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

First Draft: August 2024
This Draft: August 2025

Abstract

We introduce a class of Bayesian matrix dynamic factor models that accommodates time-varying volatility, outliers, and cross-sectional correlation in the idiosyncratic components. For model comparison, we employ an importance-sampling estimator of the marginal likelihood based on the cross-entropy method to determine: (1) the optimal dimension of the factor matrix; (2) whether a vector- or matrix-valued structure is more suitable; and (3) whether an approximate or exact factor model is favored by the data. Through a series of Monte Carlo experiments, we demonstrate the accuracy of the factor estimates and the effectiveness of the marginal likelihood estimator in correctly identifying the true model. Applications to macroeconomic and financial datasets illustrate the model's ability to capture key features in matrix-valued time series.

Keywords: Matrix-valued time series, dynamic factor models, approximate factor models, time-varying volatility, Bayesian model comparison
JEL Codes: C11, C32, C55

1 Introduction

Matrix-valued time series models have recently gained significant attention for their ability to capture complex interactions across row- and column-dimensions. In macroeconomics and finance, such data structures are common. A prominent example is macroeconomic indicators collected across multiple countries. A standard approach is to stack the matrix into a long vector and to apply standard multivariate methods, but this often overlooks important within-row and within-column dependencies. Research on statistical methods for matrix-valued time series is still evolving, and matrix factor models are becoming popular due to their ability to reduce dimensions, particularly in high-dimensional settings. Wang et al. (2019) introduce a factor model for such matrix-valued time series. Subsequent work has extended this framework in various directions: Chen et al. (2020) incorporated linear constraints to include prior knowledge, Liu and Chen (2019) developed a threshold version, and Chen et al. (2024) allowed for time-varying loadings.

We extend this literature by proposing a class of Bayesian matrix dynamic factor models (MDFMs) with several useful features for empirical macro-finance studies. First, we model factor dynamics using an autoregressive (AR) process, which enables persistent comovement and recursive forecasting.[1] Second, we build an approximate factor framework that allows for both temporal and cross-sectional correlations in the idiosyncratic components.[2] On the time dimension, we incorporate common stochastic volatility, fat-tailed errors, and COVID-19 outliers, reflecting key empirical features observed in recent macroeconomic data (e.g., Lenza and Primiceri, 2022; Carriero et al., 2024a). On the cross-sectional dimension, we use a Kronecker structure for the idiosyncratic covariance matrix to capture cross-row and cross-column correlations in the idiosyncratic components. When combined with a natural conjugate prior, this structure significantly improves computational efficiency.

An important challenge in practice is model comparison when multiple models are available. In particular, a key issue is determining the optimal dimensions of the factor matrices. Researchers may also wish to compare standard dynamic factor models using

[1] See Sargent et al. (1977), Stock and Watson (2012), Bai and Wang (2015), and Poncela et al. (2021).

[2] Approximate factor models date 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.

vectorized panels (VDFMs) with matrix dynamic factor models (MDFMs), or evaluate the performance of exact versus approximate factor models.[3] In this paper, we address these questions by adopting an importance-sampling marginal likelihood estimator of Chan and Eisenstat (2015), which uses a cross-entropy method to construct an efficient proposal density. This method offers two key advantages. First, it relies on independent draws from the importance-sampling density rather than correlated Markov Chain Monte Carlo (MCMC) samples. Second, it is computationally efficient and straightforward to implement. Specifically, this method identifies the "optimal" parameters within a given parametric density family by minimizing the Kullback-Leibler divergence to the posterior distribution. Chan and Eisenstat (2015) have shown that these "optimal" parameters correspond to the maximum likelihood estimators if we treat the posterior samples as observed data of the parameters. Since the importance-sampling density is designed to closely approximate the posterior distribution, the resulting marginal likelihood estimator is highly efficient and exhibits low variance.

Our paper builds on recent works by Yu et al. (2024) and Qin et al. (2025), who propose matrix dynamic factor models with matrix autoregressive factor evolution. While their work focuses on modeling factor dynamics under the matrix factor framework, our approach complements this line of research by incorporating additional features crucial for macro-financial applications, including time-varying volatility, outlier adjustments, and cross-sectional correlations in the idiosyncratic components. We further adopt a fully Bayesian framework with identification restrictions to uniquely identify loadings and factors for economic interpretation, and we conduct extensive model comparison exercises to determine the factor matrix dimension, model structure (VDFM vs. MDFM), and the nature of idiosyncratic correlations (exact vs. approximate factor models).

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. Moreover, we find that the marginal likelihood estimator can correctly identify the true model.

We illustrate the usefulness of the proposed MDFMs using two datasets. The first dataset

[3]For convenience, we refer to the vector version as VDFM, where "v" stands for "vector", in contrast to "m" for "matrix".

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). Several key findings emerge from our analysis. First, models with stochastic volatility yield higher marginal likelihoods for both datasets, indicating clear benefits from incorporating time-varying volatility. The international macroeconomic panel favors an MDFM with a 1×2 factor matrix and cross-sectional correlations in the idiosyncratic components, while the Fama-French panel favors a VDFM with five factors. Second, 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). Finally, we observe high volatility in both applications during major global events, including the Great Recession, the COVID-19 pandemic, and the Russia–Ukraine war.

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 heteroskedastic time-varying 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:

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

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

where $\mathbf{A}$ and $\mathbf{B}$ are loading matrices of dimensions $n \times p_1$ and $k \times p_2$, respectively, $\mathbf{F}_t$ is a $p_1 \times p_2$ latent factor matrix. $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$ are row- and column-wise covariance matrices with dimensions $n \times n$ and $k \times k$, respectively. The $p_1p_2 \times p_1p_2$ covariance matrix $\mathbf{\Lambda}$ is a diagonal matrix: $\mathbf{\Lambda} = \text{diag}(\lambda^2_{1,1}, \ldots, \lambda^2_{p_1p_2,p_1p_2})$. The autoregressive coefficient matrices $\mathbf{H}_{\boldsymbol{\rho}_1}, \ldots, \mathbf{H}_{\boldsymbol{\rho}_q}$ capture the persistence of the factors as well as their correlation.

The bilinear form $\mathbf{A}\mathbf{F}_t\mathbf{B}$ is crucial for capturing the cross-country and cross-variable dependencies. For example, the i-th row of $\mathbf{Y}_t$ is

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

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 $\mathbf{F}_t\mathbf{B}'$, weighted by the i-th row of $\mathbf{A}$. Similarly, the j-th column of the data matrix is a linear combination of the columns of $\mathbf{A}\mathbf{F}_t$, weighted by the j-th column of $\mathbf{B}'$. Thus, $\mathbf{A}$ reflects how countries load on latent factors, while $\mathbf{B}$ reflects how indicators respond to them. The rows and columns of $\mathbf{F}_t$ can be interpreted as common factors influencing countries and indicators, respectively.

2.1.1 Important Features of the Model

Based on the matrix factor model by Wang et al. (2019), our model introduces two key features: dynamic factors and cross-sectional correlations with time-varying volatility in the idiosyncratic components.

Incorporating factor dynamics is crucial, as economic indicators often exhibit persistent deviations from trend following shocks. For instance, GDP growth may remain subdued long after a recession. The autoregressive process in (2) can effectively capture such persistence and potentially improve modeling and forecasting performance.

The other key feature of model (1)–(2) is the Kronecker-structured covariance of the vectorized error, $\text{vec}(\mathbf{E}_t)$, which introduces both flexibility and interpretability. This structure allows for cross-sectional correlation in the idiosyncratic components, making the model an approximate factor model rather than an exact one. Approximate factor models, introduced by Chamberlain and Rothschild (1983), have been widely applied in macroeconomics (e.g., Forni et al., 2001; Stock and Watson, 2005).

The Kronecker structure also separates row-wise and column-wise covariances. In particular, for any row, the conditional covariance is $\text{Cov}(\mathbf{Y}'_{i,.,t} \mid \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{r,i,i} \boldsymbol{\Sigma}_c$, for $i = 1, \ldots, n$. Similarly, for any column, the conditional covariance is $\text{Cov}(\mathbf{Y}_{.,j,t} \mid \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{c,j,j} \boldsymbol{\Sigma}_r$, $j = 1, \ldots, k$. $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$ can then be interpreted as the row-wise and column-wise covariances that are not explained by the common components.

Computationally, this structure enables efficient sampling: with conjugate priors, the conditional posteriors of $(\mathbf{A}, \boldsymbol{\Sigma}_r)$ and $(\mathbf{B}, \boldsymbol{\Sigma}_c)$ follow matrix-normal-inverse-Wishart distributions, allowing for joint updates rather than element-wise sampling for the loading matrices $\mathbf{A}$ and $\mathbf{B}$ (see Section 2.4). Finally, the model flexibly accommodates time-varying volatility and outliers via the latent scale ω_t. For homoskedasticity, one can simply set $\omega_t = 1$ for all t.

