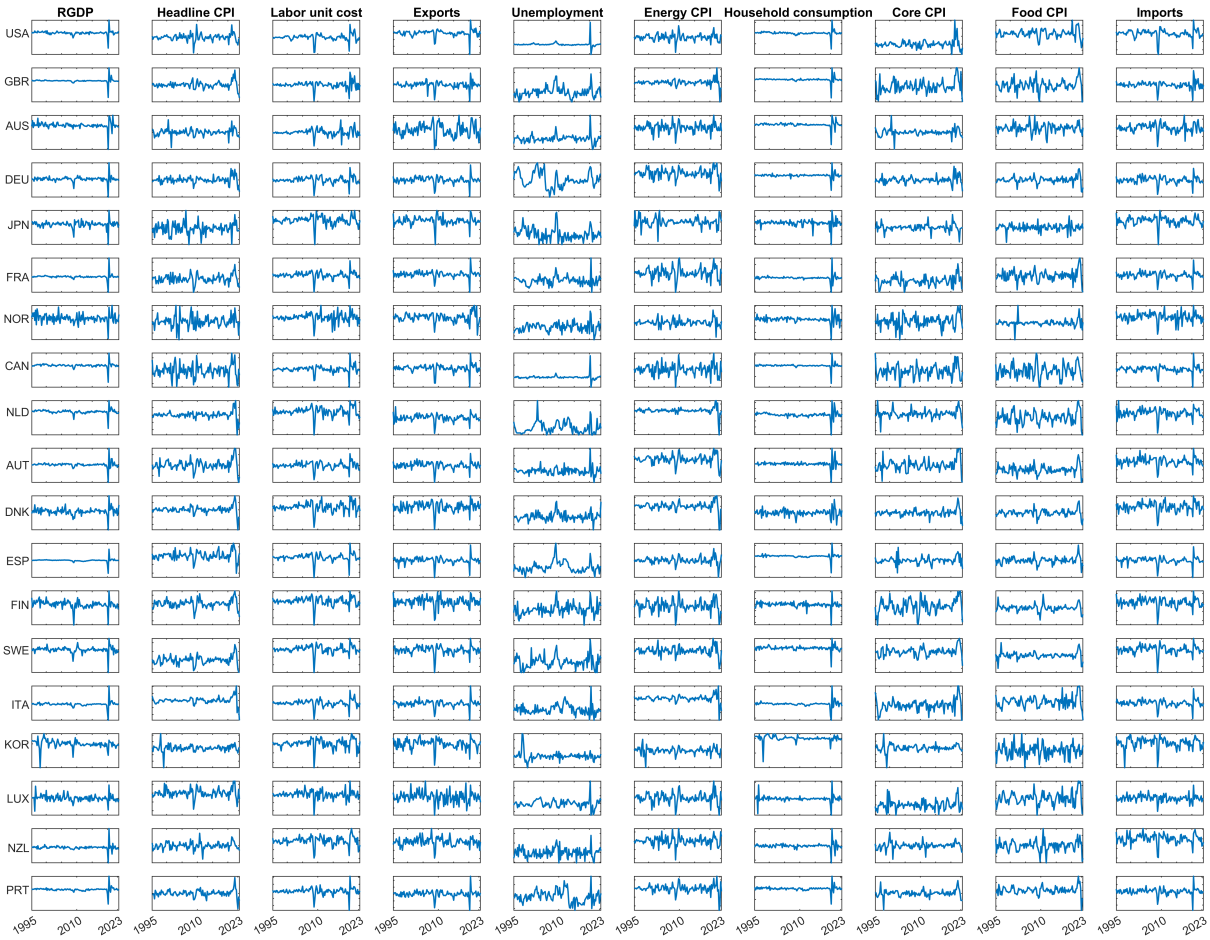

Figure 5: The ten macroeconomic indicators (by columns) for 19 countries (by rows). The horizontal axis represents time and the vertical axis represents the standardized growth rates. The ranges of the vertical values are not the same. The order of countries and indicators is the order adopted in the estimation.

We then employ our marginal likelihood estimation to determine the factor matrix dimensions. As shown in Table 1, the highest log marginal likelihood is achieved with $(p_1, p_2) = (1, 2)$. This implies that a one-factor structure for the country dimension and a two-factor structure for the indicator dimension.

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)

The latent structure of the global macroeconomy can be interpreted through the estimated row and column factor loading matrices. We sort these estimates and compute the posterior probabilities that the differences between neighboring values are greater than 0. By comparing the posterior probabilities to 0.9, we can group countries and indicators accordingly.

Figure 6 displays the bar plot of sorted estimates for $\widehat{\mathbf{A}}$.[8] The 19 countries are categorized into three groups: Japan in the first group, all European, Oceanian countries and Korea in the second, and the two North American countries in the third. These results suggest that while geographic factors influence the grouping, they are not the sole determinant, as evidenced by the integration of Oceanian and European countries into a single group.

[8]Since the factor matrix has only one row, $\mathbf{A}$ is effectively a 19×1 vector. However, we retain the notation $\mathbf{A}$ for consistency.

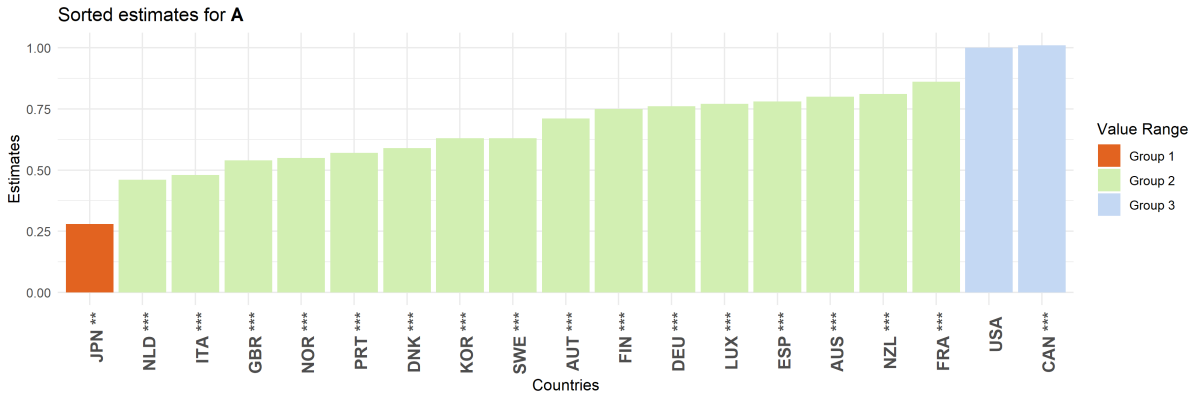

Figure 6: Bar plots of sorted estimates for loading matrix $\mathbf{A}$. The 19 countries are categorized into 3 distinct groups based on the posterior probabilities that the differences between neighboring values are greater than 0. The stars on the country labels show the significance level of the corresponding estimates. There is no significance level on USA because we fix the corresponding element in $\mathbf{A}$ to be 1.

Figure 7 contains two rows of subplots, four in total. The first row presents bar plots of sorted estimates for $\widehat{\mathbf{B}}$, while the second row shows factor estimates and their 90% credible intervals. Notably, the first factor impacts all indicators, likely representing a broad economic cycle. This factor clearly captures the 2008 Great Recession and the disruptions caused by the COVID-19 pandemic, though the early 2000s recession is less pronounced in the corresponding factor estimates.

The second factor influences all the indicators except for real GDP, starting with headline CPI. It might capture price dynamics or inflationary pressures, but do not directly affect the overall output as measured by real GDP. Figure 8 compares the four-quarter moving average of the second factor estimates with the moving average of growth rates of Brent crude oil price. The comovement between these series is evident, particularly during periods of significant events such as the 2002-2003 Iraq war and civil unrest in Venezuela, the 2008-2009 Great Recession and OPEC's production cuts, the 2014-2016 oil price collapse and the 2022 Russian invasion of Ukraine.

Based on the first column of the factor loading matrix ($\mathbf{b}_1$), the ten variables are grouped into four distinct clusters: Group 1 (unemployment), Group 2 (food CPI and core CPI), Group 3 (real GDP, energy CPI, headline CPI, and household consumption), and Group 4 (imports, labor unit cost, and exports). The second column of the factor loading

matrix ($\mathbf{b}_2$) organizes the variables into a different set of four clusters: Group 1 (core CPI, household consumption, food CPI, and unemployment), Group 2 (imports, labor unit cost, and exports), Group 3 (headline CPI), and Group 4 (energy CPI). The first column factor has stronger positive impacts on output and international trade, with less influence on prices and consumption. It negatively influences unemployment. In contrast, the second column factor has stronger positive impacts on prices, with less effect on productivity and international trade. It has statistically insignificant positive effects on core CPI and negative effects on consumption, food CPI, and unemployment.

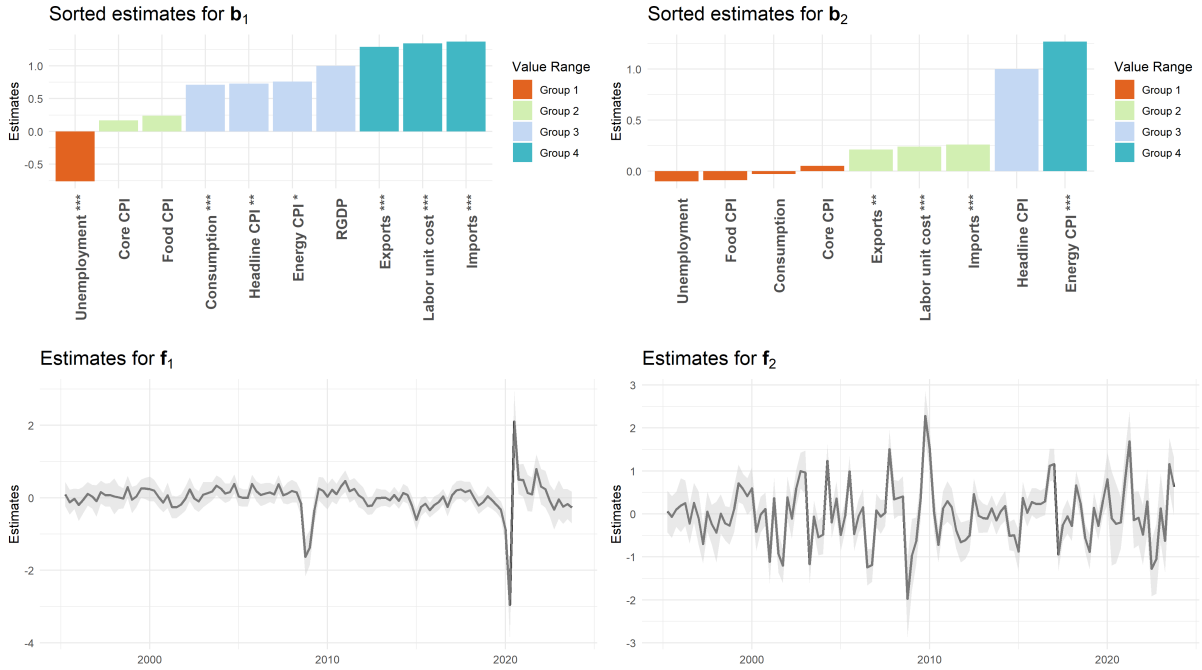

Figure 7: Bar plots of sorted estimates for loading matrix $\mathbf{B}$ and plots for factor estimates. The stars show the significance level of the corresponding estimates. According to impacts of the two column factors, the variables can be divided into 4 groups. The shaded area of plots for the factors is the 90% credible intervals.

