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

Joshua C. C. Chan[a], Wei Zhang[b]

[a] *Department of Economics, Purdue University*

[b] *Department of Economics, Johns Hopkins University*

Abstract

We develop dynamic factor models for time series in which each period's observation is a matrix, such as a panel of macroeconomic indicators observed across countries. The latent factors follow autoregressive processes, and the idiosyncratic components allow stochastic volatility, outliers, and a Kronecker-structured covariance capturing cross-row and cross-column correlation. Exploiting the matrix structure, we make these heavily parameterized models tractable in high dimensions and develop an efficient Gibbs sampler. For model comparison, we use the cross-entropy importance-sampling estimator of the marginal likelihood, which under a common criterion selects the factor dimension, the 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 the matrix structure over its vector counterpart, together with cross-sectional correlation and stochastic volatility; out of sample, the matrix model matches the vector model on point forecasts while producing sharper density forecasts.

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 Bańbura, 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 94th 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

A growing share of macroeconomic time series is matrix-valued: a set of indicators is observed for a cross-section of countries, so that each period delivers a country-by-indicator matrix. Such multi-country panels have long been used to decompose comovement into world, regional, and country-specific components (Kose et al., 2003; Del Negro and Otrok, 2008), and the workhorse tool for the task is the dynamic factor model (Stock and Watson, 2002a; Forni et al., 2000; Doz et al., 2012). Standard practice, however, first stacks the matrix into a long vector and fits a vector dynamic factor model (VDFM). This vectorization discards the matrix structure of the data, namely the dependence among indicators within a country and among countries for a given indicator. It also lets the number of loadings grow with the product of the two cross-sectional dimensions, which strains estimation as the panel widens.

Faced with this proliferation of loadings, the international business cycle literature restricts them rather than estimating them freely. In Kose, Otrok and Whiteman (2003), for example, each series loads on a world factor, a regional factor, and a country-specific factor, so the loading matrix is sparse by construction. This multilevel structure is parsimonious and lends the factors a clear economic interpretation, but the groupings must be specified in advance and the model assigns a separate factor to every country and region.

We pursue a different route, one that requires neither prespecified groupings nor a separate factor for every country. Arranging the period t observations as an $n \times k$ matrix $\mathbf{Y}_t$, we model its conditional mean as the bilinear form $\mathbf{A}\mathbf{F}_t\mathbf{B}'$, where $\mathbf{F}_t$ is a $p_1 \times p_2$ matrix of latent factors, $\mathbf{A}$ $(n \times p_1)$ holds the country-side loadings, and $\mathbf{B}$ $(k \times p_2)$ the indicator-side loadings. Equivalently, this matrix dynamic factor model (MDFM) is a VDFM whose $nk \times p_1 p_2$ loading matrix is the Kronecker product $\mathbf{B} \otimes \mathbf{A}$. This restriction cuts the number of loadings from $nk\,p_1 p_2$ to $np_1 + kp_2$, which sharpens inference when the sample is short relative to the cross-section. It also adds a factor structure along the indicator dimension that the multilevel scheme lacks. This parsimony, however, comes at the cost of flexibility: a country's exposures across indicators are tied together through the shared $\mathbf{A}$ and $\mathbf{B}$. Whether the restriction is warranted is an empirical question, which we address by formal Bayesian model comparison.

The idea of exploiting the matrix structure in this way originates in the statistical factor-model literature. Wang et al. (2019) introduce a matrix factor model in which the row and column loadings enter through this same bilinear form, later extended with linear

constraints, thresholds, and time-varying loadings (Chen et al., 2020; Liu and Chen, 2019; Chen et al., 2024). These models reduce the number of free loadings but omit two features essential for macroeconomic data: they are static, so they cannot produce iterated multi-step forecasts, and their idiosyncratic errors are homoskedastic, ruling out the time-varying volatility, heavy tails, and outliers pervasive in such 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 limitations.

This paper makes two methodological contributions. First, we introduce a class of MDFMs with dynamic factors and empirically realistic idiosyncratic errors. The latent factors $\mathbf{F}_t$ follow autoregressive processes, which capture persistent comovement and support recursive multi-step forecasting. Along the time dimension, we allow stochastic volatility (Carriero et al., 2016; Kastner et al., 2017), which matters in international data (Del Negro and Otrok, 2008); heavy-tailed errors (Jacquier et al., 2004); and outlier components for large, infrequent extreme observations such as those around the COVID-19 pandemic (Lenza and Primiceri, 2022; Carriero et al., 2024; Chan et al., 2026). Along the cross-section, their covariance has a second Kronecker structure $\mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r$, distinct from the loading restriction $\mathbf{B} \otimes \mathbf{A}$ and built from full cross-row and cross-column covariance matrices.

The model is heavily parameterized, but the matrix structure keeps estimation tractable. Exploiting the Kronecker form of the likelihood, we develop an efficient Gibbs sampler that scales to the dimensions of 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 loading matrices, and sample the constrained loadings with the fast hyperplane-truncation algorithm of Cong et al. (2017).

Taking these models to the data involves four specification choices: the number of factors, the matrix versus vector formulation, homoskedastic versus time-varying idiosyncratic errors, and a diagonal versus Kronecker idiosyncratic covariance. The matrix factor literature settles only the first, by eigenvalue-ratio criteria; our second contribution resolves all four under one criterion, the marginal likelihood. We compute it with the cross-entropy importance-sampling estimator of Chan and Eisenstat (2015), which draws independently from a calibrated importance density rather than from correlated Markov chain Monte Carlo samples, placing these otherwise separate questions on a common footing.

Monte Carlo experiments confirm that the estimator reliably selects the data-generating model across all four specification choices, and that the factor estimates approach their true values as the sample grows. We illustrate the MDFM on a quarterly OECD panel of 19 countries and 10 indicators. Three findings emerge. First, the matrix structure is strongly supported: the best MDFM raises the log marginal likelihood by more than $4,700$ points over the best VDFM. Second, the restriction sharpens inference: the MDFM and the VDFM recover very similar factors (a global real-activity factor and a price factor whose smoothed path tracks oil-price growth), yet the MDFM estimates the loadings far more precisely, with credible intervals roughly half as wide. Third, within the matrix class the data favor both a Kronecker-structured idiosyncratic covariance and common stochastic volatility, whose estimated path spikes during the Asian financial crisis, the Great Recession, and the COVID-19 pandemic. A comparison with the multilevel factor model of Kose, Otrok and Whiteman (2003) reinforces these findings: the MDFM recovers the same global business cycle and cross-country dependence structure with far fewer latent factors.

In a recursive out-of-sample exercise we benchmark the MDFM against the two models applied researchers have used to date: the VDFM and the multilevel model of Kose, Otrok and Whiteman (2003). The three models deliver similar point forecasts. The MDFM holds a small but statistically significant advantage over the VDFM at the one-quarter horizon, the multilevel model is marginally more accurate at that horizon, and the differences essentially vanish at longer horizons. The MDFM's advantage lies in density forecasts: the joint log predictive score of the full 190-series panel favors it by more than 40 points against the VDFM and by more than 55 points against the multilevel model at every horizon, a gain driven by its sharper loading estimates and its modeling of the cross-sectional dependence that both competitors ignore. The fully Bayesian treatment yields posterior predictive densities that widen during the 2008–2009 financial crisis and the 2020 pandemic, a dimension of forecast quality central to the growth-at-risk agenda (Adrian et al., 2019). A further comparison with the static matrix factor model of Wang et al. (2019) is reported in the Online Supplemental Materials.

This paper connects several literatures. It adds a matrix-valued formulation with flexible idiosyncratic dynamics to the tradition of Bayesian dynamic factor and large Bayesian time-series models (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, 2026). It also

relates 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 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: the former frequentist, with iterative eigendecomposition and eigenvalue-ratio selection; the latter Bayesian, with a Kronecker-structured idiosyncratic covariance. Our framework differs in two respects that mirror our contributions: neither accommodates the time-varying volatility, heavy tails, and outlier components we introduce, and neither provides a unified comparison across competing specifications.

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 the Monte Carlo experiments. Section 5 reports the empirical application to an OECD macroeconomic panel, including in-sample comparisons with the VDFM and the multilevel model, 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

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

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

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 $p_1p_2 \times p_1p_2$ factor innovation covariance is $\boldsymbol{\Lambda} = \text{diag}(\lambda^2_{1,1}, \ldots, \lambda^2_{p_1,p_2})$. $\mathbf{H}_{\boldsymbol{\rho}_l}$, $l = 1, \ldots, q$, are autoregressive coefficient matrices.

The bilinear form $\mathbf{A}\mathbf{F}_t\mathbf{B}'$ captures cross-country and cross-variable dependencies. The ith row of $\mathbf{Y}_t$ is $\mathbf{Y}_{i,\cdot,t} = \mathbf{A}_{i,\cdot}\mathbf{F}_t\mathbf{B}' + \mathbf{E}_{i,\cdot,t}$: 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 ith row of $\mathbf{A}$. Similarly, the jth column has a common component formed by linear combinations of columns of $\mathbf{A}\mathbf{F}_t$ with weights from the jth 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 $\text{Cov}(\mathbf{Y}'_{i,\cdot,t} \,|\, \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{r,i,i} \boldsymbol{\Sigma}_c$, and for any column j, $\text{Cov}(\mathbf{Y}_{\cdot,j,t} \,|\, \mathbf{A}, \mathbf{F}_t, \mathbf{B}) = \omega_t \sigma^2_{c,j,j} \boldsymbol{\Sigma}_r$. $\boldsymbol{\Sigma}_r$ 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, 2026; Giglio et al., 2025).

The Kronecker structure also greatly reduces the number of free parameters relative to an unrestricted idiosyncratic covariance.[1] 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.[2] We implement three popular specifications under the conditionally Gaussian framework in (1) through the scalar latent state ω_t.

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

$$h_t = \phi h_{t-1} + u^h_t, \qquad u^h_t \sim \mathcal{N}(0, \sigma^2_h), \tag{3}$$

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

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

[1] 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.

[2] 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 ($o_t \geqslant 2$). 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 the 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 $\mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r$, 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:

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

Beyond reducing the number of loadings, the matrix structure sharpens their convergence rates. 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 α-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.

These gains, however, come at a cost. The Kronecker structure imposes restrictions on how countries and indicators interact in generating common variation and residual dependence. In the unrestricted VDFM, each country-indicator pair has its own loading vector, allowing, for example, a given country to respond differently to a common shock through different indicators. By contrast, the MDFM assumes that factor loadings are separable,

so that the response of an indicator in a given country is the product of a country-specific component and an indicator-specific component. Similarly, the Kronecker covariance restricts residual dependence to be separable into a country component and an indicator component. These restrictions rule out arbitrary country-indicator-specific interactions: for example, a single country's inflation cannot respond to a global shock in a way that deviates from the product of that country's overall exposure and the indicator's typical response. Thus, the MDFM trades flexibility for parsimony and interpretability.

Whether these restrictions are supported by the data is ultimately an empirical question. The Kronecker structure is testable using the randomized specification tests proposed by He et al. (2024); in this paper, we treat the VDFM-versus-MDFM choice by Bayesian model comparison (Section 3), on the same footing as the factor-matrix dimension and the 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.

An important direction for future research is to develop matrix factor models that relax these separability restrictions while retaining the efficiency gains of the matrix representation. One possibility is to allow the loading matrix or residual covariance to be approximated by a sum of Kronecker products, for example,

$$\mathbf{M} = \sum_{r=1}^{R} \mathbf{B}_r \otimes \mathbf{A}_r,$$

which would accommodate multiple channels through which countries and indicators interact while preserving substantial dimension reduction. Other extensions could combine a Kronecker component with a low-rank or sparse correction to capture localized country-indicator relationships. Such approaches would provide a flexible continuum between the fully separable MDFM and the unrestricted VDFM, and we leave them for future research.

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) 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; $\mathbf{H}_{\boldsymbol{\rho}_l}$ is diagonal for $l = 1, \ldots, q$; and $\mathrm{Cov}(\mathbf{u}_t) = \mathbf{I}_{p_1 p_2}$.

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: $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ are identified only up to a scalar, since for any $m \in \mathbb{R} \setminus \{0\}$,

$\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r = (m\boldsymbol{\Sigma}_c) \otimes (m^{-1}\boldsymbol{\Sigma}_r)$. We fix the scale by normalizing the $(1,1)$ element of $\boldsymbol{\Sigma}_c$ 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; $\mathbf{H}_{\boldsymbol{\rho}_l}$ is a diagonal matrix for $l = 1, \ldots, q$; 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 $\mathbf{A}\mathbf{F}_t\mathbf{B}'$ 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 Appendix J.2.

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 $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$:

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

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 $\mathrm{Cov}(\mathbf{A}_{i,\cdot}, \mathbf{A}_{j,\cdot} \mid \mathbf{\Sigma}_r) = \mathbf{\Sigma}_{r,i,j}\mathbf{V}_{\mathbf{A}'}$, so that the prior correlation between the ith and jth rows of $\mathbf{A}$ along any direction in the factor space equals $\mathbf{\Sigma}_{r,i,j}/\sqrt{\mathbf{\Sigma}_{r,i,i}\mathbf{\Sigma}_{r,j,j}}$. 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 $\mathbf{\Sigma}_c$. Computationally, the conjugate prior in (6) yields a matrix-normal-inverse-Wishart conditional posterior for $(\mathbf{A}', \mathbf{\Sigma}_r)$ and $(\mathbf{B}', \mathbf{\Sigma}_c)$, 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 $\mathcal{S}_q = \{\boldsymbol{\rho} \in \mathbb{R}^q : 1 - \sum_{s=1}^{q} \rho_s z^s \neq 0 \text{ for all } |z| \leqslant 1\}$ 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 $\mathbf{H}_{\boldsymbol{\rho}}$ have truncated normal priors,

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

where $\mathbb{1}(\cdot)$ 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 $t = 1, \ldots, T$; (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 $t = 1, \ldots, T$.

