

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

Joshua C. C. Chan<sup>a</sup>, Wei Zhang<sup>b</sup>

<sup>a</sup>*Department of Economics, Purdue University*

<sup>b</sup>*Department of Economics, Johns Hopkins University*

---

## Abstract

We introduce a class of Bayesian dynamic factor models for matrix-valued time series, with autoregressive factor dynamics and idiosyncratic components that allow stochastic volatility, outliers, and a Kronecker-structured covariance capturing cross-row and cross-column correlation. Exploiting the matrix structure, we make these richly parameterized models tractable in high dimensions and develop an efficient Gibbs sampler for estimation. For model comparison, we propose a unified approach based on the cross-entropy importance-sampling estimator of the marginal likelihood, which under a common criterion selects the factor dimension, a vector versus matrix structure, and the idiosyncratic specification. Monte Carlo experiments confirm that the estimator reliably recovers the true model. In an application to an OECD macroeconomic panel of 190 time series, the data favor both cross-sectional correlation and stochastic volatility, and the model delivers statistically significant out-of-sample forecast gains over a static matrix factor benchmark.

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

*JEL Codes:* C11, C32, C55

---

---

\*We would like to thank Marta Bañbura, Domenico Giannone, Michele Lenza, Elena Bobeica, Catalina Martínez Hernández, Danilo Leiva-León, and Carlos Montes-Galdón for many constructive suggestions. We are also grateful for insightful discussions with participants at the DG-Economics Internal Seminar at the European Central Bank, the 94<sup>th</sup> SEA Annual Meeting, the 2025 Örebro Workshop on Macro and Financial Econometrics, and the 2025 CFE-CMStatistics Conference. All remaining errors are our own.

## 1. Introduction

Dynamic factor models are a workhorse for summarizing the comovement in large macroeconomic panels, and they underpin much of modern empirical macroeconomics and forecasting (Stock and Watson, 2002a; Forni et al., 2000; Bernanke et al., 2005; Doz et al., 2012). An increasing share of these panels is naturally matrix-valued: a set of macroeconomic indicators is observed for a cross-section of countries over time, so that each period delivers an indicator-by-country matrix. Such multi-country panels are a leading application of factor analysis, where a small number of common factors summarize the world, regional, and country-specific sources of macroeconomic comovement (Kose et al., 2003; Del Negro and Otrok, 2008). The standard practice is to stack the matrix into a long vector and fit a vector dynamic factor model or a large Bayesian vector autoregression (Bańbura et al., 2010; Koop, 2013; Carriero et al., 2019; Chan, 2023). This vectorization discards the matrix structure of the data—the strong dependence among indicators within a country and among countries for a given indicator—and the number of loadings grows with the product of the two cross-sectional dimensions, which strains estimation as the panel widens.

A parsimonious alternative is to exploit the matrix structure directly. Wang et al. (2019) introduce a matrix factor model in which row and column loadings enter through a bilinear form, subsequently extended with linear constraints (Chen et al., 2020), thresholds (Liu and Chen, 2019), and time-varying loadings (Chen et al., 2024). Modeling the row and column loadings separately sharply reduces the number of free loadings, making the matrix structure attractive in high dimensions. These models, however, omit two features essential for macroeconomic applications. First, they are static: the factor dynamics are left unspecified, so they cannot produce iterated multi-step forecasts. Second, the idiosyncratic errors are assumed homoskedastic, ruling out the time-varying volatility, heavy tails, and outliers pervasive in macroeconomic data (Cogley and Sargent, 2005; Justiniano and Primiceri, 2008; Stock and Watson, 2016; Chan, 2023). We take up the matrix structure as a parsimony device and build a fully Bayesian dynamic factor model that addresses both.

This paper makes two methodological contributions. First, we introduce a class of Bayesian dynamic factor models for matrix-valued time series, designed around the empirical features of macroeconomic and financial data. The latent factors follow autoregressive processes, which capture the persistent comovement characteristic of common

components and support recursive multi-step forecasting.<sup>1</sup> We allow rich dynamics and dependence in the idiosyncratic components. Along the time dimension, we incorporate stochastic volatility (Carriero et al., 2016, Kastner et al., 2017), heavy-tailed errors (Jacquier et al., 2004), and outlier components to absorb large, infrequent extreme observations such as those around the COVID-19 pandemic (Carriero et al., 2024, Chan et al., 2025). Along the cross-section, we adopt a Kronecker structure built from two full covariance matrices that captures the cross-row and cross-column correlations left unexplained by the common components.

Bringing these features together yields a flexible but heavily parameterized model, and the matrix structure is what keeps estimation tractable. Exploiting the Kronecker form of the likelihood, we develop an efficient Gibbs sampler that scales to the dimensions encountered in macroeconomic panels. To identify the factors and loadings, we extend the lower-triangular scheme of Bai and Wang (2015) to the matrix setting by anchoring both the row and column loading matrices, and we sample the constrained loadings using the fast hyperplane-truncation algorithm of Cong et al. (2017).

The second contribution is a unified approach to model comparison. The matrix factor literature has relied on eigenvalue-ratio criteria to select the number of factors. Empirical practice, however, raises broader questions of model structure: whether the data support a matrix factor structure relative to a vector dynamic factor model (VDFM),<sup>2</sup> whether the idiosyncratic component is homoskedastic or exhibits time-varying volatility, and whether weak cross-sectional correlation is present. We address these questions through the marginal likelihood, which we compute for each model using the cross-entropy importance-sampling estimator of Chan and Eisenstat (2015). This estimator draws independently from a calibrated importance density—obtained by minimizing, within a given parametric family, the Kullback–Leibler divergence to the posterior—rather than from correlated Markov chain Monte Carlo samples. Casting the choice of the factor dimension, the matrix-versus-vector structure, and the idiosyncratic specification as a Bayesian model comparison renders these distinct questions directly comparable under a common criterion (Geweke and Zhou, 1996; Chan and Eisenstat, 2018; Fong and Holmes, 2020).

---

<sup>1</sup>See, e.g., Sargent et al. (1977), Stock and Watson (2012), Poncela et al. (2021) for autoregressive factor dynamics in the vector dynamic factor literature.

<sup>2</sup>We denote a matrix dynamic factor model by MDFM and the vector counterpart by VDFM; the “m” and “v” distinguish the matrix and vector formulations.

We demonstrate the effectiveness of the cross-entropy importance-sampling estimator at distinguishing among competing specifications in a series of Monte Carlo experiments. Across all three comparison dimensions—the factor dimension, the matrix-versus-vector structure, and the idiosyncratic specification—the estimator reliably assigns the highest marginal likelihood to the data-generating model. The simulations also confirm that the factor estimates are close to their true values, with accuracy improving in longer samples. We illustrate the usefulness of the proposed MDFM in an application with an OECD macroeconomic panel of 19 countries and 10 indicators, a total of 190 time series. The data decisively favor specifications with both cross-sectional correlation and stochastic volatility—features absent from the static matrix factor model—and the estimated loadings display clear grouping patterns across countries and indicators. The estimated volatility paths spike during the Asian financial crisis, the Great Recession, and the COVID-19 pandemic, underscoring the empirical relevance of the very features that distinguish the MDFM from earlier matrix factor models.

In an out-of-sample forecasting exercise benchmarked against the static matrix factor model of [Wang et al. \(2019\)](#) with its eigenvalue-ratio rank-selection rule, the proposed MDFM with the Kronecker idiosyncratic covariance and stochastic volatility delivers statistically significant reductions in pooled root mean squared forecast error of 2.5% at the one-quarter horizon and 4.3% at the one-year horizon ([Marcellino et al., 2006](#)). The advantage is broadest at the policy-relevant short and medium horizons: the MDFM improves on the benchmark in 27 of the 29 pooled indicator-country series at  $h = 1$  and 28 of 29 at  $h = 4$ , with the largest gains in real-activity indicators (real GDP, household consumption, and unemployment improve by roughly 6–10%) and in the continental European, Nordic, and Korean economies whose business cycles are tightly tied to the global factor. Beyond point forecasts, the fully Bayesian treatment yields posterior predictive densities that integrate over parameter, factor, and volatility uncertainty—a dimension of forecast quality emphasized in the Bayesian macroeconomic forecasting literature ([Carriero et al., 2015](#)) and central to the growth-at-risk agenda ([Adrian et al., 2019](#)). The predictive variance widens visibly during the 2008–2009 financial crisis and the 2020 COVID-19 contraction.

This paper connects several literatures. It contributes to the tradition of Bayesian dynamic factor and large Bayesian time-series models for macroeconomic and financial datasets ([Aguilar and West, 2000](#); [West, 2003](#); [Lopes and West, 2004](#); [Del Negro and](#)

Otrok, 2008; Bańbura et al., 2010; Huber et al., 2020; Chan and Qi, 2025), to which it adds a matrix-valued formulation with flexible idiosyncratic dynamics. It is also related to the growing statistical literature on matrix and tensor factor models (Wang et al., 2019; Chen et al., 2022; Yu et al., 2022; He et al., 2024), which is predominantly frequentist and focused on estimation and asymptotic theory rather than the Bayesian inference, forecasting, and model comparison that we emphasize.

Two recent works are closest to ours. Yu et al. (2024) and Qin et al. (2025) both propose dynamic matrix factor models with matrix autoregressive factor evolution; Yu et al. (2024) take a frequentist approach with iterative eigendecomposition and eigenvalue-ratio factor selection, while Qin et al. (2025) adopt a Bayesian approach with a Kronecker-structured idiosyncratic covariance. Our framework differs in two key respects that mirror our two contributions. First, in modeling: neither accommodates the time-varying volatility, heavy tails, and outlier components that we introduce and that are empirically important for macroeconomic data. Second, in model comparison: we develop a unified approach that compares a wide range of competing specifications, which neither paper provides.

The rest of this paper is organized as follows. Section 2 specifies the model, identification, priors, and Bayesian estimation. Section 3 introduces the marginal likelihood estimator. Section 4 presents results from the Monte Carlo experiments. Section 5 reports the empirical application to an OECD macroeconomic panel and the out-of-sample forecasting exercise. Section 6 concludes.

## 2. A Matrix Dynamic Factor Model with Stochastic Volatility

Building on the framework of Wang et al. (2019), we introduce a dynamic factor model for matrix-valued time series. Following the spirit of the approximate factor model of Chamberlain and Rothschild (1983), we allow cross-row and cross-column correlations in the idiosyncratic components. We then incorporate time-varying volatility and outlier adjustments, present identification restrictions, and outline the Bayesian priors and posterior simulation. We close the section by sketching a heteroskedastic extension.

### 2.1. The Model

Consider an  $n \times k$  data matrix  $\mathbf{Y}_t$  observed at time  $t$ . To fix ideas, suppose  $\mathbf{Y}_t$  is a panel of macroeconomic indicators where rows index  $n$  countries and columns index  $k$  variables.

The matrix dynamic factor model is

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

$$\text{vec}(\mathbf{F}_t) = \mathbf{H}_{\rho_1} \text{vec}(\mathbf{F}_{t-1}) + \cdots + \mathbf{H}_{\rho_q} \text{vec}(\mathbf{F}_{t-q}) + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}), \quad (2)$$

where  $\mathbf{A}$  ( $n \times p_1$ ) and  $\mathbf{B}$  ( $k \times p_2$ ) are the row and column loading matrices,  $\mathbf{F}_t$  is a  $p_1 \times p_2$  latent factor matrix,  $\boldsymbol{\Sigma}_r$  ( $n \times n$ ) and  $\boldsymbol{\Sigma}_c$  ( $k \times k$ ) are the row- and column-wise idiosyncratic covariance matrices, and the  $p_1 p_2 \times p_1 p_2$  factor innovation covariance is  $\boldsymbol{\Lambda} = \text{diag}(\lambda_{1,1}^2, \dots, \lambda_{p_1 p_2, p_1 p_2}^2)$ .  $\mathbf{H}_{\rho_l}$ ,  $l = 1, \dots, q$ , are autoregressive coefficient matrices.

The bilinear form  $\mathbf{A}\mathbf{F}_t\mathbf{B}'$  captures cross-country and cross-variable dependencies. The  $i$ -th row of  $\mathbf{Y}_t$  is  $\mathbf{Y}_{i,t} = \mathbf{A}_{i,\cdot} \mathbf{F}_t \mathbf{B}' + \mathbf{E}_{i,t}$ : the common component  $\mathbf{A}_{i,\cdot} \mathbf{F}_t \mathbf{B}'$  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 has common component formed by linear combinations of columns of  $\mathbf{A}\mathbf{F}_t$  with weights from the  $j$ -th column of  $\mathbf{B}'$ . Hence rows of  $\mathbf{A}$  describe how countries load on latent factors, columns of  $\mathbf{B}'$  describe how indicators respond to latent factors, and rows (columns) of  $\mathbf{F}_t$  are interpreted as latent factors driving the country (indicator) dimension.

### 2.1.1. Important features of the model

Relative to the static matrix factor model of Wang et al. (2019), our model adds three features tailored to macroeconomic and financial data: autoregressive factor dynamics, a Kronecker-structured idiosyncratic covariance that captures cross-row and cross-column correlation, and idiosyncratic errors that accommodate stochastic volatility and outliers.

*Factor dynamics.* Modeling factor dynamics is particularly important in macroeconomic applications, where latent state variables typically follow stable laws of motion that embody business-cycle persistence. The autoregressive specification in (2) captures persistent comovement and provides a natural framework for recursive forecasting. Posterior simulation from the resulting dynamic system delivers predictive densities as standard byproducts of estimation; we exploit this in the forecasting exercise of Section 5, which compares the predictive performance of our specification against the static matrix factor benchmark.

*Kronecker idiosyncratic covariance.* The Kronecker covariance of  $\text{vec}(\mathbf{E}_t)$  separates row-wise and column-wise residual correlations. Specifically, for any row  $i$ , the conditional co-

variance is  $\text{Cov}(\mathbf{Y}'_{i,\cdot,t} | \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma_{r,i,i}^2 \boldsymbol{\Sigma}_c$ , and for any column  $j$ ,  $\text{Cov}(\mathbf{Y}_{\cdot,j,t} | \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma_{c,j,j}^2 \boldsymbol{\Sigma}_r$ .  $\boldsymbol{\Sigma}_r$  and  $\boldsymbol{\Sigma}_c$  thus capture row-wise and column-wise residual covariances not explained by the common component, making the model an approximate factor model in the sense of Chamberlain and Rothschild (1983) (see also Stock and Watson, 2005; Barigozzi and Hallin, 2024; Giglio et al., 2025).

The Kronecker structure also greatly reduces the number of free parameters relative to an unrestricted idiosyncratic covariance.<sup>3</sup> This dimensionality reduction is critical for the tractability of large Bayesian multivariate models and is analogous to the use of Kronecker-structured priors in large Bayesian VARs (Chan, 2020). The restriction is stronger than the approximate factor framework of Chamberlain and Rothschild (1983), which only requires bounded eigenvalues of the idiosyncratic covariance. In return, it yields conjugate matrix-normal–inverse-Wishart posteriors for  $(\mathbf{A}, \boldsymbol{\Sigma}_r)$  and  $(\mathbf{B}, \boldsymbol{\Sigma}_c)$ , enabling joint rather than element-wise sampling of loadings and covariance matrices (Section 2.4).

*Stochastic volatility and outliers.* For homoskedasticity, we can set  $\omega_t = 1$  for all  $t$ . It is also well known that allowing for time-varying volatility is essential in modeling macroeconomic and financial data, especially after the COVID-19 pandemic.<sup>4</sup> We implement three popular specifications under the conditionally Gaussian framework in (2) through the scalar latent state  $\omega_t$ .

**Specification 1. Common stochastic volatility.** Following Carriero et al. (2015, 2016), let  $\omega_t = e^{h_t}$  with log-volatility  $h_t$  following a stationary AR(1) process with zero mean:

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

for  $t = 2, \dots, T$ , with  $|\phi| < 1$  and  $h_1 \sim \mathcal{N}(0, \sigma_h^2 / (1 - \phi^2))$ .

**Specification 2. Explicit outliers.** Following Stock and Watson (2016),  $\omega_t = o_t^2$  where  $o_t$  follows a mixture distribution distinguishing regular observations ( $o_t = 1$ ) from

---

<sup>3</sup>An unrestricted  $nk \times nk$  covariance matrix contains  $nk(nk + 1)/2$  parameters. For the OECD panel considered below ( $n = 19$ ,  $k = 10$ ,  $T = 115$ ), this is 18,145 parameters—more than 150 times the sample size—whereas the Kronecker specification  $\boldsymbol{\Sigma}_u = \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$  requires only  $n(n + 1)/2 + k(k + 1)/2 - 1 = 244$ , while still allowing both within-row and within-column residual dependence.

<sup>4</sup>For evidence in vector autoregressions, see Cross and Poon (2016), Chan and Eisenstat (2018), and Chan (2023); for factor models, Aguilar and West (2000); Chib et al. (2006); Kastner et al. (2017); Li and Scharth (2022).

outliers ( $o_t \geq 2$ ). The outlier probability  $p_o$  has a beta prior. This specification has become standard for handling the extreme observations of macroeconomic series during the onset of COVID-19 pandemic.

**Specification 3. Fat-tailed innovations.** Following [Jacquier et al. \(2004\)](#),  $\omega_t = q_t^2$  with  $q_t^2 \sim \mathcal{IG}(l/2, l/2)$ . The marginal distribution of  $\text{vec}(\mathbf{E}_t)$  is then a multivariate  $t$  with zero mean, scale matrix  $\Sigma_c \otimes \Sigma_r$ , and degrees of freedom  $l$ .

### 2.1.2. Relation to vectorized factor models

A natural benchmark for the model in (1)–(2) is the standard VDFM applied to the vectorized panel:

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

where  $\mathbf{y}_t$  is the  $nk \times 1$  vectorized panel,  $\mathbf{M}$  is an  $nk \times p$  loading matrix,  $\mathbf{f}_t$  is a  $p \times 1$  vector of factors, and  $\mathbf{H}_\rho$  is the factor AR coefficient. We set  $p = p_1 \times p_2$  for direct comparison with the MDFM.

The MDFM is a restricted version of (4) in which the factor loadings exhibit a Kronecker product structure:

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

This restriction reduces the loading parameter count from  $nk \times p_1 p_2$  in the unrestricted VDFM to  $np_1 + kp_2$  in the MDFM, with corresponding improvements in estimation accuracy. Under principal components analysis applied to the vectorized model, [Bai \(2003\)](#) establishes that each row of  $\mathbf{B} \otimes \mathbf{A}$  converges at rate  $\min\{\sqrt{T}, nk\}^{-1}$ , which is dominated by  $T^{-1/2}$  in the empirically relevant regime where  $T \ll nk$ . The matrix structure delivers sharper rates by exploiting the fact that each column of  $\mathbf{Y}_t$  is informative about  $\mathbf{A}$  and each row about  $\mathbf{B}$ . [Chen and Fan \(2023, Theorem 1\)](#) show via  $\alpha$ -PCA that  $\mathbf{A}$  is recovered at rate  $\min\{n, Tk\}^{-1/2}$  and  $\mathbf{B}$  at rate  $\min\{k, Tn\}^{-1/2}$ , so the effective sample size for  $\mathbf{A}$  scales with  $Tk$  rather than  $T$  alone, and symmetrically for  $\mathbf{B}$ . This efficiency gain is valuable in the macro applications we target, where  $T$  is short relative to  $nk$ .