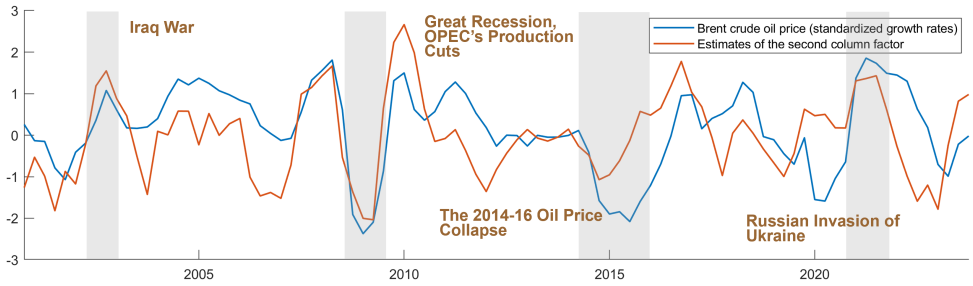

Figure 8: Yearly moving average of standardized growth rates of Brent crude oil price and the second column factor estimates.

Figure 9 presents the estimates for stochastic volatility. The shaded area represents the standard deviation. The high volatility around 1997 reflects the turbulence of the Asian financial crisis, particularly in Japan and Korea. Expectedly, increased volatility is also observed during the Great Recession and the COVID-19 pandemic.

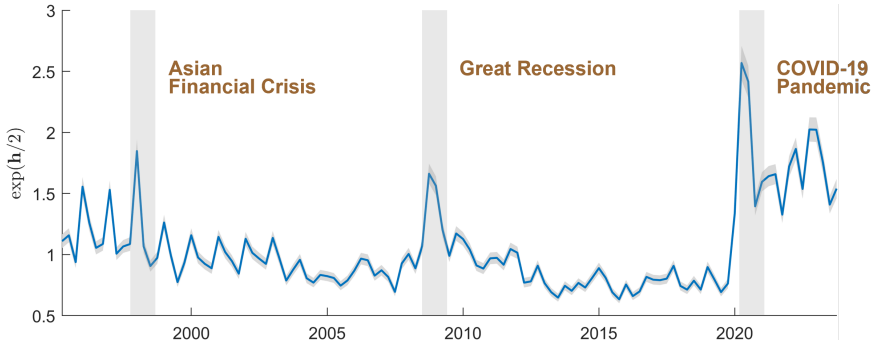

Figure 9: Estimates for stochastic volatility: $\widehat{\omega}_t = \exp(\widehat{h}_t/2)$

Figure 10-11 are heatmaps of estimates for column-wise and row-wise covariance matrix in the idiosyncratic component. From Figure 10, it is obvious that headline CPI is positively correlated with its disaggregated components: energy CPI, core CPI and food CPI. This is in line with the conclusion in Stock and Watson (2005). In addition, unemployment is negatively correlated with real GDP, labor unit cost, consumption, core CPI and imports. Labor unit cost is positively correlated to exports and imports.

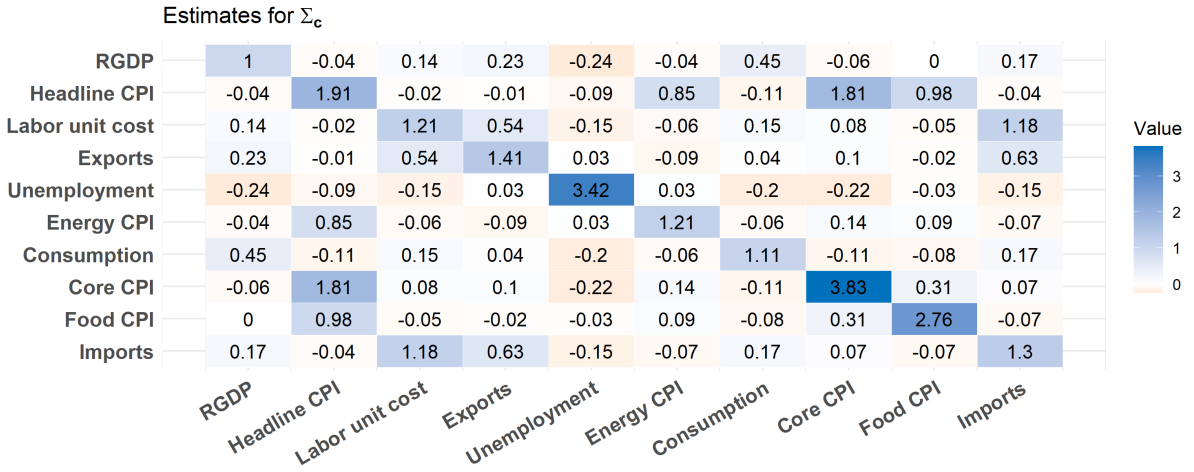

Figure 10: Heatmap of estimates for $\mathbf{\Sigma}_c$

In Figure 11, we can see that idiosyncratic risks for countries in European Union are correlated, including Germany, France, Norway, Netherlands, Austria, Denmark, Spain, Finland, Sweden, Luxembourg, Italy and Portugal. UK is weakly correlated to EU as well.

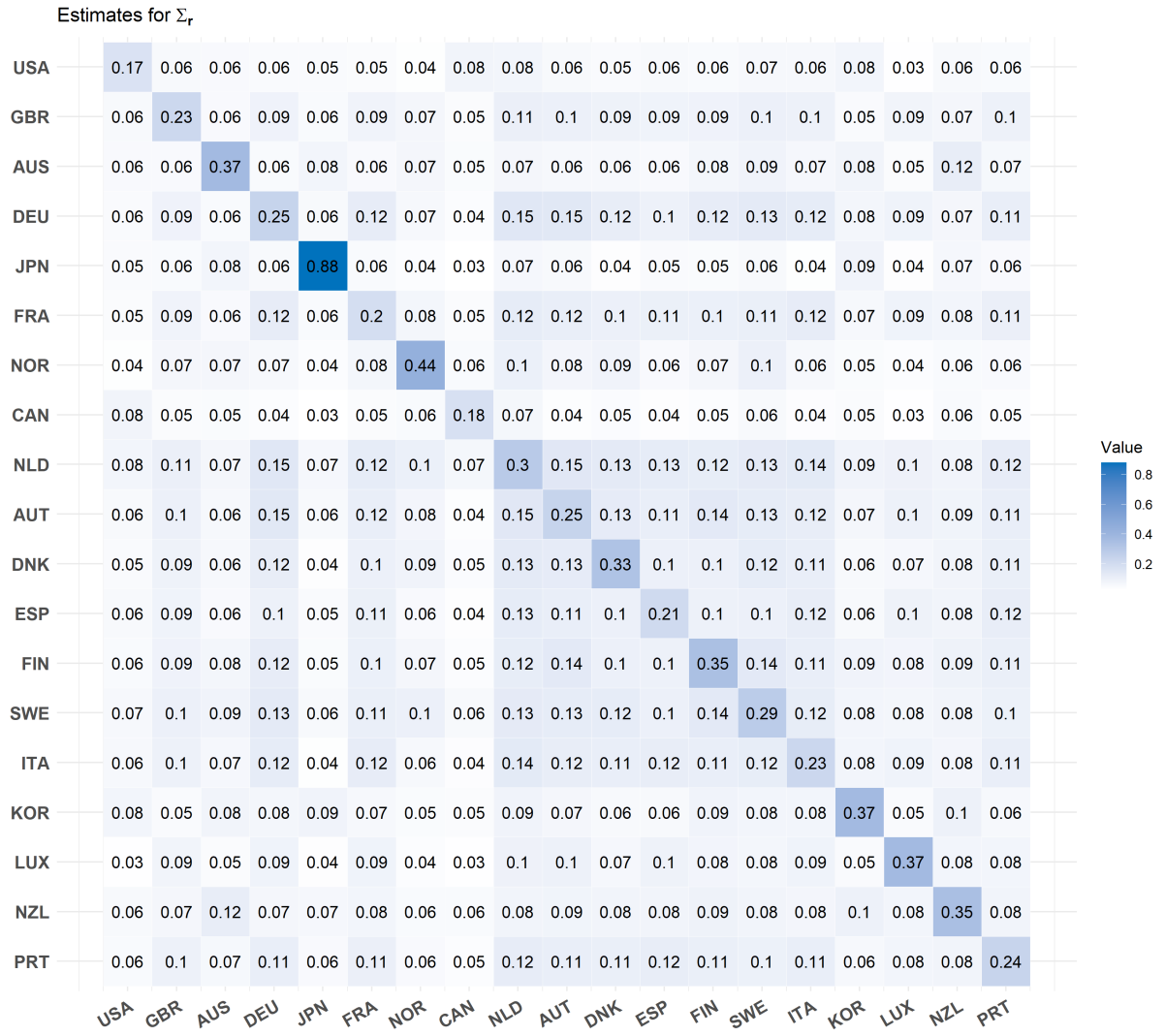

Figure 11: Heatmap of estimates for $\mathbf{\Sigma}_r$

5.2 Fama-French 10 × 10 Panel

In this application, we investigate the dynamic matrix factor model for the Fama-French return series, which was studied by Wang et al. (2019), Yu et al. (2022) and He et al. (2024). The data include monthly returns of 100 portfolios, structured in a 10 by 10 matrix according to ten levels of sizes (market equity) and ten levels of ratio of book

equity to market equity (BE/ME).[9] The return series span from January 1990 to June 2024 (414 observations).[10] Following Chang et al. (2023), we impute the missing values by the weighted averages of the three previous months, i.e., set $y_{i,j,t} = 0.5y_{i,j,t-1} + 0.3y_{i,j,t-2} + 0.2y_{i,j,t-3}$ for missing $y_{i,j,t}$.