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,

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

with closed-form expressions for the moments given in Appendix B of the Online Supplemental Materials. 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 uses the acceptance-rejection Metropolis-Hastings sampler proposed by Chan (2017), sampling fat-tailed innovation states uses the inverse-gamma posterior, and sampling the outlier components uses a discrete grid sampling algorithm. Full algorithmic details, including closed-form expressions for $\widehat{\mathbf{A}}'$, $\widehat{\mathbf{B}}'$, $\mathbf{K}_{\mathbf{A}'}$, $\mathbf{K}_{\mathbf{B}'}$, $\widehat{\mathbf{S}}_r$, $\widehat{\mathbf{S}}_c$, 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,

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

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 $(\mathbf{A}', \mathbf{\Sigma}_r)$ and $(\mathbf{B}', \mathbf{\Sigma}_c)$ 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.[3]

The obstacle to wider use of marginal likelihood is computational: the integral $p(\mathbf{y}) = \int p(\mathbf{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$ 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,

$$\hat{p}_{IS}(\mathbf{y}) = \frac{1}{N} \sum_{n=1}^{N} \frac{p(\mathbf{y} \mid \boldsymbol{\theta}_n)\, p(\boldsymbol{\theta}_n)}{g(\boldsymbol{\theta}_n)}, \qquad \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N \overset{\text{iid}}{\sim} g(\boldsymbol{\theta}). \tag{9}$$

The estimator is unbiased and consistent whenever g dominates $p(\mathbf{y} \mid \cdot)p(\cdot)$, 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 follow 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 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

[3]The closest Bayesian treatment is Qin et al. (2025), who select factor dimensions and lag orders via fractional Bayes factors under diffuse priors.

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 $j = 1, \ldots, p_1$, $k = 1, \ldots, p_2$, $\mathbf{\Sigma}_c = 0.3\mathbf{I}_k$, $\mathbf{\Sigma}_r = 0.5\mathbf{I}_n$, 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 $(p_1, p_2) \in \{(3, 2), (5, 5)\}$.[4] 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.

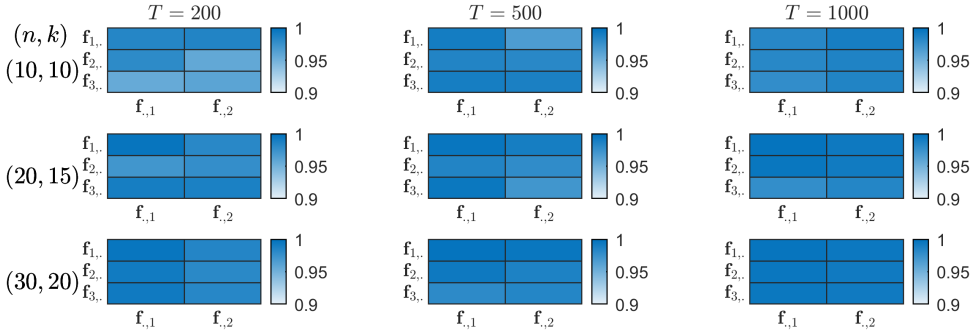


Figure 1: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 3$, $p_2 = 2$.

[4]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.

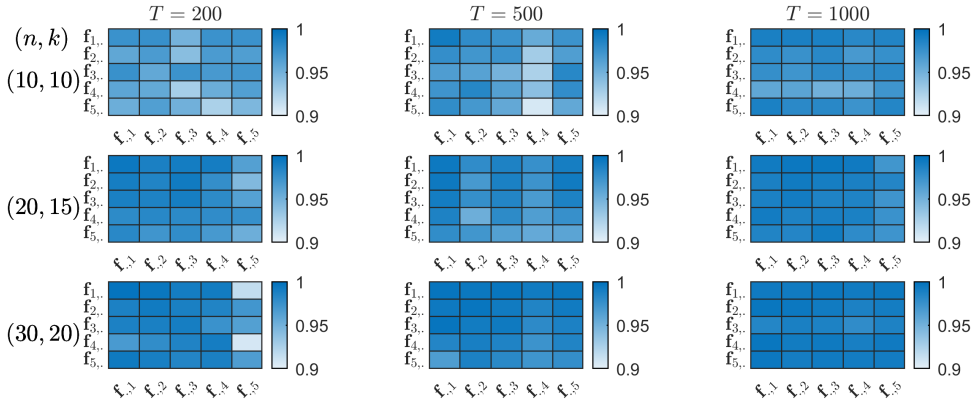


Figure 2: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 5$, $p_2 = 5$.

4.2. Identifying the dimension of the factor matrix

To evaluate the estimator's ability to identify the factor matrix dimension, we estimate log marginal likelihoods over a grid of (p_1, p_2) values using four datasets from Section 4.1, all with $T = 500$: Datasets 1 and 2 have true dimension $(p_1, p_2) = (3, 2)$ with panel sizes $(n, k) = (10, 10)$ and $(20, 15)$, respectively; Datasets 3 and 4 have true dimension $(5, 5)$ with the same two panel sizes. For Datasets 1 and 2 the grid runs over $p_1, p_2 \in \{1, \ldots, 5\}$; for Datasets 3 and 4, over $p_1, p_2 \in \{3, \ldots, 7\}$.

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.

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.

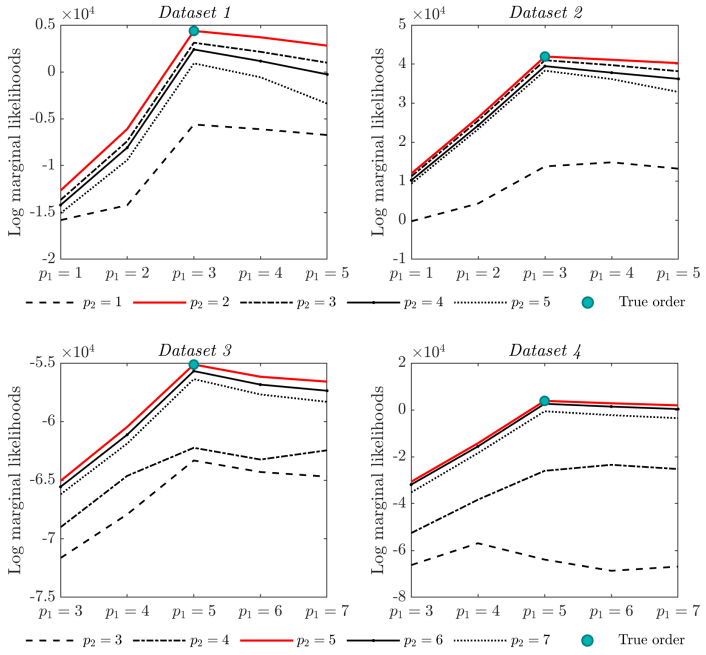


Figure 3: Log marginal likelihood estimates for the four datasets. From top left to bottom right, the panels correspond to $(n, k, p_1, p_2) = (10, 10, 3, 2)$, $(20, 15, 3, 2)$, $(10, 10, 5, 5)$, and $(20, 15, 5, 5)$. Each line represents a different value of p_2, and the blue dot marks the true value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

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 $k_f = 1, \ldots, 6$. When the true model is a VDFM, we estimate MDFMs with $(p_1, p_2) \in \{(1, 1), (1, 2), (2, 1), (2, 2)\}$.

Figure 4 reports the results. In all six cases, the true model achieves the highest marginal likelihood. A noteworthy pattern in the top row: when the true model is an 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.

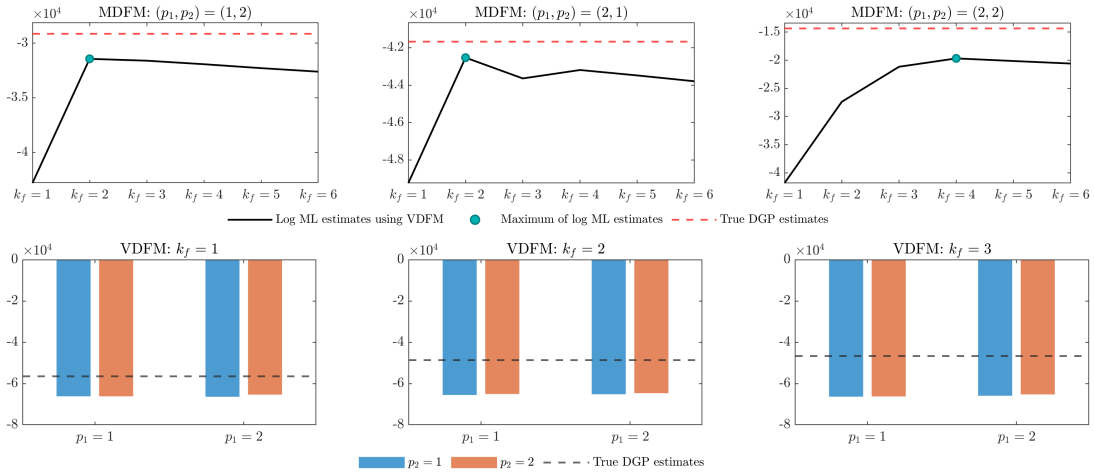


Figure 4: Log marginal likelihood estimates under misspecified factor structures. Top row: the true model is an MDFM; the black line traces VDFM marginal likelihoods across $k_f = 1, \ldots, 6$ factors, and the red dashed line marks the log marginal likelihood estimate of the true DGP. The left, middle, and right panels correspond to MDFMs with $(p_1, p_2) = (1, 2), (2, 1), (2, 2)$. Bottom row: the true model is a VDFM; bars correspond to competing MDFMs, and the black dashed line marks the log marginal likelihood estimate of the true DGP. 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 covariance, no stochastic volatility), **MDFM-cross** (Kronecker covariance $\mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r$, no stochastic volatility), and **MDFM-sv** (diagonal covariance with common stochastic volatility). For each true specification we generate 100 datasets with $(n, k, T) = (20, 20, 100)$ and $(p_1, p_2) = (2, 2)$: the free elements of $\mathbf{A}$ and $\mathbf{B}$ are drawn from $\mathcal{N}(0, 0.3^2)$, $\mathbf{\Sigma}_r \sim \mathcal{IW}(n + 2, \mathbf{I}_n)$, $\mathbf{\Sigma}_c \sim \mathcal{IW}(k + 2, \mathbf{I}_k)$, the log-volatility process has AR(1) coefficient 0.97 and innovation variance 0.1, and the factor innovation variance is $0.1\mathbf{I}_{p_1 p_2}$ with AR coefficients drawn from $\mathcal{U}(0.8, 0.9)$.

We then estimate the log marginal likelihood under all three specifications for each dataset. Figure 5 reports boxplots of the differences relative to the true model: in all three settings the true model achieves the highest marginal likelihood, with the largest margin when the truth is the most parsimonious MDFM-diag, whose alternatives are over-specified.

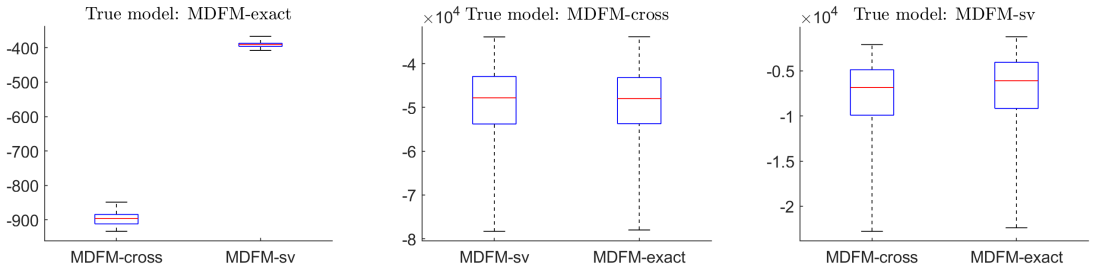


Figure 5: Boxplots of differences in log marginal likelihoods relative to the true model: MDFM-diag (left), MDFM-cross (middle), and MDFM-sv (right). A negative value indicates that the true model is favored.

5. An Application to a Large OECD Macroeconomic Panel

We illustrate the usefulness of the MDFM on a quarterly OECD macroeconomic panel. 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 model's predictive performance.[5]

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] A second application to the Fama–French 10×10 size-by-book-to-market portfolio panel, with comparable structure but a different empirical setting, is reported in Appendix K 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.

5.1. Model Selection

In this subsection, we compare log marginal likelihood estimates across MDFM specifications and a set of unrestricted VDFMs applied to the vectorized panel, where the candidate models differ in the number of factors, the idiosyncratic covariance structure, and the treatment of volatility. This exercise selects the preferred specification within each model class and quantifies the evidence for the matrix structure itself. Posterior inference is obtained via the Gibbs sampler of Section 2.4, with 10,000 posterior draws retained after a burn-in of 5,000.

Table 1 reports log marginal likelihood estimates for VDFMs (Panel A) and MDFMs (Panel B). 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. There are three interesting findings. First, the data strongly favor common stochastic volatility: in both model classes the SV variants dominate their constant-volatility analogues by thousands of log points. Second, the Kronecker idiosyncratic covariance matters even more than stochastic volatility. At $(p_1, p_2) = (1, 2)$, replacing the diagonal covariance with the Kronecker specification raises the log marginal likelihood by almost 6,900 log points, nearly four times the subsequent gain from adding stochastic volatility. The preferred factor dimensions reveal why: within the MDFM-diag class the marginal likelihood increases in both p_1 and p_2 and peaks at the largest factor matrix considered, $(3, 3)$ — under a diagonal covariance, additional factors are conscripted to absorb residual cross-sectional dependence. Once the Kronecker covariance is included, the preferred specification drops to $(1, 2)$: the covariance matrix absorbs the local dependence, and the factor matrix retains only the pervasive cycles. Third, the matrix structure is decisively favored over the vector alternative: the best MDFM attains a log marginal likelihood of $-16{,}140$, more than 4,700 log points above the best VDFM ($-20{,}862$), which requires five factors. The optimal 1×2 factor matrix — one country dimension and two indicator dimensions — indicates that variation along the indicator dimension is richer than variation along the country dimension in this panel. We hereafter estimate the MDFM-cross-sv with $(p_1, p_2) = (1, 2)$ and refer to it simply as the MDFM.