Whether the data actually support the Kronecker structure is, of course, an empirical question. [He et al. \(2024\)](#) propose a family of randomized specification tests for the matrix structure; we instead adopt a Bayesian model comparison approach using marginal likelihoods (Section 3), which has the advantage of treating the VDFM-versus-MDFM

choice on the same footing as the choice of factor-matrix dimension and idiosyncratic covariance structure. The Monte Carlo experiments in Section 4.3 confirm that the marginal likelihood estimator reliably distinguishes the two specifications in either direction.

## 2.2. Identification

The model in (1)–(2) is not identified without further restrictions. The matrix factor literature commonly resolves this by estimating only the column spaces of the loading matrices, which are uniquely identifiable without further restrictions (Wang et al., 2019; Chen et al., 2020, 2021; Yu et al., 2022).

The Bayesian dynamic factor model literature has adopted a range of identification strategies that pin down loadings and factors uniquely. The most common is the positive lower-triangular (PLT) restriction introduced by Geweke and Zhou (1996) and developed further for dynamic factor and stochastic volatility models by Aguilar and West (2000), West (2003) and Lopes and West (2004); the formal identification result for dynamic factor models is established in Bai and Wang (2015), and Bai and Ng (2013) provide the corresponding identification analysis for principal components estimators, including a triangular-block restriction (PC2) that is the frequentist analog of the lower-triangular Bayesian scheme used here. Closely related anchoring strategies that combine sign and zero restrictions on selected loadings to identify regional or group factors include Del Negro and Otrok (2008) and the related international business cycle literature. Alternative schemes that avoid ordering altogether include the order-invariant approach of Chan et al. (2018), which restricts the factor space rather than the ordering of individual variables; and the generalized lower-triangular (GLT) representation of Frühwirth-Schnatter et al. (2024a,b), which treats the leading-variable allocation as a parameter to be inferred. See also Kaufmann and Schumacher (2019); Leung and Drton (2016) for related order-invariant proposals.

In this paper we adopt the lower-triangular identification of Bai and Wang (2015), extending it to the matrix factor setting by anchoring both row and column loading matrices. The scheme pins down not only the column spaces but also the levels and signs of the loadings and factors, delivering economic interpretability for the rows and columns of  $\mathbf{F}_t$ . Proofs of the identification results below are given in Appendix A of the Online Supplemental Materials.

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

**Assumption 2.** Factor series are independent of each other;  $\mathbf{H}_{\rho_l}$  is diagonal for  $l = 1, \dots, q$ ; and  $\text{Cov}(\mathbf{u}_t) = \mathbf{I}_{p_1 p_2}$ .

**Assumption 3.** One of  $\mathbf{A}$  and  $\mathbf{B}$  is a lower-triangular matrix with ones on the diagonal, while the other is a lower-triangular matrix with strictly positive diagonal elements.

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

A second identification issue arises from the Kronecker structure of the idiosyncratic covariance:  $\Sigma_c$  and  $\Sigma_r$  are identified only up to a scalar, since for any  $m \in \mathbb{R} \setminus \{0\}$ ,  $\Sigma_c \otimes \Sigma_r = (m\Sigma_c) \otimes (m^{-1}\Sigma_r)$ . We fix the scale by normalizing the (1, 1) element of  $\Sigma_c$  to one.

An alternative identification scheme places the unit-diagonal restriction on both  $\mathbf{A}$  and  $\mathbf{B}$  symmetrically and instead allows the factor innovation covariance  $\text{Cov}(\mathbf{u}_t)$  to be a positive diagonal matrix rather than the identity. Formally:

**Assumption 4.** Factor series are independent of each other;  $\mathbf{H}_{\rho_l}$  is a diagonal matrix for  $l = 1, \dots, q$ ; and  $\text{Cov}(\mathbf{u}_t)$  is a positive definite diagonal matrix.

**Assumption 5.** The factor loading matrices  $\mathbf{A}$  and  $\mathbf{B}$  are both lower-triangular matrices with ones on the diagonal.

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

The lower triangular structure has an important interpretive implication: the first row of  $\mathbf{Y}_t$  responds to the first row factor only, the second row of  $\mathbf{Y}_t$  responds to the first two row factors only, and so on; similarly for the columns. This implies that shocks to the first row factor are shocks to the first row of the data matrix (or the first country), shocks to the second row factor are shocks to the second row of the data matrix orthogonal to the first, and so forth. The leading variables thus anchor the substantive content of each factor, and the ordering of rows and columns of  $\mathbf{Y}_t$  determines the labeling of the rows and columns of  $\mathbf{F}_t$ .

Whether this ordering is consequential depends on the use of the model. Identification matters for the substantive labeling of factors and loadings but not for forecasting, which depends on the loadings and factors through the product  $\mathbf{A}\mathbf{F}_t\mathbf{B}^\top$  and is invariant to

rotations of the factor representation. Practitioners concerned about ordering dependence can verify the robustness of their forecasting results via permutation experiments across alternative orderings; we report such an experiment for our application in Section 5.5.

Three limitations of our specification are worth noting. First, Assumption 4 treats the elements of  $\mathbf{F}_t$  as having AR( $q$ ) dynamics. Allowing for cross-factor dependencies via a vector or matrix autoregression for  $\mathbf{F}_t$  would generalize the dynamics, but combined with the time-varying volatility, outlier adjustments, and Kronecker idiosyncratic covariance we adopt, it would complicate identification and posterior computation. Second, the common stochastic volatility specification in (3) imposes that all series share a single time-varying scale; allowing factor-specific or block-specific volatilities is a natural extension, which we briefly discuss in Section 2.5. Third, our triangular identification gains interpretability at the cost of requiring the practitioner to choose an ordering of rows and columns. Adopting order-invariant identification strategies (see, e.g., Chan et al. (2018); Kaufmann and Schumacher (2019); Leung and Drton (2016); Frühwirth-Schnatter et al. (2024b,a)) is a promising direction for future research.

### 2.3. Priors

We use natural conjugate priors for the transposed loadings  $\mathbf{A}'$  and  $\mathbf{B}'$  jointly with inverse-Wishart priors for  $\Sigma_r$  and  $\Sigma_c$ :

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

where  $\mathbf{V}_{\mathbf{A}'}$  and  $\mathbf{V}_{\mathbf{B}'}$  are  $p_1 \times p_1$  and  $p_2 \times p_2$  scale matrices, respectively. This conjugate structure has both a substantive and a computational advantage. Substantively, the prior implies  $\text{Cov}(\mathbf{A}_{i,\cdot}, \mathbf{A}_{j,\cdot} \mid \Sigma_r) = \Sigma_{r,i,j} \mathbf{V}_{\mathbf{A}'}$ , so that the prior correlation between the  $i$ -th and  $j$ -th rows of  $\mathbf{A}$  along any direction in the factor space equals  $\Sigma_{r,i,j} / \sqrt{\Sigma_{r,i,i} \Sigma_{r,j,j}}$ . In other words, countries with more correlated idiosyncratic components are a priori expected to have similar factor-loading patterns; an analogous interpretation holds for the columns of  $\mathbf{B}$  through  $\Sigma_c$ . Computationally, the conjugate prior in (6) yields a matrix-normal-inverse-Wishart conditional posterior for  $(\mathbf{A}', \Sigma_r)$  and  $(\mathbf{B}', \Sigma_c)$ , which we exploit for joint sampling of the loading matrices and the row/column covariances; see Section 2.4 below. In practice,  $\mathbf{S}_r$  and  $\mathbf{S}_c$  are taken to be diagonal, reflecting the prior belief that idiosyncratic risks are uncorrelated, while the data are allowed to inform any off-diagonal correlations.

For parsimony, we consider AR(1) dynamics for each factor entry; the extension to AR( $q$ ) with  $q > 1$  is straightforward, replacing the truncation  $|\rho| < 1$  with truncation to the stationarity region  $\mathcal{S}_q = \{\boldsymbol{\rho} \in \mathbb{R}^q : 1 - \sum_{s=1}^q \rho_s z^s \neq 0 \text{ for all } |z| \leq 1\}$  and the scalar variance in the prior for initial values below with the solution to the discrete Lyapunov equation. Under AR(1), the diagonal elements of  $\mathbf{H}_\rho$  have truncated normal priors,

$$\rho_{j,k} \sim \mathcal{N}(\rho_{j,k,0}, V_{\rho_{j,k}}) \mathbf{1}(|\rho_{j,k}| < 1), \quad j = 1, \dots, p_1, \quad k = 1, \dots, p_2,$$

where  $\mathbf{1}(\cdot)$  denotes the indicator function. The factor innovation variances are  $\lambda_{j,k}^2 \sim \mathcal{IG}(\nu_{\lambda_{j,k}}, S_{\lambda_{j,k}})$ , and the initial value  $f_{j,k,1}$  of each factor process is assigned the stationary prior

$$f_{j,k,1} \mid \rho_{j,k}, \lambda_{j,k}^2 \sim \mathcal{N}\left(0, \frac{\lambda_{j,k}^2}{1 - \rho_{j,k}^2}\right),$$

which is the unconditional variance of the stationary AR(1) process.

#### 2.4. Bayesian Estimation

Posterior draws are obtained by sequentially sampling from (1)  $p(\mathbf{A}', \boldsymbol{\Sigma}_r \mid \mathbf{Y}, \mathbf{B}, \mathbf{F}, \boldsymbol{\Sigma}_c)$ ; (2)  $p(\mathbf{B}', \boldsymbol{\Sigma}_c \mid \mathbf{Y}, \mathbf{A}, \mathbf{F}, \boldsymbol{\Sigma}_r)$ ; (3)  $p(\text{vec}(\mathbf{F}_t) \mid \mathbf{Y}_t, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\omega}^2, \boldsymbol{\rho})$  for  $t = 1, \dots, T$ ; (4)  $p(\lambda_{j,k}^2 \mid \mathbf{f}_{j,k}, \rho_{j,k})$ ; (5)  $p(\rho_{j,k,l} \mid \mathbf{f}_{j,k}, \lambda_{j,k}^2)$ ; and (6)  $p(\omega_t \mid \mathbf{A}, \mathbf{F}_t, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r)$  for  $t = 1, \dots, T$ .

The conjugate prior in (6) delivers matrix-normal-inverse-Wishart conditional posteriors for  $(\mathbf{A}', \boldsymbol{\Sigma}_r)$  and  $(\mathbf{B}', \boldsymbol{\Sigma}_c)$ , which permit joint sampling of the loading matrices and the corresponding row/column covariances rather than element-wise updates. Specifically,

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

with closed-form expressions for the moments given in [Appendix B](#) of the Online Appendix. The triangular identification restrictions on  $\mathbf{A}$  and  $\mathbf{B}$  amount to linear constraints of the form  $\mathbf{M}_{\mathbf{A}'} \text{vec}(\mathbf{A}') = \mathbf{a}_0$  and  $\mathbf{M}_{\mathbf{B}'} \text{vec}(\mathbf{B}') = \mathbf{b}_0$  on the conditional posterior; we sample efficiently under these constraints using the algorithm of [Cong et al. \(2017\)](#), which generates unrestricted draws from the matrix-normal distribution and then projects them onto the constrained subspace. Sampling  $\boldsymbol{\Sigma}_c$  subject to the restriction  $\sigma_{c,1,1} = 1$  uses the algorithm of [Nobile \(2000\)](#). Sampling the dynamic factors  $\mathbf{F}_t$  uses the precision sampler proposed by [Chan and Jeliazkov \(2009\)](#), sampling the autoregressive coefficients  $\rho_{j,k}$  uses a Metropolis-Hastings sampler, sampling the stochastic volatility using the acceptance-

rejection Metropolis-Hastings sampler proposed by Chan (2017), sampling fat-tailed innovation states with the Inverse-Gamma posterior, and sampling the outlier components using a discrete grid sampling algorithm. Full algorithmic details, including closed-form expressions for  $\widehat{\mathbf{A}}'$ ,  $\widehat{\mathbf{B}}'$ ,  $\mathbf{K}_{\mathbf{A}'}$ ,  $\mathbf{K}_{\mathbf{B}'}$ ,  $\widehat{\Sigma}_r$ ,  $\widehat{\Sigma}_c$ , the constrained sampling steps, and the full Markov chain are given in Appendix B of the Online Supplemental Materials.

### 2.5. An Extension: Heteroskedastic Time-Varying Volatility

A more flexible alternative is to give each idiosyncratic error its own stochastic volatility, in the spirit of Cogley and Sargent (2005) for VARs, replacing the Kronecker-structured covariance of  $\text{vec}(\mathbf{E}_t)$  with a diagonal one,

$$\text{vec}(\mathbf{E}_t) \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t), \quad \mathbf{D}_t = \text{diag}(e^{h_{1,1,t}}, e^{h_{2,1,t}}, \dots, e^{h_{n,k,t}}), \quad (8)$$

where each log-volatility  $h_{i,j,t}$  follows an independent stationary AR(1) as in (3). This accommodates substantially richer volatility patterns at the cost of  $nk$  latent processes in place of the single common one. The trade-off is twofold: being diagonal, (8) removes all cross-row and cross-column correlation in the idiosyncratic component, which can be empirically restrictive, and it precludes the conjugate MNIW sampling of  $(\mathbf{A}', \Sigma_r)$  and  $(\mathbf{B}', \Sigma_c)$  that drives our baseline’s efficiency.

## 3. Model Comparison via Marginal Likelihood

With several models plausible, one main issue facing practitioners is the lack of tools to compare these models. We use marginal likelihoods to address three comparison questions jointly: the factor dimensions  $(p_1, p_2)$ , vector versus matrix dynamic factor model, and the idiosyncratic covariance structure (diagonal versus Kronecker, with or without time-varying volatility). The marginal likelihood is the natural Bayesian criterion here: it integrates over the parameter space rather than relying on point estimates, penalizes over-parameterization automatically, and delivers Bayes factors that quantify the relative evidence between any two models.

The tools most common in the factor-model literature answer a narrower question. Information criteria (Bai and Ng, 2002) and the eigenvalue-ratio criteria standard in the matrix factor literature (Wang et al., 2019; Chen et al., 2020; He et al., 2024) select the factor dimension within a fixed specification, operating on the unconditional second

moments of the data. They do not extend to the non-nested comparisons in this paper—vector versus matrix, exact versus approximate idiosyncratic covariance, homoskedastic versus stochastic-volatility errors—which alter the likelihood through dynamic and covariance structure rather than eigenvalue rank. The marginal likelihood places all of these on one scale.<sup>5</sup>

The obstacle to wider use of marginal likelihood is computational: the integral  $p(\mathbf{y}) = \int p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$  is high-dimensional and hard to evaluate in dynamic models. We compute it with the cross-entropy importance-sampling estimator of [Chan and Eisenstat \(2015\)](#). With  $p(\boldsymbol{\theta})$  the prior and  $g(\boldsymbol{\theta})$  an importance density,

$$\hat{p}_{IS}(\mathbf{y}) = \frac{1}{N} \sum_{n=1}^N \frac{p(\mathbf{y} | \boldsymbol{\theta}_n) p(\boldsymbol{\theta}_n)}{g(\boldsymbol{\theta}_n)}, \quad \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N \stackrel{\text{iid}}{\sim} g(\boldsymbol{\theta}). \quad (9)$$

The estimator is unbiased and consistent whenever  $g$  dominates  $p(\mathbf{y} | \cdot)p(\cdot)$ , and its efficiency hinges on how closely  $g$  approximates the posterior. If we use the posterior density as the importance density, then the associated importance sampling estimator has zero variance. However, in practice, the posterior density cannot be used as its normalizing constant is exactly the unknown quantity we aim to estimate. The cross-entropy method makes it convenient to find an importance density within a parametric family that is “closest” to the posterior density by minimizing the Kullback-Leibler divergence between them. [Chan and Eisenstat \(2015\)](#) show the minimizing hyperparameters in the parametric family coincide with the maximum likelihood estimator obtained by treating the posterior draws as data for these hyperparameters. Calibrating  $g$  thus reduces to the following step: we run the Gibbs sampler of [Section 2.4](#), and then obtain optimal hyperparameters given the posterior draws by obtaining the maximum likelihood estimates. Because the importance draws are i.i.d. from  $g$ , numerical standard errors of the marginal-likelihood estimator follows directly; and because the calibrated  $g$  closely approximates the posterior, that variance is small. We further reduce variance by integrating out the latent factors with the Kalman filter before applying (9). The importance-sampling family for each parameter block, the integrated-likelihood expression, and the Kalman-filter

---

<sup>5</sup>The closest Bayesian treatment is [Qin et al. \(2025\)](#), who select factor dimensions and lag orders via fractional Bayes factors ([O’Hagan, 1995](#)) under diffuse priors; we instead use proper priors and compute the full-sample marginal likelihood directly. [He et al. \(2023\)](#) develop a test for the one-way (vector) versus two-way (matrix) choice; our framework handles that choice alongside the covariance and volatility structure on the same scale.

recursions are given in [Appendix C](#) of the Online Supplemental Materials.

## 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. We then evaluate whether the marginal likelihood estimator can correctly identify the true model structure – the dimension of the factor matrix, the vector-versus-matrix specification, and the structure of idiosyncratic covariance.

### 4.1. Performance of factor estimators

The data are generated from (1)–(2) with  $q = 1$ . Free parameters in  $\mathbf{A}$  and  $\mathbf{B}$  are drawn from  $\mathcal{U}(0, 1)$ ,  $\rho_{j,k}$  from  $\mathcal{U}(0.8, 0.9)$  for  $j = 1, \dots, p_1$ ,  $k = 1, \dots, p_2$ ,  $\Sigma_c = 0.3\mathbf{I}_k$ ,  $\Sigma_r = 0.5\mathbf{I}_n$ , and  $\lambda_{j,k}^2 = 1$ . We consider sample sizes  $(n, k) \in \{(10, 10), (20, 15), (30, 20)\}$ , observation lengths  $T \in \{200, 500, 1000\}$ , and factor matrix dimensions  $(p_1, p_2) \in \{(3, 2), (5, 5)\}$ .<sup>6</sup> We use the posterior mean as the point estimate and project the true factors onto the estimates, reporting adjusted  $R^2$  values for each factor series.

Figures 1–2 display the adjusted  $R^2$  values. Each row corresponds to a different sample size and each column to a different  $T$ . The minimum of the color axis is set to 0.9 (the smallest adjusted  $R^2$  we obtain across all designs is 0.91); detailed numerical values are reported in [Appendix D](#) of the Online Supplemental Materials. Two patterns are visible. First, factor estimates closely track their true values across all designs. Second, accuracy improves with  $T$  and with  $(n, k)$ , in line with the inferential theory for vectorized factor models ([Bai, 2003](#)) and static matrix factor models ([Chen and Fan, 2023](#)). Smaller factor matrix dimensions  $(p_1, p_2) = (3, 2)$  deliver more accurate estimates than  $(5, 5)$ , as expected.

---

<sup>6</sup>For  $(p_1, p_2) = (3, 2)$ , we use a Gibbs chain of 10,000 iterations after 5,000 burn-in. For  $(p_1, p_2) = (5, 5)$ , we extend the chain to 20,000 iterations after 10,000 burn-in. With larger factor matrices, proper initialization is important; estimates from a VDFM (1,000 posterior draws after 1,000 burn-in) work well as starting values. Geweke statistics confirm chain convergence in all designs.