To account for market conditions, we follow Wang et al. (2019), Yu et al. (2022), and He et al. (2024) and subtract the monthly excess market return from each series. We then standardize the data by subtracting the mean and dividing by the standard deviation.

Figure 12 shows the standardized market-adjusted return series of the portfolios. The rows in Figure 12 correspond to the ten levels of sizes and the columns represent the ten levels of the BE/ME ratios. Note that the ranges of the vertical values for these subplots are not the same.

To simplify the interpretation, we reorder the size and book-to-market ratio in the data matrix as shown in Figure 12. For the rows, we put small-cap portfolios (SMALL) first, followed by medium-cap (ME5), and then large-cap portfolios (BIG). This ordering reflects the typical characteristics of these portfolios: small-cap stocks generally exhibit higher volatility and risk. By positioning small-cap portfolios at the top, the matrix emphasizes the highest risk factor first. Medium-cap stocks, which have moderate risk and return profiles, are placed second, serving as a transitional category between small and large caps. Large-cap stocks, which are more stable with lower risk, are placed the third. The lower triangular structure in the row loading matrix $\mathbf{A}$ suggests that the factors influence the portfolios in a hierarchical manner. The factor loading on the first (smallest) portfolios influences all portfolios, the second factor loading affects only those after the second, and so on.

While one might consider ordering the sizes from the smallest to the largest directly, it would offer less distinct differentiation of the factors. In addition, our arrangement also enhances the efficiency of model comparison. To determine the optimal dimension of the factor matrix, we estimate the marginal likelihoods for models with increasing dimensions, explained in Section 3.[11] If the data favor lower risk factors, the model can quickly capture this, owing to the larger size difference between the first three rows in

[9]The data is available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[10]We do not include data earlier than 1990 because there are many missing values in the early years.

[11]We increase p_1 from 1 until the marginal likelihood estimates no longer increase.

the rearranged matrix.

Similarly, for the columns, we prioritize high book-to-market ratio (HiBM) first, followed by medium (BM5) and then low book-to-market (LoBM) ratios. The rationale for this ordering is analogous to that used for the rows, so further details are omitted here.

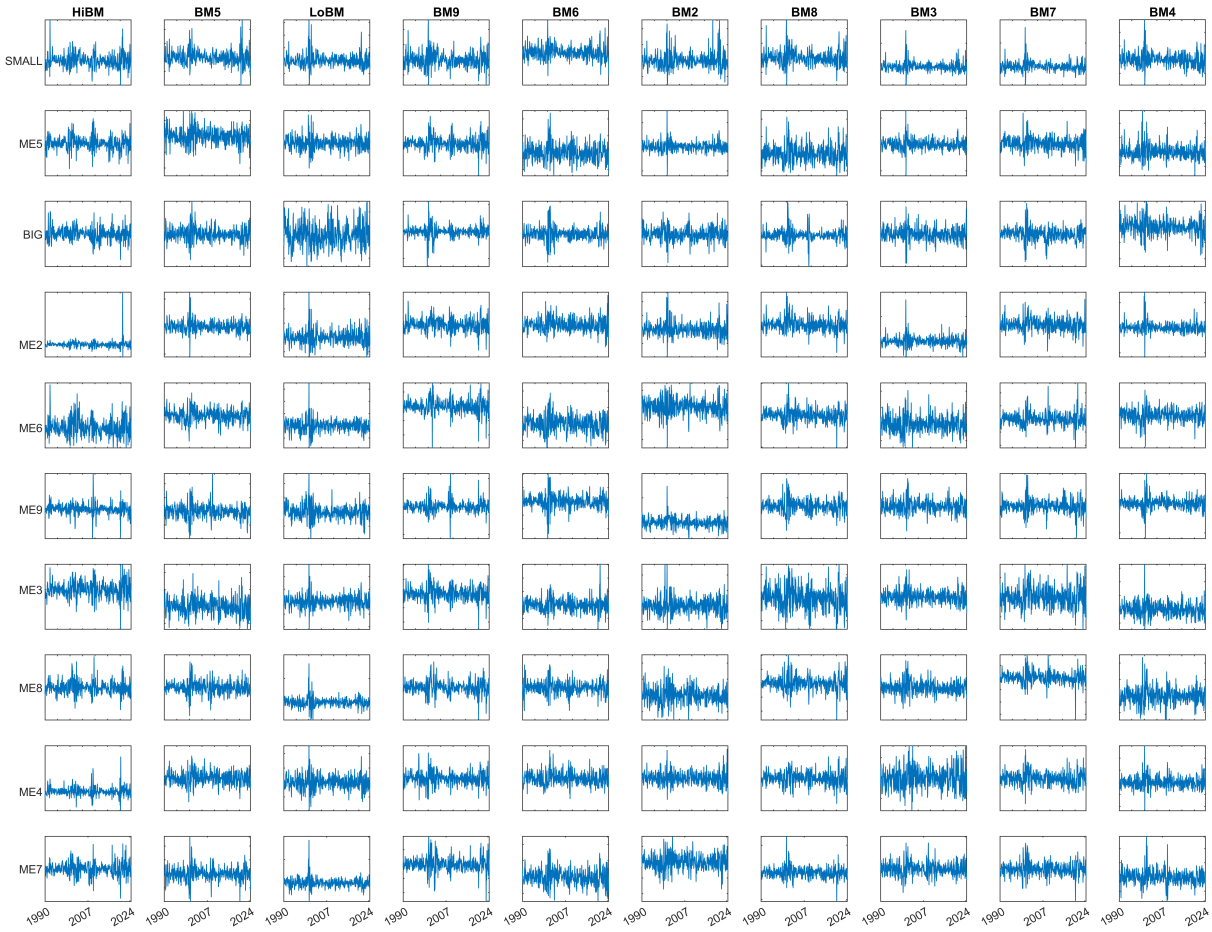

Figure 12: The return series of the portfolios constructed on different levels of sizes (rows) and book equity to market equity ratio (columns). Note that we have rearranged the order of rows and columns. The horizontal axis represents time and vertical axis represents the standardized monthly returns. The ranges of the vertical values are not the same.

Table 2 shows the estimates for log marginal likelihoods and their standard deviations. The marginal likelihood estimates suggest that $(p_1, p_2) = (2, 3)$. This indicates that a two-factor structure for size and a three-factor structure for book-to-market is more favored by the data.

Table 2: Estimates for log marginal likelihoods

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

Figure 13 shows the estimates for loading matrix $\mathbf{A}$. In specific, the two subplots correspond to $\widehat{\mathbf{A}}_{.,1}$ (left) and $\widehat{\mathbf{A}}_{.,2}$ (right). The left subplot clearly demonstrates that the high-risk factor exerts a strong influence on small portfolios, with its impact gradually decreasing as portfolio size increases, eventually turning negative for large-size portfolios. In contrast, the right subplot reveals that the moderate risk factor predominantly affects medium-size portfolios, with its influence tapering off as the portfolio size shifts either smaller or larger.

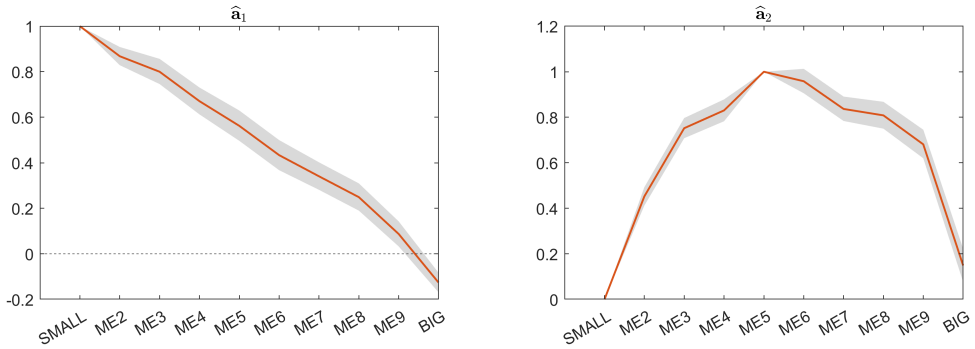

Figure 13: Estimates for row loadings: $\widehat{\mathbf{A}}$

Figure 14 shows the estimates for the three columns of loading matrix $\mathbf{B}$. The left subplot is the loadings associated with the high book-to-market ratio factor. It increases steadily from portfolios with low book-to-market ratios to those with high ratios. This aligns with the idea that value stocks (high book-to-market ratio) are more sensitive to this factor. The second factor appears to have the greatest influence on medium book-to-market

portfolios (BM5 to BM7), with diminishing effects on portfolios as the book-to-market-ratio either decreases or increases. The right subplot shows a reversed-check pattern of the loadings associated with the low book-to-market ratio factor. This implies that growth stocks (low book-to-market ratio) are more sensitive to this factor.

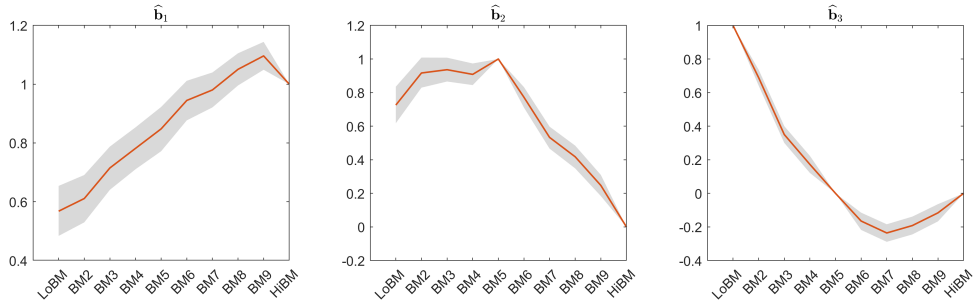

Figure 14: Estimates for column loadings: $\widehat{\mathbf{B}}$

Figure 15 shows estimated posterior densities, histograms of posterior draws, priors, as well as the posterior estimates of autoregressive coefficients ($\boldsymbol{\rho}$) for the factor evolution process. All the six posterior densities have little mass on value 0, and the posterior estimates are around 0.2 or 0.3. This indicates that the factors show mild but significant serial correlations.

