Matrix DFM and Matrix TPRF: A Nowcasting Framework for Mixed-Frequency Data in the Euro Area

Matteo Barigozzi[†] **Massimiliano Marcellino**[‡] **Davide Delfino**[‡]

[†] Alma Mater Studiorum – Università di Bologna

[‡] Bocconi University

Abstract

This paper develops a pseudo real-time nowcasting framework for mixed-frequency macroeconomic panels that are naturally matrix-valued, with predictors observed as a two-way array (e.g., countries $\times$ indicators) over time and a low-frequency target such as quarterly GDP growth. We make two contributions. First, we propose a *Matrix Mixed-Frequency Three-Pass Regression Filter* (Matrix MF–TPRF), which extends proxy-driven targeted factor extraction to matrix panels, preserving the two-way structure and isolating forecast-relevant common variation. Second, we provide a refinement and a systematic nowcasting application based on a *Dynamic Matrix Factor Model* (DMFM) estimated by quasi-maximum likelihood via an EM algorithm, accommodating general missingness and ragged-edge releases within a state-space framework.

Using the EA macroeconomic data, we construct a unified pseudo real-time evaluation for euro-area GDP growth in Germany, France, Italy, and Spain, explicitly replicating the asynchronous release calendar of monthly and quarterly indicators. We benchmark the two matrix-based approaches against their standard vector counterparts (MF–TPRF and MF–DFM) under identical information sets and missingness patterns. The results indicate that exploiting the matrix structure improves both nowcast accuracy and within-quarter updating, with the largest gains around the COVID recession and the subsequent recovery. Improvements are particularly pronounced for Spain and France across the evaluation sample, while Germany and Italy benefit mainly in the post-COVID period, consistent with the view that matrix-based methods are most valuable when cross-country comovement is strong and cross-sectional information becomes especially informative for real-time GDP tracking[1].

Keywords: (Matrix) Three Pass Regression Filter; Dynamic (Matrix) Factor Model; Mixed Frequency; Expectation-Maximization Algorithm; Kalman Smoother; Missing Values; Nowcasting.

Contents

[1]TO REVIEW: considerations just based on the thesis results

1 Introduction

Recent macroeconometric research is increasingly developing methods that exploit the richness of available information in a real-time decision-making environment, where timely assessments can meaningfully shape policy both in normal times and during crises. On the one hand, central banks and forecasters observe large panels of indicators spanning multiple sectors and countries and released at heterogeneous frequencies; on the other hand, key macroeconomic aggregates such as GDP are published with substantial delays and typically at lower frequency. This mismatch generates three intertwined challenges. First, *dimensionality*: the predictive content of hundreds of series must be summarized in a way that remains statistically stable and economically interpretable. Second, *missingness*: due to asynchronous information flows indicators are released at different frequencies and with heterogeneous publication lags, yielding mixed-frequency and ragged-edge datasets in which the information set changes within the quarter and is unbalanced by construction. Moreover as pointed out by Zhu, T. Wang, and Samworth (2022) in high dimensionality settings is not surpising to expect more occurrence of incomplete observations. Third, *fragility under disruptions*: large shocks—most notably the COVID-19 crisis and the subsequent recovery—can alter comovement patterns and complicate both estimation and real-time updating, ultimately requiring central banks to expand and adapt their forecasting toolbox.

Factor methods provide a natural response to the dimensionality problem by summarizing large datasets through a small number of latent forces driving common variation. This approach is then a desirable one in case of non-negligible cross-sectional and time-series correlations. In this spirit, diffusion index models, as introduced by Stock and Watson (2002) and Bai and Ng (2006), offer a convenient forecasting strategy: a few factors estimated from a high-dimensional panel are used to predict a target variable, achieving substantial dimensionality reduction while preserving the dominant cross-sectional and time-series comovements. However, two limitations are particularly relevant in the empirical environment considered here. First, the combination of mixed frequencies and ragged-edge releases generates complex missing-data patterns that standard PCA-based implementations do not accommodate seamlessly. Naive interpolation or temporal aggregation may distort the underlying covariance structure and, in turn, bias factor estimates. This has motivated likelihood-based approaches that embed the factor model in a state-space representation and handle missing observations via Kalman filtering and EM-type estimation.Second, when the information set has an inherent multi-way structure, and when data are naturally collected as a matrix—for instance, indicators observed for multiple countries over time—vectorizing the data by stacking countries and variables may obscure economically meaningful interactions across dimensions, inflate the parameter space, and reduce interpretability. This consideration motivates matrix-variate (and, more generally, tensor-variate) factor models, which preserve the data's native structure while delivering parsimonious representations of cross-country and cross-indicator comovement.

Building on this perspective, the work develops and studies two complementary matrix-based approaches. The first is a new matrix-valued extension of the Three-Pass Regression Filter (TPRF) by Kelly and Pruitt (2015)—and, more specifically, of its mixed-frequency variant by Hepenstrick and Marcellino (2016)—which we refer to as the Matrix Mixed-Frequency Three-Pass Regression Filter (Matrix MF–TPRF). The second is a matrix-valued counterpart of the mixed-frequency Dynamic Factor Model in the spirit of Bańbura and Modugno (2010), yielding a Dynamic Matrix Factor Model (DMFM) in which latent factors evolve over time. The two frameworks differ in estimation philosophy but are aligned in scope: Matrix MF–TPRF extracts predictive components through three OLS-based steps driven by external proxies, whereas DMFM is estimated by quasi-maximum likelihood via an EM algorithm with Kalman smoothing, following Matteo Barigozzi and Trapin (2025). Crucially, both are designed to operate under ragged-edge data releases and general missingness: in Matrix MF–TPRF missing observations are handled through an imputation-and-initialization scheme, while in DMFM the Kalman smoother embedded in EM delivers coherent latent-state estimates even when the panel is incomplete.