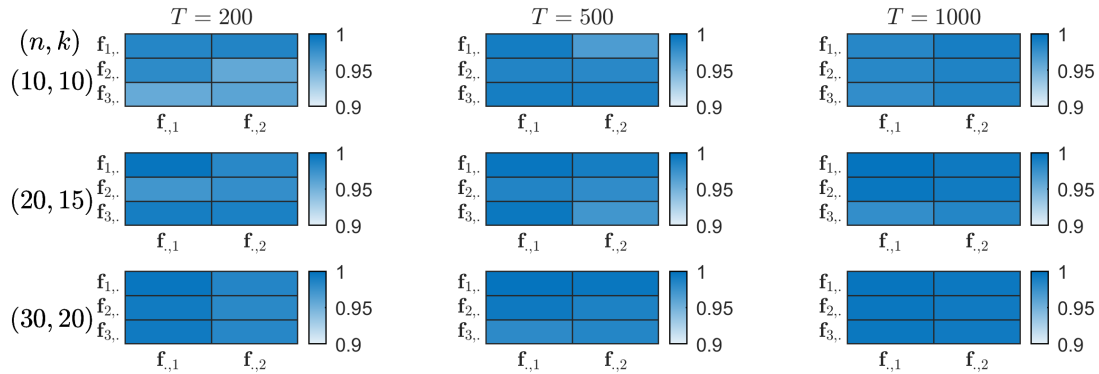


Figure 1: Adjusted  $R^2$  from regressing the true factors on the estimates:  $p_1 = 3, p_2 = 2$ .

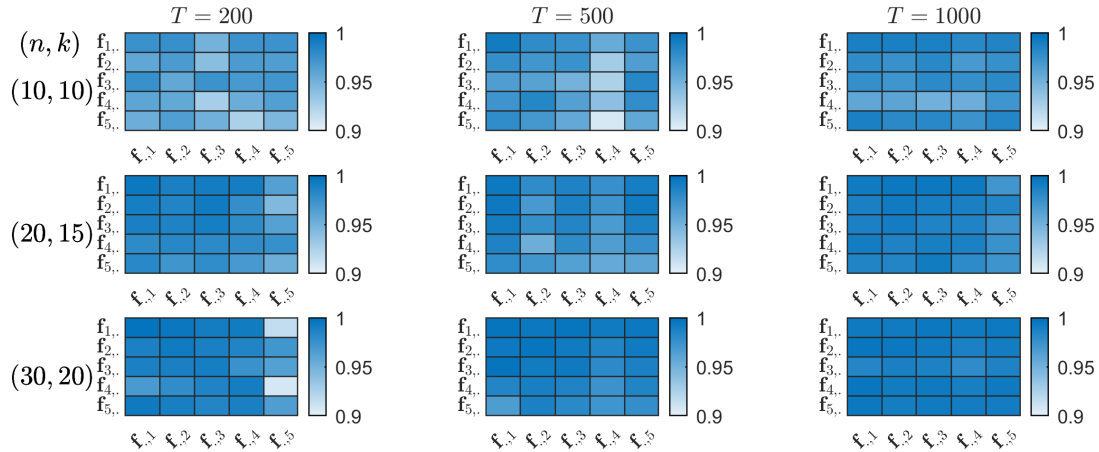


Figure 2: Adjusted  $R^2$  from regressing the true factors on the estimates:  $p_1 = 5, p_2 = 5$ .

#### 4.2. Identifying the dimension of the factor matrix

To evaluate the marginal likelihood estimator's ability to identify the correct factor matrix dimension, we estimate log marginal likelihoods over a grid of  $(p_1, p_2)$  values using 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  $(p_1, p_2) = (3, 2)$ , we estimate models with  $p_1$  and  $p_2$  ranging from 1 to 5; for  $(5, 5)$ , we estimate models with  $p_1$  and  $p_2$  ranging from 3 to 7.

Figure 3 reports the log marginal likelihood estimates. In all four datasets, the estimator correctly identifies the true dimension — the maximum log marginal likelihood is achieved at the true  $(p_1, p_2)$ . In addition, the estimates exhibit a consistent pattern: monotonically increasing as we approach the true dimension and monotonically decreasing afterwards. For example, when the true dimension is  $(3, 2)$ ,  $\log \hat{p}(\mathbf{y} \mid p_1 = 1) < \log \hat{p}(\mathbf{y} \mid p_1 = 2) < \log \hat{p}(\mathbf{y} \mid p_1 = 3)$  and  $\log \hat{p}(\mathbf{y} \mid p_1 = 3) > \log \hat{p}(\mathbf{y} \mid p_1 = 4) > \log \hat{p}(\mathbf{y} \mid p_1 = 5)$ , with an analogous pattern for  $p_2$ .

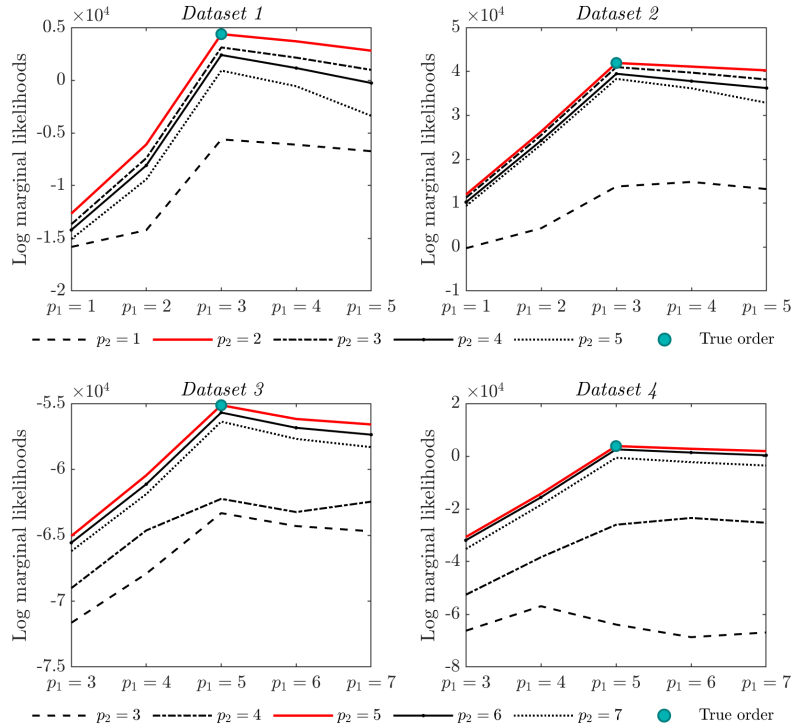


Figure 3: Log marginal likelihood estimates for the four datasets. From top left to bottom right, the panels correspond to  $(n, k, p_1, p_2) = (10, 10, 3, 2)$ ,  $(20, 15, 3, 2)$ ,  $(10, 10, 5, 5)$ , and  $(20, 15, 5, 5)$ . Each line represents a different value of  $p_2$ , and the blue dot marks the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

### 4.3. Distinguishing competing model structures

We next examine whether the marginal likelihood estimator can distinguish between competing model *structures*: vector versus matrix dynamic factor models (VDFM versus MDFM), and exact versus approximate factor models (diagonal versus Kronecker idiosyncratic covariance, with and without stochastic volatility).

*VDFM versus MDFM.* Following Section 4.1, we generate data from MDFMs with factor dimensions  $(p_1, p_2) \in \{(1, 2), (2, 1), (2, 2)\}$  and from VDFMs with  $k_f \in \{1, 2, 3\}$  factors.

For each true model, we estimate the corresponding alternative specifications and compare log marginal likelihoods. When the true model is an MDFM, we estimate VDFMs with  $k_f = 1, \dots, 6$ . When the true model is a VDFM, we estimate MDFMs with  $(p_1, p_2) \in \{(1, 1), (1, 2), (2, 1), (2, 2)\}$ .

Figures 4–5 report the results. In all six cases, the true model achieves the highest marginal likelihood. A noteworthy pattern in Figure 4: when the true model is MDFM, the best-performing VDFM is the one whose number of factors equals the total number of MDFM factors — e.g.,  $k_f = 2$  for MDFMs with  $(1, 2)$  and  $(2, 1)$ , and  $k_f = 4$  for  $(2, 2)$ . The intuition is that a VDFM with  $p_1 \times p_2$  factors can in principle span the same factor space as an MDFM with dimensions  $(p_1, p_2)$ , but does so with  $nk \times p_1 p_2$  loadings rather than the  $np_1 + kp_2$  loadings of the MDFM, leading to weaker fit when the matrix structure is true.

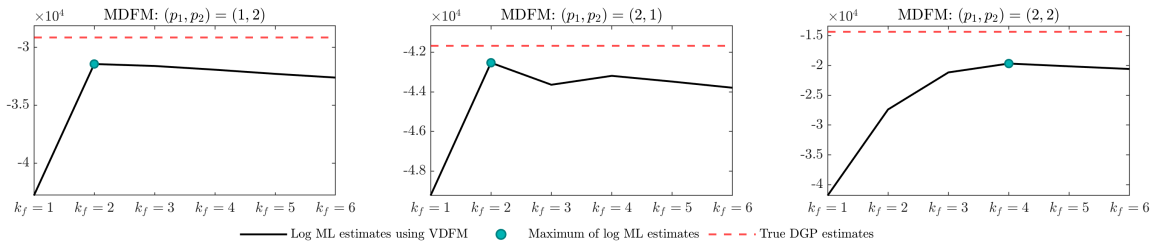


Figure 4: Log marginal likelihood estimates under VDFM specifications when the true model is an MDFM. The black line traces VDFM marginal likelihoods across  $k_f = 1, \dots, 6$  factors; the red dashed line marks the log marginal likelihood estimate of the true DGP: MDFM. The left, middle, and right panels correspond to MDFMs with  $(p_1, p_2) = (1, 2), (2, 1), (2, 2)$ .

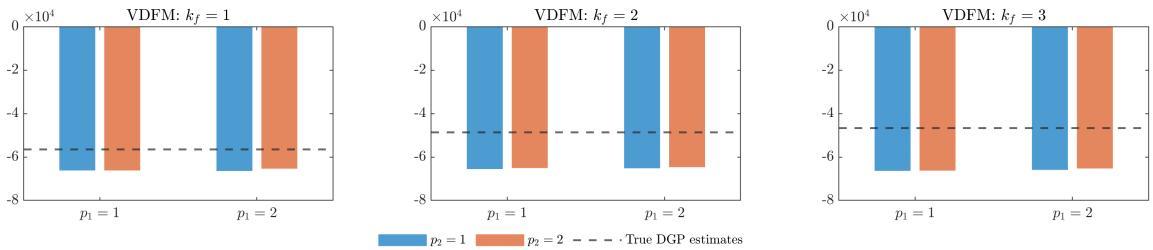


Figure 5: Log marginal likelihoods under MDFM specifications when the true model is a VDFM. The black dashed line marks the log marginal likelihood estimate of the true DGP: VDFM; bars correspond to competing MDFMs. The left, middle, and right panels correspond to VDFMs with  $k_f = 1, 2, 3$ .

*Diagonal versus Kronecker idiosyncratic covariance, with and without SV.* We next assess whether the estimator can distinguish three idiosyncratic-covariance specifications:

- **MDFM-diag:** diagonal idiosyncratic covariance, no stochastic volatility.

- **MDFM-cross:** Kronecker idiosyncratic covariance ( $\Sigma_c \otimes \Sigma_r$ ), no stochastic volatility.
- **MDFM-sv:** diagonal idiosyncratic covariance with common stochastic volatility.

We generate 100 datasets of dimension  $(n, k, T) = (20, 20, 100)$  with factor dimension  $(p_1, p_2) = (2, 2)$ . The free elements of  $\mathbf{A}$  and  $\mathbf{B}$  are drawn from  $\mathcal{N}(0, 0.3^2)$ . Row and column covariance matrices are drawn from  $\Sigma_r \sim \mathcal{IW}(n+2, \mathbf{I}_n)$  and  $\Sigma_c \sim \mathcal{IW}(k+2, \mathbf{I}_k)$ . The log-volatility process has AR(1) coefficient 0.97 and innovation variance 0.1; the factor innovation variance is  $0.1\mathbf{I}_{p_1 p_2}$ , with AR coefficients drawn from  $\mathcal{U}(0.8, 0.9)$ .

For each dataset, we estimate the log marginal likelihood under each of the three specifications and compute differences relative to the true model. Figure 6 reports boxplots of these differences. In all three settings, the true model achieves the highest marginal likelihood, indicating that the estimator reliably distinguishes the structure of the idiosyncratic component. The differentiating power is largest when the true model is MDFM-diag (the most parsimonious of the three), since the alternatives correspond to over-specified models in that case.

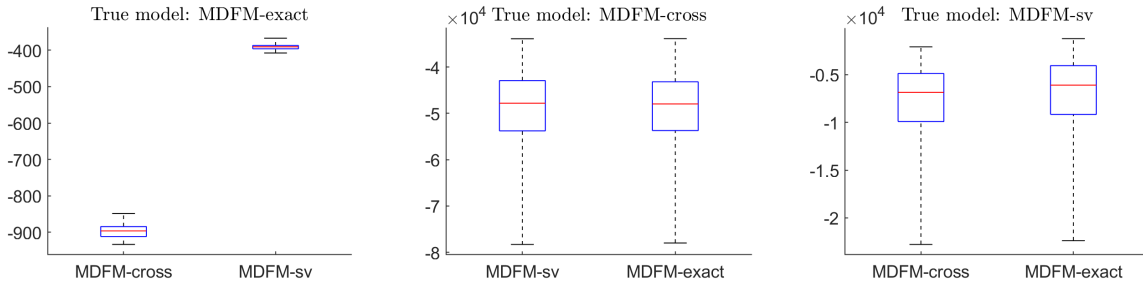


Figure 6: Boxplots of differences in log marginal likelihoods relative to the true model: MDFM-diag (left), MDFM-cross (middle), and MDFM-sv (right). A negative value indicates that the true model is favored.

## 5. An Application to a Large OECD Macroeconomic Panel

We illustrate the MDFM on a quarterly OECD macroeconomic panel covering 19 advanced economies and 10 indicators from 1995.Q1 to 2023.Q3. The application is designed to showcase three features of the model relative to existing alternatives: the economic interpretability of factors under triangular identification; the empirical relevance of the Kronecker idiosyncratic covariance structure for capturing residual cross-cell dependence;

and the predictive performance of iterated forecasts from a coherent dynamic model relative to direct projections from a static matrix factor benchmark.<sup>7</sup>

### 5.1. Data

The dataset comprises 10 indicators for 19 developed economies from North America, Europe, Asia, and Oceania, observed from 1995.Q1 to 2023.Q3 (115 quarters). The indicators are real GDP, headline CPI, labor unit cost, exports, imports, unemployment, energy CPI, household consumption, core CPI, and food CPI. Each series is rendered stationary via first or logarithmic differencing and standardized to zero mean and unit variance. Detailed variable definitions and transformation codes are given in [Appendix E](#) of the Online Supplemental Materials.

For countries, we place the United States first as the largest global economy, followed by the United Kingdom, Australia, Germany, and Japan as anchors of their respective regional economies. For indicators, we order real GDP first as the headline measure of real activity, followed by headline CPI as the principal inflation measure, followed by labor unit cost for its productivity interpretation.

### 5.2. Model Selection and Estimation

Posterior inference is obtained via the Gibbs sampler described in [Section 2.4](#). 10,000 posterior draws are used for inference after discarding 5,000 burn-in draws.

*Model selection.* [Tables 1](#) and [2](#) report log marginal likelihood estimates (with numerical standard errors in parentheses) for VDFMs and MDFMs. We use the labels VDFM-diag and VDFM-sv for vector models with diagonal idiosyncratic covariance, without and with common stochastic volatility, and analogously MDFM-diag, MDFM-cross, and MDFM-cross-sv for matrix models with diagonal, Kronecker, and Kronecker-plus-SV specifications.

---

<sup>7</sup>A second application to the Fama–French  $10 \times 10$  size-by-book-to-market portfolio panel, with comparable structure but a different empirical setting, is reported in [Appendix H](#) of the Online Supplemental Materials. In that application, we find that a vector DFM is favored by the data compared to a matrix DFM, based on the marginal likelihood estimates.

Table 1: Log marginal likelihood estimates using VDFMs. Numerical standard errors in parentheses. The largest estimate in each panel is in bold.

VDFM-diag				VDFM-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)
$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 5$	$k = 6$	$k = 7$	$k = 8$
<b>-22944</b>	-23001	-23107	-23280	<b>-20862</b>	-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. Numerical standard errors in parentheses. The largest estimate in each panel is in bold; the overall maximum is highlighted in red.

	MDFM-diag			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	<b>-17930</b>	-17950	-16144	<b>-16140</b>	-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	<b>-23012</b>	-18051	-18058	-18150	-16336	-16386	-16464
	(0.3)	(0.6)	(0.7)	(0.3)	(0.8)	(0.6)	(0.7)	(0.7)	(1.1)

There are three interesting findings. First, the data strongly favor the inclusion of common stochastic volatility: in both vector and matrix models, the SV variants dominate their constant-volatility analogues by thousands of log points (right panels of Tables 1 and 2). Second, the data favor Kronecker-structured idiosyncratic covariance over the diagonal alternative within the matrix framework. Third, the matrix structure dominates the vector representation: the best MDFM achieves a log marginal likelihood of  $-16,140$ , more than 4,700 log points above the best VDFM (at  $-20,862$ ). The optimal MDFM has a  $1 \times 2$  factor matrix, indicating one country dimension and two indicator dimensions, whereas the best VDFM requires five factors. The  $1 \times 2$  shape of  $\mathbf{F}_t$  implies that variation along the indicator dimension is richer than variation along the country dimension in this panel. We hereafter estimate the MDFM with  $(p_1, p_2) = (1, 2)$  and refer to it simply as the MDFM.

### 5.3. In-Sample Results

*Factor series and the price–oil link.* Figure 7 displays posterior estimates of the two factor series with 90% credible bands, alongside a comparison of the price factor to Brent crude oil price growth.

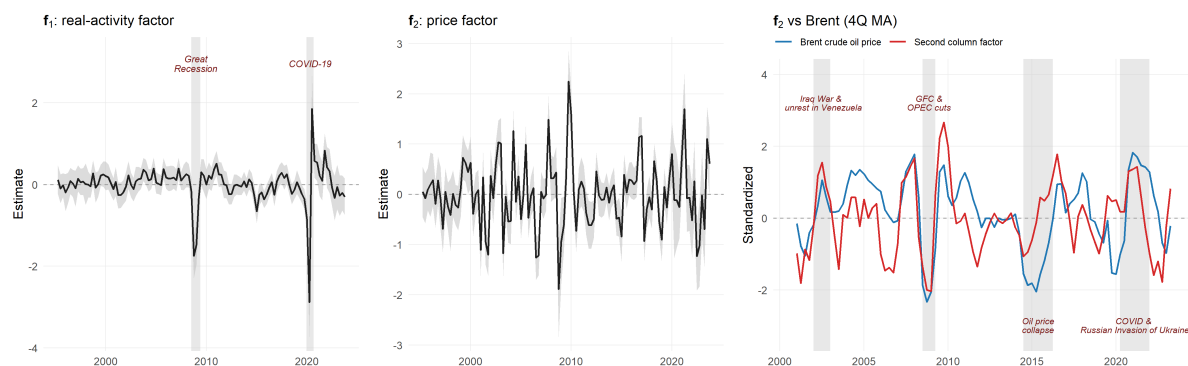


Figure 7: Left: real-activity factor  $\mathbf{f}_1$  with shaded recession episodes. Middle: price factor  $\mathbf{f}_2$  with 90% credible band. Right: four-quarter moving averages of standardized Brent crude oil price growth (blue) and the second column factor (red); shaded bands mark major oil-market episodes (war-premium spike during the 2003 Iraq War and unrest in Venezuela; the 2008–2009 oil-price collapse and OPEC production cuts; the 2014–2016 oil-price collapse driven by U.S. shale supply and weak global demand; and the COVID-19 demand shock and 2022 Russian invasion of Ukraine).