Table 1: Log marginal likelihood estimates. Numerical standard errors in parentheses. The largest estimate within each specification is in bold; the overall maximum is highlighted in red.

Panel A: VDFMs

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

Panel B: MDFMs

	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	**−17930**	−17950	−16144	**−16140**	−16164
	(0.2)	(0.3)	(0.7)	(0.3)	(0.4)	(0.8)	(0.6)	(0.4)	(0.5)
$p_1 = 2$	−26310	−24607	−23221	−18007	−17979	−18068	−16260	−16265	−16337
	(0.1)	(0.4)	(0.4)	(0.4)	(0.4)	(1.4)	(0.5)	(0.6)	(1.0)
$p_1 = 3$	−26164	−24469	**−23012**	−18051	−18058	−18150	−16336	−16386	−16464
	(0.3)	(0.6)	(0.7)	(0.3)	(0.8)	(0.6)	(0.7)	(0.7)	(1.1)

5.2. In-Sample Results and Comparisons with Kose et al. (2003) and the VDFM

Factors, loadings and common volatility. Figure 6 displays the posterior estimates of the two factor series and the common stochastic volatility. The factors admit a natural economic interpretation. The real-activity factor $\mathbf{f}_1$ (top left) is dominated by a sharp contraction during the 2008–2009 Great Recession and a much larger spike in 2020 at the onset of the COVID-19 pandemic. The second factor (top right) captures a global price cycle closely tied to energy markets: its four-quarter moving average tracks four-quarter Brent crude oil price growth throughout the sample (bottom left), with peaks and troughs coinciding with the major oil-market episodes shaded in the figure and a posterior correlation between the two smoothed series of approximately 0.66. The common volatility $\widehat{\omega}_t$ (bottom right) isolates three episodes of elevated global uncertainty — the 1997–1998 Asian financial crisis, the Great Recession, and the COVID-19 pandemic, with the COVID-19 spike by far the largest and volatility remaining elevated through 2023.

Figure 7 reports the country loadings $\widehat{\mathbf{A}}$ and indicator loadings $\widehat{\mathbf{B}}$ with 90% credible intervals.[6] The country loadings reveal substantial heterogeneity in factor exposure that

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

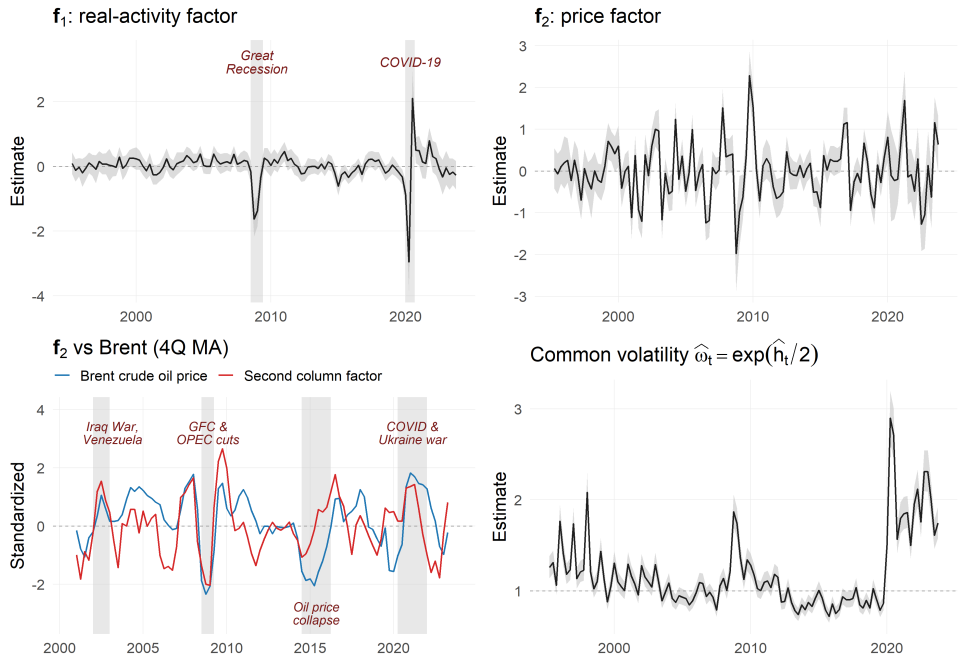


Figure 6: Posterior estimates of the latent states. Top left: real-activity factor $\mathbf{f}_1$ with shaded recession episodes. Top right: price factor $\mathbf{f}_2$ with 90% credible band. Bottom left: four-quarter moving averages of standardized Brent crude oil price growth (blue) and the second column factor (red); shaded bands mark major oil-market episodes. Bottom right: posterior mean of the common stochastic volatility, $\widehat{\omega}_t = \exp(\widehat{\mathbf{h}}_t/2)$, with 90% credible band.

does not map cleanly onto geography:[7] Japan has the smallest loading (0.29), a middle group loads between 0.48 and 0.65, a higher group between 0.73 and 0.86, and Canada and the United States load at 1.00. Geography matters but is not decisive — Japan and Korea fall into different groups despite their proximity, while Australia and New Zealand cluster with continental Europe rather than Asia. The indicator loadings separate the ten variables into four economic roles: imports, exports, and labor unit cost load most strongly on the real-activity factor (b_1 between 1.3 and 1.4); RGDP and consumption load on b_1 alone; unemployment is the only counter-cyclical indicator ($b_1 \approx -0.6$); energy and headline CPI load primarily on the price factor (b_2 around 1.3 and 1.0); and core and food CPI load weakly on both, leaving most of their variation idiosyncratic. The two-dimensional view also shows that the two factors are approximately orthogonal in their effect on indicators.

[7] We sort the posterior loading estimates and compute the posterior probabilities that adjacent values differ. Countries and indicators are grouped whenever the probability that two neighboring loadings differ exceeds 0.9.

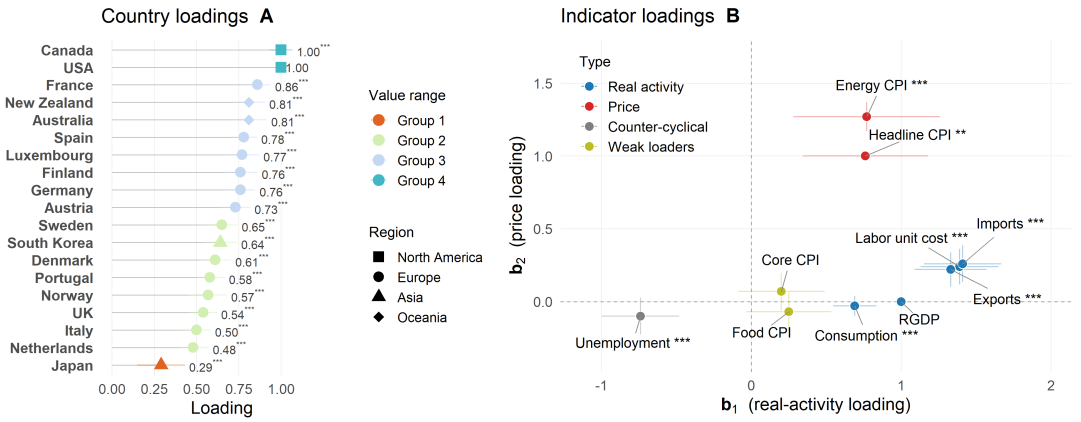


Figure 7: Loading estimates with 90% credible intervals. Left: country loadings $\widehat{\mathbf{A}}$, sorted ascending, colored by value-range group and shaped by region. The numeric value of each loading appears to the right of its credible interval, with significance stars as superscripts. The USA loading is fixed to one for identification. Right: indicator loadings $\widehat{\mathbf{B}}$ in the (b_1, b_2) plane, colored by economic role.

Comparison with Kose et al. (2003). A widely used model for international macroeconomic panels is the multilevel dynamic factor model proposed by Kose, Otrok and Whiteman (2003) (hereafter KOW), which specifies global, regional, and country-specific factors to capture business cycles at different levels of aggregation. Beyond the computational considerations discussed in Section 2, the key distinction between KOW and the MDFM is that in KOW common variation is represented through global, regional, and country-specific factors, while the idiosyncratic component is assumed independent across series. As a result, it has no mechanism to distinguish common variation across different types of indicators or to model residual dependence separately across countries and indicators. This limitation is particularly relevant for our panel, which combines real and nominal variables whose common variation is driven by different economic forces. In contrast, the MDFM exploits the matrix structure of the data and separately identifies real activity and price factors, while the Kronecker covariance captures the remaining dependence separately along the country and indicator dimensions. The comparison below illustrates the empirical implications of this separation.

To make the comparison transparent, we estimate a comparable KOW specification with one world factor and one country factor for each economy.[8] The first implication concerns

[8]For a fair comparison, the model has the same common stochastic volatility, unit-loading identification restrictions, and priors for shared parameters and states. Details are given in Appendix G of the Online Supplemental Materials.

the interpretation of the global factor. Figure 8 compares the KOW world factor with the MDFM real-activity factor. The two series are highly correlated (0.79), sharing the major downturns of the Global Financial Crisis and the COVID-19 recession. However, the KOW world factor is not purely a real-activity cycle: the MDFM real-activity factor alone explains 63% of its variation, the MDFM price factor alone explains 35%, and the two factors jointly explain 86%. With only one global dimension, the KOW specification estimates a combination of distinct common forces that the MDFM separates.

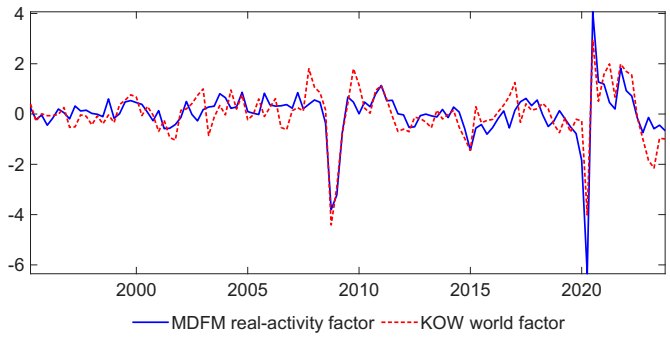


Figure 8: The MDFM real-activity factor (blue solid) and the world factor of the multilevel model (red dashed), correlated at 0.79 with coincident troughs in 2008–2009 and 2020.

At the country level, the two models recover similar patterns of cross-country dependence. Figure 9 shows that both identify a tightly integrated continental European block (correlations around 0.4–0.57), a distinct North American cluster, and Japan as relatively isolated, with the off-diagonal elements of the two dependence matrices correlating at 0.74 across the 171 country pairs. The main difference is Norway, which appears negatively correlated with most other countries in the KOW representation, even though its raw series comove positively with the rest of the panel.[9] The pattern reflects the missing price dimension. As discussed, the world factor is a blend of the two global cycles and gives each series one loading, so it cannot capture responses to the price cycle that differ in sign across countries. Norway, as the only major net oil exporter, turns out to be the only country with positive exposure to the price factor: the correlation coefficient between the MDFM price factor and Norway's country factor is $+0.13$, against a mean of -0.21 for the oil importers; its factor is also the only one negatively correlated with

[9]Indicator by indicator, the average correlation between Norway's standardized series and the corresponding series of the other eighteen economies is positive for all ten indicators (e.g., 0.47 for real GDP).

the real-activity factor (−0.13, against an average of 0.35). The price factor thus enters the country factors with opposite signs for Norway and for the rest, which is what makes Norway's factor negatively correlated with the others. In the MDFM, the price cycle has its own global factor, and Norway's correlations in $\mathbf{R}_r$ are weak but positive.

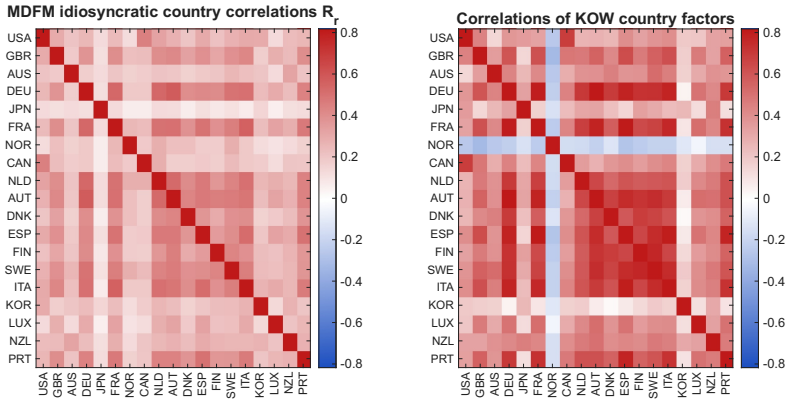


Figure 9: MDFM idiosyncratic country correlations $\mathbf{R}_r$ (left) and correlations of the estimated KOW country factors (right); their off-diagonal elements correlate at 0.74 across the 171 country pairs.

Beyond the cross-country dependence, $\mathbf{\Sigma}_c$ in the MDFM captures the cross-indicator dependence. As shown in Figure 10, the indicator correlations $\mathbf{R}_c$ (obtained by standardizing $\widehat{\mathbf{\Sigma}}_c$ to unit diagonal) concentrate in three economically meaningful blocks: RGDP and consumption remain correlated (0.47); the four price indices form a tight cluster (pairwise correlations 0.43–0.67); and labor unit cost, imports, and exports form a third block largely disconnected from the price indicators, suggesting real-side comovement, for instance, a favorable terms-of-trade shock that raises real wages relative to productivity, appreciates the exchange rate, and draws in imports.