Moreover, beyond providing these layers of flexibility the proposed models and estimation strategies naturally support a pseudo real-time forecasting exercise for EA country-specific GDP growth. In a matrix setting, nowcasting can explicitly exploit information shared across countries and across groups of indicators, allowing the common component to reflect both domestic and cross-country dynamics within the main EA economies, namely Germany, France, Italy, and Spain. The empirical analysis systematically compares, for each country, the nowcasts produced by the matrix-based models with those obtained from their standard vector-based counterparts, namely MF–TPRF and MF–DFM. The results suggest that preserving the matrix structure can improve real-time performance, both in terms of accuracy and in the ability to react to new information releases, including during periods of instability such as the COVID-19 shock and the subsequent recovery.

The remainder of this introduction motivates the use of matrix-valued representations for euro-area macroeconomic panels, provides a brief overview of the related methodological literature, and outlines

Table 1: Illustration of a matrix-valued time series

Country	GDP Growth	Unemployment Rate	...	Industrial Production
Germany	$X_{t,11}$	$X_{t,12}$	$\dots$	$X_{t,1p_2}$
France	$X_{t,21}$	$X_{t,22}$	$\dots$	$X_{t,2p_2}$
Italy	$X_{t,31}$	$X_{t,32}$	$\dots$	$X_{t,3p_2}$
Spain	$X_{t,41}$	$X_{t,42}$	$\dots$	$X_{t,4p_2}$

the two modelling tools studied in this thesis—Matrix MF–TPRF and DMFM—together with the pseudo real-time nowcasting design used in the empirical evaluation.

Modelling Matrix-Valued Time Series. Many datasets in macroeconomics and finance are naturally organized as sequences of two-dimensional panels evolving over time. Recent empirical work has exploited matrix-valued time series to study, for instance[2], country-by-indicator macroeconomic databases Cen and Lam (2025), portfolio-by-asset return panels L. Yu et al. (2022), and bilateral flow matrices (E. Y. Chen and R. Chen 2019). Treating each observation as a natural matrix, rather than vectorizing it, preserves the underlying two-way structure and can improve both interpretability and parsimony. This perspective has spurred a growing literature on matrix (and tensor) factor models. The foundational contribution is the Matrix Factor Model of D. Wang, Liu, and R. Chen (2019), which sparked a stream of developments spanning matrix autoregressive dynamics (e.g., R. Chen, Xiao, and Yang (2021), Billio et al. (2023)), matrix panel regression frameworks (e.g., Kapetanios, Serlenga, and Shin (2021)), and matrix factor-augmented regression models (e.g., Cai et al. (2025), B. Chen, Han, and Q. Yu (2025)). Parallel work has refined the estimation and identification of matrix factor structures (e.g., L. Yu et al. (2022)), and recent contributions have pushed the frontier toward richer specifications, including time-varying parameters (e.g., B. Chen, E. Y. Chen, et al. (2024)), tensor extensions of principal components (e.g., Babii, Ghysels, and Pan (2022)), and state-space formulations that explicitly incorporate dynamics through Dynamic Matrix Factor Models, estimated, for instance, via quasi-maximum likelihood and EM-type procedures (e.g., Matteo Barigozzi and Trapin (2025)). In this work, the matrix structure is particularly appealing in the EA context, where countries and groups of indicators represent distinct yet economically connected layers of comovement.

To illustrate the data structure, consider the monthly observation $\mathbf{X}_t \in \mathbb{R}^{p_1 \times p_2}$ in Table 1, with countries

[2] TO REVIEW: not the first empirical examples for such exercises... what could we cite here?

Table 2: Key methodological references

Component	Vector models		Matrix models	
	MF–TPRF	**MF–DFM**	**Matrix MF–TPRF**	**DMFM**
Variable sel.	[a]	[a]	[a]	[a]
Factor-number sel.	[b]	[b]	[c]	[c]
Estimation ref.	Hepenstrick and Marcellino (2016)	Bańbura and Modugno (2014)	*This thesis*	Matteo Barigozzi and Trapin (2025)

Notes: [a] Variable selection follows Bai and Ng (2008) and De Mol, Giannone, and Reichlin (2008) (common to all models). [b] Vector factor-number selection uses Xiong and Pelger (2023) and Ahn and Horenstein (2013). [c] Matrix factor-number selection uses Cen and Lam (2025) and L. Yu et al. (2022).

in rows and indicators in columns. Vectorization is always feasible, but it may obscure interactions across the two dimensions and inflate the parameter space; by contrast, matrix-based approaches exploit the two-way geometry while allowing each series to feature its own missing-data pattern.

Building on this literature, we consider two complementary approaches tailored to mixed-frequency and ragged-edge environments. The first is the *Matrix MF–TPRF*, a matrix-valued mixed-frequency extension of the TPRF framework of Kelly and Pruitt (2015) and Hepenstrick and Marcellino (2016), which extracts *targeted* predictive factors for quarterly GDP by leveraging a low-dimensional proxy span while retaining the two-way structure of $\mathbf{X}_t$. The second is a *Dynamic Matrix Factor Model (DMFM)*, which generalizes vector DFMs, as proposed by Bańbura and Modugno (2010), by modelling $\mathbf{X}_t$ through row/column loading spaces and a dynamically evolving latent factor matrix, estimated by QML via an EM algorithm with Kalman smoothing (Matteo Barigozzi and Trapin 2025). While Matrix MF–TPRF is regression-based and explicitly targets forecast-relevant variation, DMFM is state-space based and delivers internally consistent latent-state estimates under general missing-data patterns. For ease of reference, Table 2 maps the four models considered in this thesis to their core building blocks and the corresponding methodological references.