The real-activity factor  $\mathbf{f}_1$  traces the global business cycle. Two episodes dominate the series: a sharp contraction during the 2008–2009 Great Recession and a much larger spike in 2020 corresponding to the COVID-19 pandemic. The early-2000s slowdown is visible but quantitatively modest, consistent with the recession being concentrated in the United States and the United Kingdom rather than synchronized across the panel.

The price factor  $\mathbf{f}_2$  has its strongest external corroboration in the right-hand panel of Figure 7: a four-quarter moving average of  $\mathbf{f}_2$  tracks four-quarter Brent crude oil price growth throughout the sample, with aligned peaks and troughs at 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. The posterior correlation between the two smoothed series is approximately 0.66.

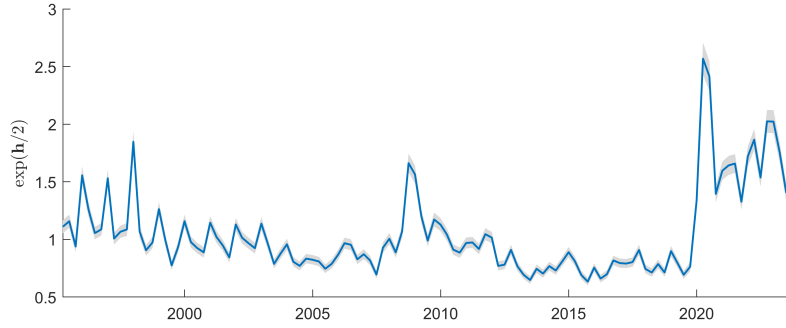


Figure 8: Posterior mean of common stochastic volatility,  $\hat{\omega}_t = \exp(\hat{\mathbf{h}}_t/2)$ , with 90% credible band.

*Loadings.* The latent structure of the global macroeconomy can be read off the row and column loading matrices. Figure 9 reports posterior estimates side by side: the country loadings  $\hat{\mathbf{A}}$  on the left and the indicator loadings  $\hat{\mathbf{B}}$  on the right, with 90% credible intervals from  $1.645 \times$  posterior standard deviations on each axis.<sup>8</sup>

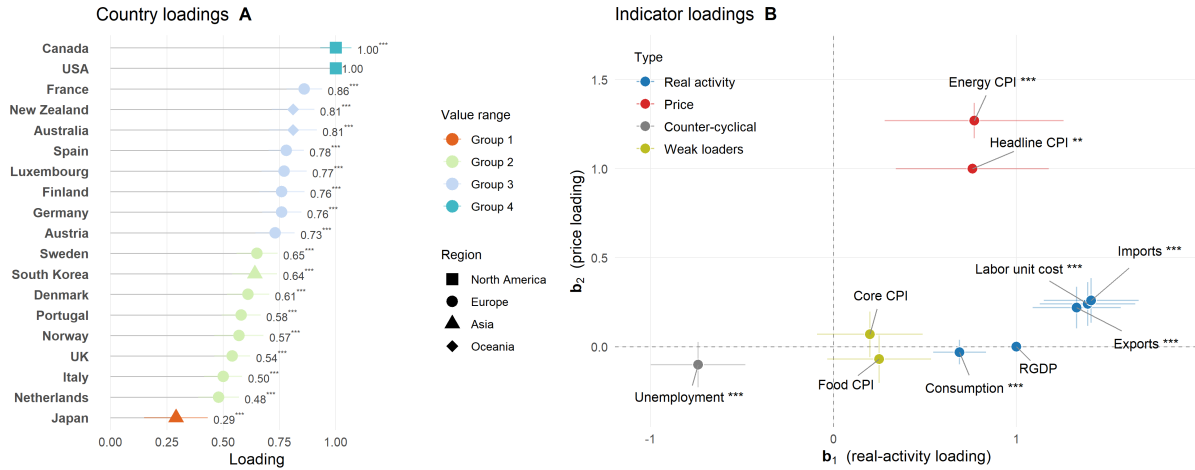


Figure 9: Loading estimates with 90% credible intervals. Left: country loadings  $\hat{\mathbf{A}}$ , sorted ascending, colored by value-range group and shaped by region. The numeric value of each loading appears to the right of its credible interval, with significance stars as superscripts. The USA loading is fixed to one for identification. Right: indicator loadings  $\hat{\mathbf{B}}$  in the  $(\mathbf{b}_1, \mathbf{b}_2)$  plane, colored by economic role.

The country loadings reveal a striking heterogeneity in exposure to the factors that does not map cleanly onto geography.<sup>9</sup> Japan stands alone at the bottom of the distribution

<sup>8</sup>Since the factor matrix has only one row,  $\mathbf{A}$  is effectively a  $19 \times 1$  vector; we retain the matrix notation for consistency.

<sup>9</sup>We sort these loading 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

with a loading of 0.29. A mid-tier of European economies plus Korea (NLD, ITA, GBR, NOR, PRT, DNK, KOR, SWE) loads between 0.48 and 0.65. A higher-tier group containing the continental European core and Oceania (AUT, DEU, FIN, LUX, ESP, AUS, NZL, FRA) loads between 0.73 and 0.86. Canada and the United States sit at 1.00. It is clear that geographic proximity matters, as the two North American economies cluster together at the top, and continental Europe is concentrated in the upper-mid range, although it is not the only factor. Japan and Korea fall in different groups despite their proximity, and Australia and New Zealand are grouped together with continental European economies rather than with Asia.

The indicator loadings sort the ten variables into four economic roles. Imports, labor unit cost, and exports load most strongly on the real-activity factor  $\mathbf{b}_1$  (loadings between 1.3 and 1.4) with small but positive loadings on the price factor  $\mathbf{b}_2$ . RGDP and consumption load positively on  $\mathbf{b}_1$  alone. Unemployment is the only counter-cyclical indicator, with a large negative  $\mathbf{b}_1$  loading of approximately  $-0.6$  and a  $\mathbf{b}_2$  loading indistinguishable from zero. Energy CPI and headline CPI load heavily on  $\mathbf{b}_2$  (loadings around 1.3 and 1.0 respectively), with smaller but positive  $\mathbf{b}_1$  components. Core CPI and food CPI are weak loaders on both factors: their 90% credible intervals overlap the origin on at least one axis, indicating that most of their variation is left to the idiosyncratic component. The two-dimensional view in the right panel of Figure 9 shows that the real-activity and price factors are approximately orthogonal in their effect on indicators. Indicators with large  $\mathbf{b}_1$  have small  $\mathbf{b}_2$ , and vice versa.

*The stochastic volatility and idiosyncratic correlations.* The common stochastic volatility component  $\hat{\omega}_t$  (Figure 8) isolates three episodes of elevated common uncertainty: the 1997–1998 Asian financial crisis (driven particularly by Japan and Korea), the 2008–2009 Great Recession, and the COVID-19 pandemic starting in 2020. The COVID-19 spike is by far the largest, consistent with the recent literature documenting the importance of stochastic volatility and outlier modeling for post-2020 macroeconomic data. Importantly, volatility remains elevated through 2023 rather than returning to its pre-2020 baseline.

In addition, there exist significant correlations among countries and indicators in idiosyncratic components. Figures 10 and 11 display the corresponding correlation matrices  $\mathbf{R}_c$

---

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.

and  $\mathbf{R}_r$ , obtained by standardizing  $\Sigma_c$  and  $\Sigma_r$  to unit diagonal. Correlations are a more interpretable summary than raw covariances because they are not inflated by the very different residual variances across rows and columns.

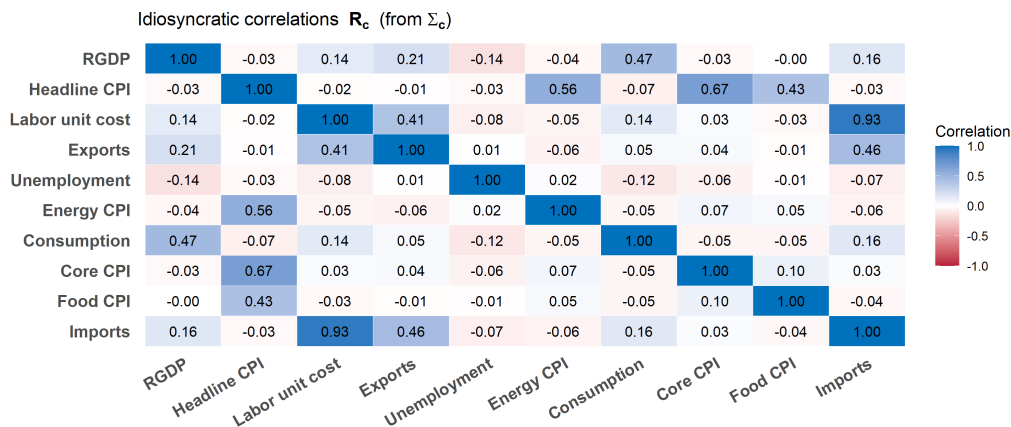


Figure 10: Idiosyncratic correlation matrix  $\mathbf{R}_c$  across indicators, standardized from  $\hat{\Sigma}_c$ .

Consider first the indicator correlations in  $\mathbf{R}_c$ . The idiosyncratic correlation is not diffuse but concentrated in a few interpretable blocks. First, real activity retains a clear correlation, with RGDP and consumption correlated at 0.47. In addition, the four price indices form a tight block among themselves (headline-core 0.67, headline-energy 0.56, headline-food 0.43). A third block links unit labor costs (ULC), imports, and exports. Notably, this block is unrelated to the price block, indicating that this idiosyncratic correlation is driven by real-side forces rather than by prices. A possible explanation is a terms-of-trade gain that at the same time raises real wages relative to productivity, strengthens the currency and draws in imports.

The idiosyncratic country correlations in  $\mathbf{R}_r$  line up along regional lines. The clearest grouping is an European block, i.e., France, Germany, Italy, the Netherlands, Austria, Spain, Portugal, and Finland, with correlations of about 0.4–0.57. The tightest pairs are neighbors and major trading partners: France–Italy (0.57), Germany–Austria (0.57), France–Germany (0.54), and Spain–Italy (0.54). Sweden, although it keeps its own currency, tracks this group closely as well, reflecting its trade ties with the European Union. The United Kingdom and Luxembourg attach to the group more loosely – the UK at 0.36–0.43 and Luxembourg at 0.29–0.35. The US and Canada form their own tight pair (0.49), yet the transatlantic link is somewhat weaker – the US correlates with European economies at around 0.13–0.34. Japan is almost an island: its correlation with every

other OECD economy in the dataset stays below 0.15. Korea and Australia are among the most weakly connected economies after Japan, peaking at 0.30 and 0.33 respectively; Australia’s only sizeable tie is to New Zealand (0.33), a small Australasian pair that mirrors the US–Canada link.

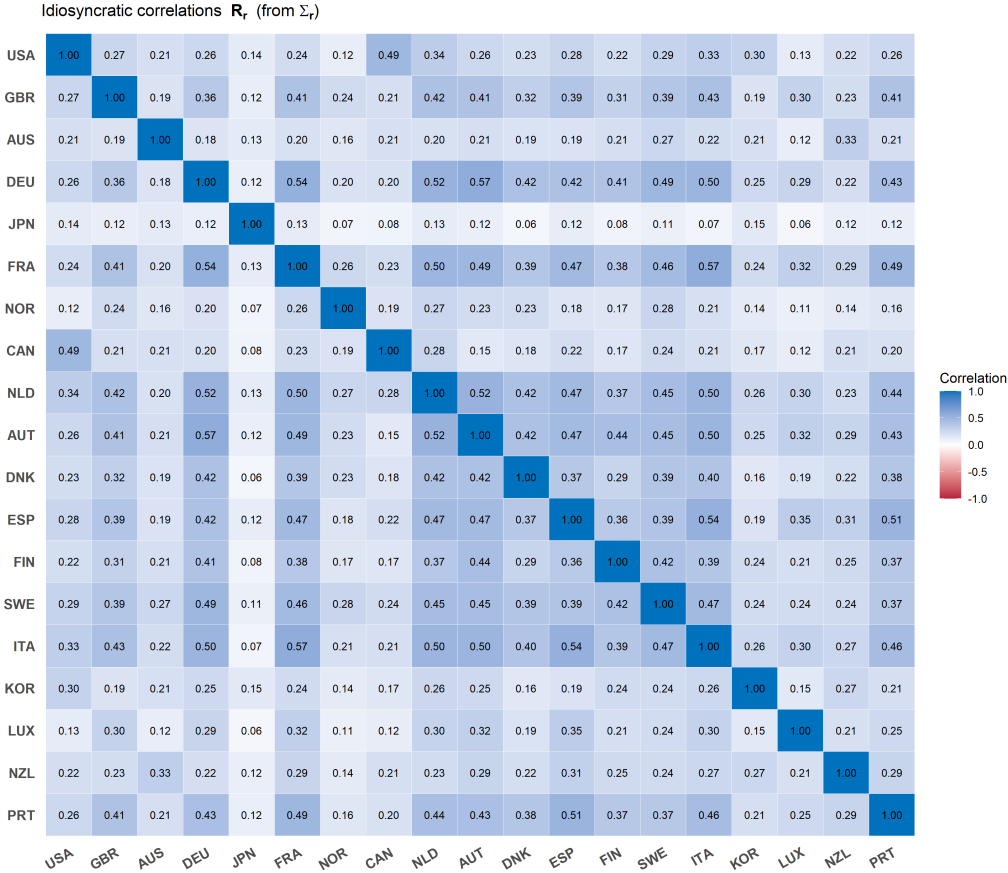


Figure 11: Idiosyncratic correlation matrix  $\mathbf{R}_r$  across countries, standardized from  $\hat{\Sigma}_r$ .

#### 5.4. Out-of-Sample Forecasting

We assess predictive performance of the MDFM via a recursive out-of-sample exercise where at each period  $t$  of the evaluation sample, we use only data up to that period to estimate the models and compute the  $h$ -step-ahead forecasts for period  $t + h$ . The forecast evaluation period starts from 2007Q4 till the end of the sample. The benchmark throughout is the static matrix factor model of Wang et al. (2019), hereafter MFM.

*MFM dimension selection and forecast construction.* The MFM selects factor dimensions  $(\hat{p}_1, \hat{p}_2)$  via an eigenvalue-ratio rule applied to the lag- $h_0$  autocovariance matrices  $\hat{\mathbf{M}}_1$  and

$\widehat{\mathbf{M}}_2$ .<sup>10</sup> Applied to our OECD panel, the rule selects  $\widehat{p}_1 = 1$  and  $\widehat{p}_2 = 1$ , agreeing with our marginal likelihood selection on the row dimension but selecting one fewer column factor. The selection is robust to the lag parameter  $h_0$ : across  $h_0 \in \{1, 2, 3, 4\}$  the rule uniformly returns (1, 1), in-sample RMSE under the auto-selected dimensions varies by less than 0.2%, and the loading subspaces drift by less than 0.08 (Table [Appendix F.1](#) in the Online Supplemental Materials). The eigenvalue-ratio rule is known to under-select the number of rows and columns in moderate- $T$  panels when the true value is small (Wang et al., 2019, Table 4), and forcing  $(p_1, p_2) = (1, 2)$  in the MFM uniformly improves its in-sample RMSE by 6–7 percentage points across all  $h_0$  values, providing evidence independent of our Bayesian framework that a second column factor may carry genuine signal. We nonetheless use the auto-selected MFM as the out-of-sample benchmark to maintain a fair comparison that respects the MFM’s native model-selection rule.

The two methods differ in the construction of multi-step forecasts. The MDFM produces *iterated* forecasts as a direct byproduct of the specified dynamics of latent states: each MCMC draw of the factor matrix  $\mathbf{F}_t$  and the volatility path is propagated forward through the model’s state and observation equations given the parameter draw, yielding draws of  $\mathbf{Y}_{t+h}$  that integrate over parameter, factor, and volatility uncertainty. The MFM forecasts are direct projections in the spirit of [Stock and Watson \(2002b\)](#). At each origin  $t$  we extract the scalar factor, then estimate horizon-specific OLS projections of  $Y_{i,j,t+h}$  on a constant and factor estimate  $f_t$  for each (country, indicator) pair  $(i, j)$  and horizon  $h$ .<sup>11</sup>

We report results at three horizons:  $h = 1$ ,  $h = 4$ , and  $h = 8$  (one-quarter, one-year, and two-year horizons). Per-cell RMSFE ratios and Diebold–Mariano test results are reported in Figure [Appendix G.1](#) in the Online Supplemental Materials. Table 3 summarizes the pooled performance by indicator (pooled across countries) and by country (pooled across indicators).

---

<sup>10</sup> $\widehat{\mathbf{M}}_1 = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_h^{(1)} [\widehat{\boldsymbol{\Sigma}}_h^{(1)}]^\top$ ,  $\widehat{\mathbf{M}}_2 = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_h^{(2)} [\widehat{\boldsymbol{\Sigma}}_h^{(2)}]^\top$ , where  $\widehat{\boldsymbol{\Sigma}}_h^{(1)}$  and  $\widehat{\boldsymbol{\Sigma}}_h^{(2)}$  are row- and column-wise sample autocovariance matrices at lag  $h$ , and  $h_0$  is a user-specified upper lag. Details of a pre-forecasting analysis of MFM is provided in [Appendix F](#).

<sup>11</sup>Direct projection is the natural choice because the MFM does not specify a dynamic law of motion for the latent factor. Although density forecasts can be constructed for direct projections using bootstrap or simulation-based methods, such forecasts require additional assumptions and implementation choices. In this paper we focus on comparing the forecasting performance of MDFM and MFM as specified.

Table 3: Out-of-sample RMSFE ratios MDFM / MFM, pooled across cells. The left panel pools across the 19 OECD countries within each indicator. The middle and right panels pool across the 10 indicators within each country. Pooled RMSFEs are computed from squared errors summed over all cells and all expanding-window forecast origins before taking the ratio. Values below 1 indicate that MDFM outperforms MFM and are shown in **bold**. Diebold–Mariano significance based on the pooled loss differential with Bartlett HAC variance estimation: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The bottom row reports the fully pooled ratio across all  $19 \times 10 = 190$  cells.