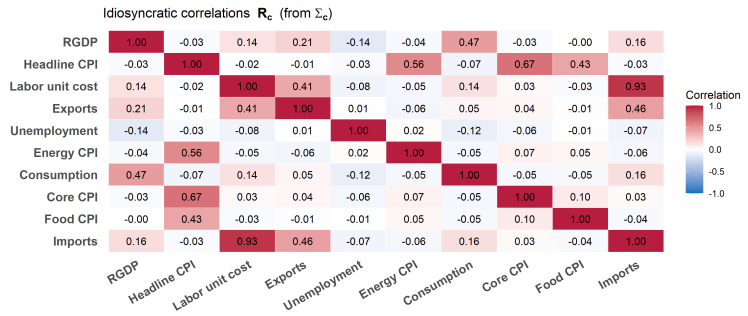


Figure 10: Idiosyncratic correlation matrix $\mathbf{R}_c$ across indicators, standardized from $\widehat{\mathbf{\Sigma}}_c$.

Comparison with the VDFM. The comparison with KOW showed that the MDFM recovers the main economic features identified by a multilevel factor model while providing a more interpretable decomposition of real and nominal comovement. The comparison with the unrestricted VDFM addresses a different question: does the Kronecker restriction in the MDFM sacrifice any common variation, and if not, what does it gain in return?

To isolate the effect of the Kronecker restriction, we compare matched specifications: both models assume diagonal, homoskedastic idiosyncratic errors, so that the loading structure is the only difference. In particular, we compare the MDFM-diag with the preferred factor dimensions, $(p_1, p_2) = (1, 2)$, against the VDFM-diag with $k_f = 5$, the preferred number of factors within the vector class. The unrestricted model estimates 950 loading parameters, whereas the MDFM estimates only 39.

Since the two models are identified under different rotations, we compare the subspaces spanned by their posterior-mean factor estimates. The first two canonical correlations between the two MDFM factors and the five VDFM factors are 1.00 and 0.99, and regressing each MDFM factor on the VDFM factors yields R^2 values of 0.99 (Figure Appendix H.1). Thus, despite using far fewer loading parameters, the MDFM extracts factors that lie almost exactly within the factor space of the unrestricted VDFM. Because loading matrices are not directly comparable across rotations, we align each VDFM posterior draw $\mathbf{M}^{(s)}$ to the MDFM loading space by least squares,

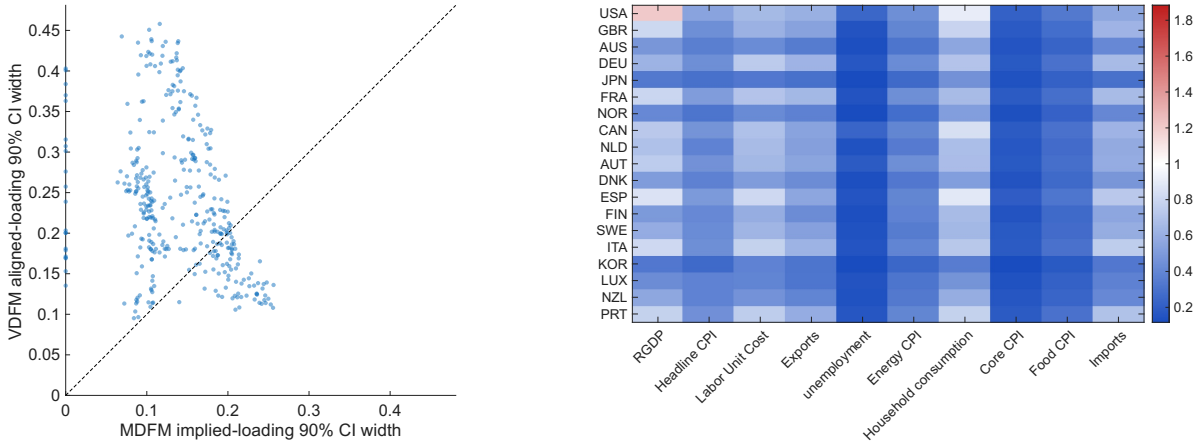
$$\tilde{\boldsymbol{L}}^{(s)} = \mathbf{M}^{(s)}\left(\mathbf{M}^{(s)\top}\mathbf{M}^{(s)}\right)^{-1}\mathbf{M}^{(s)\top}\bar{\boldsymbol{L}}, \qquad \bar{\boldsymbol{L}} = S^{-1}\sum_{s=1}^{S}\mathbf{B}^{(s)} \otimes \mathbf{A}^{(s)},$$

before comparing posterior dispersion.[10] As a fully rotation-invariant check, we also compare the time-averaged posterior standard deviation of the fitted common component, $C_{ij,t}^{(s)} = \left(\mathbf{A}^{(s)}\mathbf{F}_t^{(s)}\mathbf{B}^{(s)\top}\right)_{ij}$ for the MDFM and $\left(\mathbf{M}^{(s)}\mathbf{f}_t^{(s)}\right)_{ij}$ for the VDFM.

The gains are substantial (Figure 11). The median posterior standard deviation of the MDFM loadings is only 56% of that under the VDFM, with 79% of loadings estimated more precisely. For the fitted common component, the median ratio falls to 41%, and the MDFM achieves smaller posterior uncertainty in 99.5% of the 190 country–indicator series. These finite-sample gains mirror the theoretical results in Section 2.1.2: by ex-

[10] Since the projection maps each VDFM draw as close as possible to a fixed target, it can only reduce the measured VDFM dispersion, making the comparison conservative from the perspective of the MDFM.

ploiting the matrix structure of the data, the MDFM estimates the common structure substantially more precisely while extracting the same pervasive cycles as the unrestricted VDFM.



(a) Loading credible-interval widths (b) Common-component posterior-sd ratio

Figure 11: Posterior sharpness of the MDFM relative to the VDFM ($k_f = 5$, matched specifications). (a) 90% credible-interval widths of the MDFM implied loadings $\mathbf{B} \otimes \mathbf{A}$ (horizontal) against the aligned VDFM loadings (vertical); points above the 45° line are sharper under the MDFM. (b) Ratio of posterior standard deviations of the common component by country and indicator; blue marks cells where the MDFM is more precise (ratio below one); the scale is symmetric about one.

5.3. Out-of-Sample Forecasting

We assess the predictive performance of the MDFM in a recursive out-of-sample exercise: at each period t of the evaluation sample we use only data up to t to estimate the models and compute h-step-ahead forecasts for period $t + h$. The evaluation period runs from 2007Q4 to the end of the sample, and we report results at $h = 1$, $h = 4$, and $h = 8$ (one-quarter, one-year, and two-year horizons). We compare the MDFM with the two models examined in Section 5.2: the KOW specification and the VDFM with common stochastic volatility and $k_f = 5$ factors, the specification preferred by the marginal likelihood.[11]

We forecast recursively: at each forecast origin t we re-estimate each model on data through t and simulate the posterior predictive distribution of $\mathbf{Y}_{t+h}$ by propagating each

[11] A third comparison, against the static matrix factor model of Wang et al. (2019), is reported in Appendix I of the Online Supplemental Materials, with implementation details in Appendix F. The MDFM's factor dynamics deliver statistically significant pooled RMSFE improvements over the static benchmark of 2.5% at $h = 1$ and 4.3% at $h = 4$; by $h = 8$ both forecasts have largely reverted to the unconditional mean and the advantage disappears.

posterior draw forward through the state and observation equations. Point forecasts $\widehat{Y}^{m}_{i,j,t+h|t}$ are posterior predictive means. We use the root mean squared forecast error (RMSFE) to evaluate point forecasts and the average log predictive likelihood to evaluate density forecasts. For model m, the root mean squared forecast error of cell (i, j) and the average log predictive likelihood of country i are

$$\mathrm{RMSFE}^{m}_{i,j}(h) = \sqrt{\frac{1}{T_1 - T_0 + 1} \sum_{t=T_0}^{T_1} \left(\widehat{Y}^{m}_{i,j,t+h|t} - Y_{i,j,t+h}\right)^2},$$

$$\mathrm{ALPL}^{m}_{i}(h) = \frac{1}{T_1 - T_0 + 1} \sum_{t=T_0}^{T_1} \log p^{m}\left(\mathbf{Y}_{i,\cdot,t+h} \mid \mathbf{Y}_{1:t}\right),$$

where the forecast origins run from T_0 (2007Q3) to $T_1 = T - 8$, so that all horizons are evaluated on the same set of origins, and $p^{m}(\mathbf{Y}_{i,\cdot,t+h} \mid \mathbf{Y}_{1:t})$ is the joint posterior predictive density of the ith row of $\mathbf{Y}_{t+h}$, that is, of country i's k indicators. We report RMSFE ratios of the MDFM to each competing model, pooled across the 19 countries within each indicator, across the 10 indicators within each country, and across all 190 cells, by averaging the squared errors over the corresponding cells before taking the ratio. Analogously to ALPL^{m}_{i}, we also report the average log predictive likelihood of the joint density of indicator (column) j across the 19 countries, $j = 1, \ldots, k$, and of the full panel; we compare models by the ALPL difference. Values below one for the RMSFE ratio and positive ALPL differences favor the MDFM. Significance is assessed by the tests of Diebold and Mariano (1995) and Amisano and Giacomini (2007) applied to the per-origin differentials.

Point forecasts. Panels A and C of Table 2 report the RMSFE ratios of the MDFM to the KOW model and to the VDFM. At $h = 1$ the pooled ratio is 1.021 against KOW and 0.974 against the VDFM: KOW is more accurate than the MDFM by about 2%, and the MDFM is more accurate than the VDFM by a comparable margin, with the largest gains in real activity (real GDP and household consumption improve by 12–13%). As the forecasts revert toward the unconditional mean at longer horizons, the point forecasts of the three models become essentially indistinguishable, with all pooled ratios between 0.994 and 1.001.

Density forecasts. The MDFM's advantage lies in the density forecasts (Panels B and D of Table 2). Against KOW, the joint log score of the full panel favors the MDFM by

more than 55 points at every horizon (58.7, 66.0, and 57.8 at $h = 1$, 4, and 8), which is an even larger margin than against the VDFM (42.6, 45.1, and 45.0, overwhelmingly significant). In addition, the joint predictive density for most indicators favors the MDFM at every forecast horizon. The exceptions are labor unit cost at $h = 1$ and $h = 4$ and unemployment at $h = 1$, though neither difference is statistically significant. Among countries, KOW's densities for Japan are better at the longer horizons, as Japan has its own country factor and is nearly uncorrelated with the rest of the panel (Figure 9). Two forces drive the MDFM's advantage. First, both competitors treat the idiosyncratic errors as conditionally independent, ignoring the cross-country and cross-indicator residual correlation that the MDFM models. Second, the KOW and VDFM predictive densities additionally inherit the greater loading uncertainty documented in Section 5.2. As a result, the MDFM's joint predictive distributions remain better calibrated even at the two-year horizon.

6. Conclusion

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. An efficient Gibbs sampler exploits the matrix structure of the likelihood, and the cross-entropy importance-sampling estimator of Chan and Eisenstat (2015) selects the factor dimension, the vector versus matrix structure, and the idiosyncratic specification under a common criterion. In an application to a 19-country, 10-indicator OECD panel, the data decisively favor the matrix structure. Moreover, the MDFM recovers essentially the same factors as the unrestricted VDFM while estimating the loadings with credible intervals roughly half as wide; it reproduces the global business cycle of the multilevel model of Kose, Otrok and Whiteman (2003) with far fewer latent factors, and delivers markedly better density forecasts than both at every horizon.

Table 2: Out-of-sample forecast comparison: RMSFE ratios and average log predictive likelihood (ALPL) differences, MDFM versus KOW (Panels A–B) and MDFM versus VDFM (Panels C–D)

Panel A: RMSFE MDFM / KOW

Indicator	h=1	h=4	h=8
RGDP	1.012	1.001	1.000***
Headline CPI	1.005	1.004***	1.002***
Labor unit cost	1.038***	1.000	1.002***
Exports	1.039***	**0.998**	1.002***
Unemployment	1.012*	1.002*	1.001***
Energy CPI	**0.995**	1.003***	1.001***
Household consumption	1.022***	1.001***	1.000
Core CPI	1.006	**0.999**	1.000**
Food CPI	1.002	**0.998****	1.001***
Imports	1.045***	**0.997***	1.002***
Country	h=1	h=4	h=8
USA	1.028***	1.004**	1.001***
GBR	**0.999**	1.001	1.000**
AUS	**0.990**	1.004**	1.001***
DEU	1.029**	1.001	1.001***
JPN	1.014	**0.998**	1.000
FRA	1.029**	**0.999**	1.001***
NOR	**0.988**	**0.999**	1.001***
CAN	1.016	1.003**	1.002***
NLD	1.029**	**0.999**	1.001***
AUT	1.044***	1.000	1.001***
DNK	1.032**	1.000	1.001***
ESP	1.036***	1.000	1.001***
FIN	1.032**	**0.996****	1.002***
SWE	1.032**	**0.999**	1.001***
ITA	1.027**	**0.999**	1.000***
KOR	1.029***	1.002	1.001***
LUX	1.020	1.002	1.001***
NZL	1.028*	1.002	1.001*
PRT	1.006	**0.999**	1.001***
All (pooled)	1.021***	1.000	1.001***