Empirics. Beyond the methodological innovations proposed in this work, we develop a pseudo real-time nowcasting application to assess the forecasting performance of each model and to compare matrix- and vector-based formulations. We use the EA–MD–QD dataset of Matteo Barigozzi, Lippi, and Luciani (2021), which provides a long monthly/quarterly euro-area macroeconomic panel (January 2000–October 2025). The empirical focus is on Germany, France, Italy, and Spain, and the target variable is country-specific quarterly GDP growth.

The empirical design is built to be both comparable across models and flexible along several dimensions.

First, we consider alternative information sets (small, medium, and large) composed of quarterly GDP and a set of monthly and quarterly predictors selected consistently across specifications. In particular, variable selection follows the targeted-screening approach in Bai and Ng (2008) and Giannone, Reichlin, and Small (2008), which helps retain indicators that are most informative for factor extraction and forecasting. To ensure consistency across vector and matrix formulations, we adopt Xiong and Pelger (2023) and Ahn and Horenstein (2013) for factor-number selection and missing-value initialization in the vector case, and their matrix counterparts in Cen and Lam (2025) and L. Yu et al. (2022) (see Table 2). Second, we estimate and evaluate four competing models—two vector benchmarks (MF–TPRF and MF–DFM) and their matrix-based counterparts (Matrix MF–TPRF and DMFM)—to isolate the role of preserving the two-way (country $\times$ indicator) structure. Third, we implement a pseudo real-time vintage construction: for each evaluation date, the ragged-edge pattern implied by the release calendar is replicated via a masking procedure that removes not-yet-released observations, and nowcasts are updated as new monthly information becomes available within the quarter. The evaluation sample spans 2017Q1–2025Q3. At each vintage, models are re-estimated recursively. The number of factors (and, where relevant, the lag order) is selected within an estimation window, so that the forecasting exercise remains data-driven and comparable across approaches. This setup directly addresses the main empirical objective of the thesis: assessing whether matrix-based factor methods that exploit cross-country information can improve country-level GDP nowcasting relative to standard vector factor models, both in terms of accuracy and real-time responsiveness.

The empirical analysis is organized around two research questions. (*i*) Does preserving the matrix structure improve country-level nowcast accuracy relative to vector models, and which of the competing approaches performs best across countries and information sets? (*ii*) To what extent do cross-country information and high-frequency releases contribute to real-time GDP tracking, and how do these contributions vary across countries and over time?

Overall, the results indicate that matrix-based specifications (Matrix MF–TPRF and DMFM) can deliver more accurate and more timely nowcasts than their vector counterparts (MF–TPRF and MF–DFM), with gains that are more pronounced for countries whose business-cycle dynamics are strongly tied to euro-area comovement. At the same time, vector TPRFs and DFMs remain a competitive and computationally lighter benchmark.[3]

[3] TO REVIEW: Consideration just based on thesis results

The main contribution of this thesis is to provide a systematic empirical assessment of recent matrix-based factor methods in a pseudo real-time nowcasting environment by: (*i*) extending proxy-based targeted factor extraction to matrix panels (Matrix MF–TPRF); and (*ii*) delivering, to the best of our knowledge, the first structured comparison of vector versus matrix factor models for euro-area GDP nowcasting under realistic ragged-edge data releases.

Notation. Notation is introduced locally when first used and is kept consistent across chapters. For ease of reference, Appendix A collects the common notation, as well as model-specific notation for the Kalman filter/smoother and for each factor-model specification.

Reminder of the Paper. The remainder of this thesis is structured as follows. For each modelling framework, we present both the standard vector specification and its matrix-based counterpart. Chapter 2 introduces the (MF–)TPRF-type models and the Chapter 3 provides the theory of Matrix MF–TPRF. Chapter 4 reviews the MF–DFM, while Chapter 5 presents the DMFM. Each methodological chapter includes a dedicated subsection on missing-data handling. Chapter 6 introduces the dataset and the pseudo real-time design, evaluates and compares nowcasting performance, and discusses the main empirical findings. A full replication package is provided in `R` at `https://github.com/dolpolo/Nowcasting_EA_DMFM`. Finally, Chapter . . . summarizes the key results and outlines avenues for further research.

2 Mixed-Frequency Three-Pass Regression Filter (MF–TPRF)

The Three-Pass Regression Filter (TPRF), introduced by Kelly and Pruitt (2015), is an OLS-based approach designed to extract a small number of *targeted factors* for forecasting a scalar variable from a high-dimensional predictor panel. The distinctive feature of TPRF is that factor extraction is guided by a low-dimensional proxy vector, which anchors the procedure to the forecast-relevant component of the data. This proxy-driven targeting helps mitigate a common empirical issue in large panels: strong comovements that are statistically dominant may be weakly related to the forecasting objective, so that standard unsupervised methods (e.g. principal components) can capture variation that is largely irrelevant for predicting the target. Because it relies exclusively on linear regressions, TPRF is computationally light, easy to implement, and particularly attractive in applications where the dimensionality of predictors is large.

A limitation of the standard TPRF setup is that it is formulated for single-frequency, balanced panels. In macroeconomic nowcasting, however, datasets are typically mixed-frequency and characterized by *ragged-edge* patterns, since monthly indicators are released asynchronously and the most recent observations are often missing at the end of the sample. To accommodate these features, Hepenstrick and Marcellino (2016) extend the original TPRF to a mixed-frequency environment (MF–TPRF), allowing a quarterly target to be nowcast using a monthly predictor panel while preserving the key proxy-based logic of the three-pass construction.