2.1.2 Time-varying Volatility and Outlier Adjustment

It is useful to allow for time-varying volatility in modeling macroeconomic data in empirical macroeconomics or finance.[4] 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

[4]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).

(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:

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

for $t = 2, ..., T$, where it is assumed $|\phi| < 1$ and the initial state h_1 is assumed to have a Gaussian prior: $h_1 \sim \mathcal{N}(0, \sigma_h^2/(1-\phi^2))$.

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 Factor Models

A natural benchmark to model (1)-(2) is a typical VDFM defined as follows:

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

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 the MDFM. The matrix $\mathbf{H}_\rho$ is a k-dimensional diagonal matrix consisting of autoregressive coefficients for the evolution process of these factors.

The MDFM can be viewed as a restricted version of the VDFM in (4), where the factor loadings possess a Kronecker product structure. Specifically, model (1) can be written as:

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

This structure formulation significantly reduces the dimensionality of the parameter space. In particular, model (1) requires estimating only $np_1 + kp_2$ loading parameters, compared to $nk \times p_1 p_2$ parameters in the unrestricted DFM in (4).

However, if the Kronecker structure imposed by (1) is not supported by the data, the VDFM may provide a better fit. He et al. (2024) propose a family of randomized specification tests to formally test for the presence of this matrix structure. In this paper, we adopt a Bayesian model comparison approach by estimating the marginal likelihoods of competing models, as detailed in Section 3.

2.2 Identification

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 the literature of 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 provided in the Supplemental Appendix Section 1.

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, as specified in Assumption 4–5.

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

Another identification issue in (1)–(2) is that with the structure in the covariance matrix for the vectorized error, 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.

An important implication of the lower triangular structure of the factor loading matrices, as specified in Assumption 3 and 5, is that the order of rows and columns in data matrix

$\mathbf{Y}_t$ "define" the rows and columns of the factor matrix. In specific, the first row of the data matrix is the first row of the factor matrix plus the idiosyncratic error, and so forth. The same logic applies to the columns. The hierarchical structure adopted here thereby provides an additional layer of interpretability by linking factors to specific rows and columns of the data matrix. As a result, the ordering of the data plays a pivotal role in interpreting both the factor matrix and the loading matrices.

While our model (1)–(2) offers a parsimonious and interpretable framework for modeling matrix-valued time series, it is important to acknowledge its limitations. One key assumption (Assumption 2) is that the elements of the factor matrix are i.i.d. In practice, this assumption may not hold. One way to extend the model is to allow for such dependencies via a vector or matrix autoregressive process for $\mathbf{F}_t$. Considering that incorporating these additional features alongside time-varying volatility and cross-sectional dependence in the idiosyncratic components would increase model complexity and pose further identification challenges, we opt for a framework that strikes a balance between flexibility and complexity, and we leave such generalizations to future research.

2.3 Priors

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

$$\begin{aligned} \mathbf{\Sigma}_r &\sim \mathcal{IW}(\nu_r, \mathbf{S}_r), \quad (\mathrm{vec}(\mathbf{A}') \mid \mathbf{\Sigma}_r) \sim \mathcal{N}(\mathrm{vec}(\mathbf{A}_0'), \mathbf{\Sigma}_r \otimes \mathbf{V}_{\mathbf{A}'}), \\ \mathbf{\Sigma}_c &\sim \mathcal{IW}(\nu_c, \mathbf{S}_c), \quad (\mathrm{vec}(\mathbf{B}') \mid \mathbf{\Sigma}_c) \sim \mathcal{N}(\mathrm{vec}(\mathbf{B}_0'), \mathbf{\Sigma}_c \otimes \mathbf{V}_{\mathbf{B}'}). \end{aligned} \tag{6}$$

where $\mathbf{V}_{\mathbf{A}'}$ and $\mathbf{V}_{\mathbf{B}'}$ are $p_1 \times p_1$ and $p_2 \times p_2$ symmetric matrices, respectively.

Intuitively, this prior means that if a country's idiosyncratic variance is large, the prior allows more uncertainty in its corresponding element in $\mathbf{A}$. In addition, if the idiosyncratic components of two countries are correlated, the prior treats it as plausible that these two countries might respond to latent factors in correlated ways, while still allowing the data to overturn this if needed. To see this more concretely, consider the covariance between the i-th and j-th countries (rows) in $\mathbf{A}$:

$$\mathrm{Cov}(\mathbf{A}_{i,\cdot}, \mathbf{A}_{j,\cdot} \mid \mathbf{\Sigma}_r) = \mathbf{\Sigma}_{r,i,j} \mathbf{V}_{\mathbf{A}'},$$

where $\mathbf{\Sigma}_{r,i,j}$ denote the (i,j)-th element of $\mathbf{\Sigma}_r$. Therefore, the corresponding prior correlation of the loadings is $\mathbf{\Sigma}_{r,i,j}/\sqrt{\mathbf{\Sigma}_{r,i,i}\mathbf{\Sigma}_{r,j,j}}$, implying that countries with more correlated idiosyncratic components are encouraged to have more correlated factor loadings. An analogous interpretation holds for the columns (indicators) through $\mathbf{\Sigma}_c$ and $\mathbf{B}$.

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.

The autoregressive coefficient, i.e., the diagonal elements of $\mathbf{H}_{\boldsymbol{\rho}}$, is assumed to have a truncated normal prior on the interval $(-1,1)$:

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

where $\mathbb{1}(\cdot)$ denotes the indicator function.

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

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) $p(\mathbf{B}', \mathbf{\Sigma}_c \mid \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r)$; (3) $p(\text{vec}(\mathbf{F}_t) \mid \mathbf{Y}_t, \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c, \boldsymbol{\omega}^2, \boldsymbol{\rho})$, $t = 1, \ldots, T$; (4) $p(\lambda_{j,k}^2 \mid \mathbf{f}_{j,k}, \rho_{j,k})$, $j = 1, \ldots, p_1, k = 1, \ldots, p_2$; (5) $p(\boldsymbol{\rho}_{j,k,l} \mid \mathbf{f}_{j,k}, \lambda_{j,k}^2)$, $j = 1, \ldots, p_1, k = 1, \ldots, p_2, l = 1, \ldots, q$; (6) $p(\omega_t \mid \mathbf{A}, \mathbf{F}_t, \mathbf{B}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r)$, $t = 1, \ldots, T$.

With the natural conjugate prior, we can fully exploit the structure of the data matrix and the Kronecker form of the idiosyncratic components to directly sample the loading matrices, rather than sampling their free elements iteratively, thereby improving computational efficiency. To sample these loadings with identification restrictions outlined in Section 2.2, we adopt the approach proposed by Cong et al. (2004). To sample the covariance matrix $\mathbf{\Sigma}_c$ with the first element fixed at 1, we use the algorithm proposed by Nobile (2000). Sampling the latent factors and parameters such as $\boldsymbol{\lambda}$, $\boldsymbol{\rho}$, as well as stochastic volatility and outliers, is relatively straightforward and is discussed in detail in Supple-

mental Appendix Section 2. In the following, we focus on the implementation of Step 1–2, which involves sampling $(\mathbf{A}', \mathbf{\Sigma}_r \mid \cdot)$ and $(\mathbf{B}', \mathbf{\Sigma}_c \mid \cdot)$ under parameter restrictions.

Step 1. Sampling from $(\mathbf{A}', \mathbf{\Sigma}_r \mid \mathbf{Y}, \mathbf{B}, \mathbf{F}, \mathbf{\Sigma}_c)$

We sample $(\mathbf{A}', \mathbf{\Sigma}_r)$ conditional on the latent factors and other parameters from a normal-inverse-Wishart distribution:

$$(\mathbf{A}', \mathbf{\Sigma}_r \mid \cdot) \sim \mathcal{NIW}(\widehat{\mathbf{A}}', \mathbf{K}_{\mathbf{A}'}^{-1}, \widehat{\nu}_r, \widehat{\mathbf{S}}_r), \tag{7}$$

where

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

We can sample the normal-inverse-Wishart distribution in (7) in two steps. First, we sample from the inverse Wishart distribution for $\mathbf{\Sigma}_r$ marginally. Then given $\mathbf{\Sigma}_r$, we can sample from the normal distribution of $(\mathbf{A}' \mid \cdot)$.