Panel B: ALPL, MDFM − KOW

Indicator	h=1	h=4	h=8
RGDP	**8.46*****	**9.90*****	**8.36*****
Headline CPI	**3.20*****	**3.35*****	**2.98*****
Labor unit cost	**2.20*****	**1.01**	**1.18***
Exports	**3.90*****	**3.21*****	**3.64*****
Unemployment	**0.14**	**2.69**	**0.49**
Energy CPI	**4.37*****	**3.57*****	**2.86****
Household consumption	**3.48****	**9.04**	**2.55***
Core CPI	**0.81****	**1.27****	**1.57*****
Food CPI	**2.42*****	**2.70*****	**2.73*****
Imports	**4.57*****	**4.17*****	**4.65*****
Country	h=1	h=4	h=8
USA	**0.25**	**0.51**	−0.21
GBR	**1.37*****	**1.08**	**1.19*****
AUS	**1.21*****	**1.88***	**1.48*****
DEU	**0.83***	**0.73**	**1.21*****
JPN	−0.62	−1.14**	−1.54***
FRA	**0.40**	**0.61**	**0.78**
NOR	**1.34*****	**1.19****	**1.60*****
CAN	**0.65**	**1.82**	**0.82**
NLD	**1.02****	**0.71**	**1.28*****
AUT	**0.65**	**0.48**	**1.03*****
DNK	**0.63**	**0.18**	**0.99***
ESP	**1.04**	**0.93**	**1.81*****
FIN	**0.73**	**0.39**	**1.54*****
SWE	−0.30	−0.63	**0.11**
ITA	**0.10**	−0.61	**0.62**
KOR	**0.85*****	**0.48**	**0.66*****
LUX	**1.10*****	**1.10***	**1.63*****
NZL	**1.00*****	**0.99****	**1.19*****
PRT	**0.72**	**0.29**	**1.09***
All (joint)	**58.68*****	**66.00*****	**57.80*****

Panel C: RMSFE MDFM / VDFM

Indicator	h=1	h=4	h=8
RGDP	**0.869*****	**0.991*****	1.000
Headline CPI	1.020**	1.004***	1.001***
Labor unit cost	1.036***	**0.994*****	1.001***
Exports	1.011	**0.991*****	1.001***
Unemployment	1.031	**0.995****	1.000
Energy CPI	**0.994**	1.001	1.000
Household consumption	**0.875*****	**0.990*****	1.000**
Core CPI	1.019**	**0.998**	1.000
Food CPI	1.029***	**0.991*****	1.002***
Imports	1.030*	**0.990*****	1.001***
Country	h=1	h=4	h=8
USA	**0.963**	**0.997**	1.001**
GBR	**0.957**	**0.992****	1.001**
AUS	**0.991**	**0.996**	1.000
DEU	**0.970**	**0.995****	1.001**
JPN	1.009	**0.995****	1.000
FRA	**0.958**	**0.995****	1.001***
NOR	**0.984**	**0.995****	1.000
CAN	**0.948***	1.000	1.001*
NLD	**0.970**	**0.985****	1.000
AUT	**0.987**	**0.993*****	1.001*
DNK	1.016	**0.990****	1.000
ESP	**0.959**	**0.994*****	1.001
FIN	**0.990**	**0.993*****	1.000
SWE	**0.983**	**0.995****	1.000
ITA	**0.935**	**0.992****	1.000
KOR	1.010	**0.994***	1.001
LUX	**0.985**	**0.993****	1.001**
NZL	**0.988**	**0.998**	1.001**
PRT	**0.952**	**0.992****	1.000
All (pooled)	**0.974*****	**0.994*****	1.000***

Panel D: ALPL, MDFM − VDFM

Indicator	h=1	h=4	h=8
RGDP	**1.69*****	**1.67*****	**1.34*****
Headline CPI	**1.73*****	**1.86*****	**1.86*****
Labor unit cost	−0.86	−0.72	**0.02**
Exports	**0.44**	**0.47**	**1.05*****
Unemployment	−0.16	**0.61**	**0.70***
Energy CPI	**0.58***	**0.28**	**0.10**
Household consumption	**1.68*****	**0.89**	**0.70**
Core CPI	**0.39**	**0.87*****	**1.06*****
Food CPI	**1.05*****	**1.14*****	**1.27****
Imports	**0.95****	**1.05****	**1.74*****
Country	h=1	h=4	h=8
USA	**0.85**	**0.24**	**0.65**
GBR	**1.68*****	**1.15**	**1.57****
AUS	**1.50*****	**1.71*****	**2.13*****
DEU	**1.63*****	**1.36**	**2.10*****
JPN	−0.15	−0.56	−0.51
FRA	**1.07**	**0.86**	**1.47***
NOR	**1.39*****	**1.57*****	**2.07*****
CAN	**1.23***	**0.74**	**1.46*****
NLD	**1.06**	**1.04**	**1.71*****
AUT	**1.04***	**1.06**	**1.78*****
DNK	**0.74**	**0.98**	**1.89*****
ESP	**0.64**	**0.82**	**1.74*****
FIN	**1.18**	**1.30**	**2.36*****
SWE	**0.67**	**0.38**	**1.43****
ITA	−0.18	−0.73	**0.62**
KOR	**1.93*****	**1.72*****	**1.99*****
LUX	**1.73*****	**1.91****	**2.59*****
NZL	**1.48*****	**1.64*****	**2.08*****
PRT	**0.74**	**0.64**	**1.29***
All (joint)	**42.64*****	**45.10*****	**45.01*****

Notes: Panels A and C report pooled RMSFE ratios of the MDFM to the competing model; values below 1 favor the MDFM and are shown in **bold**. The indicator block pools across the 19 OECD countries within each indicator and the country block pools 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, and the last row reports the fully pooled ratio across all $19 \times 10 = 190$ cells. Panels B and D report differences in average log predictive likelihoods (average log predictive Bayes factors) for the joint density of each indicator across the 19 countries and of each country across the 10 indicators; positive values favor the MDFM and are shown in **bold**, and the last row reports the joint density of the full 190-series panel. Significance is based on Diebold–Mariano (Panels A and C) and Amisano–Giacomini (Panels B and D) tests on the per-origin differentials with Bartlett HAC variance estimation: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

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.

Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. Journal of Business & Economic Statistics 25, 177–190.

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.

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

Bańbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. Journal of Applied Econometrics 25, 71–92.

Barigozzi, M., Hallin, M., 2026. The dynamic, the static, and the weak: factor models and the analysis of high-dimensional time series. Journal of Time Series Analysis 47, 201–219.

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., 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 51, 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.

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., 2026. Bayesian model comparison for large Bayesian VARs after the COVID-19 pandemic. Journal of Econometrics 256, 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., Jeliazkov, 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., 2026. Large Bayesian matrix autoregressions. Journal of Econometrics 256, 105955.

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

Chen, B., Chen, E.Y., Bolivar, S., Chen, R., 2024. Time-varying matrix factor models. arXiv preprint arXiv: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 115, 775–793.

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 .

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

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.

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 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.

Geweke, J., Zhou, G., 1996. Measuring the Pricing Error of the Arbitrage Pricing Theory. Review of Financial Studies 9, 557–587.

Giglio, S., Xiu, D., Zhang, D., 2025. Test assets and weak factors. Journal of Finance 80, 259–319.

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.

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.

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

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

Leung, D., Drton, M., 2016. Order-Invariant Prior Specification in Bayesian Factor Analysis. Statistics and Probability Letters 111, 60–66.

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 14, 41–67.

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

Qin, L., Wang, Y., Zhu, Y., Shia, B.C., 2025. Bayesian dynamic matrix factor models. Journal of Business & Economic Statistics 43, 1170–1182.

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., 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 $\mathbf{C}\mathbf{F}_t\mathbf{D}'$.

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

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

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

Let

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

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,

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

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,

$$\begin{bmatrix} b_{11} & 0 & ... & 0 \\ b_{21} & b_{22} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{p_2 1} & b_{p_2 2} & ... & b_{p_2 p_2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & ... & b_{np_2} \end{bmatrix} \begin{bmatrix} d_{11} & ... & d_{1p_2} \\ \vdots & \ddots & \vdots \\ d_{p_2 1} & ... & d_{p_2 p_2} \end{bmatrix} = \begin{bmatrix} b^*_{11} & 0 & ... & 0 \\ b^*_{21} & b^*_{22} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b^*_{p_1 1} & b^*_{p_1 2} & ... & b^*_{p_2 p_2} \\ \vdots & \vdots & \ddots & \vdots \\ b^*_{n1} & b^*_{n2} & ... & b^*_{np_2} \end{bmatrix} \quad \text{(Appendix A.3)}$$

For (Appendix A.3) to hold, we must have $d_{ij} = 0$ for any i, j such that $i < j$, or $\mathbf{D}^{-1}$ is lower triangular given the assumption that $b_{ii} \neq 0$, $b^*_{ii} \neq 0$, for $i = 1, ..., p_2$.

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

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

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

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

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

Therefore

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

The observation equation after the rotation becomes

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

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 C and D are lower triangular matrices and the diagonal elements of $\mathbf{C}$ is ones.

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

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

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., $\mathbf{M}\boldsymbol{\Lambda}_t\mathbf{M}' = \boldsymbol{\Lambda}_t$, where $\boldsymbol{\Lambda}_t$ is a diagonal matrix. Given that the diagonal elements in $\boldsymbol{\Lambda}_t$ must be larger than 0, this requires that $m_{i,j}$ for all $i > j$ must be zero, for $\mathbf{M}\boldsymbol{\Lambda}_t\mathbf{M}'$ to only have non-zero terms on its diagonal and match $\boldsymbol{\Lambda}_t$. 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

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

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

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

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

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

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

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

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

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

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

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

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

inverse-Wishart distribution:

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

where

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

With the constraints for identification, we cannot directly sample from the above normal-inverse-Wishart distribution. Here we outline the sampling scheme for $\mathbf{A}'$ with the structure constraints. To that end, we first represent the restrictions as a system of linear restrictions. For example, for $\mathbf{A}'$, we represent the restrictions that $\mathbf{A}$ is a lower triangular matrix with ones on the diagonal using $\mathbf{M}_{\mathbf{A}'} \text{vec}(\mathbf{A}') = \mathbf{a}_0$. Assuming $n > p_1$, $\mathbf{M}_{\mathbf{A}'} = (m_{i,j})$ is a $p_1(p_1+1)/2 \times np_1$ selection matrix, and $\mathbf{a}_0$ is a $p_1(p_1+1)/2 \times 1$ vector consisting of ones and zeros. Then we apply Algorithm 2 in Cong et al. (2017) or Algorithm 1 in Chan and Qi (2026) to efficiently sample $(\text{vec}(\mathbf{A}') | \cdot) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}'), \mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})$ such that $\mathbf{M}_{\mathbf{A}'} \text{vec}(\mathbf{A}') = \mathbf{a}_0$. In particular, one can first sample $\text{vec}(\mathbf{A}_u')$ from the unconstrained conditional posterior distribution in Step 1, and then return

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

which can be realized by the following four steps:

(1) Compute $\mathbf{C} = \mathbf{C}_{\mathbf{\Sigma}_r^{-1}} \otimes \mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$, where $\mathbf{C}_{\mathbf{\Sigma}_r^{-1}}$ is the lower Cholesky factor of $\mathbf{\Sigma}_r^{-1}$, and $\mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$ is the lower Cholesky factor of $\mathbf{K}_{\mathbf{A}'}$;

(2) Solve $\mathbf{C}\mathbf{C}'\mathbf{U} = \mathbf{M}_{\mathbf{A}'}'$ for $\mathbf{U}$;

(3) Solve $\mathbf{M}_{\mathbf{A}'}\mathbf{U}\mathbf{V} = \mathbf{U}'$ for $\mathbf{V}$;

(4) Return $\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}_u') + \mathbf{V}'(\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'} \text{vec}(\mathbf{A}_u'))$.

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

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

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

where

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

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

We sample $(\mathbf{B}, \mathbf{\Sigma}_c | \cdot)$ in two steps. First, we sample $\mathbf{\Sigma}_c$ marginally from $(\mathbf{\Sigma}_c | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r) \sim \mathcal{IW}(\nu_c + Tn, \widehat{\mathbf{S}}_c)$ 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 $(\text{vec}(\mathbf{B}') | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{B}}'), \mathbf{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}'}^{-1})$, which can be done using the algorithm described in step 1.

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

(1) Exchange row/column 1 and n in the matrix $\widehat{\mathbf{S}}_c$. Denote this matrix as $\widehat{\mathbf{S}}_c^{Trans}$.

(2) Construct a lower triangular matrix $\mathbf{\Delta}$ such that

- δ_{ii} equal to the square root of $\chi^2_{\widehat{\nu}_c + 1 - i}$ for $i = 1, \ldots, n-1$;
- $\delta_{nn} = (l_{nn})^{-1}$, where l_{nn} is the (n,n)th element in the Cholesky decomposition of $(\widehat{\mathbf{S}}_c^{Trans})^{-1}$, denoted as $\boldsymbol{L}$;
- δ_{ij} equal to $\mathcal{N}(0,1)$ random variates, $i > j$.

(3) Set $\mathbf{\Sigma}_c = (\boldsymbol{L}^{-1})'(\mathbf{\Delta}^{-1})' \mathbf{\Delta}^{-1} \boldsymbol{L}^{-1}$.

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

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

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

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

where

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

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

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

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

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

$$\widetilde{\mathbf{f}}_{j,k} = \widetilde{\mathbf{F}}_{j,k}\boldsymbol{\rho}_{j,k}+\mathbf{u}_{j,k}, \quad \mathbf{u}_{j,k}\sim\mathcal{N}(\mathbf{0},\lambda_{j,k}^2\mathbf{I}_{T-q}), \qquad \text{(Appendix B.4)}$$

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

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

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

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

6. 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.

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

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

an acceptance-rejection Metropolis-Hastings step. Samplers for ϕ and σ_h^2 are standard and we omit the details in this paper.