Indicator	$h=1$	$h=4$	$h=8$	Country	$h=1$	$h=4$	$h=8$	Country	$h=1$	$h=4$	$h=8$
RGDP	<b>0.911***</b>	<b>0.926***</b>	1.060	USA	<b>0.974***</b>	<b>0.970**</b>	1.028	DNK	<b>0.976***</b>	<b>0.933***</b>	<b>0.999</b>
Headline CPI	<b>0.997</b>	<b>0.996</b>	<b>0.989</b>	GBR	<b>0.971***</b>	<b>0.973**</b>	<b>0.976</b>	ESP	<b>0.954***</b>	<b>0.955***</b>	<b>0.990</b>
Labor unit cost	<b>0.985***</b>	<b>0.958***</b>	1.018*	AUS	<b>0.983*</b>	<b>0.965*</b>	<b>0.986</b>	FIN	<b>0.956***</b>	<b>0.945***</b>	1.039***
Exports	<b>0.982***</b>	<b>0.955***</b>	1.057***	DEU	<b>0.965***</b>	<b>0.943***</b>	1.008	SWE	<b>0.966***</b>	<b>0.961**</b>	1.057*
Unemployment	<b>0.942***</b>	<b>0.941***</b>	<b>0.947*</b>	JPN	1.000	<b>0.975*</b>	1.016	ITA	<b>0.957***</b>	<b>0.928***</b>	1.028
Energy CPI	1.026***	1.008	1.027***	FRA	<b>0.982*</b>	<b>0.953***</b>	1.046***	KOR	<b>0.996</b>	<b>0.978*</b>	1.023
Household consumption	<b>0.898***</b>	<b>0.929***</b>	<b>0.964</b>	NOR	<b>0.973**</b>	<b>0.951***</b>	<b>0.981</b>	LUX	<b>0.975*</b>	<b>0.967**</b>	1.006
Core CPI	<b>0.982</b>	<b>0.982</b>	<b>0.981</b>	CAN	<b>0.985</b>	<b>0.988</b>	1.063***	NZL	<b>0.991</b>	<b>0.981</b>	1.020
Food CPI	<b>0.967***</b>	<b>0.905***</b>	1.008	NLD	<b>0.967***</b>	<b>0.918***</b>	<b>0.986</b>	PRT	<b>0.983**</b>	<b>0.960***</b>	<b>0.987</b>
Imports	<b>0.972***</b>	<b>0.953***</b>	1.071***	AUT	<b>0.976**</b>	<b>0.937***</b>	1.034**				
<i>All cells pooled: <math>h=1</math>: <b>0.975***</b>, <math>h=4</math>: <b>0.957***</b>, <math>h=8</math>: 1.016***</i>											

The MDFM substantially outperforms MFM at the 1-quarter and 1-year horizons in pooled RMSFE, with statistically significant improvements of 2.5% at  $h = 1$  and 4.3% at  $h = 4$ . In particular, at  $h = 1$ , 27 out of the 29 pooled indicator-country ratios fall below one. At  $h = 4$ , MDFM outperforms MFM even more broadly and the only exception is energy CPI for which MFM outperforms MDFM insignificantly. The MDFM’s advantage attenuates and partially reverses at the 2-year horizon. The fully-pooled  $h = 8$  ratio rises to 1.016, indicating that MDFM is marginally less accurate than MFM in pooled terms at the 2-year horizon, where the MFM’s gains are significant at the 1% level for exports, imports and energy CPI among indicators, and for France, Canada, and Finland among countries.

In Figure 12, we show the predictive density for real GDP growth in the US and Germany at two horizons. The top panels show 1-quarter-ahead forecasts and realizations; the bottom panels show 1-year-ahead forecasts. The light blue shade represent the 5%-95% interval of the predictive density, while the darker blue represent the 25%-75% interval. The full G7 comparison is reported in Figure Appendix G.4 of the Online Supplement Materials.

The incorporation of time-varying volatility lets the bands respond to changing risk: they widen during volatile episodes—modestly at the 1-quarter horizon and markedly in the 1-year panels, where they fan out around the 2008–09 crisis and the 2020–22 period. In

addition, it is clear that the 1-year panels expose a finite-sample weakness of the MFM direct projection: after the 2008–09 contraction the MFM forecast for the US jumps to about 5% in early 2010, which is well ahead of the realized recovery, with a subtler version for Germany. The MDFM’s iterated forecasts instead propagate the factor through the AR dynamics, yielding forecasts that better match the slow realized recovery.

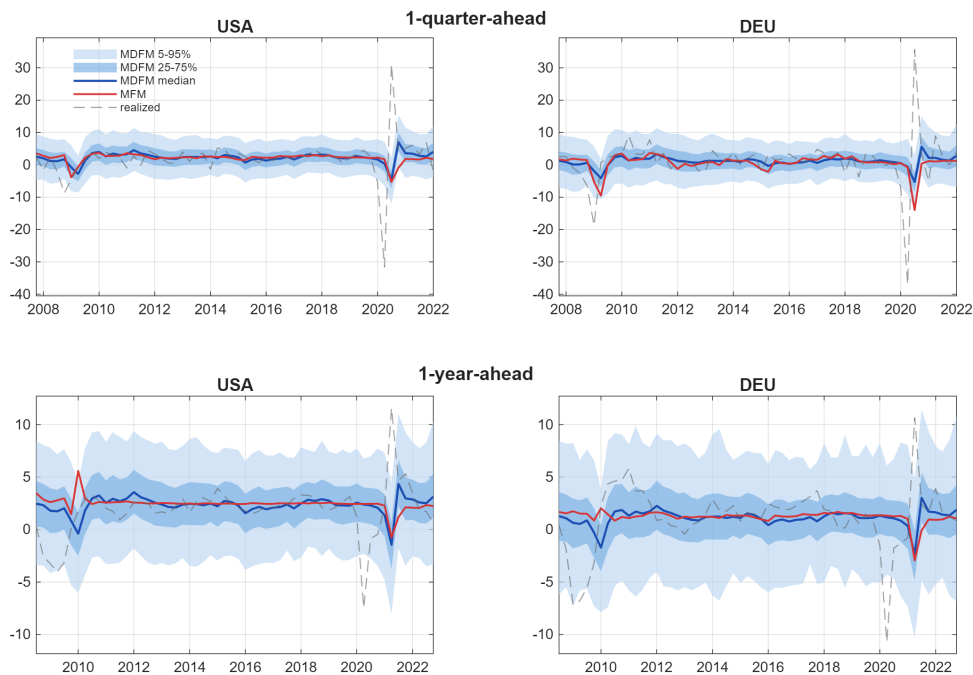


Figure 12: Posterior predictive fan charts for real GDP growth (annualized percent change) in the United States and Germany. The top row shows 1-quarter-ahead forecasts; the bottom row shows 4-quarter (1-year) average forecasts, with both MDFM posterior predictive draws and realized growth averaged across  $h = 1, \dots, 4$ . Shaded bands show 25%–75% (dark blue) and 5%–95% (light blue) MDFM posterior predictive intervals; the solid blue line is the MDFM posterior median; the solid red line is the MFM direct-projection point forecast; the dashed grey line shows realized values.

### 5.5. Robustness to data ordering

The lower-triangular identification scheme adopted in Section 2.2 fixes the location of the latent factor matrix at the cost of conditioning the loadings on the chosen ordering of rows and columns. To assess whether this choice materially affects the out-of-sample evidence, we re-run the entire expanding-window pipeline under seven alternative orderings of the OECD panel: full reversal of rows and columns, alphabetical, three random permutations with fixed seeds, a rows-only random permutation (columns held at baseline), and a columns-only random permutation (rows held at baseline). For each ordering we permute the predictive draws back to baseline coordinates before computing forecast

losses. Table 4 reports the pooled RMSFE across the panel for each ordering. It is clear that the pooled RMSFE varies by less than six percent across the eight orderings at every horizon. The pooled forecasting performance of the MDFM is therefore practically invariant to the ordering of the OECD panel.

Table 4: Pooled out-of-sample RMSFE across the OECD panel for the MDFM under eight orderings of rows and columns. RMSFE is pooled over all  $19 \times 10$  cells and all expanding-window forecast origins.

Ordering	$h = 1$	$h = 4$	$h = 8$
Baseline (economic ordering)	1.131	0.656	0.431
Reverse rows and columns	1.087	0.646	0.428
Alphabetical	1.110	0.641	0.419
Random #1 (seed 1001)	1.123	0.638	0.412
Random #2 (seed 1002)	1.117	0.654	0.430
Random #3 (seed 1003)	1.081	0.645	0.424
Rows random, columns baseline	1.137	0.673	0.446
Rows baseline, columns random	1.124	0.649	0.424
Range across orderings (%)	5.2	5.5	8.3

## 6. Concluding Remarks and Future Research

We have proposed a class of Bayesian dynamic factor models for high-dimensional matrix-valued time series, with autoregressive factor dynamics, stochastic volatility and outlier components, and a Kronecker-structured idiosyncratic covariance that captures cross-row and cross-column correlation. An efficient Gibbs sampler exploits the Kronecker structure to sample the loading matrices and row/column covariances jointly. For model comparison, we adapt the cross-entropy importance-sampling estimator of [Chan and Eisenstat \(2015\)](#) to select the factor-matrix dimension, distinguish vector from matrix specifications, and identify the structure of the idiosyncratic component. Applied to a 19-country, 10-indicator OECD panel, the model recovers a two-factor structure—a real-activity factor and a price factor that comoves with global oil prices—with idiosyncratic risks clustering within the European Union and volatility spiking during the Asian financial crisis, the Great Recession, and the COVID-19 pandemic. Out of sample, it delivers statistically significant forecast gains over a static matrix factor benchmark.

Several extensions are natural directions for future work. First, replacing the AR(1) dynamics for each cell of  $\mathbf{F}_t$  with a matrix autoregressive process would allow cross-cell dependence in factor evolution, at the cost of additional identification and computational challenges. Second, a structural matrix factor model that imposes economic restrictions on  $\mathbf{A}$ ,  $\mathbf{B}$ , or  $\mathbf{F}_t$  would enable analysis of cross-country shock transmission. Third, sparse priors on the loadings could automate the selection of relevant factor cells. We leave these to future research.

## References

- Adrian, T., Boyarchenko, N., Giannone, D., 2019. Vulnerable Growth. *American Economic Review* 109, 1263–1289.
- Aguilar, O., West, M., 2000. Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics* 18, 338–357.
- Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–171.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bai, J., Ng, S., 2013. Principal Components Estimation and Identification of Static Factors. *Journal of Econometrics* 176, 18–29. doi:[10.1016/j.jeconom.2013.03.007](https://doi.org/10.1016/j.jeconom.2013.03.007).
- Bai, J., Wang, P., 2015. Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics* 33, 221–240.
- Bañbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. *Journal of Applied Econometrics* 25, 71–92.
- Barigozzi, M., Hallin, M., 2024. The dynamic, the static, and the weak: factor models and the analysis of high-dimensional time series. *Journal of Time Series Analysis* .
- Bernanke, B.S., Boivin, J., Elias, P., 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics* 120, 387–422. doi:[10.1162/0033553053327452](https://doi.org/10.1162/0033553053327452).
- Carriero, A., Clark, T.E., Marcellino, M., 2015. Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics* 30, 46–73.
- Carriero, A., Clark, T.E., Marcellino, M., 2016. Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics* 34, 375–390.
- Carriero, A., Clark, T.E., Marcellino, M., 2019. Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics* 212, 137–154.

- Carriero, A., Clark, T.E., Marcellino, M., 2024. Capturing macro-economic tail risks with Bayesian vector autoregressions. *Journal of Money, Credit and Banking* 56, 1099–1127.
- Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* , 1281–1304.
- Chan, J., León-González, R., Strachan, R.W., 2018. Invariant Inference and Efficient Computation in the Static Factor Model. *Journal of the American Statistical Association* 113, 819–828. doi:[10.1080/01621459.2017.1287080](https://doi.org/10.1080/01621459.2017.1287080).
- 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.
- Chan, J.C., Yu, X., Zhang, W., 2025. Bayesian model comparison for large bayesian vars after the covid-19 pandemic. *Journal of Econometrics* , 106072.
- Chan, J.C.C., 2020. Large Bayesian VARs: A flexible Kronecker error covariance structure. *Journal of Business & Economic Statistics* 38, 68–79.
- Chan, J.C.C., 2023. Comparing stochastic volatility specifications for large Bayesian VARs. *Journal of Econometrics* 235, 1419–1446.
- Chan, J.C.C., Eisenstat, E., 2015. Marginal likelihood estimation with the cross-entropy method. *Econometric Reviews* 34, 256–285.
- Chan, J.C.C., Eisenstat, E., 2018. Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics* 33, 509–532.
- Chan, J.C.C., Jeliaskov, I., 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.C., Qi, Y., 2025. Large Bayesian matrix autoregressions. *Journal of Econometrics* Forthcoming.
- Chang, J., Han, Y., Yu, R., Zhang, B., 2023. Modelling matrix-variate time series via dynamic tensor factor models. Working paper .
- Chen, B., Chen, E.Y., Bolivar, S., Chen, R., 2024. Time-varying matrix factor models. arXiv preprint [arXiv:2404.01546](https://arxiv.org/abs/2404.01546) .
- Chen, E.Y., Fan, J., 2023. Statistical inference for high-dimensional matrix-variate factor models. *Journal of the American Statistical Association* 118, 1038–1055.
- Chen, E.Y., Tsay, R.S., Chen, R., 2020. Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association* .
- Chen, R., Xiao, H., Yang, D., 2021. Autoregressive models for matrix-valued time series. *Journal of Econometrics* 222, 539–560.

- Chen, R., Yang, D., Zhang, C.H., 2022. Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association* 117, 94–116.
- Chib, S., Greenberg, E., 1994. Bayes inference in regression models with ARMA (p, q) errors. *Journal of Econometrics* 64, 183–206.
- Chib, S., Nardari, F., Shephard, N., 2006. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics* 134, 341–371.
- Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics* 8, 262–302.
- Cong, Y., Chen, B., Zhou, M., 2017. Fast simulation of hyperplane-truncated multivariate normal distributions. *Bayesian Analysis* 12, 1017–1037.
- Cross, J., Poon, A., 2016. Forecasting structural change and fat-tailed events in Australian macroeconomic variables. *Economic Modelling* 58, 34–51.
- Del Negro, M., Otrok, C., 2008. Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles. Federal Reserve Bank of New York Staff Reports .
- Doz, C., Giannone, D., Reichlin, L., 2012. A quasi-maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics* 94, 1014–1024.
- Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. *Journal of Finance* 47, 427–465.
- Fong, E., Holmes, C.C., 2020. On the marginal likelihood and cross-validation. *Biometrika* 107, 489–496.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: identification and estimation. *Review of Economics and Statistics* 82, 540–554.
- Frühwirth-Schnatter, S., Hosszejni, D., Lopes, H.F., 2024a. Sparse Bayesian Factor Analysis When the Number of Factors is Unknown. *Bayesian Analysis* doi:[10.1214/24-BA1423](https://doi.org/10.1214/24-BA1423). advance publication.
- Frühwirth-Schnatter, S., Hosszejni, D., Lopes, H.F., 2024b. When It Counts — Econometric Identification of the Basic Factor Model Based on GLT Structures. *Econometrics* 12, 1–29. doi:[10.3390/econometrics12040026](https://doi.org/10.3390/econometrics12040026).
- Geweke, J., Zhou, G., 1996. Measuring the Pricing Error of the Arbitrage Pricing Theory. *Review of Financial Studies* 9, 557–587. doi:[10.1093/rfs/9.2.557](https://doi.org/10.1093/rfs/9.2.557).
- Giglio, S., Xiu, D., Zhang, D., 2025. Test assets and weak factors. *Journal of Finance* 80, 259–319.
- He, Y., Kong, X., Trapani, L., Yu, L., 2023. One-way or two-way factor model for matrix sequences? *Journal of Econometrics* 235, 1981–2004.

- He, Y., Kong, X., Yu, L., Zhang, X., Zhao, C., 2024. Matrix factor analysis: from least squares to iterative projection. *Journal of Business & Economic Statistics* 42, 322–334.
- Huber, F., Pfarrhofer, M., Piribauer, P., 2020. A Multi-Country Dynamic Factor Model with Stochastic Volatility for Euro Area Business Cycle Analysis. *Journal of Forecasting* 39, 911–926. doi:[10.1002/for.2667](https://doi.org/10.1002/for.2667).
- Jacquier, E., Polson, N.G., Rossi, P.E., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122, 185–212.
- Justiniano, A., Primiceri, G.E., 2008. The time-varying volatility of macroeconomic fluctuations. *American Economic Review* 98, 604–641.
- Kastner, G., Frühwirth-Schnatter, S., Lopes, H.F., 2017. Efficient Bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics* 26, 905–917.
- Kaufmann, S., Schumacher, C., 2019. Bayesian Estimation of Sparse Dynamic Factor Models with Order-Independent and Ex-Post Mode Identification. *Journal of Econometrics* 210, 116–134. doi:[10.1016/j.jeconom.2018.11.008](https://doi.org/10.1016/j.jeconom.2018.11.008).
- Koop, G.M., 2013. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics* 28, 177–203.
- Kose, M.A., Otrok, C., Whiteman, C.H., 2003. International business cycles: world, region, and country-specific factors. *American Economic Review* 93, 1216–1239.
- Leung, D., Drton, M., 2016. Order-Invariant Prior Specification in Bayesian Factor Analysis. *Statistics and Probability Letters* 111, 60–66. doi:[10.1016/j.spl.2016.01.006](https://doi.org/10.1016/j.spl.2016.01.006).
- Li, M., Scharth, M., 2022. Leverage, asymmetry, and heavy tails in the high-dimensional factor stochastic volatility model. *Journal of Business & Economic Statistics* 40, 285–301.
- Liu, X., Chen, E.Y., 2019. Helping effects against curse of dimensionality in threshold factor models for matrix time series. arXiv preprint arXiv:1904.07383 .
- Lopes, H.F., West, M., 2004. Bayesian model assessment in factor analysis. *Statistica Sinica* , 41–67.
- Marcellino, M., Stock, J.H., Watson, M.W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* 135, 499–526.
- Nobile, A., 2000. Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics* 99, 335–345.
- O’Hagan, A., 1995. Fractional Bayes Factors for Model Comparison. *Journal of the Royal Statistical Society, Series B* 57, 99–118.

- Poncela, P., Ruiz, E., Miranda, K., 2021. Factor extraction using Kalman filter and smoothing: this is not just another survey. *International Journal of Forecasting* 37, 1399–1425.
- Qin, L., Wang, Y., Zhu, Y., Shia, B.C., 2025. Bayesian dynamic matrix factor models. *Journal of Business & Economic Statistics* , 1–13.
- Sargent, T.J., Sims, C.A., et al., 1977. Business cycle modeling without pretending to have too much a priori economic theory, in: *New Methods in Business Cycle Research*. volume 1, pp. 145–168.
- Stock, J.H., Watson, M.W., 2002a. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.
- Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.
- Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467 .
- Stock, J.H., Watson, M.W., 2012. Dynamic factor models, in: *The Oxford Handbook of Economic Forecasting*. Oxford University Press, pp. 35–59.
- Stock, J.H., Watson, M.W., 2016. Core inflation and trend inflation. *Review of Economics and Statistics* 98, 770–784.
- Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics* 208, 231–248.
- West, M., 2003. Bayesian Factor Regression Models in the “Large  $p$ , Small  $n$ ” Paradigm, in: Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., West, M. (Eds.), *Bayesian Statistics 7*. Oxford University Press, pp. 733–742.
- Yu, L., He, Y., Kong, X., Zhang, X., 2022. Projected estimation for large-dimensional matrix factor models. *Journal of Econometrics* 229, 201–217.
- Yu, R., Chen, R., Xiao, H., Han, Y., 2024. Dynamic matrix factor models for high dimensional time series. arXiv preprint arXiv:2407.05624 .