With the constraints on the structure of $\mathbf{A}'$ for identification, we cannot directly sample from the above normal distribution. Here we outline the sampling scheme for $\mathbf{A}'$ with the 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 $(\text{vec}(\mathbf{A}') \mid \cdot) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}'), \mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})$ 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. It is important to note that given the Kronecker structure in the covariance matrix of the posterior, i.e., $(\mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})$, we are able to sample from the matrix normal distribution, and thus facilitate the computation. In particular, we first sample a $p_1 \times n$ matrix of independent samples from a standard normal distribution, denoted as $\mathbf{Z}$. Then let

$$\mathbf{A}_u' = \widehat{\mathbf{A}}' + \mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}^{-1'} \mathbf{Z} \mathbf{C}_{\mathbf{\Sigma}_r}', \tag{8}$$

where $\mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$ and $\mathbf{C}_{\mathbf{\Sigma}_r}$ are the Cholesky decomposition of $\mathbf{K}_{\mathbf{A}'}$ and $\mathbf{\Sigma}_r$. One can show that $(\mathbf{A}'_u \mid \cdot) \sim \mathcal{MN}(\widehat{\mathbf{A}}', \mathbf{K}_{\mathbf{A}'}^{-1}, \mathbf{\Sigma}_r)$, where $\mathcal{MN}$ denotes the matrix normal distribution. Therefore, $(\text{vec}(\mathbf{A}'_u) \mid \cdot) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}'), \mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})$.

Given the unrestricted $\mathbf{A}'_u$, we 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 $\mathbf{CC}'\mathbf{U} = \mathbf{M}'_{\mathbf{A}'}$ for $\mathbf{U}$;
(3) Solve $\mathbf{M}_{\mathbf{A}'}\mathbf{UV} = \mathbf{U}'$ for $\mathbf{V}$;
(4) Return $\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}'_u) + \mathbf{V}'(\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}'_u))$.

It can be shown that $\mathbf{A}'$ follows the distribution $(\mathbf{A}' \mid \cdot, \mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}') = \mathbf{a}_0)$.

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

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

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 $(\mathbf{B}', \mathbf{\Sigma}_c \mid \cdot)$ in two steps. First, we sample $\mathbf{\Sigma}_c$ marginally from $(\mathbf{\Sigma}_c \mid \cdot) \sim \mathcal{IW}(\widehat{\mathbf{S}}_c, \widehat{\nu}_c)$ 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 $\boldsymbol{\Sigma}_c = (\boldsymbol{L}^{-1})'(\boldsymbol{\Delta}^{-1})'\boldsymbol{\Delta}^{-1}\boldsymbol{L}^{-1}$.

(4) Exchange the row/column 1 and n of $\boldsymbol{\Sigma}_c$ back.

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

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

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

2.5 An Extension: Heteroskedastic Time-Varying Volatility

A more flexible way to model time-varying volatility is to incorporate heteroskedastic stochastic volatility processes for each of the variables included, 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:

$$\text{vec}(\mathbf{E}_t) \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t), \tag{9}$$

where $\mathbf{D}_t = \text{diag}(\mathrm{e}^{h_{1,1,t}}, \mathrm{e}^{h_{2,1,t}}, \ldots, \mathrm{e}^{h_{n,k,t}})$ is a diagonal matrix. The log-volatility follows a stationary AR(1) process with 0 mean similar to (3).

Compared to the specification in (1), this extension contains nk stochastic volatility processes, which can accommodate more complex volatility patterns. However, these volatility processes are assumed independent and there will be no cross-sectional correlation for rows and columns in the idiosyncratic components, which can be unrealistic in many practical applications. Moreover, this comes at a cost of more intensive posterior computations since natural conjugate priors cannot be applied to (9). Ultimately, there is always a trade-off between model complexity and computational burden, and the choice of specification depends on the application and its specific goals.

3 Model Comparison Using the Marginal Likelihood

When multiple models are available, a major challenge for practitioners is the lack of tools for comparing these models. In this section, we employ an importance-sampling estimator of the marginal likelihood to conduct model comparison. In particular, we are interested in determining the optimal dimension of the factor matrix, selecting between the VDFM and the MDFM, and distinguishing between exact factor models and approximate factor models.

3.1 A Bayesian Approach to Model Comparison

A natural Bayesian approach of model comparison involves computing the marginal likelihood for each available model and selecting the model that yields the highest marginal likelihood. Using marginal likelihoods for model comparison has several advantages. First, it accounts for model complexity and avoids overfitting by integrating over the entire parameter space, rather than replying on point estimates. This penalizes over-parameterized models, as more complex models spread probability mass over a larger parameter space, leading to lower marginal likelihoods unless justified by the data. Second, marginal likelihood is a consistent model selection criterion–if the true data-generating process is included in the set of candidate models, the marginal likelihood will asymptotically favor the correct model as more data becomes available. Moreover, Bayes factors, which are based on the ratio of marginal likelihoods between competing models, provide a direct measure of relative evidence in favor of one model over another.

Despite its strong theoretical foundation and automatic penalty for overfitting, marginal likelihood is often criticized for its high computational cost, especially in high-dimensional settings. For this reason, several alternative methods have been proposed. For example, Lopes and West (2004) introduced a reversible jump MCMC algorithm for moving between models that allows movement between models with different numbers of factors, while Lee and Song (2002) developed a path sampling approach to compute the Bayes factor efficiently. 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.

3.2 Estimating the Marginal Likelihood via Importance Sampling

In this paper, instead of computing the marginal likelihood directly, we estimate it using an importance-sampling estimator, which significantly reduces computational complexity while maintaining accuracy. Specifically, let $\boldsymbol{\theta}$ denote the unknown parameters, $g(\boldsymbol{\theta})$ be the importance sampling density, and $p(\boldsymbol{\theta})$ be the prior distribution. The marginal likelihood is estimated as follows:

$$\hat{p}_{IS}(\mathbf{y}) = \frac{1}{N}\sum_{n=1}^{N}\frac{p(\mathbf{y} \mid \boldsymbol{\theta}_n)p(\boldsymbol{\theta}_n)}{g(\boldsymbol{\theta}_n)}, \tag{10}$$

where $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n$ are independent draws from the importance sampling density $g(\boldsymbol{\theta})$. It is clear that this estimator is unbiased and consistent, assuming g dominates $p(\mathbf{y} \mid \cdot)p(\cdot)$.

The efficiency of the estimator depends critically on the choice of g. Ideally, if g is close to the posterior distribution, the estimator will have low variance. In this paper, we employ the cross-entropy method proposed by Chan and Eisenstat (2015) to find the optimal density within a given parametric family of distributions by minimizing the Kullback-Leibler divergence of the posterior distribution from the importance sampling density. This approach has two major advantages. First, using the importance-sampling density is convenient as it generates independent draws instead of correlated MCMC draws. Second, since the importance density is close to the posterior, the estimator exhibits low variance, requiring only a few thousand samples for accurate estimation. Specifically, Chan and Eisenstat (2015) show that, for a given parametric family of densities, the optimal hyperparameters correspond to the maximum likelihood estimators when posterior samples are treated as observed data.

To further facilitate 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 on integrating the factors and the choice of the importance sampling density are provided in Supplemental Appendix Section 3.

3.3 The Choice of the Importance Sampling Density

After integrating out the factors, the importance density is denoted as

$$\begin{aligned} f(\boldsymbol{\theta};\mathbf{v}) &= f(\mathbf{A},\mathbf{B},\boldsymbol{\Sigma}_r,\boldsymbol{\Sigma}_c,\boldsymbol{\rho},\boldsymbol{\lambda},\boldsymbol{\omega};\mathbf{v}) \\ &= f(\mathbf{A};\overline{\mathbf{A}},\overline{\mathbf{D}}_{\mathbf{A}})\cdot f(\mathbf{B};\overline{\mathbf{B}},\overline{\mathbf{D}}_{\mathbf{B}})\cdot f(\boldsymbol{\Sigma}_r;\nu_r,\Psi_r)\cdot \\ &\quad f(\boldsymbol{\Sigma}_c;\nu_c,\Psi_c)\cdot f(\boldsymbol{\rho};\overline{\boldsymbol{\rho}},\overline{\mathbf{D}}_{\boldsymbol{\rho}})\cdot f(\boldsymbol{\lambda};\nu_\lambda,S_\lambda)\cdot f(\boldsymbol{\omega};\mathbf{v}_{\boldsymbol{\omega}}). \end{aligned} \tag{11}$$

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 $f(\boldsymbol{\Sigma}_r;\nu_r,\Psi_r)$, where ν_c and ν_r are degrees of freedom, while Ψ_c and Ψ_r are scale matrices. 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. Moreover, we use inverse-gamma distributions for $f(\boldsymbol{\lambda};\nu_\lambda,S_\lambda)$.

Recall that we consider three specifications for time-varying volatility, $\boldsymbol{\omega}$. In the first specification, the log-volatility $\mathbf{h}$ follows an AR(1) process with an autoregressive coefficient ϕ and corresponding innovations σ_h^2. We use a normal distribution for $\mathbf{h}$, a truncated normal on $(-1,1)$ for ϕ and an inverse-gamma distribution for σ_h^2.