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

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

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

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

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

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

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

Appendix C. Estimating Marginal Likelihoods

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

Appendix C.1. Integrated Likelihood

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

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

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

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

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

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

where

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

and

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

It can be very costly to compute the inverse of the covariance matrix $\overline{\mathbf{D}}_{\mathbf{y}}$. Therefore, here we use the 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:

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

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

Therefore, the integrated likelihood at time t is:

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

where

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

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

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

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

In order to obtain the maximum likelihood estimators for the parameters in the 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

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

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

A Wishart distribution is defined as:

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

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

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

The log-likelihood function is therefore

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

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

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

where two results are used

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

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

the degree of freedom δ

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

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 δ. We then use 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

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

The second derivative is

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

Maximum likelihood estimators for parameters for normal distributions and inverse gamma distributions are straightforward to obtain, so 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 from which we downloaded the specific variable to the variable name in the table. The second column of Table 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	Δx
Unemployment	Δx
Headline CPI	Δx
Energy CPI	Δx
Food CPI	Δx
Core CPI	Δx
Imports	$\Delta \log(x)$
Exports	$\Delta \log(x)$

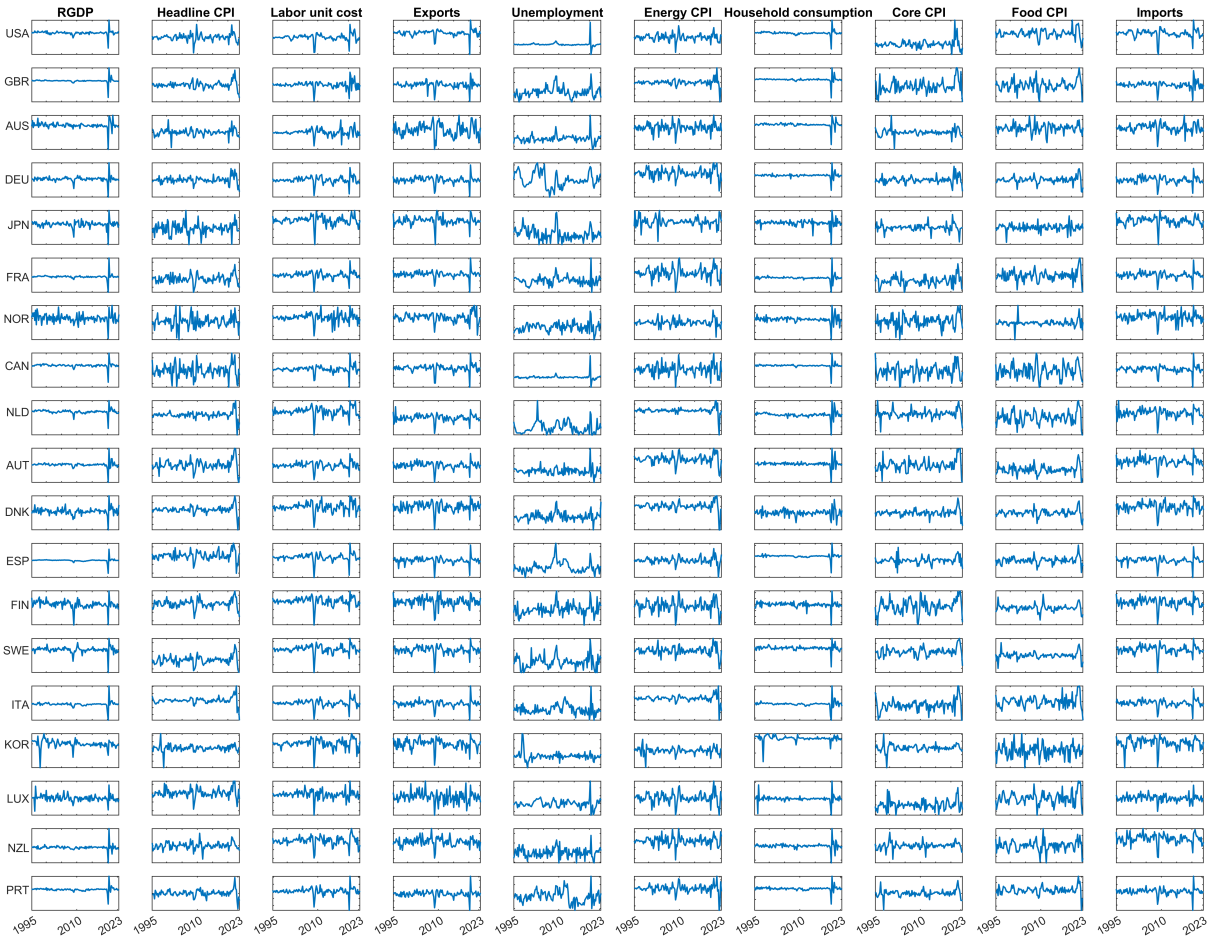


Figure Appendix E.1: The ten macroeconomic indicators (by columns) for 19 countries (by rows). The horizontal axis represents time and the vertical axis represents the standardized growth rates. The ranges of the vertical values are not the same. The order of countries and indicators is the order adopted 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 span($\mathbf{A}$) and span($\mathbf{B}$) are recovered as the leading eigenspaces of two symmetric positive semidefinite matrices constructed from lagged sample autocovariances:

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

where $\widehat{\mathbf{\Sigma}}_h^{(1)}$ and $\widehat{\mathbf{\Sigma}}_h^{(2)}$ 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 $\mathrm{Cov}(\mathbf{F}_t, \mathbf{F}_{t-h}) \neq 0$ for $h \geqslant 1$, 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: $\widehat{p}_d = \arg\min_{i \leqslant P_d/2} \lambda_{d,i+1}/\lambda_{d,i}$ for $d = 1, 2$, where $\lambda_{d,1} \geqslant \lambda_{d,2} \geqslant \cdots$ 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.

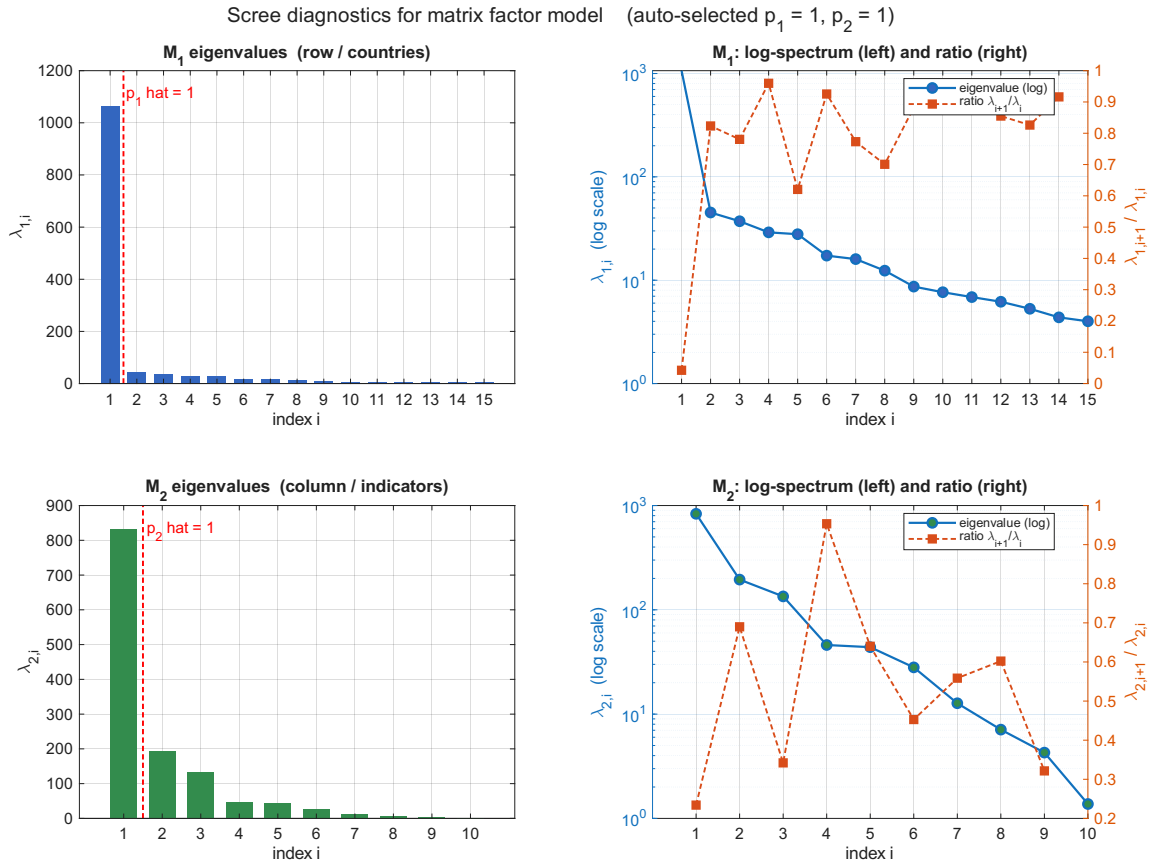


Figure Appendix F.1: Scree diagnostics for the Wang et al. (2019) estimator on the OECD panel. Top row: $\widehat{\mathbf{M}}_1$ eigenvalues (row / countries) in linear scale (left) and log scale with the eigenvalue-ratio diagnostic (right). Bottom row: same diagnostics for $\widehat{\mathbf{M}}_2$ (column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

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	$\widehat{p}_1$	$\widehat{p}_2$	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 $(\widehat{p}_1, \widehat{p}_2) = (1, 1)$ 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. Finally, 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 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 $(\widehat{p}_1, \widehat{p}_2) = (1, 1)$. Figure Appendix F.2 reports the per-cell ratio of in-sample RMSE between the two methods, arranged as a 19×10 heatmap with countries on the rows and indicators on the columns.

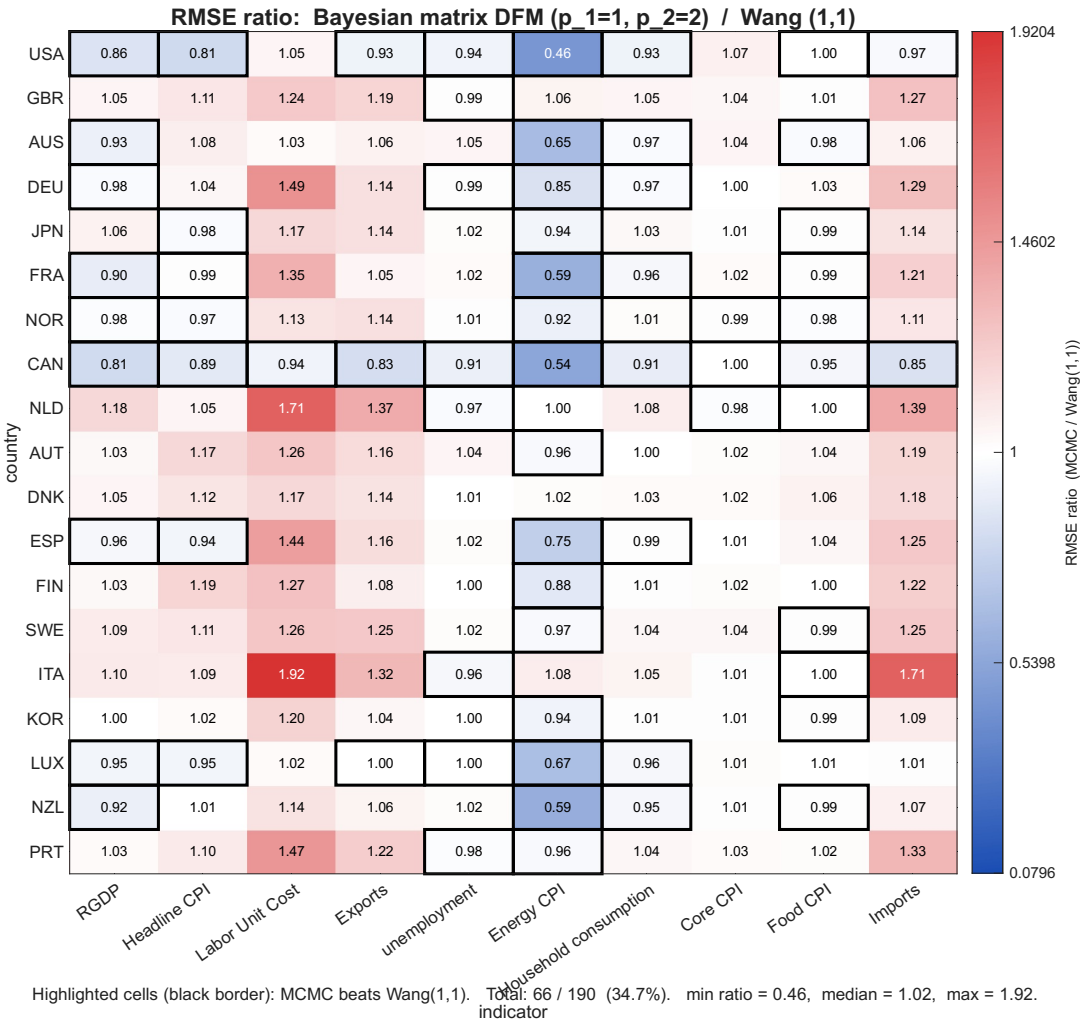


Figure Appendix F.2: Per-cell ratio of in-sample RMSE, Bayesian MDFM-cross-sv with $(p_1, p_2) = (1, 2)$ relative to Wang et al. (2019) with $(\widehat{p}_1, \widehat{p}_2) = (1, 1)$. Black borders mark cells where the Bayesian model achieves lower RMSE.

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 $s \leqslant t$. For each forecast horizon h and each (country, indicator) pair (i,j), we then estimate the horizon-specific direct projection

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

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. The Multilevel Dynamic Factor Model Used in Section 5

For the comparison, we estimate a two-level multilevel specification in the spirit of KOW: each series loads on a world factor w_t common to all series and on a factor $g_{i,t}$ specific to country i,

$$y_{ijt} = \lambda_{ij}^{w}\,w_t + \lambda_{ij}^{c}\,g_{i,t} + e_{ijt}, \qquad e_{ijt} \sim \mathcal{N}\left(0,\,\omega_t\,\sigma_{ij}^2\right), \tag{Appendix G.1}$$

$$w_t = \rho^{w} w_{t-1} + u_t^{w}, \qquad u_t^{w} \sim \mathcal{N}(0,\sigma_w^2), \tag{Appendix G.2}$$

$$g_{i,t} = \rho_i^{c}\,g_{i,t-1} + u_{i,t}^{c}, \qquad u_{i,t}^{c} \sim \mathcal{N}(0,\sigma_{g,i}^2), \qquad i = 1,\dots,n, \tag{Appendix G.3}$$

where the factor innovations and the idiosyncratic shocks are mutually independent at all leads and lags, so that all comovement — across countries and across indicators — is channeled through the factors, and $\omega_t = \exp(h_t)$ is the same common stochastic volatility process as in the MDFM and the VDFM. The specification is identified by unit-loading anchors that parallel the identification of the MDFM: the world-factor loading of U.S. real GDP and the loading of each country's real GDP on its own country factor are fixed to one. The model is implemented as a VDFM with $1+n$ factors and a sparse loading

pattern, and is estimated with the same priors and samplers as the VDFM of Section 5.1, so the three models are directly comparable.