# Online Supplemental Materials

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

### Appendix A. Proof of Propositions

#### Appendix 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 (Appendix A.1), we may identify a rotation of  $\mathbf{F}_t$ , given by  $\mathbf{C}\mathbf{F}_t\mathbf{D}'$ .

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

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

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

Let

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

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,

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p_1 1} & a_{p_1 2} & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np_1} \end{bmatrix} \begin{bmatrix} c_{11} & \dots & c_{1p_1} \\ \vdots & \ddots & \vdots \\ c_{p_1 1} & \dots & c_{p_1 p_1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{21}^* & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p_1 1}^* & a_{p_1 2}^* & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}^* & a_{n2}^* & \dots & a_{np_1}^* \end{bmatrix} \quad (\text{Appendix A.2})$$

For (Appendix A.2) 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,

$$\begin{bmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{p_2 1} & b_{p_2 2} & \dots & b_{p_2 p_2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np_2} \end{bmatrix} \begin{bmatrix} d_{11} & \dots & d_{1p_2} \\ \vdots & \ddots & \vdots \\ d_{p_2 1} & \dots & d_{p_2 p_2} \end{bmatrix} = \begin{bmatrix} b_{11}^* & 0 & \dots & 0 \\ b_{21}^* & b_{22}^* & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{p_2 1}^* & b_{p_2 2}^* & \dots & b_{p_2 p_2}^* \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1}^* & b_{n2}^* & \dots & b_{np_2}^* \end{bmatrix} \quad (\text{Appendix A.3})$$

For (Appendix A.3) to hold, we must have  $d_{i,j} = 0$  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  $i = 1, \dots, p_2$ .

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

$$\mathbf{f}_t = \mathbf{H}_\rho \mathbf{f}_{t-1} + \mathbf{u}_t, \quad (\text{Appendix A.4})$$

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

Define  $\mathbf{M} \equiv \mathbf{D} \otimes \mathbf{C}$ , multiply (Appendix A.4) by  $\mathbf{M}$  on both side, we have

$$\mathbf{M} \mathbf{f}_t = \mathbf{M} \mathbf{H}_\rho \mathbf{f}_{t-1} + \mathbf{M} \mathbf{u}_t. \quad (\text{Appendix A.5})$$

Therefore

$$\tilde{\mathbf{f}}_t = \mathbf{M}\mathbf{H}_\rho\mathbf{M}^{-1}\tilde{\mathbf{f}}_{t-1} + \mathbf{M}\mathbf{u}_t. \quad (\text{Appendix A.6})$$

The observation equation after the rotation becomes

$$\mathbf{Y}_t = \mathbf{A}\mathbf{C}^{-1}\tilde{\mathbf{F}}_t(\mathbf{D}')^{-1}\mathbf{B}' + \mathbf{E}_t. \quad (\text{Appendix A.7})$$

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  $\mathbf{C}$  and  $\mathbf{D}$  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  $i = 1, \dots, p_2$ , 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.

### *Appendix 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 ([Appendix A.2](#)). 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.

## Appendix B. Bayesian Estimation for MDFM with Stochastic Volatility

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

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

$$\text{vec}(\mathbf{F}_t) = \mathbf{H}_{\rho_1} \text{vec}(\mathbf{F}_{t-1}) + \dots + \mathbf{H}_{\rho_q} \text{vec}(\mathbf{F}_{t-q}) + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_t), \quad (\text{Appendix B.2})$$

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,  $\mathbf{H}_{\rho_l}$  is a diagonal matrix of autoregressive coefficients  $(\rho_{1,l}, \dots, \rho_{p_1 p_2, l})'$ ,  $l = 1, \dots, q$ , 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  $\boldsymbol{\Sigma}_c$ :

$$\begin{aligned} \boldsymbol{\Sigma}_r &\sim \mathcal{IW}(\nu_r, \mathbf{S}_r), & (\text{vec}(\mathbf{A}') | \boldsymbol{\Sigma}_r) &\sim \mathcal{N}(\text{vec}(\mathbf{A}'_0), \boldsymbol{\Sigma}_r \otimes \mathbf{V}_{\mathbf{A}'}), \\ \boldsymbol{\Sigma}_c &\sim \mathcal{IW}(\nu_c, \mathbf{S}_c), & (\text{vec}(\mathbf{B}') | \boldsymbol{\Sigma}_c) &\sim \mathcal{N}(\text{vec}(\mathbf{B}'_0), \boldsymbol{\Sigma}_c \otimes \mathbf{V}_{\mathbf{B}'}). \end{aligned} \quad (\text{Appendix B.3})$$

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

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

The prior variance  $\lambda_{j,k}^2$  is assumed to have an 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

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

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 $(\mathbf{A}', \boldsymbol{\Sigma}_r | \mathbf{Y}, \mathbf{B}, \mathbf{F}, \boldsymbol{\Sigma}_c)$

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

inverse-Wishart distribution:

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

where

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

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. \(2017\)](#) or Algorithm 1 in [Chan and Qi \(2025\)](#) to efficiently sample  $(\text{vec}(\mathbf{A}') | \cdot) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}'), \boldsymbol{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})$  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

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

which can be realized by the following four steps:

- (1) Compute  $\mathbf{C} = \mathbf{C}_{\boldsymbol{\Sigma}_r^{-1}} \otimes \mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$ , where  $\mathbf{C}_{\boldsymbol{\Sigma}_r^{-1}}$  is the lower Cholesky factor of  $\boldsymbol{\Sigma}_r^{-1}$ , 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 $(\mathbf{B}', \boldsymbol{\Sigma}_c | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \boldsymbol{\Sigma}_r)$

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

$$(\mathbf{B}, \boldsymbol{\Sigma}_c | \cdot) \mathcal{N}\mathcal{IW}(\hat{\mathbf{B}}', \mathbf{K}_{\mathbf{B}'}^{-1}, \hat{\nu}_c, \hat{\mathbf{S}}_c),$$

where

$$\begin{aligned} \mathbf{K}_{\mathbf{B}'} &= \mathbf{V}_{\mathbf{B}'}^{-1} + \sum_{t=1}^T \omega_t^{-1} \mathbf{F}_t' \mathbf{A}' \boldsymbol{\Sigma}_r^{-1} \mathbf{A} \mathbf{F}_t, & \hat{\mathbf{B}}' &= \mathbf{K}_{\mathbf{B}'}^{-1} \left( \mathbf{V}_{\mathbf{B}'}^{-1} \mathbf{B}'_0 + \sum_{t=1}^T \omega_t^{-1} \mathbf{F}_t' \mathbf{A}' \boldsymbol{\Sigma}_r^{-1} \mathbf{Y}_t \right) \\ \hat{\nu}_c &= \nu_c + Tn, & \hat{\mathbf{S}}_c &= \mathbf{S}_c + \mathbf{B}_0 \mathbf{V}_{\mathbf{B}'}^{-1} \mathbf{B}'_0 + \sum_{t=1}^T \omega_t^{-1} \mathbf{Y}_t' \boldsymbol{\Sigma}_r^{-1} \mathbf{Y}_t - \hat{\mathbf{B}} \mathbf{K}_{\mathbf{B}'} \hat{\mathbf{B}}'. \end{aligned}$$

We sample  $(\mathbf{B}, \boldsymbol{\Sigma}_c | \cdot)$  in two steps. First, we sample  $\boldsymbol{\Sigma}_c$  marginally from  $(\boldsymbol{\Sigma}_c | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \boldsymbol{\Sigma}_r) \sim \mathcal{IW}(\hat{\mathbf{S}}_c, \nu_c + Tn)$  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  $(\text{vec}(\mathbf{B}') | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\hat{\mathbf{B}}), \boldsymbol{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}'}^{-1})$ , 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  $\hat{\mathbf{S}}_c$ . Denote this matrix as  $\hat{\mathbf{S}}_c^{Trans}$ .
- (2) Construct a lower triangular matrix  $\boldsymbol{\Delta}$  such that
  - $\delta_{ii}$  equal to the square root of  $\chi_{\hat{\nu}_c+1-i}^2$  for  $i = 1, \dots, n-1$ ;
  - $\delta_{nn} = (l_{nn})^{-1}$ , where  $l_{nn}$  is the  $(n, n)$ -th element in the Cholesky decomposition of  $(\hat{\mathbf{S}}_c^{Trans})^{-1}$ , denoted as  $\mathbf{L}$
  - $\delta_{ij}$  equal to  $\mathcal{N}(0, 1)$  random variates,  $i > j$ .
- (3) Set  $\boldsymbol{\Sigma}_c = (\mathbf{L}^{-1})' (\boldsymbol{\Delta}^{-1})' \boldsymbol{\Delta}^{-1} \mathbf{L}^{-1}$ .
- (4) Exchange the row/column 1 and  $n$  of  $\boldsymbol{\Sigma}_c$  back.

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

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

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

where

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

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

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

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

Note that  $\boldsymbol{\rho}_{j,k}$  is a  $q \times 1$  vector:  $\boldsymbol{\rho}_{j,k} = (\rho_{j,k,1}, \dots, \rho_{j,k,q})'$ . We rewrite (2) as follows:

$$\tilde{\mathbf{f}}_{j,k} = \tilde{\mathbf{F}}_{j,k} \boldsymbol{\rho}_{j,k} + \mathbf{u}_{j,k}, \quad \mathbf{u}_{j,k} \sim \mathcal{N}(\mathbf{0}, \lambda_{j,k} \mathbf{I}_{T-q}), \quad (\text{Appendix B.4})$$

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

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

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  $\mathbf{K}_{\boldsymbol{\rho}_{j,k}} = \mathbf{V}_{\boldsymbol{\rho}_{j,k}}^{-1} + \tilde{\mathbf{F}}_{j,k}' \tilde{\mathbf{F}}_{j,k} / \lambda_{j,k}^2$ ,  $\hat{\boldsymbol{\rho}}_{j,k} = \mathbf{K}_{\boldsymbol{\rho}_{j,k}}^{-1} (\mathbf{V}_{\boldsymbol{\rho}_{j,k}}^{-1} \boldsymbol{\rho}_{j,k,0} + \tilde{\mathbf{F}}_{j,k}' \tilde{\mathbf{f}}_{j,k} / \lambda_{j,k}^2)$ . The proposed value  $\boldsymbol{\rho}_{j,k}^*$  is accepted with probability

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

## 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  $\hat{\mathbf{h}}$  and  $\mathbf{K}_{\mathbf{h}}$ , respectively. Then we use  $\mathcal{N}(\hat{\mathbf{h}}, \mathbf{K}_{\mathbf{h}}^{-1})$  as the proposal distribution, and sample  $\mathbf{h}$  using an acceptance-rejection Metropolis-Hasting step. Samplers for  $\phi$  and  $\sigma_h^2$  are standard and we omit the details in this paper.

### 5.2 The explicit outlier components: sampling from $(\mathbf{o}, p_{\mathbf{o}} | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r)$

We follow Stock and Watson (2016) to discretize the support of  $o_t$  to simplify estimation.

Specifically, we use a grid with points at  $1, 2, 3, \dots, 20$ . 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 $(q_t^2 | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r)$ , $t = 1, \dots, T$

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

## Appendix C. Estimating Marginal Likelihoods

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

### Appendix C.1. Integrated Likelihood

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

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

System (Appendix C.1) can be rewritten as follows

$$\begin{aligned} \mathbf{y} &= (\mathbf{I}_T \otimes \mathbf{A} \otimes \mathbf{B})\mathbf{f} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes (\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r)), \\ \mathbf{f} \mid \boldsymbol{\rho}, \boldsymbol{\Omega} &\sim \mathcal{N}\left(\mathbf{0}, [\mathbf{H}'_\rho(\mathbf{I}_T \otimes \boldsymbol{\Omega})^{-1}\mathbf{H}_\rho]^{-1}\right). \end{aligned} \tag{Appendix C.2}$$

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

$$\mathbf{y} \mid \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho} \sim \mathcal{N}(\bar{\mathbf{y}}, \bar{\mathbf{D}}_{\mathbf{y}}), \tag{Appendix C.3}$$

where

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

and

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

It can be very costly to compute the inverse of the covariance matrix  $\bar{\mathbf{D}}_{\mathbf{y}}$ . Therefore, here we use Kalman filter. In particular, it is not difficult to show that the marginal

distribution for  $\mathbf{f}_t \equiv \text{vec}(\mathbf{F}_t)$  is as follows:

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

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

Therefore, the integrated likelihood at time  $t$  is:

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

where

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

### Appendix 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

$$\begin{aligned} f(\boldsymbol{\theta}; \mathbf{v}) &= f(\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\rho}; \mathbf{v}) \\ &= f(\mathbf{A}; \bar{\mathbf{A}}, \bar{\mathbf{D}}_{\mathbf{A}}) \cdot f(\boldsymbol{\Sigma}_c; \Psi_c, \nu_c) \cdot f(\boldsymbol{\Sigma}_r; \Psi_r, \nu_r) \cdot f(\boldsymbol{\lambda}; \nu_\lambda, S_\lambda) \cdot f(\boldsymbol{\rho}; \bar{\boldsymbol{\rho}}, \bar{\mathbf{D}}_\rho). \end{aligned} \tag{Appendix C.4}$$

In terms of the parameteric family, we use Gaussian density for  $f(\mathbf{A}; \bar{\mathbf{A}}, \bar{\mathbf{D}}_{\mathbf{A}})$ , where  $\bar{\mathbf{A}}$  and  $\bar{\mathbf{D}}$  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  $f(\boldsymbol{\rho}; \bar{\boldsymbol{\rho}}, \bar{\mathbf{D}}_\rho)$ , where  $\bar{\boldsymbol{\rho}}$  and  $\bar{\mathbf{D}}_\rho$  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 distribution given the posterior samples, and then compute the degree of freedom and scale matrix of the inverse Wishart distribution using *Lemma 1*.

*Lemma 1:*  $\boldsymbol{\Sigma}$  follows an inverse Wishart distribution if  $\mathbf{K} \equiv \boldsymbol{\Sigma}^{-1}$  follows a Wishart

distribution, formally expressed as

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_d(\delta - d + 1, \boldsymbol{\Psi}^{-1}) \Leftrightarrow \mathbf{K} = \boldsymbol{\Sigma}^{-1} \sim \mathcal{W}_d(\delta, \boldsymbol{\Psi}), \quad (\text{Appendix C.5})$$

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

A Wishart distribution is defined as:

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

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

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

The log-likelihood function is therefore

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

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

$$\frac{d \log(f(\mathbf{K}_1, \dots, \mathbf{K}_M \mid \boldsymbol{\Psi}, \delta))}{d \boldsymbol{\Psi}} = -\frac{M \delta}{2} \boldsymbol{\Psi}^{-1} + \frac{1}{2} \boldsymbol{\Psi}^{-1} \sum_{m=1}^M \mathbf{K}_m \boldsymbol{\Psi}^{-1}, \quad (\text{Appendix C.6})$$

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 ([Appendix C.6](#)) we obtain a function of the MLE of  $\boldsymbol{\Psi}$  with respect to

the degree of freedom  $\delta$

$$\widehat{\Psi}^{mle} = \frac{1}{M\delta} \sum_{m=1}^M \mathbf{K}_m. \quad (\text{Appendix C.7})$$

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  $\delta$ . 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 (Appendix C.7) is

$$\begin{aligned} \frac{\partial \log f(\mathbf{K}_1, \dots, \mathbf{K}_M | \delta)}{\partial \delta} &= -\frac{dM}{2}(\log 2 + 1) + \frac{Md}{2} \log \delta - \frac{M}{2} \log \left| M^{-1} \sum_m \mathbf{K}_m \right| \\ &\quad - \frac{M}{2} \psi_d \left( \frac{\delta}{2} \right) + \frac{1}{2} \sum_m \log |\mathbf{K}_m|. \end{aligned} \quad (\text{Appendix C.8})$$

The second derivative is

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

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

## Appendix D. Additional Simulation Results

Table Appendix D.1: 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 Appendix D.2: 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				
(10, 10)	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
	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				
(20, 15)	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
	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				
(30, 20)	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
	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				

## Appendix E. Data: Multinational Macroeconomic Panel

Table [Appendix E.1](#) 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 [Appendix E.1](#) is the stationarity transformation for each variable.

Table Appendix E.1: List of variables

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

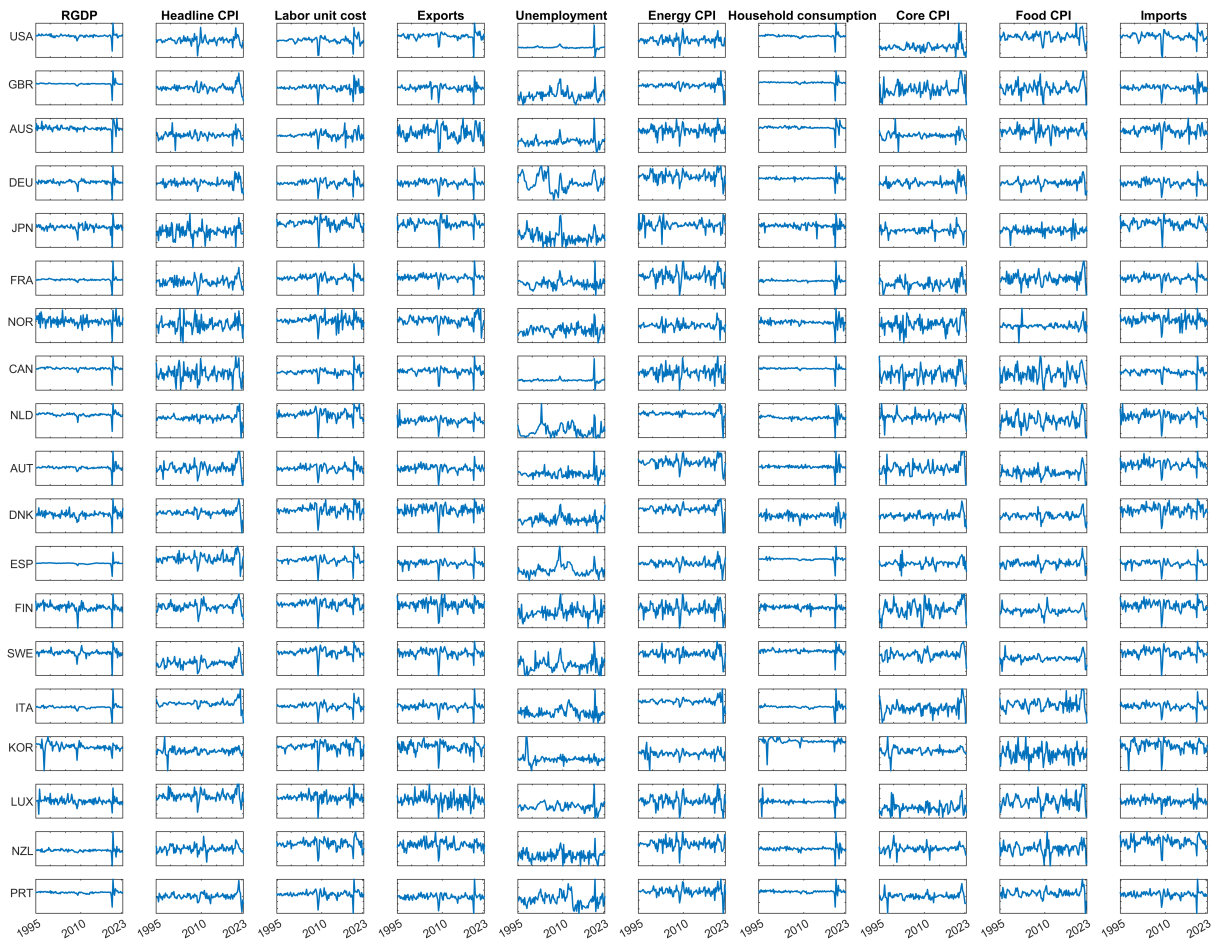


Figure Appendix E.1: The ten macroeconomic indicators (by columns) for 19 countries (by rows). The horizontal axis represents time and the vertical axis represents the standardized growth rates. The ranges of the vertical values are not the same. The order of countries and indicators is the order adopted in the estimation.