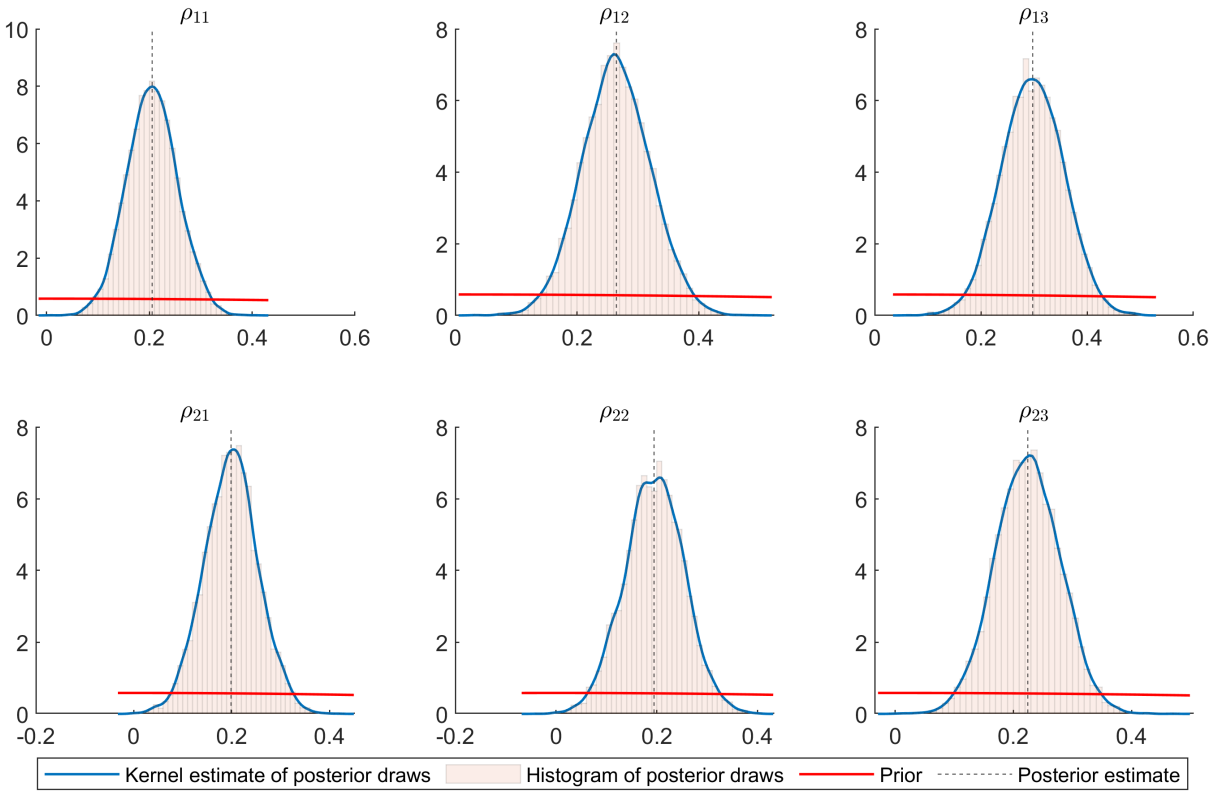

Figure 15: Posterior densities, histograms of posterior draws, priors and posterior estimates for autogressive coefficients $\boldsymbol{\rho}$

A correlation analysis for the factor estimates shows that there exists contemporary correlation among factors. Table 3 presents the correlation coefficients along with their significance level. The correlations between $\mathbf{f}_{1,2}$ and $\mathbf{f}_{1,3}$, $\mathbf{f}_{1,2}$ and $\mathbf{f}_{2,3}$, and $\mathbf{f}_{1,3}$ and $\mathbf{f}_{2,3}$ are not weak and statistically significant. This suggests that relaxing the independence restrictions among the factors may be beneficial. Additionally, a principal component analysis of the six factor series indicates that five principal components account for 96% of the variation among the six factors, implying some redundancy and the potential for further model simplification.

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

Figure 16 shows the estimates and standard deviations of stochastic volatility for stock returns over time. Clearly, the volatility of stock returns exhibits considerable variation throughout the observed period. Notably, the volatility peaks around February 2000, just one month before the onset of the dot-com bubble burst. Additionally, significant spikes in volatility are observed during the 2008 financial crisis and the COVID-19 pandemic.

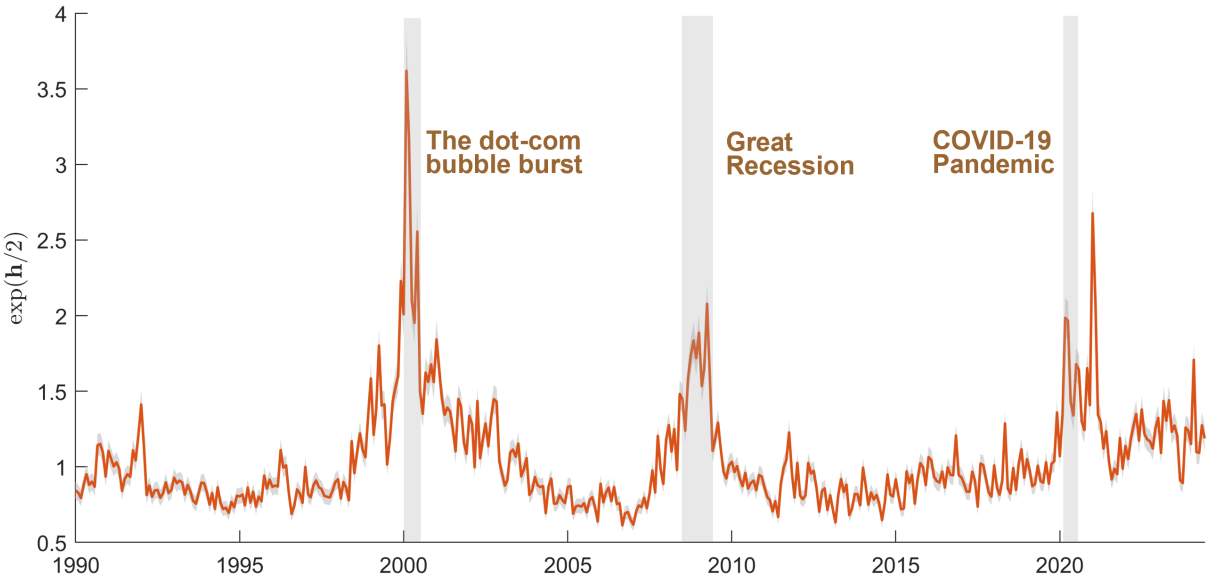

Figure 16: Estimates and standard errors of stochastic volatility: $\exp(\mathbf{h}/2)$

Compared to the macroeconomic application, here the correlation in idiosyncratic component is not significant, as shown in Figure 17-18. This implies that conditional on the common components, there is no significant contemporary correlations among these

portfolios.

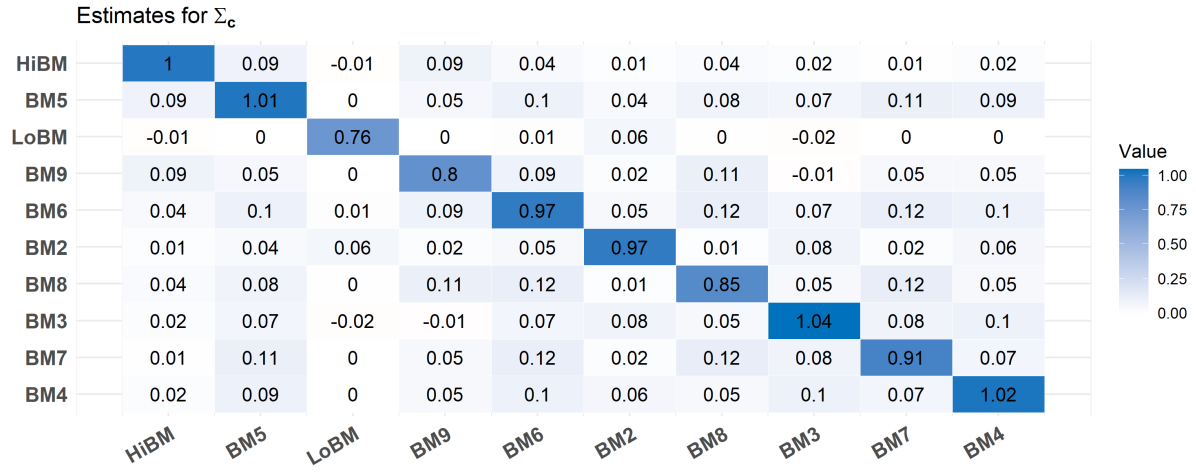

Figure 17: Heatmap of estimates for $\mathbf{\Sigma}_c$

Figure 18: Heatmap of estimates for $\mathbf{\Sigma}_r$

6 Conclusion and Future Research

In this paper we propose a new series of dynamic factor models designed for high-dimensional matrix-valued time series, incorporating both time-varying volatility and cross-sectional correlation in the idiosyncratic components. We develop a MCMC method for Bayesian estimation and introduce an importance-sampling estimator for marginal

likelihood to determine the dimension of the factor matrices through Bayesian model comparison.

To illustrate the usefulness of our model, we conduct Monte Carlo simulations to evaluate the properties of the factor estimates and the performance of the marginal likelihood estimator in correctly identify the true dimensions of the factor matrices. The application of our model to macroeconomic and Fama-French panels demonstrates its capability to unveil interesting features within high-dimensional time series.

Our research can be extended in two promising directions. First, the application of our model to macroeconomics and financial economics holds potential for uncovering the spillover effects of macroeconomic or technological shocks, examining trade networks, providing insights into asset pricing models, and improving forecasting accuracy.

In addition, matrix-valued time series can be viewed as a specific example of tensor time series. Future research should extend our model to dynamic tensor factor models and explore the possibilities offered by even higher-dimensional data structures.

References

Aguilar, O. and M. West (2000): "Bayesian dynamic factor models and portfolio allocation," *Journal of Business & Economic Statistics*, 18, 338–357.

Ahn, S. C. and A. R. Horenstein (2013): "Eigenvalue ratio test for the number of factors," *Econometrica*, 81, 1203–1227.

Alessi, L. and M. Kerssenfischer (2019): "The response of asset prices to monetary policy shocks: stronger than thought," *Journal of Applied Econometrics*, 34, 661–672.

Amengual, D. and M. W. Watson (2007): "Consistent estimation of the number of dynamic factors in a large N and T panel," *Journal of Business & Economic Statistics*, 25, 91–96.

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

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

Bai, J. and P. Wang (2015): "Identification and Bayesian estimation of dynamic factor models," *Journal of Business & Economic Statistics*, 33, 221–240.

Barigozzi, M. and M. Luciani (2017): "Common factors, trends, and cycles in large datasets," *arXiv preprint arXiv:1709.01445*.