Appendix H. MDFM factors versus the VDFM factor space

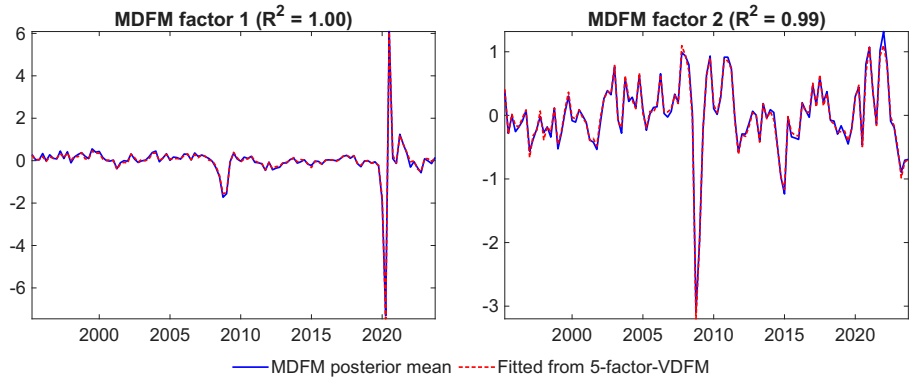


Figure Appendix H.1: MDFM factors versus the VDFM factor space (matched specifications, $k_f = 5$). Each panel plots the posterior mean of one MDFM factor (blue solid) and its fitted value from a regression on the five VDFM factors (red dashed); the R^2 values of 0.99 show that the matrix factors lie within the space spanned by the VDFM factors.

Appendix I. Extra Forecasting Results

Table Appendix I.1: Out-of-sample RMSFE ratios MDFM / MFM, pooled across cells

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

Notes: 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 the MDFM outperforms the 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.

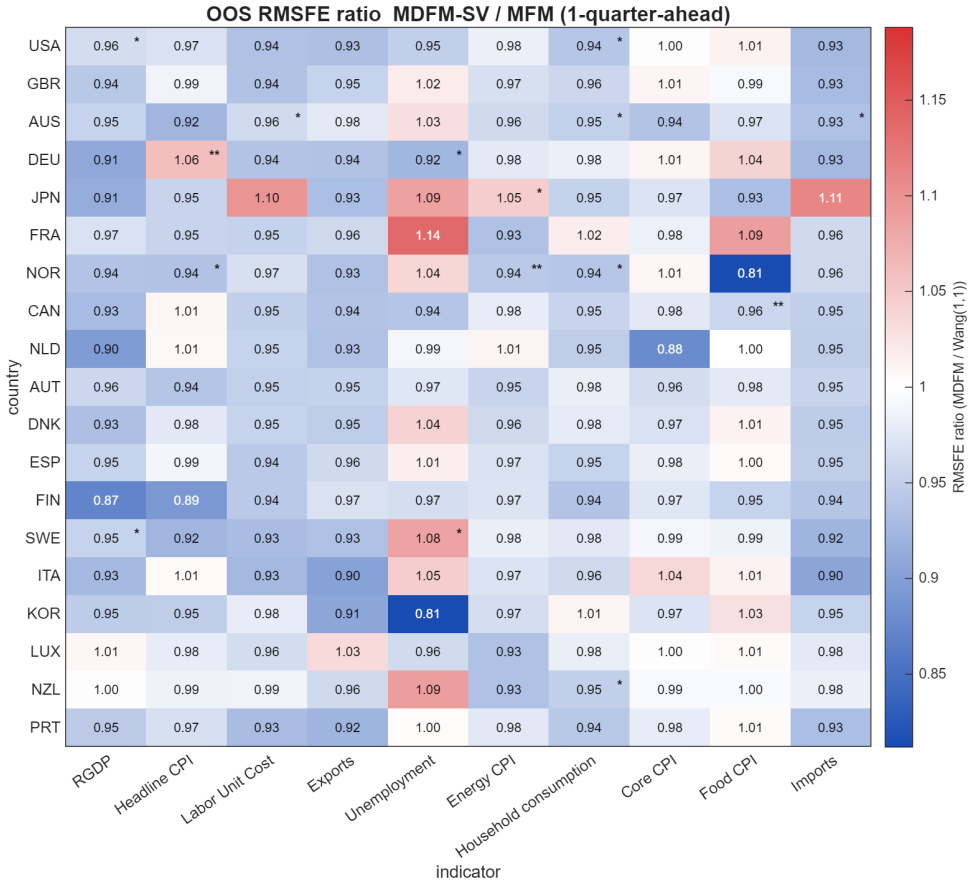


Figure Appendix I.1: Per-cell RMSFE ratios of the MDFM-SV relative to the static matrix factor benchmark, by indicator and country, at the 1-quarter-ahead horizon ($h = 1$). Entries below one indicate that the MDFM-SV forecast is more accurate; asterisks denote rejection of the Diebold–Mariano test of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

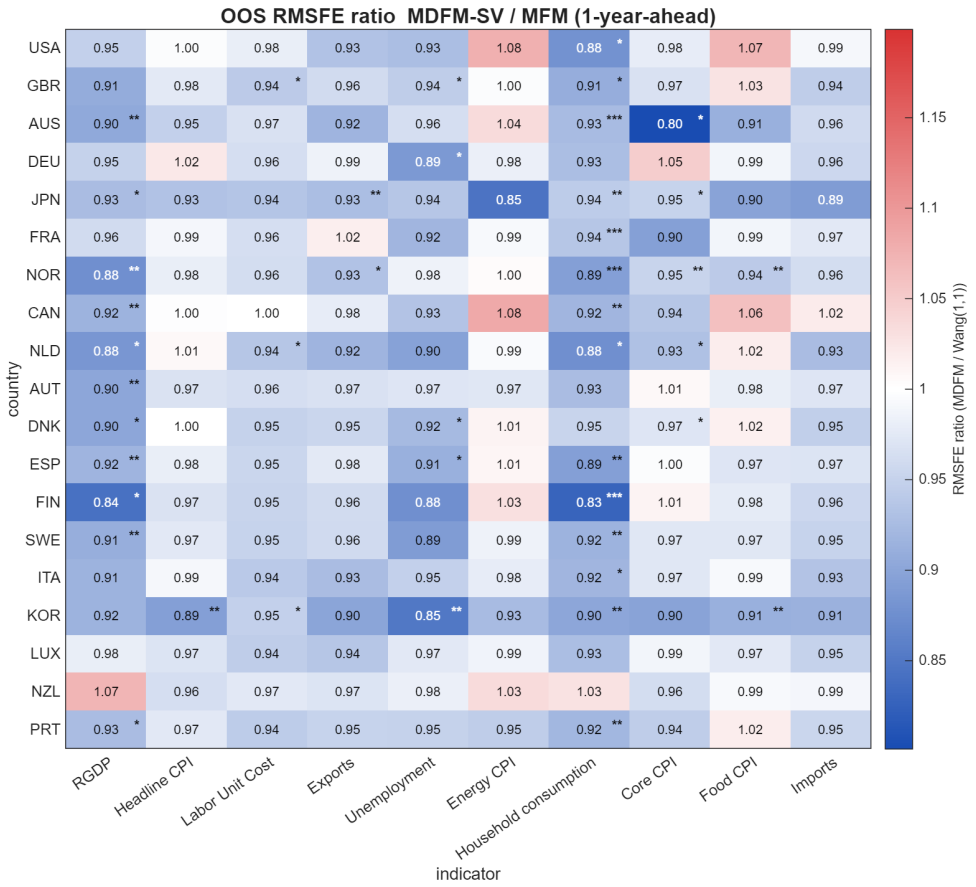


Figure Appendix I.2: Per-cell RMSFE ratios of the MDFM-SV relative to the static matrix factor benchmark, by indicator and country, at the 1-year-ahead horizon ($h = 4$). See the note to Figure Appendix I.1.

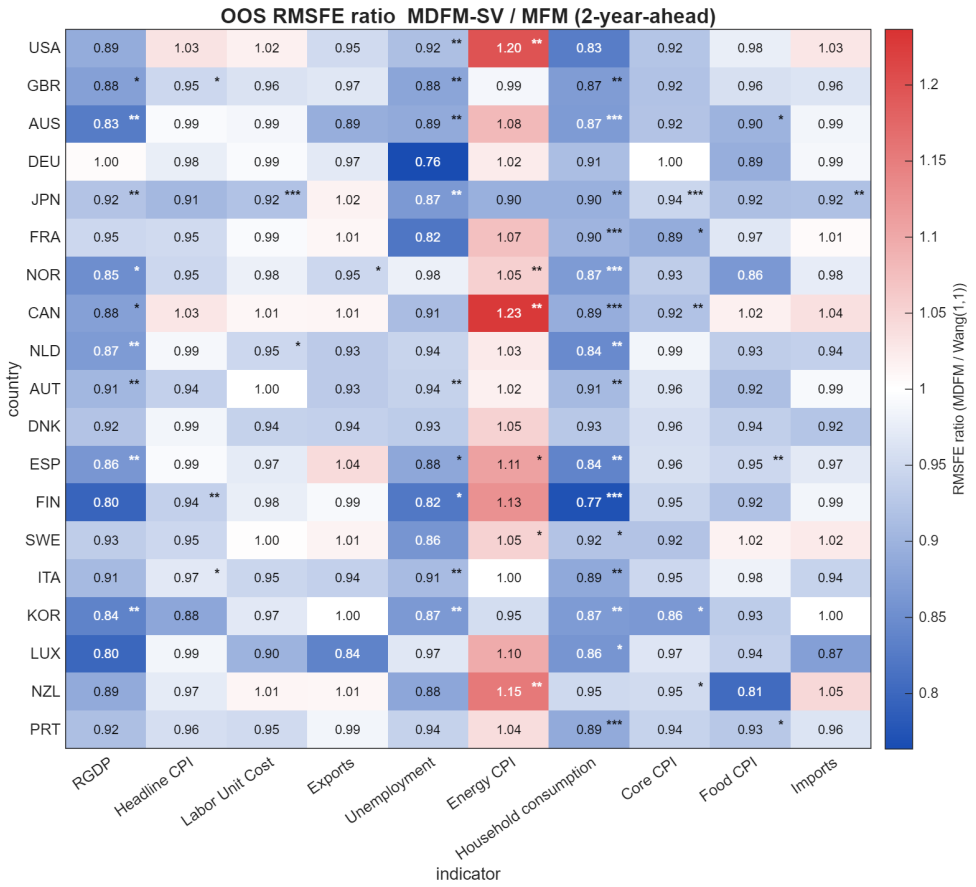


Figure Appendix I.3: Per-cell RMSFE ratios of the MDFM-SV relative to the static matrix factor benchmark, by indicator and country, at the 2-year-ahead horizon ($h = 8$). See the note to Figure Appendix I.1.

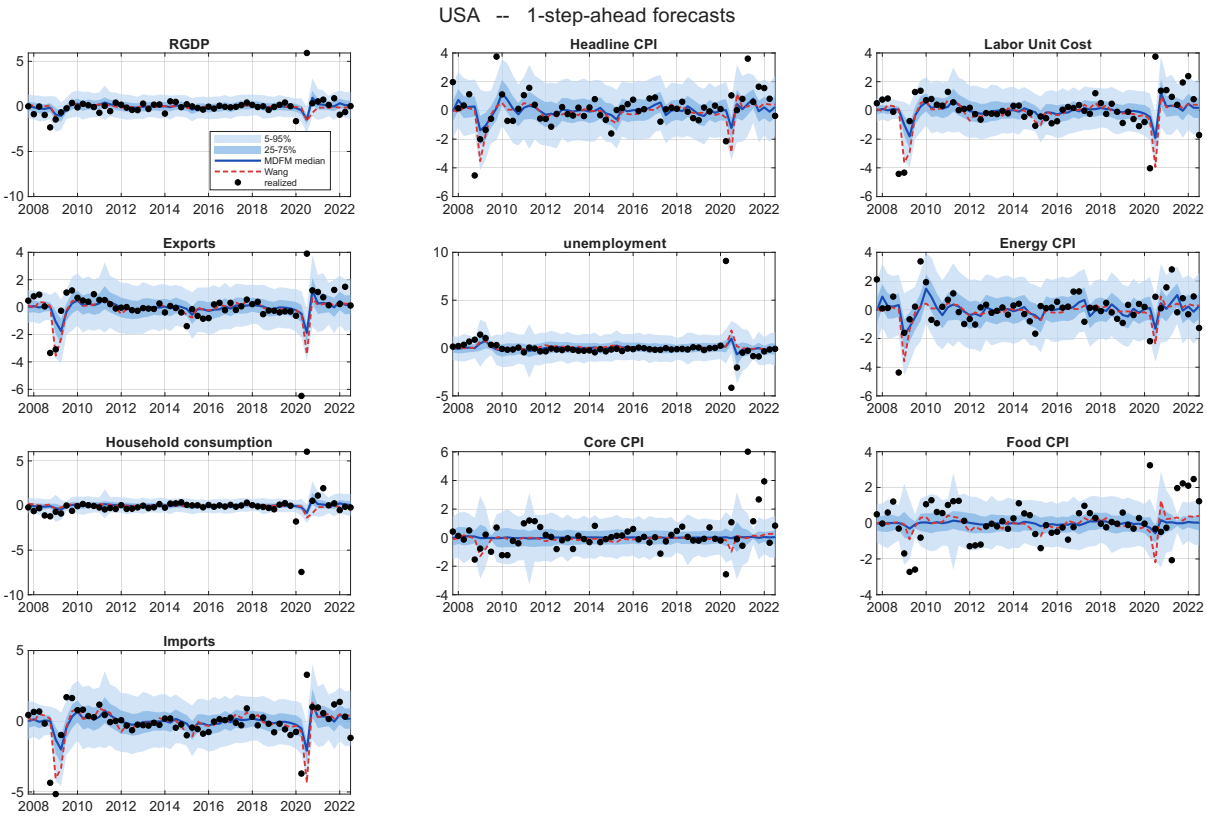


Figure Appendix I.4: Posterior predictive fan charts for real GDP growth (annualized percent change) across the G7 economies. The top block shows 1-quarter-ahead forecasts; the bottom block shows 4-quarter (1-year) average forecasts. Each panel reports the MDFM posterior median and 5–95% and 25–75% predictive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