## Appendix F. The Wang et al. (2019) Benchmark: Implementation and Comparison

This appendix documents the implementation of the Wang et al. (2019) estimator used as the frequentist benchmark throughout the paper, and reports the auxiliary diagnostics that informed the comparison reported in Section 5 of the main text. The main text emphasizes the out-of-sample forecasting comparison; this appendix collects the methodological details and the in-sample diagnostics.

*Appendix F.1. Estimator and rank-selection rule*

Wang et al. (2019) estimate the matrix factor model  $\mathbf{Y}_t = \mathbf{A}\mathbf{F}_t\mathbf{B}^\top + \mathbf{E}_t$  in a frequentist framework. The loading spaces  $\text{span}(\mathbf{A})$  and  $\text{span}(\mathbf{B})$  are recovered as the leading eigenspaces of two symmetric positive semidefinite matrices constructed from lagged sample autocovariances:

$$\widehat{\mathbf{M}}_1 = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_h^{(1)} \left[ \widehat{\boldsymbol{\Sigma}}_h^{(1)} \right]^\top, \quad \widehat{\mathbf{M}}_2 = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_h^{(2)} \left[ \widehat{\boldsymbol{\Sigma}}_h^{(2)} \right]^\top, \quad (\text{Appendix F.1})$$

where  $\widehat{\boldsymbol{\Sigma}}_h^{(1)}$  and  $\widehat{\boldsymbol{\Sigma}}_h^{(2)}$  are row- and column-wise sample autocovariance matrices at lag  $h$ , and  $h_0$  is a user-specified upper lag. The construction in equation (Appendix F.1) exploits the fact that under the matrix factor model, lagged autocovariances retain only the factor-driven signal: the common factors are persistent so  $\text{Cov}(\mathbf{F}_t, \mathbf{F}_{t-h}) \neq 0$  for  $h \geq 1$ , while the idiosyncratic errors  $\mathbf{E}_t$  are assumed serially uncorrelated and so contribute zero in expectation. We follow the paper’s default choice of  $h_0 = 1$  throughout and report a sensitivity analysis below.

The factor dimensions  $(p_1, p_2)$  are selected by the eigenvalue-ratio rule:  $\widehat{p}_d = \arg \min_{i \leq P_d/2} \lambda_{d,i+1} / \lambda_{d,i}$  for  $d = 1, 2$ , where  $\lambda_{d,1} \geq \lambda_{d,2} \geq \dots$  are the eigenvalues of  $\widehat{\mathbf{M}}_d$  and  $P_d$  is the observed dimension along axis  $d$  ( $P_1 = n = 19$ ,  $P_2 = k = 10$  in our application). Loadings are then estimated as the orthonormal eigenvectors corresponding to the leading  $\widehat{p}_d$  eigenvalues, and the factor matrix is recovered as  $\widehat{\mathbf{F}}_t = \widehat{\mathbf{A}}^\top \mathbf{Y}_t \widehat{\mathbf{B}}$ .

*Appendix F.2. Application to the OECD panel*

On our OECD panel ( $n = 19$  countries,  $k = 10$  indicators,  $T = 115$  quarters), the eigenvalue-ratio rule selects  $(\widehat{p}_1, \widehat{p}_2) = (1, 1)$ , as shown in Figure Appendix F.1. The selection of  $\widehat{p}_1$  is decisive:  $\lambda_{1,2} / \lambda_{1,1} \approx 0.04$ , three orders of magnitude smaller than the next ratio in the sequence. The selection of  $\widehat{p}_2$  is less clear-cut:  $\lambda_{2,2} / \lambda_{2,1} \approx 0.23$  versus  $\lambda_{2,3} / \lambda_{2,2} \approx 0.69$  and  $\lambda_{2,4} / \lambda_{2,3} \approx 0.34$ , with  $\lambda_{2,2}$  and  $\lambda_{2,3}$  both clearly visible above the noise floor of the  $\widehat{\mathbf{M}}_2$  spectrum. Wang et al. (2019) document in their Table 5 that the eigenvalue-ratio rule is unstable in moderate- $T$  panels, with  $\widehat{p}_2 = 1$  selected with high frequency in simulations even when the true value is 2.

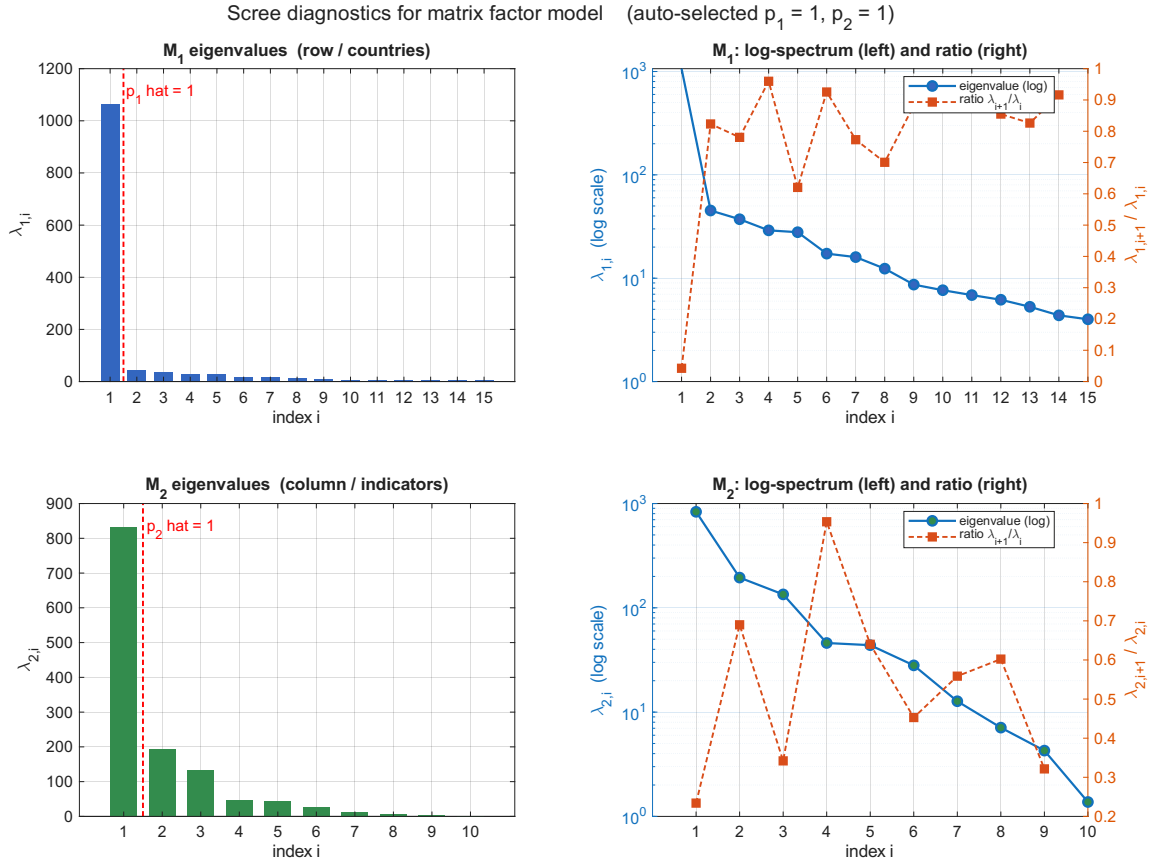


Figure Appendix F.1: Scree diagnostics for the Wang et al. (2019) estimator on the OECD panel. Top row:  $\widehat{M}_1$  eigenvalues (row / countries) in linear scale (left) and log scale with the eigenvalue-ratio diagnostic (right). Bottom row: same diagnostics for  $\widehat{M}_2$  (column / indicators). The vertical dashed lines mark the auto-selected  $\hat{p}_d$ .

### Appendix F.3. Sensitivity to the lag parameter $h_0$

Although  $h_0 = 1$  is the canonical choice, we verify that the rank selection and loading estimates are stable under alternative values. Table Appendix F.1 reports the auto-selected factor dimensions, in-sample RMSE under both the auto-selected ranks and a forced (1, 2) specification, and subspace distances of the loading estimates relative to the  $h_0 = 1$  baseline.

Table Appendix F.1: Sensitivity of the Wang et al. (2019) estimator to the lag parameter  $h_0$  in equation (Appendix F.1). Columns report the auto-selected dimensions, in-sample RMSE under auto selection, in-sample RMSE under forced  $(p_1, p_2) = (1, 2)$ , and subspace distances of the row and column loading estimates relative to  $h_0 = 1$ .

$h_0$	$\hat{p}_1$	$\hat{p}_2$	RMSE auto	RMSE (1, 2)	$D(\widehat{\mathbf{Q}}_1)/D(\widehat{\mathbf{Q}}_2)$
1	1	1	0.8280	0.7711	0.0000 / 0.0000
2	1	1	0.8272	0.7657	0.0297 / 0.0399
3	1	1	0.8273	0.7630	0.0378 / 0.0664
4	1	1	0.8264	0.7621	0.0452 / 0.0745

The auto-selected factor dimensions  $(\hat{p}_1, \hat{p}_2) = (1, 1)$  are invariant across  $h_0$ . The disagreement with our marginal-likelihood selection is therefore not an artifact of the choice of  $h_0$ . In addition, in-sample RMSE under auto selection varies by less than 0.2% across the grid, and the loading subspace distances  $D(\widehat{\mathbf{Q}}_1)$  and  $D(\widehat{\mathbf{Q}}_2)$  remain below 0.08, indicating high stability of the loading estimates with respect to  $h_0$ . Third, and most directly relevant, in-sample RMSE under the forced (1, 2) specification is uniformly lower than under auto selection, by 6–7 percentage points, across all values of  $h_0$ . This is an evidence that the OECD panel may exhibit a second column factor that the eigenvalue-ratio rule misses in a moderate- $T$  scenario.

#### Appendix F.4. In-sample fit comparison

We compare the in-sample fit of the Bayesian MDFM-cross-sv with  $(p_1, p_2) = (1, 2)$  to the Wang et al. (2019) benchmark with  $(\hat{p}_1, \hat{p}_2) = (1, 1)$ . Figure Appendix F.2 reports the per-cell ratio of in-sample RMSE between the two methods, arranged as a  $19 \times 10$  heatmap with countries on the rows and indicators on the columns.

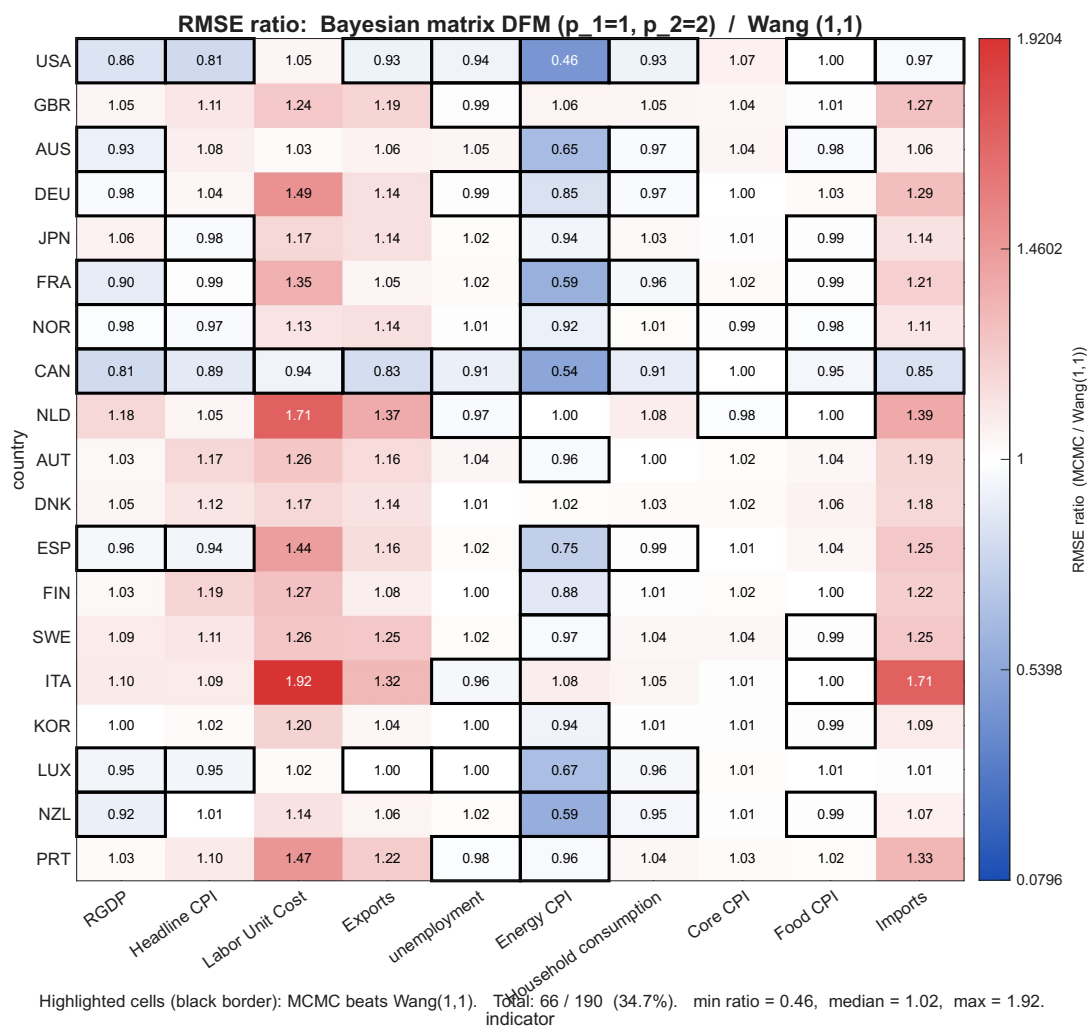


Figure Appendix F.2: Per-cell ratio of in-sample RMSE, Bayesian MDFM-cross-sv with  $(p_1, p_2) = (1, 2)$  relative to Wang et al. (2019) with  $(\hat{p}_1, \hat{p}_2) = (1, 1)$ . Black borders mark cells where the Bayesian model achieves lower RMSE.

The MDFM achieves lower RMSE on 66 of 190 cells (34.7%) with median ratio 1.02 — close to a tie. The in-sample comparison is structurally unfavorable to the more flexible model: Wang’s projection has substantially fewer effective parameters than MDFM-cross-sv with its Kronecker idiosyncratic covariance, AR(1) idiosyncratic dynamics, and common stochastic volatility, and reusing the estimation sample for evaluation flatters the parsimonious specification.

*Appendix F.5. Implementation of the forecasting benchmark*

For the out-of-sample comparison, the Wang et al. (2019) estimator is implemented as a direct projection benchmark in the spirit of Stock and Watson (2002b). At each forecast origin  $t$ , we first obtain the rank-(1, 1) loading estimates  $\widehat{\mathbf{Q}}_1$  and  $\widehat{\mathbf{Q}}_2$  from the in-sample data  $\{\mathbf{Y}_s\}_{s=1}^t$ , and construct the scalar factor estimate  $\widehat{f}_s = \widehat{\mathbf{Q}}_1^\top \mathbf{Y}_s \widehat{\mathbf{Q}}_2$  for  $s \leq t$ . For each forecast horizon  $h$  and each (country, indicator) pair  $(i, j)$ , we then estimate the horizon-specific direct projection

$$Y_{i,j,s+h} = \alpha_{i,j}^{(h)} + \beta_{i,j}^{(h)} \widehat{f}_s + \varepsilon_{i,j,s+h}^{(h)}, \quad s = 1, \dots, t - h, \quad (\text{Appendix F.2})$$

by ordinary least squares, and produce the forecast  $\widehat{Y}_{i,j,t+h} = \widehat{\alpha}_{i,j}^{(h)} + \widehat{\beta}_{i,j}^{(h)} \widehat{f}_t$ .

This is a one-factor analogue of the standard direct projection forecasting framework in the dynamic factor model literature. Direct projection is the natural choice for the Wang benchmark because the estimator does not specify dynamics for the latent factor matrix  $\mathbf{F}_t$ .

## Appendix G. Extra Forecasting Results

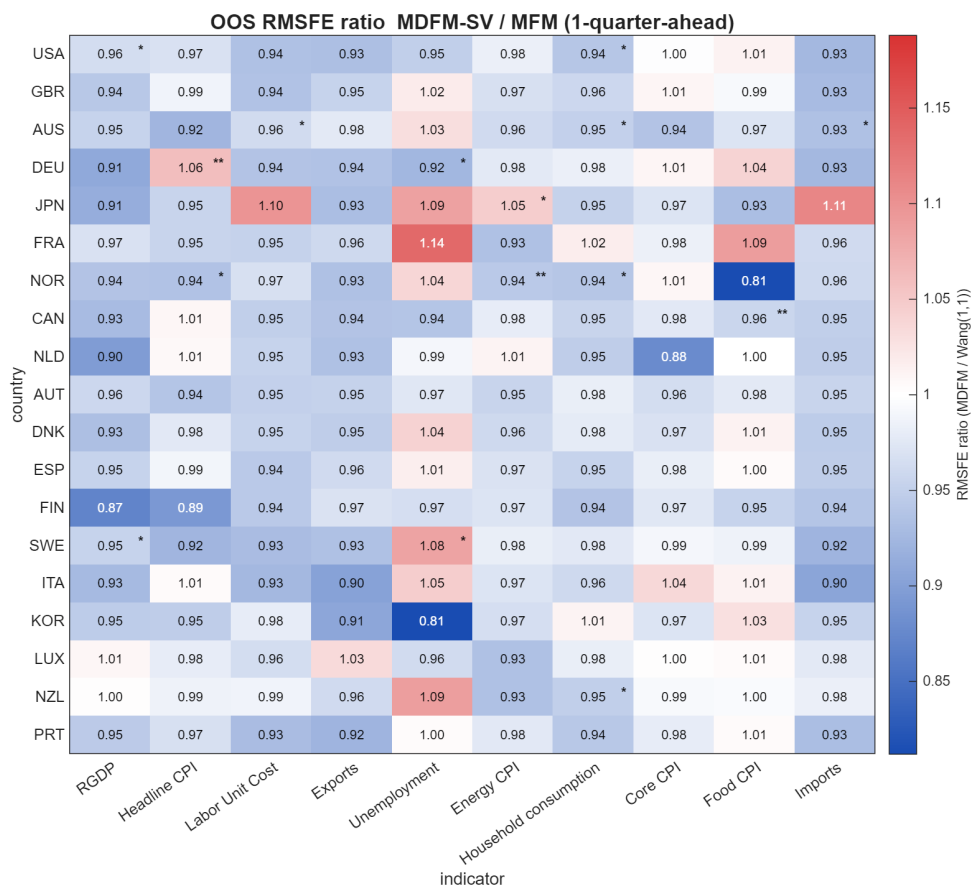


Figure Appendix G.1: Per-cell RMSFE ratios of the MDFM-SV relative to the static matrix factor benchmark, by indicator and country, at the 1-quarter-ahead horizon ( $h = 1$ ). Entries below one indicate that the MDFM-SV forecast is more accurate; asterisks denote rejection of the Diebold–Mariano test of equal predictive accuracy.