2.1 Standard Three-Pass Regression Filter (TPRF)

The Three-Pass Regression Filter (TPRF), introduced by Kelly and Pruitt (2015), constructs *targeted factors* that span the forecast-relevant component of a large predictor panel using only OLS regressions. The key insight is that, in finite samples, strong but forecast-irrelevant common movements may contaminate standard factor extraction (e.g. PCA) and weaken forecasting performance. By exploiting a low-dimensional proxy vector $\mathbf{z}_t$, TPRF targets the predictive subspace directly, providing a simple alternative to ad-hoc variable pre-selection or screening procedures (e.g. Bai and Ng (2002)).

2.1.1 The Model

For $t = 1, \ldots, T$, consider the system

$$y_{t+1} = \beta_0 + \boldsymbol{\beta}'\mathbf{F}_t + \eta_{t+1}, \qquad y_{t+1} \in \mathbb{R},\ \mathbf{F}_t \in \mathbb{R}^r,\ \boldsymbol{\beta} \in \mathbb{R}^r, \tag{1}$$

$$\mathbf{z}_t = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\omega}_t, \qquad \mathbf{z}_t \in \mathbb{R}^L,\ \boldsymbol{\Lambda} \in \mathbb{R}^{L\times r}, \tag{2}$$

$$\mathbf{x}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\mathbf{F}_t + \boldsymbol{\xi}_t, \qquad \mathbf{x}_t \in \mathbb{R}^N,\ \boldsymbol{\Phi} \in \mathbb{R}^{N\times r}. \tag{3}$$

Here η_{t+1}, $\boldsymbol{\omega}_t$, and $\boldsymbol{\xi}_t$ are mean-zero disturbances. The system is composed of three equations:

- *Target (forecasting) equation* (1). The scalar target y_{t+1} is driven by a low-dimensional set of latent forces.

  $$\mathbf{F}_t = \begin{pmatrix} \mathbf{f}_t \\ \mathbf{g}_t \end{pmatrix}, \qquad \mathbf{f}_t \in \mathbb{R}^{r_f},\ \mathbf{g}_t \in \mathbb{R}^{r_g},\ r = r_f + r_g,$$

  where $\mathbf{f}_t$ collects the *predictive* factors and $\mathbf{g}_t$ the *non-predictive* ones. The defining targeting restriction is that the target depends only on the predictive component, i.e.

  $$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_f \\ \mathbf{0} \end{pmatrix}, \qquad \boldsymbol{\beta}_f \in \mathbb{R}^{r_f},\ \mathbf{0} \in \mathbb{R}^{r_g}.$$

  Hence only $\mathbf{f}_t$ enters the forecasting relationship, while $\mathbf{g}_t$ is forecast-irrelevant for y_{t+1}, and (1) can be equivalently written as

  $$y_{t+1} = \beta_0 + \boldsymbol{\beta}_f'\mathbf{f}_t + \eta_{t+1}.$$

- *Proxy equation* (2). The proxy vector $\mathbf{z}_t$ is low-dimensional and co-moves with the predictive component of the latent space. In practice, $\mathbf{z}_t$ provides an observable anchor that helps isolate the forecast-relevant subspace from a large predictor panel. The matrix $\boldsymbol{\Lambda}$ collects proxy loadings on the latent factors (so that $\mathbf{z}_t$ is informative about $\mathbf{f}_t$ up to rotation).

  $$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_f & \mathbf{0} \end{pmatrix}, \qquad \boldsymbol{\Lambda}_f \in \mathbb{R}^{L\times r_f},\ \mathbf{0} \in \mathbb{R}^{L\times r_g}.$$

- *Predictor (panel) equation* (3). The high-dimensional predictor vector $\mathbf{x}_t$ is driven by both predictive and non-predictive common components, plus idiosyncratic noise. In empirical applications, predictors are typically transformed so that they can be treated as weakly stationary.

Partial Least Squares (PLS) is a special case of TPRF obtained when predictors are standardized and centered, implying that the first two regression passes are run without intercepts, and the proxies are selected through this automated procedure. For the estimation we consider a set of not centered and standardized target and predictors.

2.1.2 Estimation

TPRF estimates the forecast-relevant component of the latent space by using the proxy vector $\mathbf{z}_t \in \mathbb{R}^L$ to *target* the predictive variation in a large panel $\mathbf{x}_t \in \mathbb{R}^N$, via three OLS passes. In applications we set the proxy dimension equal to L, so that the extracted targeted factors satisfy $\widehat{\mathbf{f}}_t \in \mathbb{R}^L$.

- **Pass 1 (targeting): time-series regressions.**

  For each $i = 1, \dots, N$ and $t = 1, \dots, T$, run

$$x_{i,t} = \phi_{0,i} + \boldsymbol{\phi}_i' \mathbf{z}_t + \xi_{i,t}, \qquad \mathbf{z}_t \in \mathbb{R}^L,\ \boldsymbol{\phi}_i \in \mathbb{R}^L \tag{4}$$

  Stacking over t gives $\mathbf{x}_{(i)} = (x_{i,1}, \dots, x_{i,T})' \in \mathbb{R}^T$ and $Z = (\mathbf{z}_1, \dots, \mathbf{z}_T)' \in \mathbb{R}^{T \times L}$. Let $\boldsymbol{\xi}_{(i)} = (\xi_{i,1}, \dots, \xi_{i,T})' \in \mathbb{R}^T$. Then

$$\mathbf{x}_{(i)} = \phi_{0,i} \mathbf{1}_T + Z \boldsymbol{\phi}_i + \boldsymbol{\xi}_{(i)}.$$