Bhattacharya, A. and D. B. Dunson (2011): "Sparse Bayesian infinite factor models," *Biometrika*, 98, 291–306.

Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (2018): "Macroeconomic nowcasting and forecasting with big data," *Annual Review of Economics*, 10, 615–643.

Carriero, A., T. E. Clark, and M. Marcellino (2015): "Bayesian VARs: specification choices and forecast accuracy," *Journal of Applied Econometrics*, 30, 46–73.

——— (2016): "Common drifting volatility in large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390.

——— (2024a): "Capturing Macro-Economic Tail Risks with Bayesian Vector Autoregressions," *Journal of Money, Credit and Banking*, 56, 1099–1127.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2024b): "Addressing COVID-19 outliers in BVARs with stochastic volatility," *Review of Economics and Statistics*, 1–15.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica: Journal of the Econometric Society*, 1281–1304.

Chan, J. C. (2017): "The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling," *Journal of Business & Economic Statistics*, 35, 17–28.

——— (2023): "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*, 235, 1419–1446.

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

——— (2018): "Bayesian model comparison for time-varying parameter VARs with stochastic volatility," *Journal of applied econometrics*, 33, 509–532.

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*.

Chang, J., J. He, L. Yang, and Q. Yao (2023): "Modelling matrix time series via a tensor CP-decomposition," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85, 127–148.

Chen, B., E. Y. Chen, S. Bolivar, and R. Chen (2024): "Time-Varying Matrix Factor Models," *arXiv preprint arXiv:2404.01546*.

Chen, E. Y. and J. Fan (2023): "Statistical inference for high-dimensional matrix-variate factor models," *Journal of the American Statistical Association*, 118, 1038–1055.

Chen, E. Y., R. S. Tsay, and R. Chen (2020): "Constrained factor models for high-dimensional matrix-variate time series," *Journal of the American Statistical Association*.

Chen, R., H. Xiao, and D. Yang (2021): "Autoregressive models for matrix-valued time series," *Journal of Econometrics*, 222, 539–560.

Chen, R., D. Yang, and C.-H. Zhang (2022): "Factor models for high-dimensional tensor time series," *Journal of the American Statistical Association*, 117, 94–116.

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

Chib, S., F. Nardari, and N. Shephard (2006): "Analysis of high dimensional multivariate stochastic volatility models," *Journal of Econometrics*, 134, 341–371.

Cogley, T. and T. J. Sargent (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic dynamics*, 8, 262–302.

Cong, Y., B. Chen, and M. Zhou (2004): "Fast Simulation of Hyperplane-Truncated Multivariate Normal Distributions," *Bayesian Analysis*, 1.

Cross, J. and A. Poon (2016): "Forecasting structural change and fat-tailed events in Australian macroeconomic variables," *Economic Modelling*, 58, 34–51.

Fan, J., Y. Liao, and J. Yao (2015): "Power enhancement in high-dimensional cross-sectional tests," *Econometrica*, 83, 1497–1541.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001): "The Generalized Factor Model: One-Sided Estimation and Forecast," Tech. rep., mimeo.

Gao, Z. and R. S. Tsay (2022): "Modeling high-dimensional time series: A factor model with dynamically dependent factors and diverging eigenvalues," *Journal of the American Statistical Association*, 117, 1398–1414.

——— (2023): "A two-way transformed factor model for matrix-variate time series," *Econometrics and Statistics*, 27, 83–101.

Giannone, D., L. Reichlin, and L. Sala (2004): "Monetary policy in real time," *NBER macroeconomics annual*, 19, 161–200.

Hallin, M. and R. Liška (2007): "Determining the number of factors in the general dynamic factor model," *Journal of the American Statistical Association*, 102, 603–617.

He, Y., X. Kong, L. Trapani, and L. Yu (2023): "One-way or two-way factor model for matrix sequences?" *Journal of Econometrics*, 235, 1981–2004.

He, Y., X. Kong, L. Yu, X. Zhang, and C. Zhao (2024): "Matrix factor analysis: From least squares to iterative projection," *Journal of Business & Economic Statistics*, 42, 322–334.

Hoff, P. D. (2015): "Multilinear tensor regression for longitudinal relational data," *The annals of applied statistics*, 9, 1169.

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.

Kastner, G., S. Frühwirth-Schnatter, and H. F. Lopes (2017): "Efficient Bayesian inference for multivariate factor stochastic volatility models," *Journal of Computational and Graphical Statistics*, 26, 905–917.

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.

Lam, C. and Q. Yao (2012): "Factor modeling for high-dimensional time series: inference for the number of factors," *The Annals of Statistics*, 694–726.

Lee, J., S. Jo, and J. Lee (2022): "Robust sparse Bayesian infinite factor models," *Computational Statistics*, 37, 2693–2715.

Lee, S.-Y. and X.-Y. Song (2002): "Bayesian selection on the number of factors in a factor analysis model," *Behaviormetrika*, 29, 23–39.

Lenza, M. and G. E. Primiceri (2022): "How to estimate a vector autoregression after March 2020," *Journal of Applied Econometrics*, 37, 688–699.

Li, M. and M. Scharth (2022): "Leverage, asymmetry, and heavy tails in the high-dimensional factor stochastic volatility model," *Journal of Business & Economic Statistics*, 40, 285–301.

Li, Z. and H. Xiao (2021): "Multi-linear tensor autoregressive models," *arXiv preprint arXiv:2110.00928*.

Liu, X. and E. Chen (2019): "Helping effects against curse of dimensionality in threshold factor models for matrix time series," *arXiv preprint arXiv:1904.07383*.

Lopes, H. F. and M. West (2004): "Bayesian model assessment in factor analysis," *Statistica Sinica*, 41–67.

Marcellino, M., M. Porqueddu, and F. Venditti (2016): "Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility," *Journal of Business & Economic Statistics*, 34, 118–127.

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

Onatski, A. (2010): "Determining the number of factors from empirical distribution of eigenvalues," *The Review of Economics and Statistics*, 92, 1004–1016.

Poncela, P., E. Ruiz, and K. Miranda (2021): "Factor extraction using Kalman filter and smoothing: This is not just another survey," *International Journal of Forecasting*, 37, 1399–1425.

Sargent, T. J., C. A. Sims, et al. (1977): "Business cycle modeling without pretending to have too much a priori economic theory," *New methods in business cycle research*, 1, 145–168.

Stock, J. H. and M. W. Watson (2005): "Implications of dynamic factor models for VAR analysis," .

——— (2012): "Dynamic Factor Models," *The Oxford Handbook of Economic Forecasting*.

——— (2016): "Core inflation and trend inflation," *Review of Economics and Statistics*, 98, 770–784.

Thorsrud, L. A. (2020): "Words are the new numbers: A newsy coincident index of the business cycle," *Journal of Business & Economic Statistics*, 38, 393–409.

Wang, D., X. Liu, and R. Chen (2019): "Factor models for matrix-valued high-dimensional time series," *Journal of econometrics*, 208, 231–248.

Yu, L., Y. He, X. Kong, and X. Zhang (2022): "Projected estimation for large-dimensional matrix factor models," *Journal of Econometrics*, 229, 201–217.

Yu, R., R. Chen, H. Xiao, and Y. Han (2024): "Dynamic matrix factor models for high dimensional time series," *arXiv preprint arXiv:2407.05624*.

Yuan, C., Z. Gao, X. He, W. Huang, and J. Guo (2023): "Two-way dynamic factor models for high-dimensional matrix-valued time series," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85, 1517–1537.

A Proof of Propositions

A.1 Proof of *Proposition 1*

Proof of Proposition 1: Without the loss of generality, we prove one of the two cases in Proposition 1. That is, we assume $\text{Var}(\mathbf{u}_t) = \mathbf{I}_{p_1 p_2}$ and $\mathbf{A}$ is a lower-triangular matrix with ones on the diagonal, while $\mathbf{B}$ is a lower-triangular matrix with strictly positive diagonal elements.

As shown in (A.15), we may identify a rotation of $\mathbf{F}_t$, given by $\mathbf{C}\mathbf{F}_t\mathbf{D}'$.

$$\mathbf{Y}_t = \mathbf{A}\mathbf{C}^{-1}\mathbf{C}\mathbf{F}_t\mathbf{D}'(\mathbf{D}')^{-1}\mathbf{B}' + \mathbf{E}_t, \tag{A.15}$$

where $\mathbf{C}$ and $\mathbf{D}$ are $p_1 \times p_1$ and $p_2 \times p_2$ invertible matrices.

We use $\widetilde{\mathbf{F}}_t$ to denote the rotated factor matrix: $\widetilde{\mathbf{F}}_t \equiv \mathbf{C}\mathbf{F}_t\mathbf{D}'$, and we use $\widetilde{\mathbf{f}}_t$ to denote the vectorized $\mathbf{F}_t$: $\widetilde{\mathbf{f}}_t \equiv (\mathbf{D} \otimes \mathbf{C})\mathbf{f}_t$.

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & ... & 0 \\ a_{21} & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p_1 1} & a_{p_1 2} & ... & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & ... & a_{np_1} \end{bmatrix}, \mathbf{C}^{-1} = \begin{bmatrix} c_{11} & ... & c_{1p_1} \\ \vdots & \ddots & \vdots \\ c_{p_1 1} & ... & c_{p_1 p_1} \end{bmatrix}$$

Then the rotated factor loadings $\mathbf{A}\mathbf{C}^{-1}$ needs to be a lower triangular matrix with ones on the diagonal as well, that is,

$$\begin{bmatrix} 1 & 0 & ... & 0 \\ a_{21} & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p_1 1} & a_{p_1 2} & ... & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & ... & a_{np_1} \end{bmatrix} \begin{bmatrix} c_{11} & ... & c_{1p_1} \\ \vdots & \ddots & \vdots \\ c_{p_1 1} & ... & c_{p_1 p_1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & ... & 0 \\ a^*_{21} & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a^*_{p_1 1} & a^*_{p_1 2} & ... & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a^*_{n1} & a^*_{n2} & ... & a^*_{np_1} \end{bmatrix} \tag{A.16}$$

For (A.16) to hold, we must have $c_{i,j} = 0$ for any i, j such that $i < j$ and $c_{i,i} = 1$, or $\mathbf{C}^{-1}$ is lower triangular with ones on the diagonal.