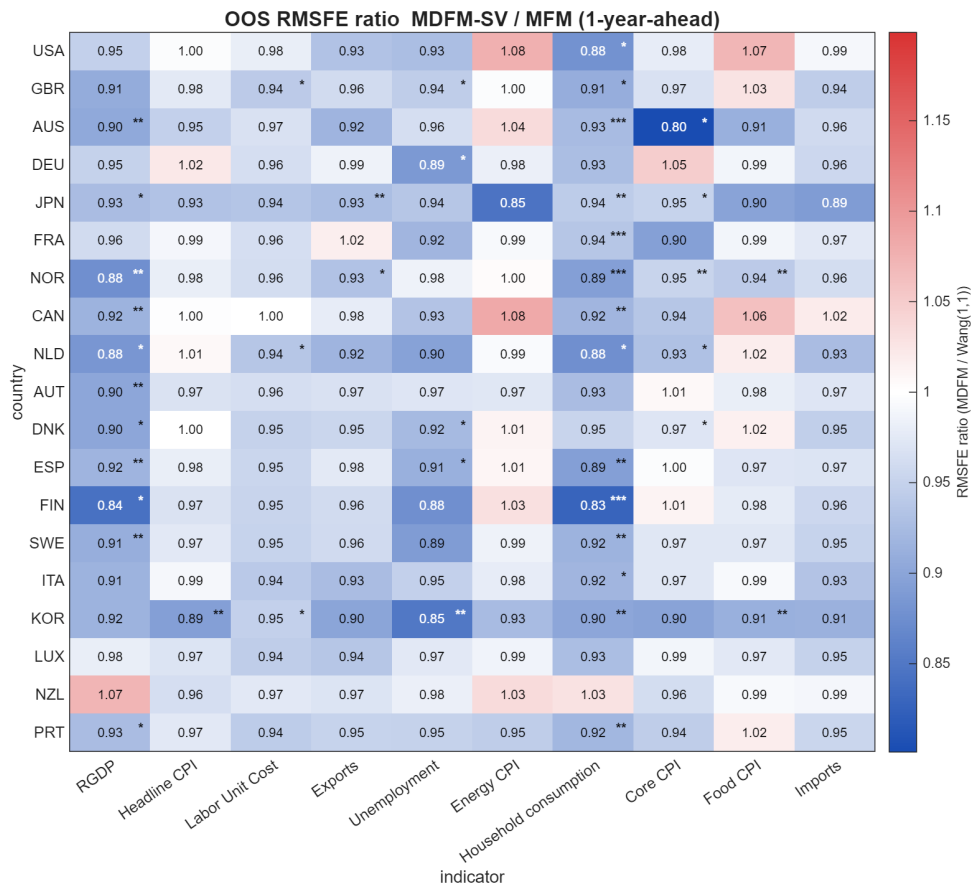


Figure Appendix G.2: Per-cell RMSFE ratios of the MDFM-SV relative to the static matrix factor benchmark, by indicator and country, at the 1-year-ahead horizon ( $h = 4$ ). See the note to Figure Appendix G.1.

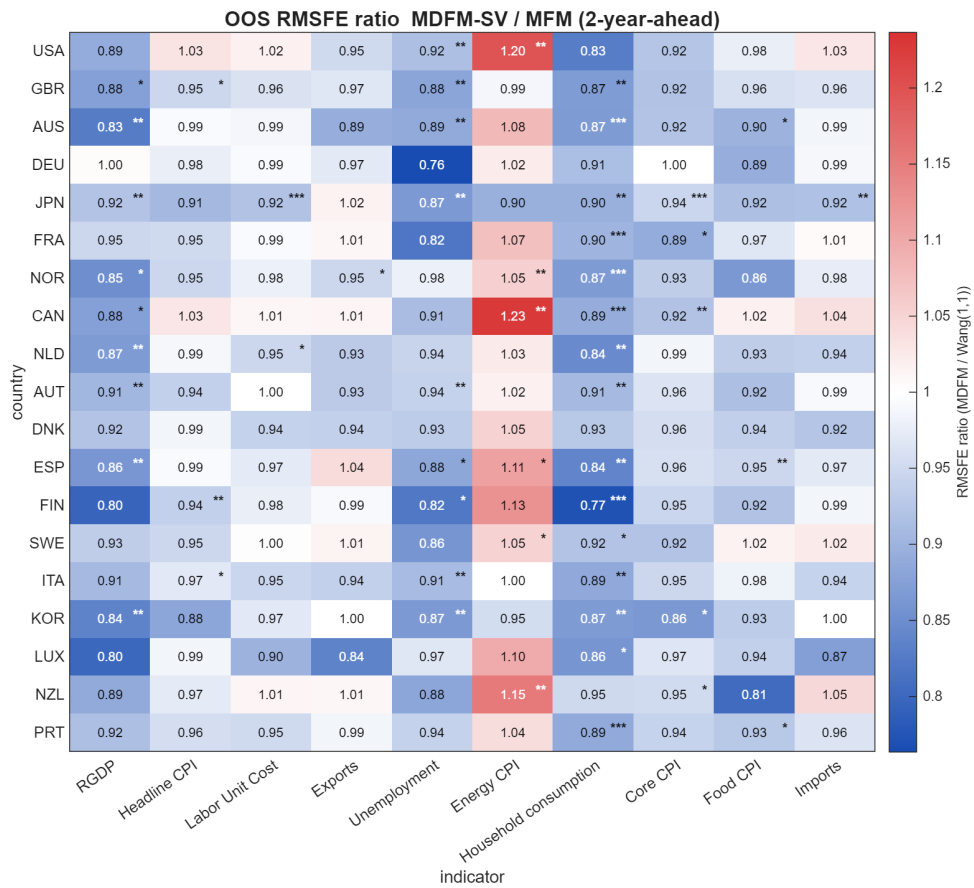


Figure Appendix G.3: Per-cell RMSFE ratios of the MDFM-SV relative to the static matrix factor benchmark, by indicator and country, at the 2-year-ahead horizon ( $h = 8$ ). See the note to Figure Appendix G.1.

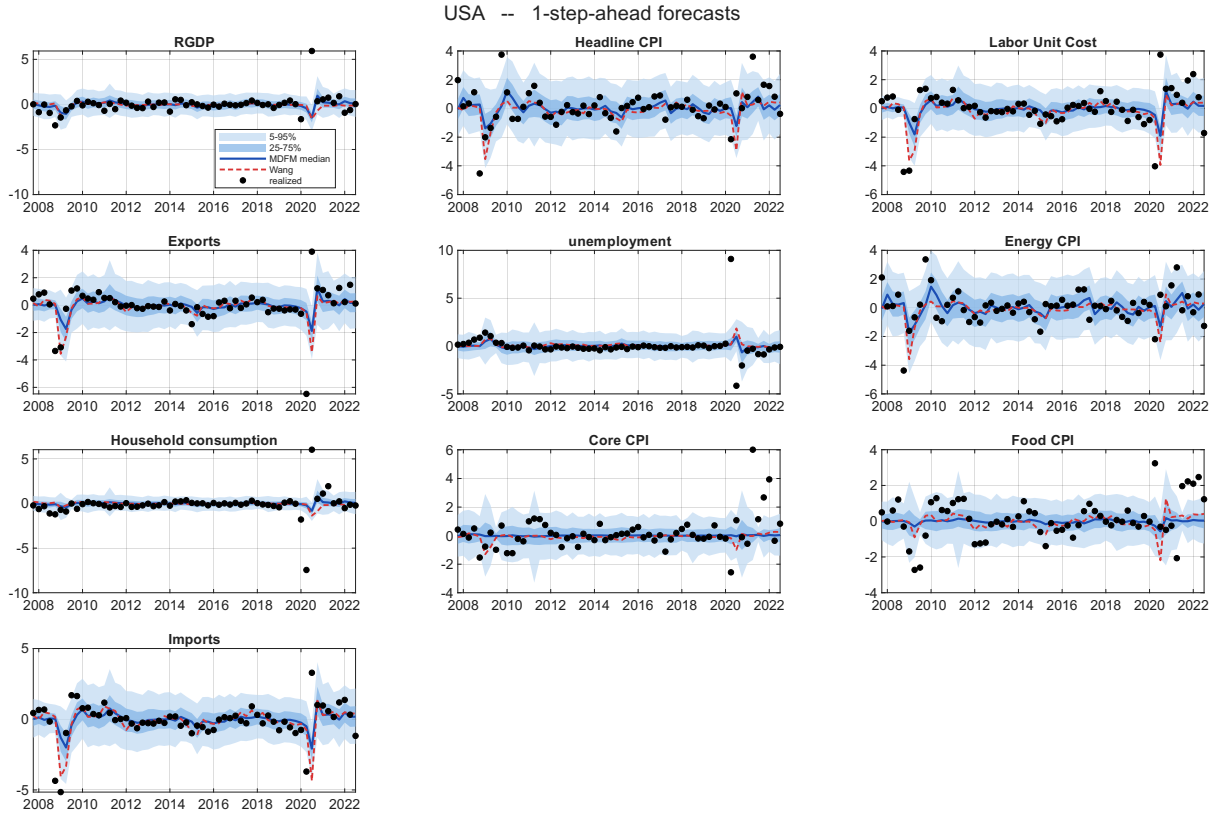


Figure Appendix G.4: Posterior predictive fan charts for real GDP growth (annualized percent change) across the G7 economies. The top block shows 1-quarter-ahead forecasts; the bottom block shows 4-quarter (1-year) average forecasts. Each panel reports the MDFM posterior median and 5–95% and 25–75% predictive bands together with the realized series.

## Appendix H. Application: Fama-French $10 \times 10$ 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).<sup>12</sup> The return series span from January 1990 to June 2024 (414 observations).<sup>13</sup> 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

<sup>12</sup>The data is available at [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

<sup>13</sup>We do not include data earlier than 1990 because there are many missing values in the early years.

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 is shown in Figure Appendix H.1.

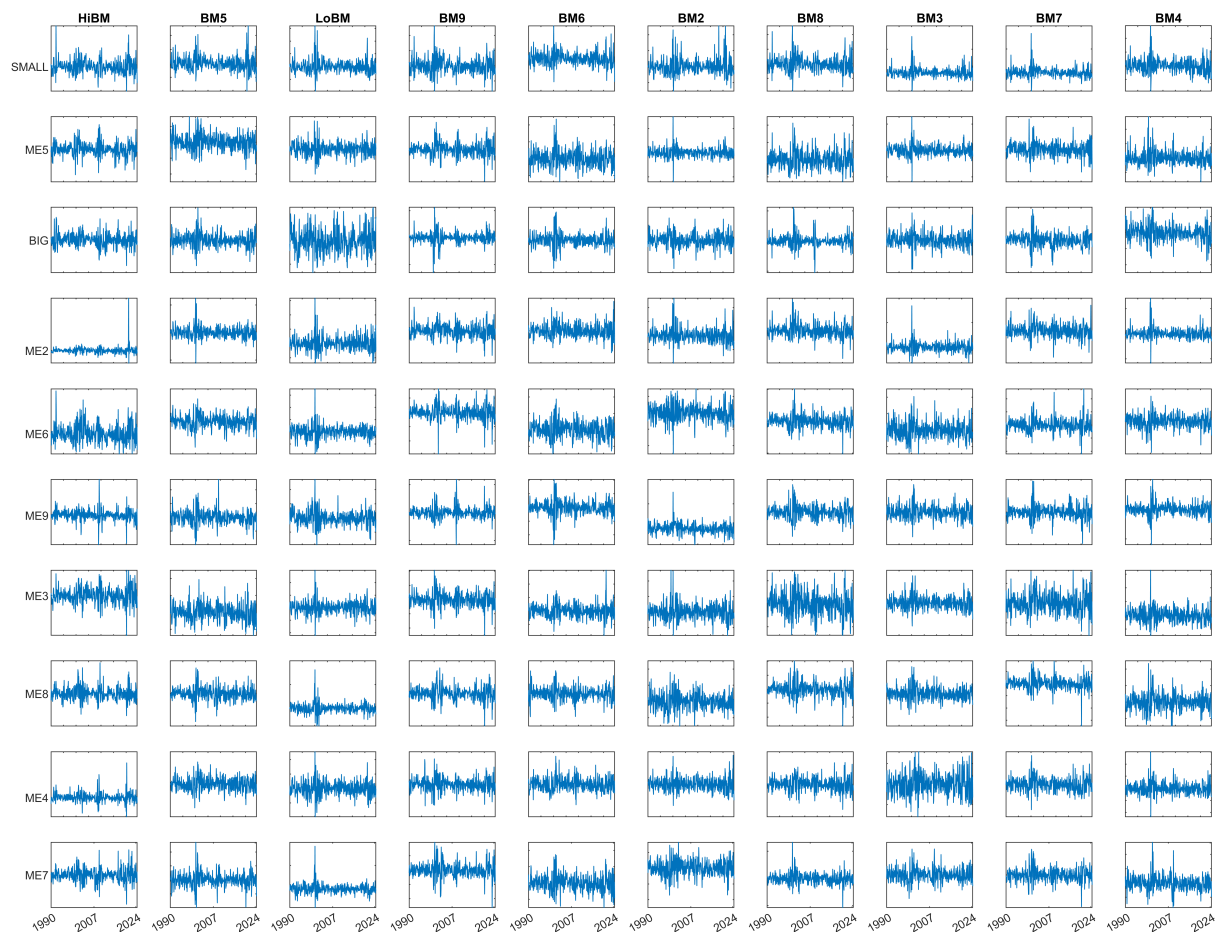


Figure Appendix H.1: The return series of the portfolios structured by different levels of sizes (rows) and book equity to market equity ratio (columns). Note that we have rearranged the order of rows and columns. The horizontal axis represents time and vertical axis represents the standardized monthly returns. The ranges of the vertical values are not the same.

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 [Appendix H.1–Appendix H.2](#) 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  $2 \times 2$  MDFM with both cross-sectional correlation and stochastic volatility 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 Appendix H.1: 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	<b>-45729</b>	-45792	-45886	<b>-42877</b>	-43007	-43155	-43328
(0.6)	(0.9)	(1.0)	(1.0)	(0.7)	(0.8)	(1.6)	(0.8)

Table Appendix H.2: 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 (0.2)	-49622 (0.3)	-49382 (0.4)	-47210 (0.2)	-46809 (0.4)	-46739 (0.4)	-43527 (1.3)	-43163 (1.4)	-43161 (1.4)
$p_1 = 2$	-49603 (0.3)	-47130 (0.4)	-46572 (0.4)	-46897 (0.4)	-46557 (0.4)	-46164 (0.7)	-43304 (2.0)	<b>-42951</b> (0.9)	-43024 (2.9)
$p_1 = 3$	-49285 (0.4)	-46847 (0.5)	<b>-46177</b> (0.9)	-46928 (0.6)	-46553 (0.9)	<b>-46087</b> (0.8)	-43406 (1.3)	-43075 (1.2)	-43027 (1.4)

Figure Appendix H.2 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 subplots correspond to  $\widehat{\mathbf{A}}_{.,1}$  (first),  $\widehat{\mathbf{A}}_{.,2}$  (second),  $\widehat{\mathbf{B}}_{.,1}$  (third) and  $\widehat{\mathbf{B}}_{.,2}$  (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.

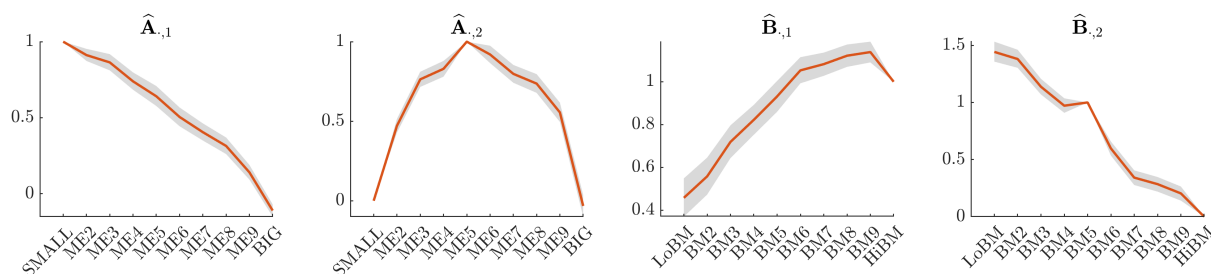


Figure Appendix H.2: Estimates for row loading matrices (the first and second panel from left) and column loading matrices (the third and fourth panel). The gray area represents the 90% credible interval.

Figure Appendix H.3 shows estimated posterior densities, histograms of posterior draws, priors, as well as the posterior estimates of autoregressive coefficients ( $\rho$ ) 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.

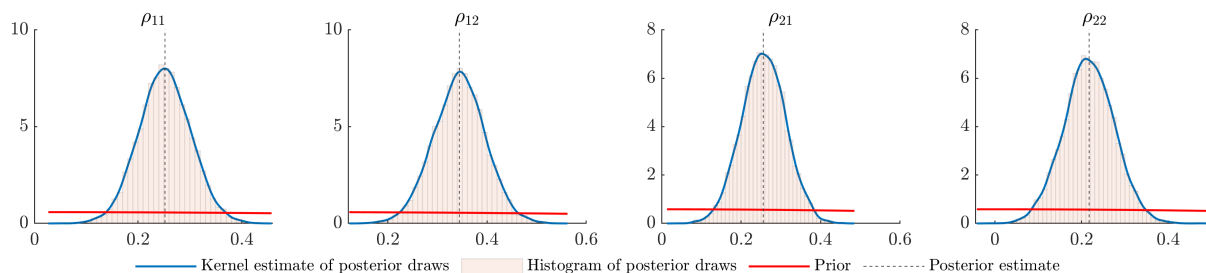


Figure Appendix H.3: Posterior densities, histograms of posterior draws, priors and posterior estimates for autoregressive coefficients  $\rho$

Figure Appendix H.4 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, significant spikes in volatility are observed during the 2008 financial crisis and the COVID-19 pandemic.



Figure Appendix H.4: Estimates and standard errors of stochastic volatility:  $\exp(\mathbf{h}/2)$