  Equivalently, stacking across i define $X = \left[\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(N)}\right] \in \mathbb{R}^{T \times N}$ and $\Xi = \left[\boldsymbol{\xi}_{(1)}, \dots, \boldsymbol{\xi}_{(N)}\right] \in \mathbb{R}^{T \times N}$. Collect intercepts and slope coefficients as $\boldsymbol{\phi}_0 = (\phi_{0,1}, \dots, \phi_{0,N})' \in \mathbb{R}^N$ and $\Phi = \left[\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N\right]' \in \mathbb{R}^{N \times L}$. Hence the multivariate regression can be written as

$$X = \mathbf{1}_T \boldsymbol{\phi}_0' + Z\Phi' + \Xi, \qquad \mathbf{1}_T \boldsymbol{\phi}_0' \in \mathbb{R}^{T \times N},\ Z\Phi' \in \mathbb{R}^{T \times N}.$$

  To include the intercept in a single regressor matrix, define $\widetilde{Z} = [\mathbf{1}_T \;\; Z] \in \mathbb{R}^{T \times (L+1)}$ and $\widetilde{\Phi}' = [\boldsymbol{\phi}_0' \;\; \Phi'] \in \mathbb{R}^{(L+1) \times N}$. Then

$$X = \widetilde{Z}\widetilde{\Phi}' + \Xi$$

and the multivariate OLS regression yields

$$\widehat{\widetilde{\Phi}}' = (\widetilde{Z}'\widetilde{Z})^{-1}\widetilde{Z}'X \in \mathbb{R}^{(L+1)\times N}.$$

The first row of $\widehat{\widetilde{\Phi}}'$ equals $\widehat{\boldsymbol{\phi}}_0' \in \mathbb{R}^{1\times N}$, while the remaining rows form $\widehat{\Phi}' \in \mathbb{R}^{L\times N}$.

- **Pass 2 (factor extraction): cross-sectional regressions.**

  Given the first-pass estimates $\widehat{\Phi} \in \mathbb{R}^{N\times L}$, extract the targeted factors $\mathbf{f}_t \in \mathbb{R}^L$ by cross-sectional OLS. For each $t = 1,\ldots,T$ and $i = 1,\ldots,N$, consider

$$x_{i,t} = \phi_{0,t} + \widehat{\boldsymbol{\phi}}_i'\mathbf{f}_t + \xi_{i,t}, \qquad \widehat{\boldsymbol{\phi}}_i \in \mathbb{R}^L,\ \mathbf{f}_t \in \mathbb{R}^L. \tag{5}$$

  Stacking over i gives $\mathbf{x}_{(t)} = (x_{1,t},\ldots,x_{N,t})' \in \mathbb{R}^N$ and $\boldsymbol{\xi}_{(t)} = (\xi_{1,t},\ldots,\xi_{N,t})' \in \mathbb{R}^N$, so that

$$\mathbf{x}_{(t)} = \phi_{0,t}\mathbf{1}_N + \widehat{\Phi}\mathbf{f}_t + \boldsymbol{\xi}_{(t)}.$$

  Equivalently, stacking across t define $X = (\mathbf{x}_{(1)},\ldots,\mathbf{x}_{(T)})' \in \mathbb{R}^{T\times N}$ and $\Xi = (\boldsymbol{\xi}_{(1)},\ldots,\boldsymbol{\xi}_{(T)})' \in \mathbb{R}^{T\times N}$. Collect the time-varying intercepts as $\boldsymbol{\phi}_0 = (\phi_{0,1},\ldots,\phi_{0,T})' \in \mathbb{R}^T$ and stack the factors as $F = (\mathbf{f}_1,\ldots,\mathbf{f}_T)' \in \mathbb{R}^{T\times L}$, where $\mathbf{f}_t \in \mathbb{R}^L$ (hence the targeted factor dimension equals the proxy dimension L). Then

$$X = \boldsymbol{\phi}_0\mathbf{1}_N' + F\widehat{\Phi}' + \Xi, \qquad \boldsymbol{\phi}_0\mathbf{1}_N' \in \mathbb{R}^{T\times N},\ \ F\widehat{\Phi}' \in \mathbb{R}^{T\times N}.$$

  To include the intercept in a single regressor matrix, define

$$\widetilde{F} = [\boldsymbol{\phi}_0\ \ F] \in \mathbb{R}^{T\times(L+1)}, \qquad \widetilde{\Phi} = [\mathbf{1}_N\ \ \widehat{\Phi}] \in \mathbb{R}^{N\times(L+1)}.$$

  Then the multivariate regression can be written as

$$X = \widetilde{F}\widetilde{\Phi}' + \Xi,$$

and the multivariate OLS regression yields

$$\widehat{\widehat{F}} = X\widetilde{\Phi}(\widetilde{\Phi}'\widetilde{\Phi})^{-1} \in \mathbb{R}^{T\times(L+1)}.$$

The first column of $\widehat{\widehat{F}}$ equals $\widehat{\boldsymbol{\phi}}_0 \in \mathbb{R}^T$, while the remaining L columns yield $\widehat{F} \in \mathbb{R}^{T\times L}$ with rows $\widehat{\mathbf{f}}_t' \in \mathbb{R}^{1\times L}$.

- **Pass 3 (forecast regression): time-series regression.**

  Using the targeted factors $\widehat{\mathbf{f}}_t \in \mathbb{R}^L$, estimate the forecasting regression, allowing for an autoregressive term in y,

$$y_t = \beta_0 + \rho\, y_{t-1} + \boldsymbol{\beta}'\widehat{\mathbf{f}}_{t-1} + \eta_t, \qquad t = 2,\ldots,T. \tag{6}$$

  where $y_t \in \mathbb{R}$, $\widehat{\mathbf{f}}_{t-1} \in \mathbb{R}^L$, and $\boldsymbol{\beta} \in \mathbb{R}^L$. Stacking over $t = 2,\ldots,T$ define

$$\mathbf{y} = (y_2,\ldots,y_T)' \in \mathbb{R}^{T-1}, \qquad \mathbf{y}_{-1} = (y_1,\ldots,y_{T-1})' \in \mathbb{R}^{T-1}, \qquad \widehat{F}_{-1} = (\widehat{\mathbf{f}}_1,\ldots,\widehat{\mathbf{f}}_{T-1})' \in \mathbb{R}^{(T-1)\times L}.$$