The optimal hyperparameters for the truncated normal and inverse-gamma distributions can be conveniently estimated via maximum likelihood estimation. However, handling $\mathbf{h}$ presents challenges. In specific, if we use a normal importance sampling density of the form $\mathcal{N}(\widehat{\mathbf{h}},\mathbf{K}_{\mathbf{h}}^{-1})$, one can obtain $\widehat{\mathbf{h}}$ and $\mathbf{K}_{\mathbf{h}}^{-1}$ analytically, where $\mathbf{K}_{\mathbf{h}}$ is a $T\times T$ full matrix. This approach has two key drawbacks. First, sampling from a Gaussian density with large full covariance matrix is computationally intensive. Second, in $\mathbf{K}_{\mathbf{h}}$ there are $T(T+1)/2$ parameters to be estimated. This means that a large number of posterior draws is needed to ensure accurate estimation.

To address these problems, we follow Chan (2023) and impose a restricted family of Gaussian density that exploits an AR process for the latent states $\mathbf{h}$, which includes the prior density of $\mathbf{h}$ in (3) as a special case. This restriction significantly reduces the parameter space from $T+T(T+1)/2$ to $2T+1$. Furthermore, this specification allows for analytical solutions for the autoregressive coefficients in the AR process, making the mean and covariance matrix of the normal importance sampling density straightforward

to obtain. We refer the reader to Chan (2023) for further details.

For the outlier components $\mathbf{o}$, we employ a discrete distribution over a predefined grid during estimation.[5] As a result, the probability of each grid point can be computed using the posterior draws of $\mathbf{o}$. We then use the beta distribution for the the outlier probability $p_{\mathbf{o}}$. For the fat-tailed innovations, we adopt an inverse-gamma distribution for q_t^2. Further details on the procedure can be found in Supplemental Appendix Section 3.

4 Monte Carlo Studies

In this section, we first assess the accuracy of the factor estimates by comparing them to their true values across datasets of varying sizes. Then, we evaluate whether the marginal likelihood estimator can accurately identify the true model.

4.1 Performance of Factor Estimators under Different Sample Sizes and Dimensions of 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 drawn from the uniform distribution, $\mathcal{U}(0, 1)$, and $\rho_{j,k}$ is drawn from $\mathcal{U}(0.8, 0.9)$ for $j = 1, \ldots, p_1$, $k = 1, \ldots, p_2$. 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 $j = 1, \ldots, p_1$, $k = 1, \ldots, p_2$. To assess the accuracy of our factor estimates, 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

[5] Following Carriero et al. (2024b), we use $(1, 20)$ as the support (grid) for $\mathbf{o}$.

[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 VDFM (1,000 posterior draws after 1,000 burn-in draws) work well

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.[7] 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 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 Supplemental Appendix Section 4.

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 VDFM and static matrix factor model.[8]

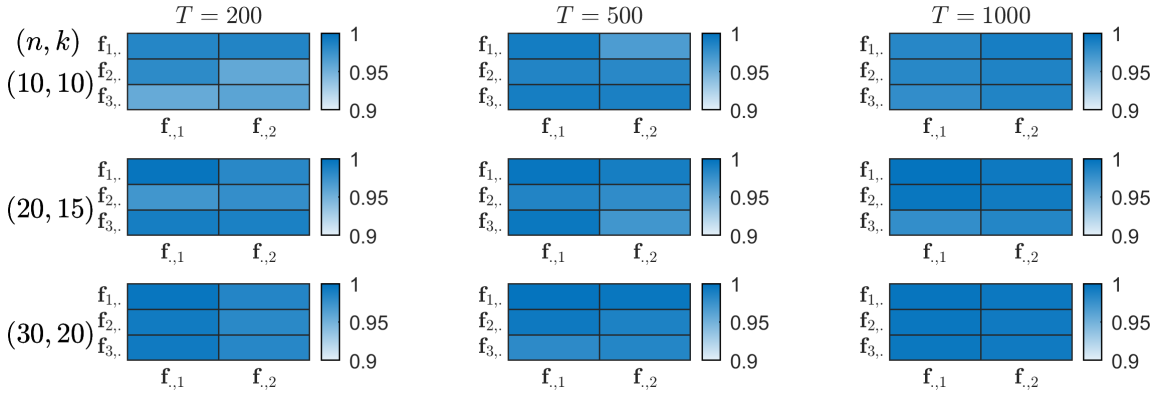

Figure 1: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 3$, $p_2 = 2$

as initial values. Geweke statistics are computed to ensure the convergence of Markov chains.

[7]For example, the first row represents $(n, k) = (10, 10)$, while the second row represents $(20, 15)$.

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

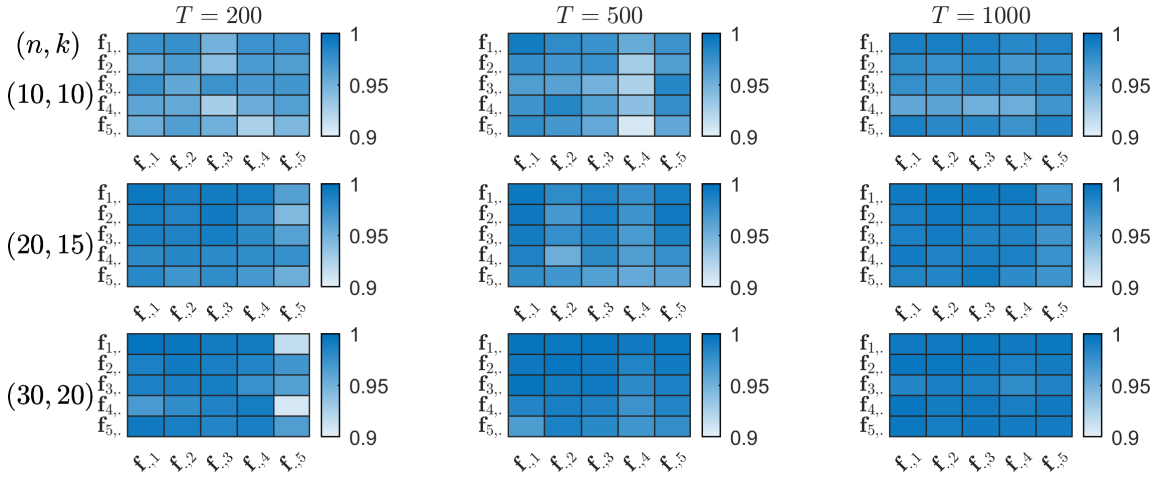

Figure 2: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 5$, $p_2 = 5$

4.2 Can the Marginal Likelihood Estimator Identify the Correct Dimension of the Factor Matrix?

To evaluate the performance of the marginal likelihood estimator in correctly identifying the true dimension of the factor matrix, we estimate log marginal likelihoods for models with a variety of combinations of (p_1, p_2). Specifically, we use four datasets from Section 4.1:

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

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 presents the estimates for log marginal likelihoods for the four datasets. 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 $\log \widehat{p}(\mathbf{y} \mid p_1 = 1) < \log \widehat{p}(\mathbf{y} \mid p_1 = 2) < \log \widehat{p}(\mathbf{y} \mid p_1 = 3)$ and $\log \widehat{p}(\mathbf{y} \mid p_1 = 3) > \log \widehat{p}(\mathbf{y} \mid p_1 = 4) > \log \widehat{p}(\mathbf{y} \mid p_1 = 5)$ is observed. Similarly, $\log \widehat{p}(\mathbf{y} \mid p_2 = 1) < \log \widehat{p}(\mathbf{y} \mid p_2 = 2)$ and $\log \widehat{p}(\mathbf{y} \mid p_2 = 2) > \log \widehat{p}(\mathbf{y} \mid p_2 = 3) > \log \widehat{p}(\mathbf{y} \mid p_2 = 4) > \log \widehat{p}(\mathbf{y} \mid p_2 = 5)$.