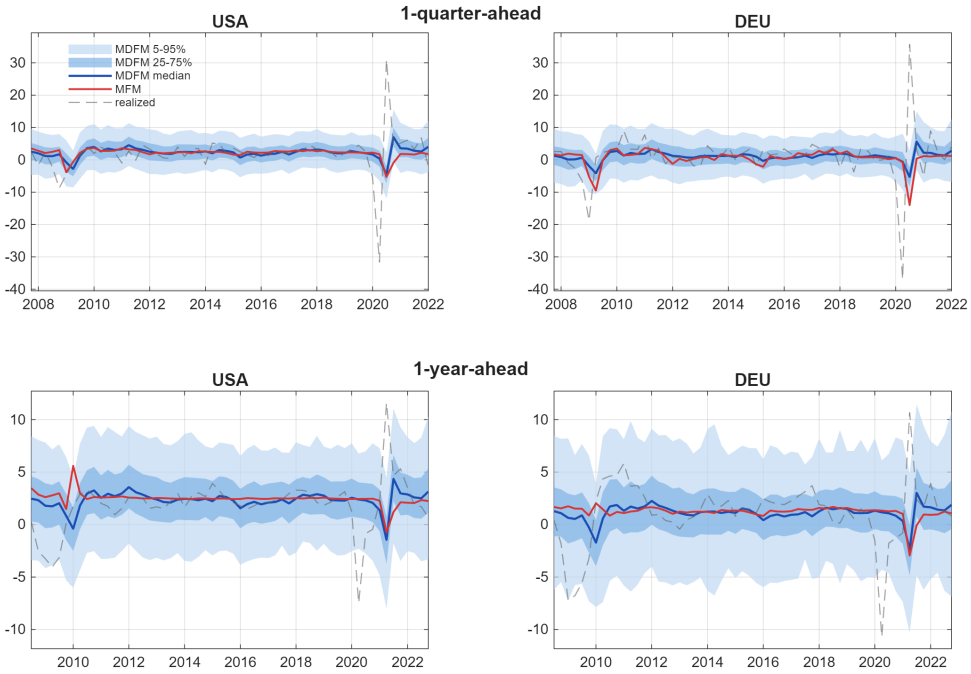


Figure Appendix I.5: Posterior predictive fan charts for real GDP growth (annualized percent change) in the United States and Germany. The top row shows 1-quarter-ahead forecasts; the bottom row shows 4-quarter (1-year) average forecasts, with both MDFM posterior predictive draws and realized growth averaged across $h = 1, \ldots, 4$. 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.

Appendix J. Robustness of the OECD Application to Data Ordering

The lower-triangular identification scheme used throughout the paper pins down the factor matrix by anchoring the leading rows and columns of the loading matrices. The substantive interpretation of the loadings, and the cell-level decomposition of forecasts into common and idiosyncratic components, therefore depend on the ordering of the data. We assess the empirical importance of this dependence on the OECD panel of Section 5.

Appendix J.1. Design

We re-run the full expanding-window OOS pipeline of Section 5 under eight orderings of the panel: the baseline economic ordering and seven alternatives. The alternatives are chosen to span both structured perturbations (full reversal of rows and columns; alphabetical sorting of country and indicator labels) and random perturbations (three random permutations of both dimensions with fixed seeds; one random permutation of

rows with columns held at baseline; one random permutation of columns with rows held at baseline). The latter two isolate the contributions of row and column ordering separately.

For each ordering, we permute the data once at the outset, run the expanding-window MDFM-cross-sv MCMC at every origin, and compute predictive means at horizons $h = 1, 2, \ldots, 8$. We then permute the predictive tensor back to baseline coordinates so that all forecast comparisons take place in a common frame. Table Appendix J.1 lists the orderings tested and identifies the country and indicator placed in the leading row and column of each.

Table Appendix J.1: Orderings tested in the permutation experiment. The leading row and column carry the identifying weight under the lower-triangular scheme.

ID	Label	Leading row	Leading column
1	Baseline (economic ordering)	USA	RGDP
2	Reverse rows and columns	PRT	Imports
3	Alphabetical	AUS	Core CPI
4	Random #1 (seed 1001)	JPN	RGDP
5	Random #2 (seed 1002)	LUX	Imports
6	Random #3 (seed 1003)	NOR	Imports
7	Rows random, columns baseline	GBR	RGDP
8	Rows baseline, columns random	USA	Food CPI

Appendix J.2. 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 Appendix J.2 reports the pooled RMSFE across the panel for each ordering. It is clear that the pooled RMSFE varies by less than nine 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 Appendix J.2: Pooled out-of-sample RMSFE across the OECD panel for the MDFM under eight orderings of rows and columns. RMSFE is pooled over all 19×10 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

Appendix K. Application: Fama–French 10×10 Panel

In this application, we investigate the usefulness of the dynamic matrix factor model on the Fama-French return series, which was studied by Wang et al. (2019), Yu et al. (2022) and He et al. (2024). The data include monthly returns of 100 portfolios, structured in a 10 by 10 matrix according to ten levels of sizes (market equity) and ten levels of ratio of book equity to market equity (BE/ME).[12] The return series span from January 1990 to June 2024 (414 observations).[13] Following Chang et al. (2023), we impute the missing values by the weighted averages of the three previous months, i.e., set $y_{i,j,t} = 0.5y_{i,j,t-1} + 0.3y_{i,j,t-2} + 0.2y_{i,j,t-3}$ for missing $y_{i,j,t}$. To account for market conditions, we follow Wang et al. (2019), Yu et al. (2022), and He et al. (2024) and subtract the monthly excess market return from each series. We then standardize the data by subtracting the mean and dividing by the standard deviation. The standardized market-adjusted return series of the portfolios is shown in Figure Appendix K.1.

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

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

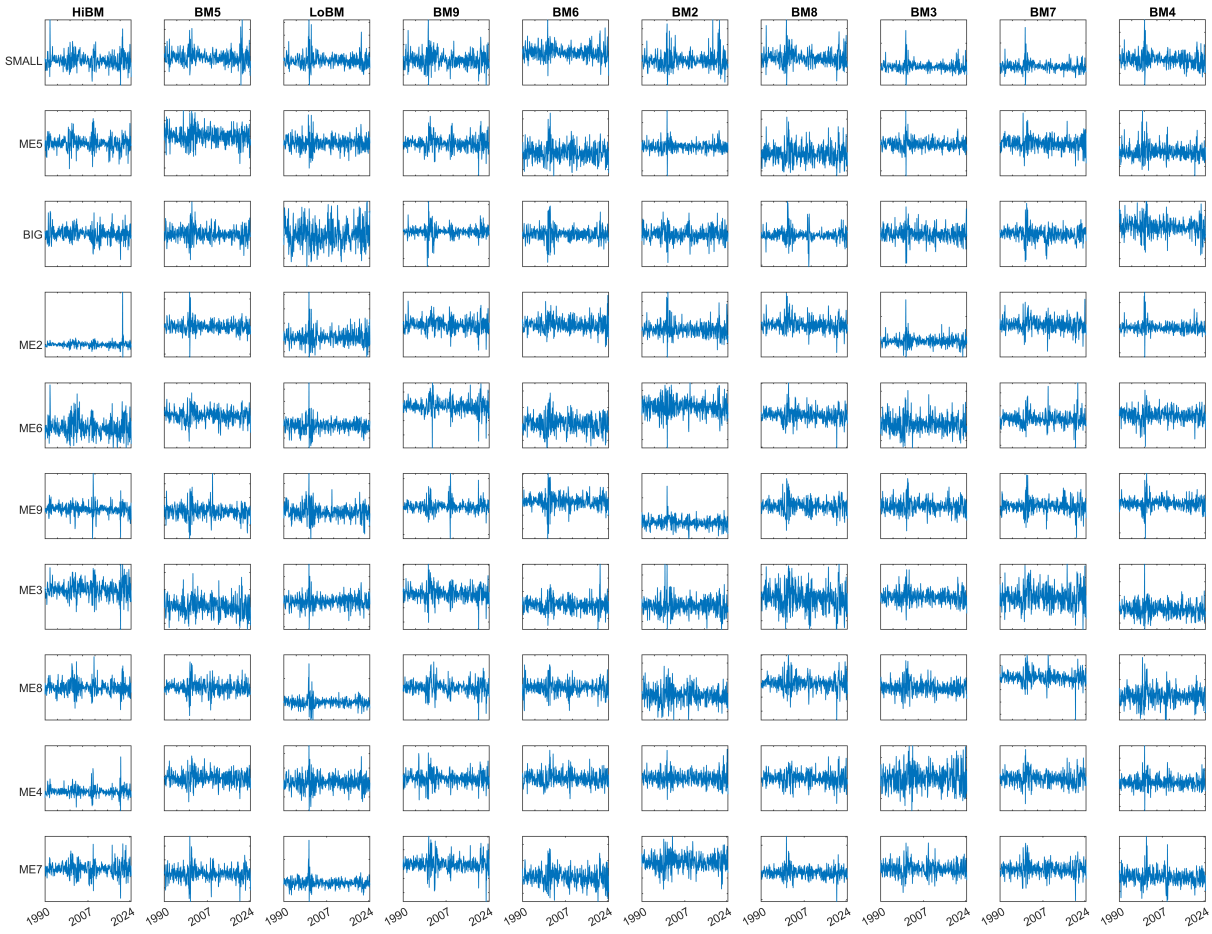


Figure Appendix K.1: The return series of the portfolios structured by different levels of sizes (rows) and book equity to market equity ratio (columns). Note that we have rearranged the order of rows and columns. The horizontal axis represents time and the vertical axis represents the standardized monthly 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.

Tables Appendix K.1 and Appendix K.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 2×2 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 K.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	**-45729**	-45792	-45886	**-42877**	-43007	-43155	-43328
(0.6)	(0.9)	(1.0)	(1.0)	(0.7)	(0.8)	(1.6)	(0.8)

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

Figure Appendix K.2 shows the estimates for row loading matrices (the first and second panel from left) and column loading matrices (the third and fourth panel). Specifically, the four panels 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 show clear 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 influences portfolios with similar sizes, with its influence tapering off as the portfolio size shifts either smaller or larger. Similar patterns hold for the BE/ME factors.

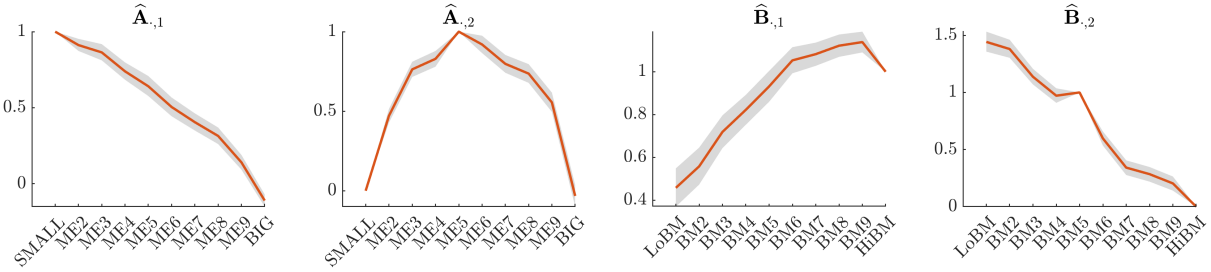


Figure Appendix K.2: Estimates for row loading matrices (the first and second panel from left) and column loading matrices (the third and fourth panel). The gray area represents the 90% credible interval.

Figure Appendix K.3 shows estimated posterior densities, histograms of posterior draws, priors, as well as the posterior estimates of autoregressive coefficients ($\boldsymbol{\rho}$) for the factor evolution process. All four posterior densities have little mass near zero, 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.

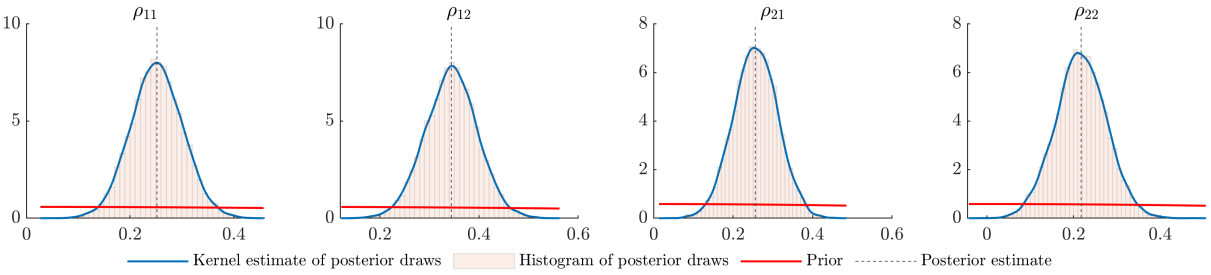


Figure Appendix K.3: Posterior densities, histograms of posterior draws, priors and posterior estimates for autoregressive coefficients $\boldsymbol{\rho}$.

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