  Let the augmented regressor matrix be

$$\widetilde{F} = [\mathbf{1}_{T-1}\ \ \mathbf{y}_{-1}\ \ \widehat{F}_{-1}] \in \mathbb{R}^{(T-1)\times(L+2)}, \qquad \widetilde{\boldsymbol{\beta}} = (\beta_0, \rho, \boldsymbol{\beta}')' \in \mathbb{R}^{L+2}.$$

  Then the multivariate OLS regression yields

$$\widehat{\widetilde{\boldsymbol{\beta}}} = (\widetilde{F}'\widetilde{F})^{-1}\widetilde{F}'\mathbf{y} \in \mathbb{R}^{L+2},$$

  whose first entry is $\widehat{\beta}_0$, the second entry is $\widehat{\rho}$, and the remaining L entries yield $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^L$.

Kelly and Pruitt (2015) also discuss how the proxy vector can be chosen. In the simplest case with a single targeted factor, they suggest setting the proxy equal to the target itself, $z_{1,t} = y_t$, referring to this specification as *target–proxy TPRF*. When multiple targeted factors are needed, they propose either theory-driven proxies or a simple automated selection scheme based on iterative residualization. Specifically, denoting the ℓ-th proxy series by $z_{\ell,t}$ for $\ell = 1,\ldots,L$ and collecting them in $\mathbf{z}_t = (z_{1,t},\ldots,z_{L,t})'$, the algorithm sets $z_{1,t} = y_t$, runs TPRF and obtains the fitted values $\widehat{y}_t^{(1)}$ and residuals $e_t^{(1)} = y_t - \widehat{y}_t^{(1)}$; then it sets $z_{2,t} = e_t^{(1)}$, re-runs TPRF using $(z_{1,t}, z_{2,t})$ as proxies and computes $e_t^{(2)} = y_t - \widehat{y}_t^{(2)}$; and it iterates until $z_{L,t} = e_t^{(L-1)}$, finally producing $\widehat{y}_t^{(L)}$ based on $(z_{1,t}, \ldots, z_{L,t})$.

Moreover, following Kelly and Pruitt (2015), we choose the number of predictive indices (3PRF factors) L via an information criterion. This is especially relevant for the *automatic-proxy* 3PRF, where proxies are constructed from the forecast target so that factor extraction depends on $\mathbf{y}$ (and, under centering/standardization and suitable normalizations, the procedure can be cast in a PLS form). This response-driven extraction makes the forecasting operator non-linear in $\mathbf{y}$, so the effective degrees of freedom typically exceed the naive count L from an OLS regression on L fixed regressors. We therefore compute generalized degrees of freedom using the unbiased *Trace of the Krylov Representation* estimator of Krämer and Sugiyama (2011) and plug them into a BIC.

2.1.3 Forecasting

The one-step-ahead forecast is then

$$\widehat{y}_{t+1} = \widehat{\beta}_0 + \widehat{\rho}\, y_t + \widehat{\boldsymbol{\beta}}' \widehat{\mathbf{f}}_t, \qquad t = 1, \ldots, T. \tag{7}$$

2.2 Mixed-Frequency Three-Pass Regression Filter (MF–TPRF)

Macroeconomic datasets typically combine a low-frequency target (e.g. quarterly GDP growth) with a large set of high-frequency indicators released at the monthly frequency, often determining a ragged edge at the end of the sample. The standard TPRF cannot handle this feature so, Hepenstrick and Marcellino (2016) propose a mixed frequency extension of the model. Let $y_\tau \in \mathbb{R}$ denote the quarterly target, let $\mathbf{z}_\tau \in \mathbb{R}^L$ be a low-dimensional vector of proxies observed at the same quarterly frequency (consistently with the idea that the first proxy can be the target itself), and let $\mathbf{x}_t \in \mathbb{R}^N$ be the monthly predictor panel. The mixed-frequency extension preserves a the model in the three-equation factor structure of the standard TPRF in (1)–(3), and adapts the estimation through the regression passes that are directly affected by the frequency mismatch. In particular, Pass 1 is implemented at the quarterly frequency after temporally aggregating the monthly predictors, while Pass 3 maps the within-quarter monthly factor estimates into quarterly forecasts through a linear U–MIDAS regression as in Ghysels, Sinko, and Valkanov 2007. Pass 2, instead, is unchanged and still extracts the targeted factors at the monthly frequency via cross-sectional regressions.

2.2.1 Estimation

The mismatch between target (and proxy), available at each quarter $\tau = 1 \dots \mathscr{T}$, with predictors frequency collected at higher frequency $t = 1 \dots T$, is handled by temporal alignment from high to low frequency. Let $\mathscr{L}$ denote the monthly lag operator. Define the aggregation operator

$$\omega(\mathscr{L}) = \omega_0 + \omega_1 \mathscr{L} + \omega_2 \mathscr{L}^2,$$

which, in our case, maps monthly series into quarterly-aligned observations. The typical choices is

$$\omega(\mathscr{L}) = 1 + \mathscr{L} + \mathscr{L}^2 \quad \text{(flows)}, \qquad \omega(\mathscr{L}) = \tfrac{1}{3}(1 + \mathscr{L} + \mathscr{L}^2) \quad \text{(stocks)}.$$

For each monthly predictor $i = 1, \dots, N$, define the quarterly-aligned series

$$x_{i,\tau}^{(m \to q)} := \omega(\mathscr{L}) x_{i,t}, \qquad \tau = 1, \dots, \mathscr{T},$$

and apply the same transformation to the proxy, if selected among the high frequency indicators,to preserve the structure of the following first-pass regression.