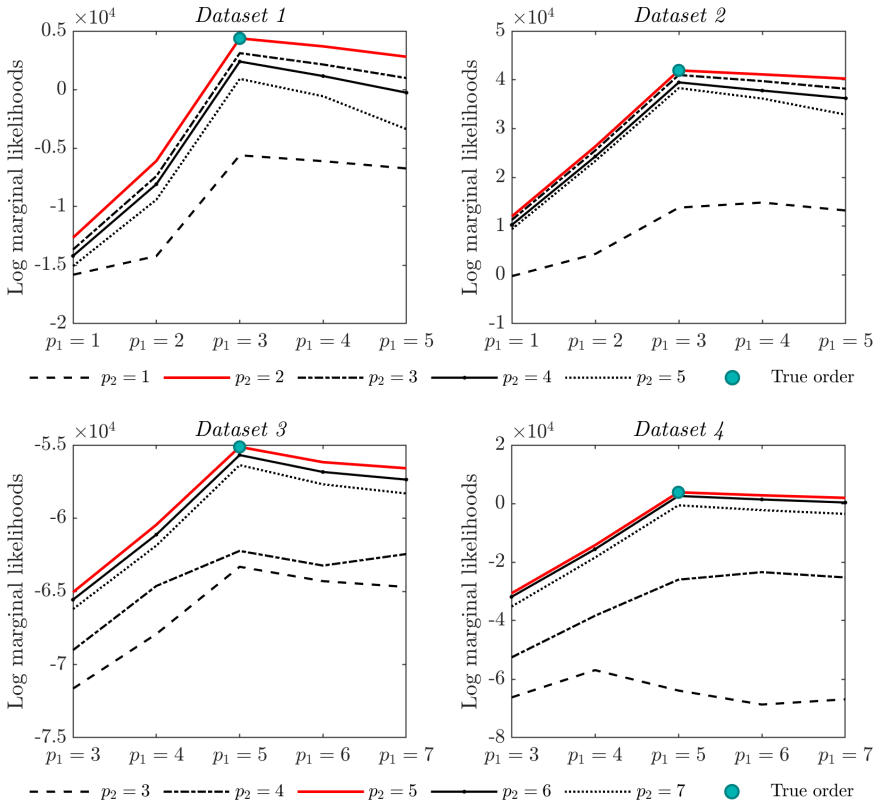

Figure 3: Log marginal likelihood estimates for the four datasets. From top left to bottom right, the panels correspond to $(n, k, p_1, p_2) = (10, 10, 3, 2)$, $(20, 15, 3, 2)$, $(10, 10, 5, 5)$, and $(20, 15, 5, 5)$. Each line represents a different value of p_2, and the blue dot represents the true value for p_1 and p_2.

4.3 Can Marginal Likelihood Distinguish VDFM from MDFM?

In this subsection, we examine the ability of the marginal likelihood estimator to correctly identify the underlying factor structure. Specifically, we generate data from two types of dynamic factor models: a VDFM defined in Equations 4 and an MDFM. When the

true model is VDFM, the free elements in loading matrix $\mathbf{M}$ are randomly drawn from a uniform distribution $\mathcal{U}(-1,1)$, and the error covariance matrices for $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\nu}_t$ are set to identity matrices. The autoregressive coefficients in the factor evolution equation are drawn from a uniform distribution $\mathcal{U}(0.7,0.95)$. When the true model is MDFM, the free elements in the two loading matrices $\mathbf{A}$ and $\mathbf{B}$ are drawn from the uniform distribution $\mathcal{U}(0,1)$, the row-wise and column-wise covariance matrices $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_r$ are $0.5\mathbf{I}_n$ and $0.3\mathbf{I}_n$, respectively. The autoregressive coefficients in the factor evolution equation are drawn from $\mathcal{U}(0.8,0.9)$. The covariance matrices for factor evolution equation are identity matrices.

First, we generate data using the MDFM with factor matrices of sizes (1×2), (2×1), and (2×2), and estimate the marginal likelihood of each (true) model. Then we compare these marginal likelihoods with those of VDFMs specified with $k_f = 1,2,\ldots,6$ factors, where k_f denotes the number of factors in the VDFM. Figure 4 shows the log marginal likelihood estimates under different model specifications. The left, middle, and right panels correspond to MDFMs with $(p_1,p_2) = (1,2)$, $(2,1)$, and $(2,2)$, respectively. The black line shows the log marginal likelihood estimates for the VDFMs, while the red dashed line indicates the marginal likelihood of the true MDFM.

The results clearly show that the marginal likelihood for the true model (MDFM) is higher than for any of the VDFM alternatives. Interestingly, the VDFM achieves its highest marginal likelihood when its number of factors matches the total number of factors in the corresponding MDFM. For instance, the VDFM performs best when $k_f = 2$ for MDFMs with $(p_1,p_2) = (1,2)$ and $(2,1)$, and when $k_f = 4$ for $(p_1,p_2) = (2,2)$.

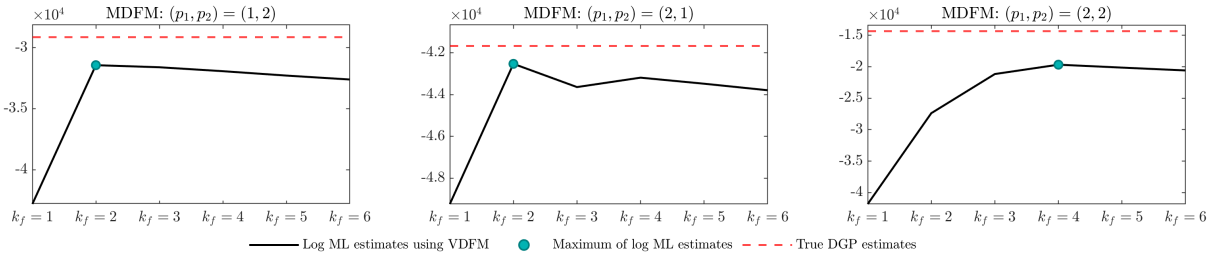

Figure 4: Log marginal likelihoods under the VDFM relative to the true model – a MDFM with a 2×2 factor matrix. A negative value indicates that the correct model is favored.

Next, we generate data from the VDFM with $k_f \in 1,2,3$ and estimate the log marginal likelihood for each true model. We compare these values with those from MDFMs with

$(p_1, p_2) \in (1,1), (1,2), (2,1), (2,2)$. Figure 5 presents the log marginal likelihood estimates under different model specifications. The left, middle, and right panels correspond to VDFMs with $k_f = 1$, 2, and 3, respectively. The black dashed line indicates the log marginal likelihood of the true VDFM, while the bars represent those of the competing MDFMs. In all three cases, the true model achieves a higher marginal likelihood than the alternatives.

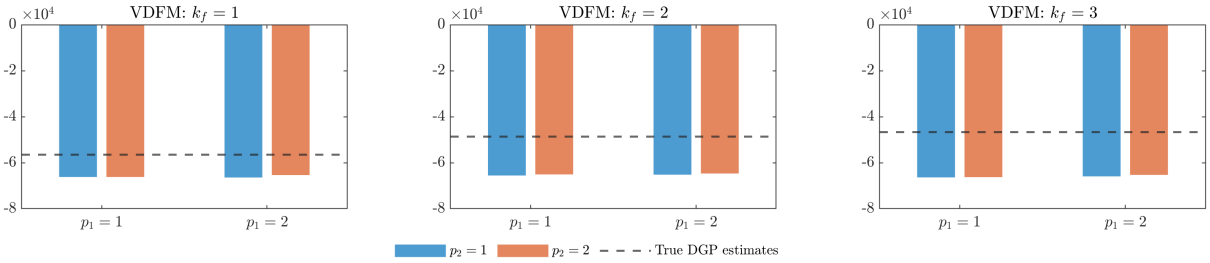

Figure 5: Log marginal likelihoods under the MDFM relative to the true models – VDFMs with one (left panel), two (middle panel) and three factors (right panel). A negative value indicates that the correct model is favored.

4.4 Can the Marginal Likelihood Estimator Distinguish Approximate DFMs from Exact DFMs?

In this subsection, we investigate whether the marginal likelihood estimator can correctly identify the structure of the idiosyncratic components. Specifically, we assess whether the estimator can distinguish an exact matrix factor model (i.e., without stochastic volatility or cross-sectional correlations) from approximate alternatives. To this end, we generate 100 datasets from the following models:

MDFM-exact: A matrix factor model with a diagonal covariance matrix and no stochastic volatility.

MDFM-cross: A matrix factor model with a Kronecker-structured covariance matrix and no stochastic volatility.

MDFM-sv: A matrix factor model with a diagonal covariance matrix and a common stochastic volatility.

The free elements in the two factor loading matrices $\mathbf{A}$ and $\mathbf{B}$ are drawn from a normal

distribution $\mathcal{N}(0, 0.3^2)$. The row-wise and column-wise covariance matrices are drawn from inverse-Wishart distributions: $\mathbf{\Sigma}_r \sim \mathcal{IW}(n+2, \mathbf{I}_n)$ and $\mathbf{\Sigma}_c \sim \mathcal{IW}(k+2, \mathbf{I}_k)$. The log-volatility follows an AR(1) process with an autoregressive coefficient of 0.97 and innovation variance 0.1. The covariance matrix in the factor evolution equation is $0.1\mathbf{I}_{p_1 p_2}$, and the AR coefficients are drawn from $\mathcal{U}(0.8, 0.9)$.