Similarly,

$$\begin{bmatrix} b_{11} & 0 & ... & 0 \\ b_{21} & b_{22} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{p_2 1} & b_{p_2 2} & ... & b_{p_2 p_2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & ... & b_{np_2} \end{bmatrix} \begin{bmatrix} d_{11} & ... & d_{1p_2} \\ \vdots & \ddots & \vdots \\ d_{p_2 1} & ... & d_{p_2 p_2} \end{bmatrix} = \begin{bmatrix} b^*_{11} & 0 & ... & 0 \\ b^*_{21} & b^*_{22} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b^*_{p_1 1} & b^*_{p_1 2} & ... & b^*_{p_2 p_2} \\ \vdots & \vdots & \ddots & \vdots \\ b^*_{n1} & b^*_{n2} & ... & b^*_{np_2} \end{bmatrix} \tag{A.17}$$

For (A.17) to hold, we must have $d_{ij} = 0$ for any i, j such that $i < j$, or $\mathbf{D}^{-1}$ is lower triangular given the assumption that $b_{ii} \neq 0$, $b^*_{ii} \neq 0$, for $i = 1, ..., p_2$.

Define $\mathbf{f}_t \equiv \text{vec}(\mathbf{F}_t)$. Consider the case $q = 1$, we rewrite (2) as follows

$$\mathbf{f}_t = \mathbf{H}_{\boldsymbol{\rho}} \mathbf{f}_{t-1} + \mathbf{u}_t, \tag{A.18}$$

where $\mathbf{H}_{\boldsymbol{\rho}}$ is a diagonal matrix with $\boldsymbol{\rho} = (\rho_{1,1,t}, ..., \rho_{p_1,p_2,t})'$ on the diagonal. $\mathbf{u}_t = (u_{1,1,t}, ..., u_{p_1,p_2,t})'$, $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \Lambda_t)$, where $\Lambda_1 = \text{diag}(\lambda^2_{1,1}/(1-\rho^2_{1,1}), ..., \lambda^2_{p_1,p_2}/(1-\rho^2_{p_1,p_2}))$ for $t = 1$, and $\Lambda_t = \text{diag}(\lambda^2_{1,1}, ..., \lambda^2_{p_1,p_2})$ for $t = 2, ..., T$.

Define $\mathbf{M} \equiv \mathbf{D} \otimes \mathbf{C}$, multiply (A.18) by $\mathbf{M}$ on both side, we have

$$\mathbf{M}\mathbf{f}_t = \mathbf{M}\mathbf{H}_{\boldsymbol{\rho}} \mathbf{f}_{t-1} + \mathbf{M}\mathbf{u}_t. \tag{A.19}$$

Therefore

$$\widetilde{\mathbf{f}}_t = \mathbf{M}\mathbf{H}_{\boldsymbol{\rho}} \mathbf{M}^{-1} \widetilde{\mathbf{f}}_{t-1} + \mathbf{M}\mathbf{u}_t. \tag{A.20}$$

The observation equation after the rotation becomes

$$\mathbf{Y}_t = \mathbf{A}\mathbf{C}^{-1} \widetilde{\mathbf{F}}_t (\mathbf{D}')^{-1} \mathbf{B}' + \mathbf{E}_t. \tag{A.21}$$

Given the condition that $\text{Var}(\mathbf{u}_t) = \mathbf{I}_{p_1 p_2}$, $\text{Var}(\mathbf{M}\mathbf{u}_t)$ should be an identity matrix as well.

That is, $\mathbf{M}\text{Var}(\mathbf{u}_t)\mathbf{M}' = \mathbf{I}_{p_1 p_2}$. Therefore, we have $\mathbf{M}\mathbf{M}' = \mathbf{I}_{p_1 p_2}$. Or $\mathbf{M}$ is an orthogonal matrix. Therefore, we have

$$\mathbf{M}\mathbf{M}' = \mathbf{I} \Leftrightarrow (\mathbf{D} \otimes \mathbf{C})(\mathbf{D} \otimes \mathbf{C})' = \mathbf{I} \Leftrightarrow (\mathbf{D}\mathbf{D}') \otimes (\mathbf{C}\mathbf{C}') = \mathbf{I},$$

which holds if and only if $\mathbf{D}\mathbf{D}' = \mathbf{I}_{p_2}$ and $\mathbf{C}\mathbf{C}' = \mathbf{I}_{p_1}$, given that C and D are lower triangular matrices and the diagonal elements of $\mathbf{C}$ is ones.

This means that $\mathbf{C}$ and $\mathbf{D}$ are orthogonal matrices. An orthogonal matrix that is lower triangular must be diagonal. Therefore, the rotation matrix $\mathbf{C}$ is an identity matrix. Given that $b_{ii} > 0$ for $i = 1, ..., p_2$, we must have that the rotation matrix $\mathbf{D}$ is also an identity matrix.

This proves that the proposed assumptions in *MDFM1* fully identify the factor matrix and the factor loading matrices.

A.2 Proof of *Proposition 2*

Proof of Proposition 2: Similar to the proof of proposition 1, the rotated factor loadings $\mathbf{C}^{-1}$ needs to be a lower triangular matrix, as shown in (A.16). Additionally, given we have ones on the diagonal of $\mathbf{A}$, $\mathbf{C}^{-1}$ needs to have ones on its diagonal as well. Similarly, $\mathbf{D}^{-1}$ needs to be a lower triangular matrix with ones on its diagonal. Therefore, the matrix $\mathbf{M}$ is a lower triangular matrix with ones on its diagonal.

Again, we need $\text{Cov}(\mathbf{M}\mathbf{u}_t) = \text{Cov}(\mathbf{u}_t)$, i.e., $\mathbf{M}\mathbf{\Lambda}_t\mathbf{M}' = \mathbf{\Lambda}_t$, where $\mathbf{\Lambda}_t$ is a diagonal matrix. Given that the diagonal elements in $\mathbf{\Lambda}_t$ must be larger than 0, this requires that $m_{i,j}$ for all $i > j$ must be zero, for $\mathbf{M}\mathbf{\Lambda}_t\mathbf{M}'$ to only have non-zero terms on its diagonal and match $\mathbf{\Lambda}_t$. Therefore, $\mathbf{M}$ must be identity matrix.

This proves that assumptions 1, 4 and 5 fully identifies the factor matrix and the factor loading matrices.

B Bayesian Estimation for MDFM with Stochastic Volatility

Recall the dynamic factor model for matrix-valued time series with stochastic volatility

$$\mathbf{Y}_t = \mathbf{A}\mathbf{F}_t\mathbf{B}' + \mathbf{E}_t, \quad \text{vec}(\mathbf{E}_t) \sim \mathcal{N}(\mathbf{0}, \omega_t\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r), \tag{B.22}$$

$$\text{vec}(\mathbf{F}_t) = \mathbf{H}_{\boldsymbol{\rho}_1}\text{vec}(\mathbf{F}_{t-1}) + \ldots + \mathbf{H}_{\boldsymbol{\rho}_q}\text{vec}(\mathbf{F}_{t-q}) + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_t), \tag{B.23}$$

where $\mathbf{A}$ is a $n \times p_1$ matrix of factor loadings, $\mathbf{B}$ is a $k \times p_2$ matrix of factor loadings, $\mathbf{F}_t$ is a $p_1 \times p_2$ latent matrix-valued time series of common factors, $\mathbf{E}_t$ is a $n \times k$ idiosyncratic component, vec($\cdot$) is a vectorizing function, $\mathbf{H}_{\boldsymbol{\rho}_l}$ is a diagonal matrix of autoregressive coefficients $(\rho_{1,l}, \ldots, \rho_{p_1p_2,l})'$, $l = 1, \ldots, q$, and $\boldsymbol{\Lambda}_t$ is a covariance matrix for the error in factor evolution process.

We use a natural conjugate prior for the transpose of factor loadings: $\mathbf{A}'$ and $\mathbf{B}'$. In addition, we use inverse-Wishart prior for $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$:

$$\begin{aligned} \boldsymbol{\Sigma}_r &\sim \mathcal{IW}(\nu_r, \mathbf{S}_r), \quad (\text{vec}(\mathbf{A}')|\boldsymbol{\Sigma}_r) \sim \mathcal{N}(\text{vec}(\mathbf{A}_0'), \boldsymbol{\Sigma}_r \otimes \mathbf{V}_{\mathbf{A}'}), \\ \boldsymbol{\Sigma}_c &\sim \mathcal{IW}(\nu_c, \mathbf{S}_c), \quad (\text{vec}(\mathbf{B}')|\boldsymbol{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\mathbf{B}_0'), \boldsymbol{\Sigma}_c \otimes \mathbf{V}_{\mathbf{B}'}). \end{aligned} \tag{B.24}$$

The autoregressive coefficient $\rho_{j,k,l}$ is assumed to have a truncated normal prior on the interval $(-1, 1)$:

$$\rho_{j,k,l} \sim \mathcal{TN}(\rho_{j,k,l,0}, V_{\rho_{j,k,l}}), \quad j = 1, ..., p_1, \quad k = 1, ..., p_2, \quad l = 1, \ldots, q.$$

The prior variance $\lambda_{j,k}^2$ is assumed to have a inverse-gamma prior: $\mathcal{IG}(\nu_{\lambda_{j,k}}, S_{\lambda_{j,k}})$. We also treat the first q factors as unknown, and use the following prior

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

For identification, we use assumptions 1, 4 and 5. We employ Markov Chain Monte Carlo (MCMC) methods to obtain a draw from the joint posterior of the latent factors and parameters of the model. Specifically, the following steps are carried out:

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

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

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

With the constraints for identification, we cannot directly sample from the above normal-inverse-Wishart distribution. Here we outline the sampling scheme for $\mathbf{A}'$ with the structure constraints. To that end, we first represent the restrictions as a system of linear restrictions. For example, for $\mathbf{A}'$, we represent the restrictions that $\mathbf{A}$ is a lower triangular matrix with ones on the diagonal using $\mathbf{M}_{\mathbf{A}'} \text{vec}(\mathbf{A}') = \mathbf{a}_0$. Assuming $n > p_1$, $\mathbf{M}_{\mathbf{A}'} = (m_{i,j})$ is a $p_1(p_1+1)/2 \times np_1$ selection matrix, and $\mathbf{a}_0$ is a $p_1(p_1+1)/2 \times 1$ vector consisting of ones and zeros. Then we apply Algorithm 2 in Cong et al. (2004) 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 $\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}_u') + \mathbf{V}'(\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'} \text{vec}(\mathbf{A}_u'))$.