- **Pass 1 (targeting at quarterly frequency).**

  Let $\tau = 1,\dots,\mathscr{T}$ denote the quarter index and $i = 1,\dots,N$ the predictor index. For each i, estimate the low-frequency (quarterly) targeting regression

$$x_{i,\tau}^{(m\to q)} = \alpha_{0,i} + \boldsymbol{\alpha}_i' \mathbf{z}_\tau + u_{i,\tau}, \qquad \mathbf{z}_\tau \in \mathbb{R}^L,\ \boldsymbol{\alpha}_i \in \mathbb{R}^L, \tag{8}$$

  and collect the estimated slope vectors in $\widehat{A} \in \mathbb{R}^{N\times L}$, whose i-th row is $\widehat{\boldsymbol{\alpha}}_i'$. As in Hepenstrick and Marcellino (2016), one may optionally set $\widehat{\boldsymbol{\alpha}}_i = \mathbf{0}$ when a joint test of $H_{0,i} : \boldsymbol{\alpha}_i = \mathbf{0}$ is not rejected, thereby excluding predictor i from the subsequent factor construction. (See Section 2.1 for the corresponding multivariate OLS form.)

- **Pass 2 (factor extraction at monthly frequency).**

  Using the first-pass estimates $\widehat{A}$, recover the targeted factors $\widehat{\mathbf{f}}_t$ from the monthly cross-section as in standard TPRF. For each month $t = 1,\dots,T$, consider the cross-sectional regression over $i = 1,\dots,N$:

$$x_{i,t} = a_{0,t} + \widehat{\boldsymbol{\alpha}}_i' \mathbf{f}_t + \xi_{i,t}, \qquad \widehat{\boldsymbol{\alpha}}_i \in \mathbb{R}^L,\ \mathbf{f}_t \in \mathbb{R}^L. \tag{9}$$

  Hence the targeted factor has dimension L, equal to the proxy dimension. Stacking over t yields

$$\widehat{F} = (\widehat{\mathbf{f}}_1,\dots,\widehat{\mathbf{f}}_T)' \in \mathbb{R}^{T\times L}.$$

- **Pass 3 (quarterly U–MIDAS forecast regression).**

  Let $m \in \{1,2,3\}$ denote the month-within-quarter position. For each quarter τ, define the within-quarter factor vectors $(\widehat{\mathbf{f}}_\tau^{(1)},\widehat{\mathbf{f}}_\tau^{(2)},\widehat{\mathbf{f}}_\tau^{(3)}) \in \mathbb{R}^L$, and stack them as

$$\widehat{\mathbf{f}}_\tau^{Q} = \begin{pmatrix} \widehat{\mathbf{f}}_\tau^{(1)} \\ \widehat{\mathbf{f}}_\tau^{(2)} \\ \widehat{\mathbf{f}}_\tau^{(3)} \end{pmatrix} \in \mathbb{R}^{3L}.$$

  The quarterly forecasting equation for $\tau = 2,\dots,\mathscr{T}$ is the linear U–MIDAS regression estimated by OLS on the quarterly sample

$$y_\tau = \beta_0 + \rho\, y_{\tau-1} + \boldsymbol{\beta}_1' \widehat{\mathbf{f}}_{\tau-1}^{(1)} + \boldsymbol{\beta}_2' \widehat{\mathbf{f}}_{\tau-1}^{(2)} + \boldsymbol{\beta}_3' \widehat{\mathbf{f}}_{\tau-1}^{(3)} + \eta_\tau, \tag{10}$$

where $\boldsymbol{\beta}_m \in \mathbb{R}^L$, $m = 1,2,3$

Lag-Length Selection in the MF–TPRF In Pass 3, the U–MIDAS forecasting regression can include multiple quarterly lags of the within-quarter factor realizations $\widehat{\mathbf{f}}_\tau^{(1)}, \widehat{\mathbf{f}}_\tau^{(2)}, \widehat{\mathbf{f}}_\tau^{(3)} \in \mathbb{R}^L$. We denote by P_F the maximum number of quarterly lags of these factor blocks retained in the regression and select it data-driven via information criteria.

For each $P_F \in \{1, \ldots, P_{F,\max}\}$, estimate

$$y_\tau = \beta_0 + \rho\, y_{\tau-1} + \sum_{p_f=1}^{P_F} \left(\boldsymbol{\beta}_{p_f}^{(1)\prime} \widehat{\mathbf{f}}_{\tau-p_f}^{(1)} + \boldsymbol{\beta}_{p_f}^{(2)\prime} \widehat{\mathbf{f}}_{\tau-p_f}^{(2)} + \boldsymbol{\beta}_{p_f}^{(3)\prime} \widehat{\mathbf{f}}_{\tau-p_f}^{(3)} \right) + \eta_\tau, \tag{11}$$

Let $\mathscr{V}$ denote the number of quarterly observations effectively used in the P_F-lag regression (e.g. $\tau = P_F + 1, \ldots, \mathscr{T}$ so that $\mathscr{V} = \mathscr{T} - P_F$), and let $k(P_F)$ be the total number of estimated parameters under lag order P_F. With an intercept, an AR(1) term, and P_F lags of the three L-dimensional factor blocks, $k(P_F) = 2 + 3LP_F$. Define

$$\mathrm{AIC}(P_F) = \log\big(\widehat{\sigma}^2(P_F)\big) + \frac{2\,k(P_F)}{\mathscr{V}}, \qquad \mathrm{BIC}(P_F) = \log\big(\widehat{\sigma}^2(P_F)\big) + \frac{k(P_F)\log \mathscr{V}}{\mathscr{V}},$$

where $\widehat{\sigma}^2(P_F)$ is the estimated residual variance from the P_F-lag regression.