We generate 100 datasets with dimensions $(n, k, T) = (20, 20, 100)$. The dimension of the factor matrix $(p_1, p_2) = (2, 2)$. For each dataset, we estimate the log marginal likelihoods under the MDFM-exact, MDFM-cross and MDFM-sv specifications. Then we subtract the log marginal likelihood of the true model from those of the competing models. A negative value indicates that the marginal likelihood is larger for the true model.

Figure 6 shows boxplots of the differences in log marginal likelihoods relative to the true model. The left, middle and right panel corresponds to cases where the true model is MDFM-exact, MDFM-cross, and MDFM-sv, respectively. In all cases, the true model achieves the highest marginal likelihood, suggesting that the estimator effectively identifies the correct structure in the idiosyncratic component.

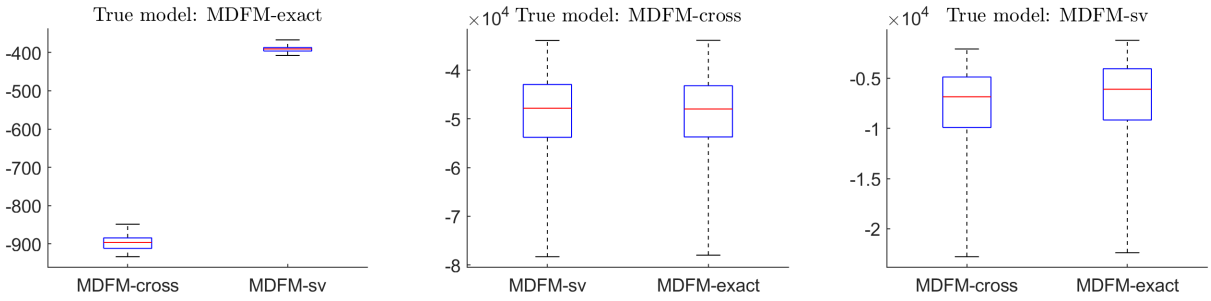

Figure 6: Boxplots of differences of log marginal likelihoods relative to the true model: MDFM-exact (left panel), MDFM-cross (middle panel) and MDFM-sv (right panel). A negative value indicates that the correct model is favored.

5 Empirical Applications

In this section, we demonstrate the usefulness of the Bayesian matrix dynamic factor models 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 optimal factor structure is 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 Supplemental Appendix Section 5.

While theoretically one could compute marginal likelihood estimates to assess different orders of countries and variables, this approach is computationally intensive due to the large number of combinations. For this reason, we prioritize the global and regional economic significance of countries and the relationships among variables when ordering the data. Particularly, in terms of countries, we order the US the first, due to its status as the largest economy and its significant influence on the global economy. The UK, Australia, Germany and Japan follow, due to their significant position in the corresponding regional economy. Among the indicators, real GDP is prioritized as it serves as the critical measure of real economic activity, following by Headline CPI, given its importance as a key inflation indicator closely tied to price dynamics. We put labor unit cost the third, since it is crucial for insights into productivity.

We then employ the method for marginal likelihood estimation introduced in Section 3 to determine the optimal factor structure. Tables 1–2 report the log marginal likelihood estimates for VDFMs and MDFMs. Comparing the VDFM without stochastic volatility to its counterpart with common stochastic volatility (the left and right panel of Table 1), we find that the data strongly favor the inclusion of stochastic volatility. A similar pattern is observed for the MDFMs (in Table 2). Additionally, the results indicate that the model with cross-sectional correlation is preferred over the exact one, suggesting that accounting for both cross-sectional correlation and time-varying volatility leads to improved model performance. Notably, the marginal likelihood for the VDFM is maximized with five factors, while the approximate MDFM achieves a better fit with a more parsimonious 1×2 factor matrix $\mathbf{F}_t$. This suggests that ignoring the matrix structure can lead to

overestimation of the number of factors in VDFM. The 1×2 factor matrix implies that the variation is larger across indicators than across countries.

Table 1: Log marginal likelihood estimates using VDFMs

LDFM-exact				LDFM-sv			
$k=1$	$k=2$	$k=3$	$k=4$	$k=1$	$k=2$	$k=3$	$k=4$
-26361	-24786	-23602	-23017	-22715	-21642	-21272	-20910
(0.1)	(0.3)	(0.9)	(1.3)	(0.2)	(0.6)	(0.7)	(1.0)
$k=5$	$k=6$	$k=7$	$k=8$	$k=5$	$k=6$	$k=7$	$k=8$
-22944	-23001	-23107	-23280	**-20862**	-21130	-21202	-21371
(1.6)	(1.6)	(2.3)	(4.3)	(2.0)	(3.2)	(4.2)	(6.0)

Table 2: Log marginal likelihood estimates using MDFMs

	MDFM-exact			MDFM-cross			MDFM-cross-sv		
	$p_2=1$	$p_2=2$	$p_2=3$	$p_2=1$	$p_2=2$	$p_2=3$	$p_2=1$	$p_2=2$	$p_2=3$
$p_1=1$	-26472	-24796	-23493	-17968	**-17930**	-17950	-16144	**-16140**	-16164
	(0.2)	(0.3)	(0.7)	(0.3)	(0.4)	(0.8)	(0.6)	(0.4)	(0.5)
$p_1=2$	-26310	-24607	-23221	-18007	-17979	-18068	-16260	-16265	-16337
	(0.1)	(0.4)	(0.4)	(0.4)	(0.4)	(1.4)	(0.5)	(0.6)	(1.0)
$p_1=3$	-26164	-24469	**-23012**	-18051	-18058	-18150	-16336	-16386	-16464
	(0.3)	(0.6)	(0.7)	(0.3)	(0.8)	(0.6)	(0.7)	(0.7)	(1.1)

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. We then group countries and indicators by comparing these posterior probabilities against a 0.9 threshold: when the probability exceeds 0.9, it indicates that the neighboring values are significantly different, so they are placed in separate groups.

Figure 7 displays the bar plot of sorted estimates for $\widehat{\mathbf{A}}$.[9]The 19 countries are grouped into four categories: (1) Japan; (2) seven European countries and Korea; (3) six other

[9]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.

European countries along with two Oceania countries; and (4) two North American countries. The fact that the two North American countries fall into the same group suggests that geographic proximity influences the grouping, but geography is clearly not the only factor. For example, Japan and Korea are placed in different groups, while Oceania and several European countries are grouped together.

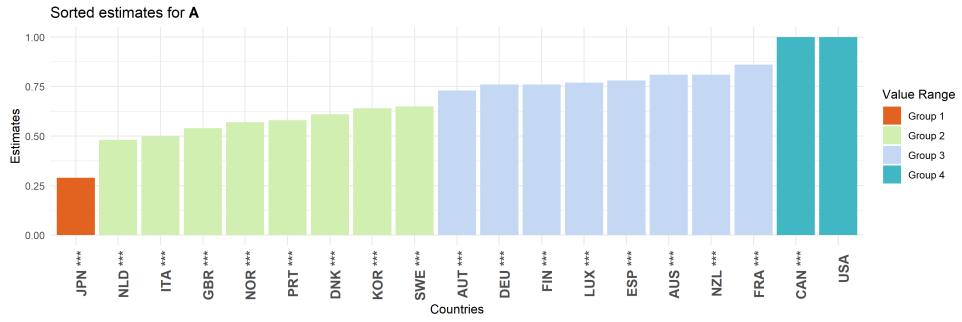

Figure 7: Bar plots of sorted estimates for loading matrix $\mathbf{A}$. The 19 countries are categorized into 3 distinct groups based on the posterior probabilities that the differences between neighboring values are greater than 0. The stars on the country labels show the significance level of the corresponding estimates. There is no significance level on USA because we fix the corresponding element in $\mathbf{A}$ to be 1.

Figure 8 contains two rows of subplots. 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 representing the real economic activity clearly capture the 2008 Great Recession and the disruptions caused by the COVID-19 pandemic, though the early 2000s recession is less pronounced. The second factor captures price dynamics, but do not directly affect the real output as measured by real GDP. Figure 9 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 the two 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 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 real activity factor has stronger impacts on

international trade and labor unit cost, while also having a positive, albeit less pronounced, impact on consumption, headline CPI, and energy CPI. However, its effects on core CPI and food CPI are not statistically significant. Additionally, it negatively affects unemployment.

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 price factor has strong positive impacts on energy prices, with less effect on labor unit cost and international trade. It has statistically insignificant positive effects on core CPI and negative effects on consumption, food CPI, and unemployment.