2. Sampling from $(\mathbf{B}', \mathbf{\Sigma}_c | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r)$

Similar to step 1, $(\mathbf{B}, \mathbf{\Sigma}_c)$ are drawn from a normal-inverse-Wishart distribution:

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

where

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

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

We sample $(\mathbf{B}, \mathbf{\Sigma}_c| \cdot)$ in two steps. First, we sample $\mathbf{\Sigma}_c$ marginally from $(\mathbf{\Sigma}_c \mid \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r) \sim \mathcal{IW}(\widehat{\mathbf{S}}_c, \nu_c + Tn)$ with the normalization restriction that $\sigma_{c,1,1} = 1$. This can be done using the algorithm in Nobile (2000) described below. Then we simulate $(\text{vec}(\mathbf{B}') \mid \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{B}}), \mathbf{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}'}^{-1})$, which can be done using the algorithm described in step 1.

The algorithm in Nobile (2000) can be realized by the following steps:

(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 $\mathbf{\Delta}$ 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 $\boldsymbol{L}$
   - δ_{ij} equal to $\mathcal{N}(0,1)$ random variates, $i > j$.
(3) Set $\mathbf{\Sigma}_c = (\boldsymbol{L}^{-1})'(\mathbf{\Delta}^{-1})'\mathbf{\Delta}^{-1}\boldsymbol{L}^{-1}$.
(4) Exchange the row/column 1 and n of $\mathbf{\Sigma}_c$ back.

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

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

Step 4. Sampling from $(\lambda_{j,k}^2|\mathbf{f}_{j,k}, \boldsymbol{\rho}_{j,k}), \quad j = 1, ..., p_1, k = 1, ..., p_2$
It is clear that $(\lambda_{j,k}^2|\mathbf{f}_{j,k}, \boldsymbol{\rho}_{j,k}) \sim \mathcal{IG}(\widehat{\nu}_{\lambda_{j,k}}, \widehat{S}_{\lambda_{j,k}})$, where $\widehat{\nu}_{\lambda_{j,k}} = \nu_{\lambda_{j,k}} + \frac{T}{2}$, and $\widehat{S}_{\lambda_{j,k}} = S_{\lambda_{j,k}} + \frac{1}{2}\left[\sum_{t=1}^{q} f_{j,k,t}^2(1 - \sum_m \rho_{j,k,m}^2) + \sum_{t=q+1}^{T}(f_{j,k,t} - \rho_{j,k,1}f_{j,k,t-1} - ... - \rho_{j,k,q}f_{j,k,q})^2\right]$.

Step 5. Sampling from $(\boldsymbol{\rho}_{j,k}|\mathbf{f}_{j,k}, \lambda_{j,k}^2), \quad j = 1, ..., p_1, k = 1, ..., p_2$
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 (2) as follows:

$$\widetilde{\mathbf{f}}_{j,k} = \widetilde{\mathbf{F}}_{j,k}\boldsymbol{\rho}_{j,k} + \mathbf{u}_{j,k}, \quad \mathbf{u}_{j,k} \sim \mathcal{N}(\mathbf{0}, \lambda_{j,k}\mathbf{I}_{T-q}), \tag{B.25}$$

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

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

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

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

5. Sampling the time-varying volatility

For clearer illustration, assume that we have only one type of time-varying volatility. The following three steps correspond to each type.

5.1 Common stochastic volatility: sampling from $(\mathbf{h}|\mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r)$

The conditional posterior for $\mathbf{h}$ is not a standard distribution. In this paper, we follow Chan (2017) for this purpose. In particular, we first obtain the mode of the log density of $(\mathbf{h}|\cdot)$ as well as the negative Hessian evaluated at the mode, denoted as $\widehat{\mathbf{h}}$ and $\mathbf{K}_{\mathbf{h}}$, respectively. Then we use $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}_{\mathbf{h}}^{-1})$ as the proposal distribution, and sample $\mathbf{h}$ using an acceptance-rejection Metropolis-Hasting step. Samplers for ϕ and σ_h^2 are standard and we omit the details in this paper.

5.2 The explicit outlier components: sampling from $(\mathbf{o}, p_{\mathbf{o}}|\mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r)$
We follow Stock and Watson (2016) to discretize the support of o_t to simplify estimation. Specifically, we use a grid with points at $1, 2, 3, ..., 20$. The likelihood can be easily evaluated at these grid points. Finally, a draw from the full conditional posterior distribution of o_t can be obtained using the inverse transform method.

The conditional distribution of p_{o_i} is a Beta distribution:

$$(p_{o_i}|\mathbf{o}_i) \sim \mathcal{B}(a_{p_{o_i}} + n_2, b_{p_{o_i}} + n_1),$$

where $n_1 = \sum_{t=1}^{T} \mathbf{I}(o_{i,t} = 1)$ is the number of "regular" periods, and $n_2 = T - \sum_{t=1}^{T} \mathbf{I}(o_{i,t} = 1)$ is the number of "outlier" periods.

5.3 Fat-tailed innovations: sampling from $(q_t^2|\mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r)$**,** $t = 1, \ldots, T$
Conditional on the factors and parameters, the posterior for q_t^2 has an inverse-gamma distribution:

$$(q_t^2|\cdot) \sim \mathcal{IG}((nk + l)/2, (s_t^2 + l)/2),$$

where $s_t^2 = \text{tr}\left[\mathbf{\Sigma}_c^{-1}(\mathbf{Y}_t - \mathbf{A}\mathbf{F}_t\mathbf{B})'\mathbf{\Sigma}_r^{-1}(\mathbf{Y}_t - \mathbf{A}\mathbf{F}_t\mathbf{B})\right]$.

C Estimating Marginal Likelihoods

This section describes the method we use to obtain integrated likelihood and the importance-sampling densities. For illustration, we consider $q = 1$.

C.1 Integrated Likelihood

Model (1)(2) can be rewritten as follows

$$\begin{aligned} &\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r), \\ &\mathbf{f}|\boldsymbol{\rho}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\mathbf{0}, \left[\mathbf{H}_{\boldsymbol{\rho}}'(\mathbf{I}_T \otimes \boldsymbol{\Omega})^{-1}\mathbf{H}_{\boldsymbol{\rho}}\right]^{-1}\right), \end{aligned} \tag{C.26}$$

System (C.26) can be rewritten as follows

$$\begin{aligned} &\mathbf{y} = (\mathbf{I}_T \otimes \mathbf{A} \otimes \mathbf{B})\mathbf{f} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes (\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r)), \\ &\mathbf{f}|\boldsymbol{\rho}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\mathbf{0}, \left[\mathbf{H}_{\boldsymbol{\rho}}'(\mathbf{I}_T \otimes \boldsymbol{\Omega})^{-1}\mathbf{H}_{\boldsymbol{\rho}}\right]^{-1}\right). \end{aligned} \tag{C.27}$$

It is easy to integrate out $\mathbf{f}$ and we can get the following likelihood

$$\mathbf{y}|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho} \sim \mathcal{N}(\bar{\mathbf{y}}, \overline{\mathbf{D}}_{\mathbf{y}}), \tag{C.28}$$

where

$$\begin{aligned} \bar{\mathbf{y}} &= \mathbb{E}\left[\mathbb{E}\left(\mathbf{y}|\mathbf{f}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right)|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right] \\ &= \mathbb{E}\left[(\mathbf{I}_T \otimes \mathbf{B} \otimes \mathbf{A})\mathbf{f}|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right] \\ &= (\mathbf{I}_T \otimes \mathbf{B} \otimes \mathbf{A})\mathbb{E}[\mathbf{f}|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}] \\ &= \mathbf{0}, \end{aligned}$$

and

$$\begin{aligned} \overline{\mathbf{D}}_{\mathbf{y}} &= \mathbb{E}\left\{\left[\mathrm{Var}\left(\mathbf{y}|\mathbf{f}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right)|\cdot\right\} + \mathrm{Var}\left(\mathbb{E}[\mathbf{y}|\mathbf{f}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}]|\cdot\right)\right] \\ &= \mathbf{I}_T \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + (\mathbf{I}_T \otimes \mathbf{B} \otimes \mathbf{A})[\mathbf{H}_{\rho}'(\mathbf{I}_T \otimes \boldsymbol{\Omega})^{-1}\mathbf{H}_{\rho}]^{-1}(\mathbf{I}_T \otimes \mathbf{B}' \otimes \mathbf{A}'). \end{aligned}$$

It can be very costly to compute the inverse of the covariance matrix $\overline{\mathbf{D}}_{\mathbf{y}}$. Therefore, here we use Kalman filter. In particular, it is not difficult the marginal distribution for $\mathbf{f}_t \equiv \text{vec}(\mathbf{F}_t)$,:

$$\begin{aligned}(\mathbf{f}_1 \mid \boldsymbol{\rho}, \boldsymbol{\lambda}) &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_1) \\ (\mathbf{f}_t \mid \boldsymbol{\rho}, \boldsymbol{\lambda}) &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_t + \mathbf{H}_{\boldsymbol{\rho}} \boldsymbol{\Lambda}_{t-1} \mathbf{H}_{\boldsymbol{\rho}}'), \quad t = 2, \ldots, T,\end{aligned}$$

where $t = 2, \ldots, T$, $\boldsymbol{\Lambda}_t = \text{diag}(\lambda_{1,1}^2, \lambda_{2,1}^2, ..., \lambda_{p_1,p_2}^2)$, and for $t = 1$, $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_{1,1}^2/(1 - \rho_{1,1}^2), \lambda_{2,1}^2/(1 - \rho_{2,1}^2), ..., \lambda_{p_1,p_2}^2/(1 - \rho_{p_1,p_2}^2))$. $\mathbf{H}_{\boldsymbol{\rho}} = \text{diag}(\rho_{1,1}, \rho_{2,1}, ..., \rho_{p_1,p_2})$

Therefore, the integrated likelihood at time t is:

$$(\mathbf{y}_t \mid \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r) \sim \mathcal{N}(\mathbf{0}, \overline{\mathbf{D}}_{\mathbf{y}_t}),$$

where

$$\begin{aligned}\overline{\mathbf{D}}_{\mathbf{y}_1} &= \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + (\mathbf{B} \otimes \mathbf{A}) \boldsymbol{\Lambda}_1 (\mathbf{B}' \otimes \mathbf{A}') \\ \overline{\mathbf{D}}_{\mathbf{y}_t} &= \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + (\mathbf{B} \otimes \mathbf{A})(\boldsymbol{\Lambda}_t + \mathbf{H}_{\boldsymbol{\rho}} \boldsymbol{\Lambda}_{t-1} \mathbf{H}_{\boldsymbol{\rho}}')(\mathbf{B}' \otimes \mathbf{A}'), \quad t = 2, \ldots, T.\end{aligned}$$