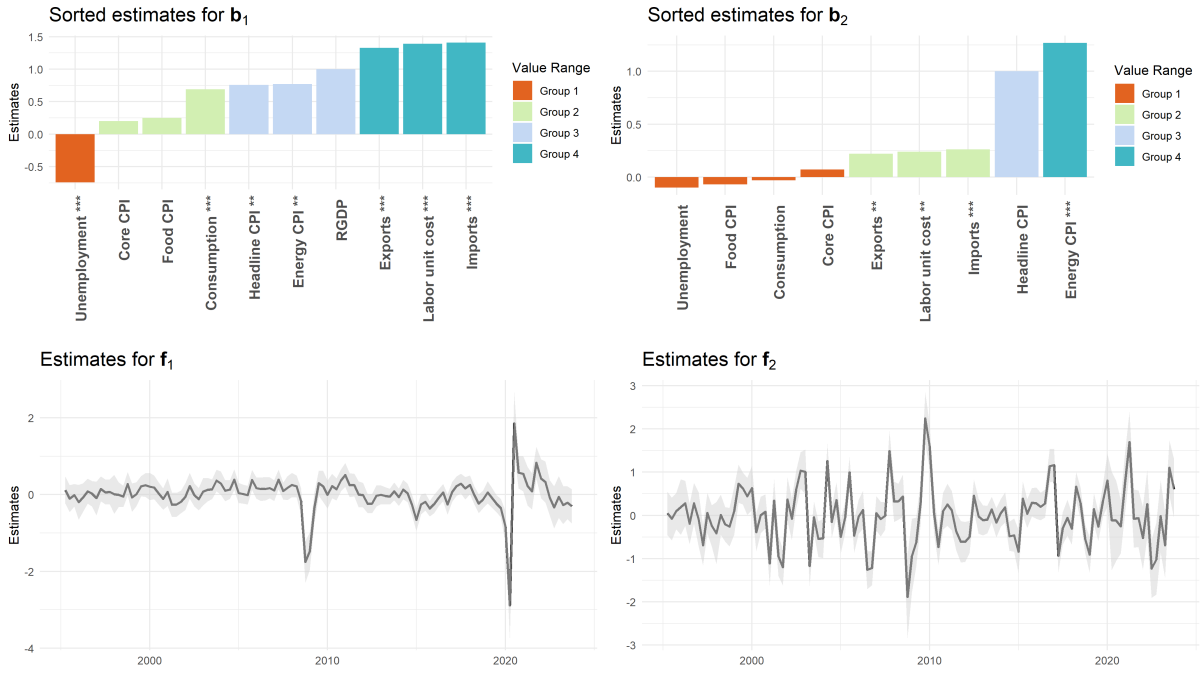

Figure 8: Bar plots of sorted estimates for loading matrix $\mathbf{B}$ and plots for factor estimates. The stars show the significance level of the corresponding estimates. According to impacts of the two column factors, the variables can be divided into 4 groups. The shaded area of plots for the factors is the 90% credible intervals.

Figure 10 presents the estimates for stochastic volatility and their standard deviations. The high volatility around 1997 reflects the turbulence of the Asian financial crisis, par-

ticularly in Japan and Korea. Expectedly, increased volatility is also observed during the Great Recession and the COVID-19 pandemic.

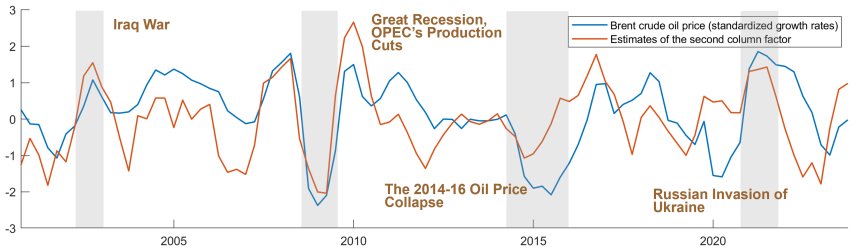

Figure 9: Yearly moving average of standardized growth rates of Brent crude oil price and the second column factor estimates.

Figure 10: Estimates for stochastic volatility: $\widehat{\omega}_t = \exp(\widehat{h}_t/2)$

Figure 11–12 are heatmaps of estimates for column-wise and row-wise covariance matrix in the idiosyncratic component. From Figure 11, it is obvious that headline CPI is positively correlated with its disaggregated components: energy CPI, core CPI and food CPI. In addition, unemployment is negatively correlated with real GDP, labor unit cost, consumption, core CPI and imports. Labor unit cost is positively correlated with both exports and imports. In Figure 12, 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.

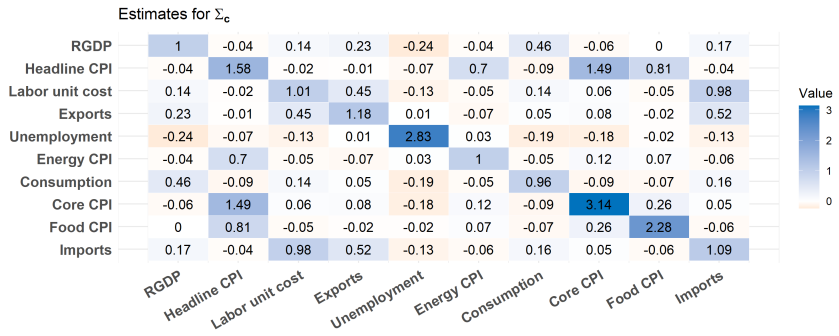

Figure 11: Heatmap of estimates for $\mathbf{\Sigma}_c$

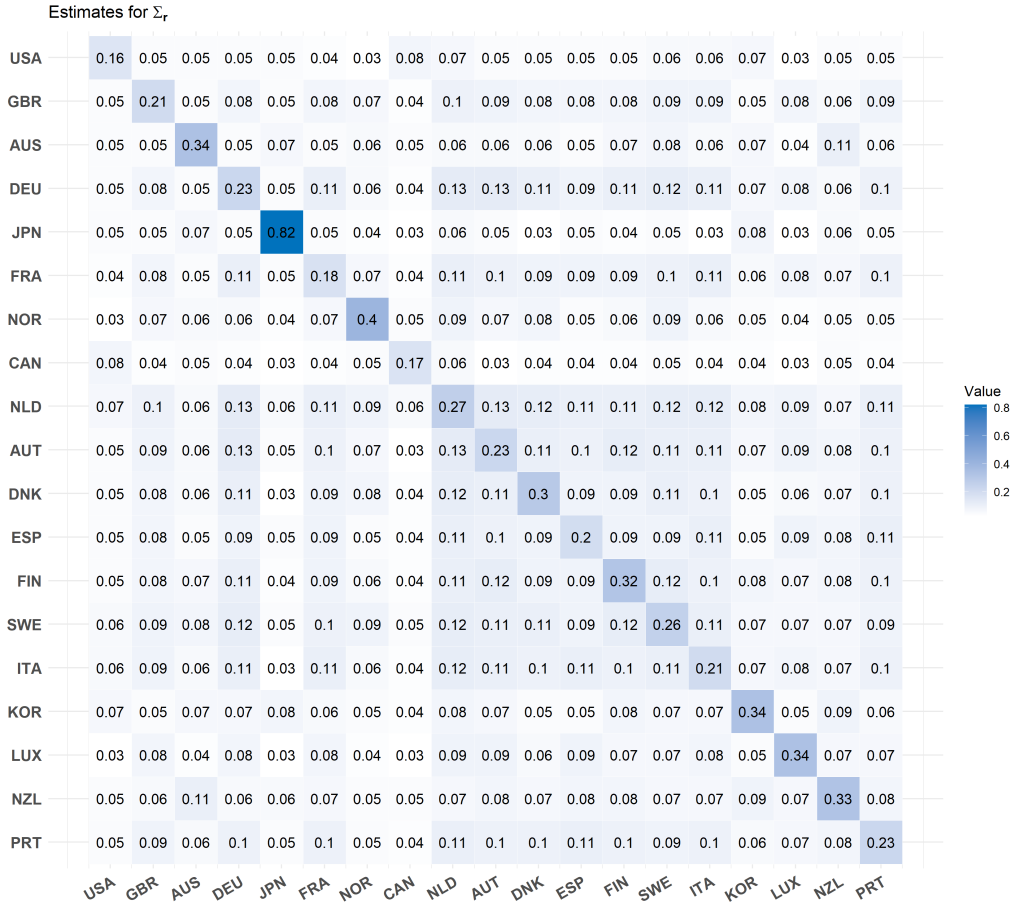

Figure 12: Heatmap of estimates for $\mathbf{\Sigma}_r$

5.2 Fama-French 10 × 10 Panel

In this application, we investigate the usefulness of the dynamic matrix factor model on 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).[10] The return series span from January 1990 to June 2024 (414 observations).[11] 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. The standardized market-adjusted return series of the portfolios can be found in Supplemental Appendix Section 5.

Empirical evidence shows that small-cap stocks tend to earn higher average returns than large-cap stocks, while value stocks (high BE/ME) tend to outperform growth stocks (low BE/ME). To account for these cross-sectional return patterns, Fama and French (1992) introduce the Small Minus Big (SMB) and High Minus Low (HML) factors to explain return variations across stocks with varying size and book-to-market ratios. Motivated by this framework, we reorder the size and BE/ME ratios in the data matrix. Specifically, portfolios are ordered by size across rows, with small-cap (SMALL) portfolios listed first, followed by medium-cap (ME5), and then large-cap (BIG) portfolios. Similarly, columns are arranged by book-to-market ratio, with high BE/ME (HiBM) portfolios on the left, followed by medium (BM5), and then low BE/ME (LoBM) portfolios.

Table 3–4 report the log marginal likelihood estimates for both VDFMs and MDFMs. The results clearly indicate a strong preference for models with stochastic volatility, regardless of whether the VDFM or MDFM is used. Additionally, a comparison between the MDFM-exact and MDFM-cross reveals that accounting for cross-sectional correlation in idiosyncratic component further improves model performance. Among the MDFM specifications, the 2×2 MDFM with both cross-sectional correlation and stochastic volatility

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

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

achieves the highest marginal likelihood (–42,951), although it remains slightly lower than that of the 4-factor VDFM (–42,943). The best overall performance is attained by the 5-factor VDFM with stochastic volatility, which yields a marginal likelihood of –42,877. These results suggest that the data favor the more flexible VDFM specification over the more structured MDFM, despite the marginal likelihood's built-in penalty for model complexity. For illustration, we use the MDFM with a factor matrix of dimensions $(p_1, p_2) = (2, 2)$ and stochastic volatility in the subsequent analysis.

Table 3: Log marginal likelihood estimates using VDFM

VDFM-exact				VDFM-sv			
$k=1$	$k=2$	$k=3$	$k=4$	$k=1$	$k=2$	$k=3$	$k=4$
-51647	-47191	-46217	-45940	-46096	-43942	-43069	-42943
(0.1)	(0.2)	(0.4)	(0.5)	(0.4)	(0.4)	(0.5)	(0.8)
$k=5$	$k=6$	$k=7$	$k=8$	$k=5$	$k=6$	$k=7$	$k=8$
-45787	**-45729**	-45792	-45886	**-42877**	-43007	-43155	-43328
(0.6)	(0.9)	(1.0)	(1.0)	(0.7)	(0.8)	(1.6)	(0.8)

Table 4: Log marginal likelihood estimates using MDFM

	MDFM-exact			MDFM-cross			MDFM-cross-sv		
	$p_2=1$	$p_2=2$	$p_2=3$	$p_2=1$	$p_2=2$	$p_2=3$	$p_2=1$	$p_2=2$	$p_2=3$
$p_1=1$	-51867	-49622	-49382	-47210	-46809	-46739	-43527	-43163	-43161
	(0.2)	(0.3)	(0.4)	(0.2)	(0.4)	(0.4)	(1.3)	(1.4)	(1.4)
$p_1=2$	-49603	-47130	-46572	-46897	-46557	-46164	-43304	**-42951**	-43024
	(0.3)	(0.4)	(0.4)	(0.4)	(0.4)	(0.7)	(2.0)	(0.9)	(2.9)
$p_1=3$	-49285	-46847	**-46177**	-46928	-46553	**-46087**	-43406	-43075	-43027
	(0.4)	(0.5)	(0.9)	(0.6)	(0.9)	(0.8)	(1.3)	(1.2)	(1.4)

Figure 13 shows the estimates for row loading matrices (the first and second panel from left) and column loading matrices (the third and fourth panel). In specific, the two sub-plots correspond to $\widehat{\mathbf{A}}_{\cdot,1}$ (first), $\widehat{\mathbf{A}}_{\cdot,2}$ (second), $\widehat{\mathbf{B}}_{\cdot,1}$ (third) and $\widehat{\mathbf{B}}_{\cdot,2}$ (fourth). Both the size (row) loadings and the BE/ME (column) loadings have shown cross-sectional patterns. Particularly, the small-cap factor exerts a strong influence on smaller portfolios, with its

impact gradually decreasing as portfolio size increases, eventually turning negative for large-size portfolios. The medium-cap factor similarly influence portfolios with similar sizes, with its influence tapering off as the portfolio size shifts either smaller or larger. Similar patterns go for BE/ME factors.

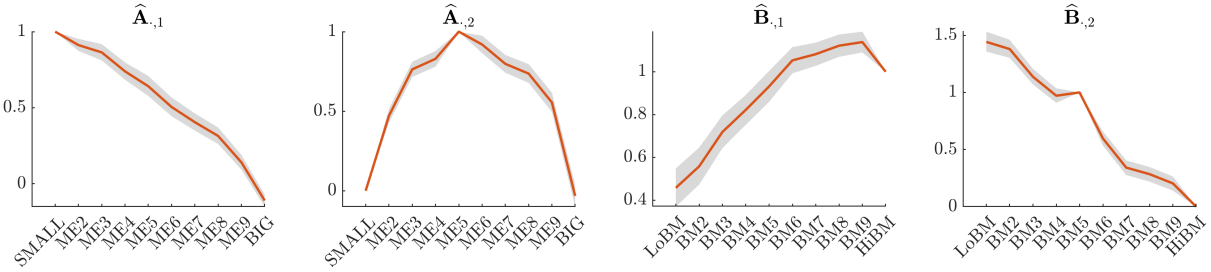

Figure 13: Estimates for row loading matrices (the first and second panel from left) and column loading matrices (the third and fourth panel). The gray area represents the 90% credible interval.

Figure 14 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 suggests that an AR process for factor evolution is supported by the data.

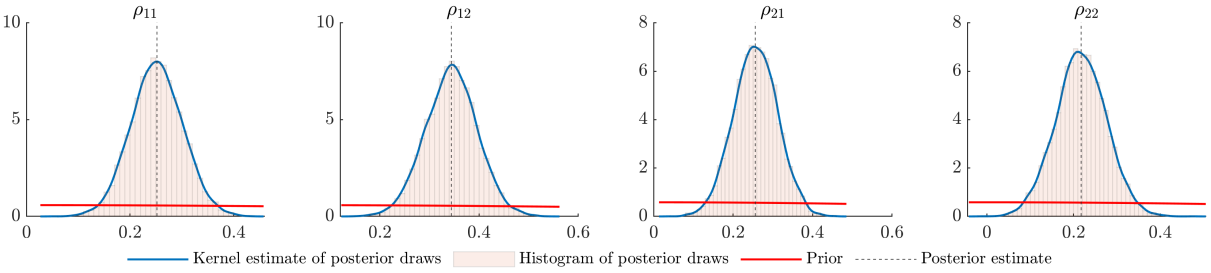

Figure 14: Posterior densities, histograms of posterior draws, priors and posterior estimates for autogressive coefficients $\boldsymbol{\rho}$

Figure 15 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, about one month before the onset of the dot-com bubble burst. Additionally, signifi-

cant spikes in volatility are observed during the 2008 financial crisis and the COVID-19 pandemic.

Figure 15: Estimates and standard errors of stochastic volatility: $\exp(\mathbf{h}/2)$

6 Conclusion and Future Research

In this paper, we propose a new class of dynamic factor models tailored to high-dimensional matrix-valued time series, incorporating cross-sectional correlations in the idiosyncratic components, time-varying volatility and outlier adjustments. We develop an MCMC algorithm for Bayesian estimation and introduce an importance-sampling estimator for the marginal likelihood to facilitate Bayesian model comparison. Monte Carlo simulations demonstrate the accuracy of the factor estimates and the ability of the marginal likelihood estimator to correctly identify the true model. Applications to macroeconomic and Fama-French panels highlight the model's ability to uncover interesting structures in high-dimensional data.

This research opens several avenues for future work. First, a structural matrix factor model could be developed to study the transmission of shocks across countries. Second, it would be valuable to assess the forecasting performance of matrix factor models compared to traditional approaches using vectorized panels. Finally, the current specification requires the number of factors to match the product of the matrix dimensions; a sparse matrix factor model could be designed to automatically select the most relevant latent factors.

References

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

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.

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

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

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

Boivin, J. and S. Ng (2006): "Are more data always better for factor analysis?" *Journal of Econometrics*, 132, 169–194.

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. (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 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., 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.

Fama, E. F. and K. R. French (1992): "The cross-section of expected stock returns," *the Journal of Finance*, 47, 427–465.

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

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

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.

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.

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.

Qin, L., Y. Wang, Y. Zhu, and B.-C. Shia (2025): "Bayesian Dynamic Matrix Factor Models," *Journal of Business & Economic Statistics*, 1–13.

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*.