C.2 Finding the Optimal Importance-sampling Densities

The next step is to find the maximum likelihood estimators for the hyperparameters in the importance-sampling density. The importance-sampling density is denoted as

$$\begin{aligned}f(\boldsymbol{\theta}; \mathbf{v}) &= f(\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\rho}; \mathbf{v}) \\ &= f(\mathbf{A}; \overline{\mathbf{A}}, \overline{\mathbf{D}}_{\mathbf{A}}) \cdot f(\boldsymbol{\Sigma}_c; \Psi_c, \nu_c) \cdot f(\boldsymbol{\Sigma}_r; \Psi_r, \nu_r) \cdot f(\boldsymbol{\lambda}; \nu_\lambda, S_\lambda) \cdot f(\boldsymbol{\rho}; \overline{\boldsymbol{\rho}}, \overline{\mathbf{D}}_{\boldsymbol{\rho}}).\end{aligned} \tag{C.29}$$

In terms of the parameteric family, we use Gaussian density for $f(\mathbf{A}; \overline{\mathbf{A}}, \overline{\mathbf{D}}_{\mathbf{A}})$, where $\overline{\mathbf{A}}$ and $\overline{\mathbf{D}}$ are the corresponding mean and covariance matrix. We use inverse Wishart densities for $f(\boldsymbol{\Sigma}_c; \nu_c, \Psi_c)$ as well as $f(\boldsymbol{\Sigma}_r; \nu_r, \Psi_r)$. We use inverse gamma density for We use the truncated normal density on the interval $(-1, 1)$ for $f(\boldsymbol{\rho}; \overline{\boldsymbol{\rho}}, \overline{\mathbf{D}}_{\boldsymbol{\rho}})$, where $\overline{\boldsymbol{\rho}}$ and $\overline{\mathbf{D}}_{\boldsymbol{\rho}}$ are the corresponding mean and covariance matrix. we use inverse-gamma distribution for $f(\boldsymbol{\lambda}; \nu_\lambda, S_\lambda)$.

In order to obtain the maximum likelihood estimators for the parameters in inverse Wishart distribution, we first use maximum likelihood estimation on the Wishart dis-

tribution given the posterior samples, and then compute the degree of freedom and scale matrix of the inverse Wishart distribution using *Lemma 1*.

Lemma 1: $\mathbf{\Sigma}$ follows an inverse Wishart distribution if $\mathbf{K} \equiv \mathbf{\Sigma}^{-1}$ follows a Wishart distribution, formally expressed as

$$\mathbf{\Sigma} \sim \mathcal{IW}_d(\delta - d + 1, \mathbf{\Psi}^{-1}) \Leftrightarrow \mathbf{K} = \mathbf{\Sigma}^{-1} \sim \mathcal{W}_d(\delta, \mathbf{\Psi}), \tag{C.30}$$

where d is the dimension of the matrix $\mathbf{\Sigma}$, δ is the degree of freedom of the Wishart distribution, and $\mathbf{\Psi}$ is the scale matrix.

A Wishart distribution is defined as:

$$f(\mathbf{K}|\mathbf{\Psi}, \delta) = \frac{|\mathbf{K}|^{\frac{\delta-d-1}{2}}}{2^{\frac{\delta d}{2}}|\mathbf{\Psi}|^{\frac{\delta}{2}}\Gamma_d\left(\frac{\delta}{2}\right)} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{K}\mathbf{\Psi}^{-1})\right\}.$$

We assume that each matrix is drawn independently from the same Wishart distribution $\mathcal{W}(\mathbf{\Psi}, \delta)$. Therefore, we can model the joint distribution as:

$$f(\mathbf{K}_1, ..., \mathbf{K}_M|\mathbf{\Psi}, \delta) = \prod_{m=1}^{M} \frac{|\mathbf{K}_m|^{\frac{\delta-d-1}{2}}}{2^{\frac{\delta d}{2}}|\mathbf{\Psi}|^{\frac{\delta}{2}}\Gamma_d\left(\frac{\delta}{2}\right)} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{K}_m\mathbf{\Psi}^{-1})\right\}.$$

The log-likelihood function is therefore

$$\begin{aligned}\log f(\mathbf{K}_1, ..., \mathbf{K}_M|\mathbf{\Psi}, \delta) = &-\frac{\delta d M}{2}\log 2 - \frac{\delta M}{2}\log|\mathbf{\Psi}| - M\log\Gamma_d\left(\frac{\delta}{2}\right) + \\ &\frac{\delta - d - 1}{2}\sum_{m=1}^{M}\log|\mathbf{K}_m| - \frac{1}{2}\text{tr}\left(\sum_{m=1}^{M}\mathbf{K}_m\mathbf{\Psi}^{-1}\right).\end{aligned}$$

The first derivative of the log-likelihood function with respect to the scale matrix $\mathbf{\Psi}$ is equal to

$$\frac{\mathrm{d}\log(f(\mathbf{K}_1, ..., \mathbf{K}_M|\mathbf{\Psi}, \delta))}{\mathrm{d}\mathbf{\Psi}} = -\frac{M\delta}{2}\mathbf{\Psi}^{-1} + \frac{1}{2}\mathbf{\Psi}^{-1}\sum_{m=1}^{M}\mathbf{K}_m\mathbf{\Psi}^{-1}, \tag{C.31}$$

where two results are used

1. $\frac{\partial|\mathbf{X}|}{\partial\mathbf{X}} = |\mathbf{X}|\mathbf{X}^{-1}$;

2. $\frac{\partial \text{tr}(\mathbf{A}\mathbf{X}^{-1})}{\partial \mathbf{X}} = -\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}$.

From equation (C.31) we obtain a function of the MLE of $\mathbf{\Psi}$ with respect to the degree of freedom δ

$$\widehat{\mathbf{\Psi}}^{mle} = \frac{1}{M\delta}\sum_{m=1}^{M}\mathbf{K}_m. \tag{C.32}$$

In order to obtain the MLE for the degree of freedom, a straightforward way is to find the first order condition and second order condition to maximize the log-likelihood function with respect to δ. We then use the Newton-type methods to find the estimate for $\widehat{\delta}$.

In particular, the first derivative of the log-likelihood function after we plug in (C.32) is

$$\begin{aligned}\frac{\partial \log f(\mathbf{K}_1, ..., \mathbf{K}_M|\delta)}{\partial \delta} =& -\frac{dM}{2}(\log 2 + 1) + \frac{Md}{2}\log\delta - \frac{M}{2}\log\left|M^{-1}\sum_m \mathbf{K}_m\right| \\ & - \frac{M}{2}\psi_d\left(\frac{\delta}{2}\right) + \frac{1}{2}\sum_m \log|\mathbf{K}_m|.\end{aligned} \tag{C.33}$$

The second derivative is

$$\frac{\partial^2 \log f(\mathbf{K}_1, ..., \mathbf{K}_M|\delta)}{\partial \delta^2} = -\frac{Md}{2\delta} - \frac{M}{4}\psi_d^{(2)}\left(\frac{1}{2}\delta\right).$$

Maximum likelihood estimators for parameters for normal distributions and inverse gamma distributions are straightforward to obtain so that we omit the details here.

D Additional Simulation Results

Table D.4: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 3$, $p_2 = 2$

(n, k)	T = 200		T = 500		T = 1000	
(10, 10)	0.98	0.98	0.99	0.97	0.98	0.99
	0.98	0.96	0.98	0.98	0.98	0.99
	0.96	0.96	0.99	0.99	0.98	0.99
Average	0.97		0.98		0.98	
(20, 15)	1.00	0.98	1.00	0.99	1.00	0.99
	0.97	0.98	0.99	0.98	1.00	0.99
	0.99	0.99	0.99	0.97	0.98	0.98
Average	0.98		0.99		0.99	
(30, 20)	1.00	0.98	1.00	1.00	1.00	0.99
	0.99	0.98	0.99	0.99	0.99	0.99
	0.99	0.98	0.98	0.98	0.99	0.99
Average	0.99		0.99		0.99	

Table D.5: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 5$, $p_2 = 5$

(n, k)	$T = 200$					$T = 500$					$T = 1000$				
	0.97	0.98	0.95	0.97	0.97	0.99	0.98	0.97	0.95	0.97	0.99	0.99	0.99	0.98	0.98
	0.96	0.97	0.94	0.97	0.97	0.98	0.97	0.97	0.93	0.97	0.98	0.97	0.98	0.97	0.98
(10, 10)	0.97	0.96	0.97	0.97	0.97	0.97	0.96	0.95	0.92	0.98	0.98	0.97	0.98	0.98	0.98
	0.96	0.96	0.93	0.95	0.96	0.97	0.98	0.96	0.94	0.98	0.96	0.96	0.95	0.95	0.97
	0.95	0.96	0.95	0.93	0.95	0.98	0.97	0.96	0.91	0.96	0.99	0.98	0.98	0.97	0.98
Average			0.96					0.96					0.98		
	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.97
	0.99	0.99	0.99	0.98	0.94	0.99	0.97	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.98
(20, 15)	0.99	0.99	0.99	0.98	0.96	0.99	0.98	0.98	0.97	0.99	0.99	0.99	0.99	0.99	0.97
	0.98	0.98	0.98	0.98	0.98	0.99	0.95	0.98	0.97	0.98	0.99	0.99	0.99	0.99	0.97
	0.98	0.97	0.98	0.97	0.95	0.98	0.97	0.96	0.96	0.96	0.99	0.98	0.99	0.98	0.97
Average			0.98					0.98					0.99		
	1.00	1.00	0.99	0.99	0.92	1.00	0.99	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	0.99	0.99	0.99	0.99	0.97	0.99	1.00	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99
(30, 20)	0.99	0.99	0.99	0.97	0.96	1.00	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.98	0.99
	0.97	0.98	0.99	0.99	0.91	0.98	0.99	0.99	0.97	0.99	1.00	0.99	0.99	0.99	0.99
	0.99	0.99	0.98	0.99	0.97	0.97	0.99	0.98	0.97	0.98	0.99	0.99	0.99	0.99	0.99
Average			0.98					0.99					0.99		

E Data: Multinational Macroeconomic Panel

Table E.6 describes the list of variables we use for the first application. We attach the link of the website we downloaded the specific variable to the variable name in the table. The second column of E.6 is the stationarity transformation for each variable.

Table E.6: List of variables

Variable	Transformation
Real GDP	No transformation
Consumption	$\Delta \log(x)$
Labor unit costs	Δx
Unemployment	Δx
Headline CPI	Δx
Energy CPI	Δx
Food CPI	Δx
Core CPI	Δx
Imports	$\Delta \log(x)$
Exports	$\Delta \log(x)$