The selected lag order is

$$\widehat{P}_{F,\mathrm{AIC}} = \arg \min_{1 \le P_F \le P_{F,\max}} \mathrm{AIC}(P_F), \qquad \widehat{P}_{F,\mathrm{BIC}} = \arg \min_{1 \le P_F \le P_{F,\max}} \mathrm{BIC}(P_F).$$

The selected $\widehat{P}_F$ determines the dynamic contribution of lagged quarterly factor blocks, conditional on the included AR(1) term in y_τ.

2.2.2 EM Algorithm for Missing Values

While the target y_τ and the proxy vector $\mathbf{z}_\tau$ are observed at quarterly frequency, the predictor panel includes both monthly and quarterly indicators. Together with asynchronous releases, this generates an unbalanced panel with ragged edges. Following Hepenstrick and Marcellino (2016), we complete the high-frequency panel by applying an EM algorithm to the *centered and standardized* predictors.[4]

For each predictor $i = 1,\dots,N$, let $\mathbf{x}_{(i)} := (x_{i,1},\dots,x_{i,T})' \in \mathbb{R}^T$ denote the monthly-calendar series and let $\mathscr{O}_i \subseteq \{1,\dots,T\}$ be the set of observed months. Define the selection matrix $A_i \in \{0,1\}^{|\mathscr{O}_i|\times T}$ such that

$$\mathbf{x}_{\text{obs},(i)} := (x_{i,t})_{t\in\mathscr{O}_i} \in \mathbb{R}^{|\mathscr{O}_i|}, \qquad \mathbf{x}_{\text{obs},(i)} = A_i\,\mathbf{x}_{(i)}.$$

Lower-frequency predictors correspond to $\mathbf{A}i$ selecting only quarter-end months (possibly with additional missingness). We initialize a balanced standardized panel $\widehat{X}^{\text{std},(0)} \in \mathbb{R}^{T\times N}$ by filling the missing entries of each standardized series $\mathbf{x}^{\text{std}}(i)$ using the procedure of Xiong and Pelger (2023): we reconstruct a consistent estimator of the covariance matrix under missingness, compute the factors, and replace missing observations with their estimated common component.

For EM iteration $n \geq 1$, let $\widehat{\mathbf{F}}^{(n-1)} \in \mathbb{R}^{T\times r}$ and $\widehat{\boldsymbol{\lambda}}_i^{(n-1)} \in \mathbb{R}^r$ denote the factor and loading estimates obtained by PCA on $\widehat{X}^{\text{std},(n-1)}$, where r is the number of principal components retained.

In the E-step, the completed standardized series for predictor i is updated as

$$\widehat{\mathbf{x}}_{(i)}^{\text{std},(n)} = \widehat{\mathbf{F}}^{(n-1)}\,\widehat{\boldsymbol{\lambda}}_i^{(n-1)} + A_i'(A_iA_i')^{-1}\left(\mathbf{x}^{\text{std}}_{\text{obs},(i)} - A_i\,\widehat{\mathbf{F}}^{(n-1)}\,\widehat{\boldsymbol{\lambda}}_i^{(n-1)}\right), \tag{12}$$

which enforces exact agreement on observed entries, $A_i\,\widehat{\mathbf{x}}_{(i)}^{\text{std},(n)} = \mathbf{x}^{\text{std}}_{\text{obs},(i)}$. Collecting $\{\widehat{\mathbf{x}}_{(i)}^{\text{std},(n)}\}_{i=1}^N$ yields the updated balanced panel $\widehat{X}^{\text{std},(n)}$. In the M-step, factors and loadings are re-estimated by PCA on $\widehat{X}^{\text{std},(n)}$, and the procedure iterates until convergence.

After convergence, the completed panel is transformed back to the original scale by re-centering and de-standardizing each series entrywise.

[4] TO REVISE...

2.2.3 Nowcasting

In the pseudo real-time exercise, each monthly vintage contains only observations released up to that date, so the information set expands within the quarter as new releases arrive. The model is re-estimated at each vintage using an expanding sample, while the factor dimension L and the lag order of the U–MIDAS regression are selected once on an initial training window.

Let $\widehat{\mathbf{f}}_\tau^{(m)} \in \mathbb{R}^L$ denote the vector of targeted factors available at the m-th month of quarter τ, for $m = 1,2,3$, and let $(\widehat{\beta}_0, \widehat{\rho}, \widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\boldsymbol{\beta}}_3)$ be the OLS estimates from the quarterly U–MIDAS regression in Pass 3. The within-quarter nowcast of y_τ is then updated sequentially as

$$
\begin{aligned}
\widehat{y}_\tau^{(1)} &= \widehat{\beta}_0 + \widehat{\rho}\, y_{\tau-1} + \widehat{\boldsymbol{\beta}}_1' \widehat{\mathbf{f}}_\tau^{(1)}, \\
\widehat{y}_\tau^{(2)} &= \widehat{\beta}_0 + \widehat{\rho}\, y_{\tau-1} + \widehat{\boldsymbol{\beta}}_1' \widehat{\mathbf{f}}_\tau^{(1)} + \widehat{\boldsymbol{\beta}}_2' \widehat{\mathbf{f}}_\tau^{(2)}, \\
\widehat{y}_\tau^{(3)} &= \widehat{\beta}_0 + \widehat{\rho}\, y_{\tau-1} + \widehat{\boldsymbol{\beta}}_1' \widehat{\mathbf{f}}_\tau^{(1)} + \widehat{\boldsymbol{\beta}}_2' \widehat{\mathbf{f}}_\tau^{(2)} + \widehat{\boldsymbol{\beta}}_3' \widehat{\mathbf{f}}_\tau^{(3)}.
\end{aligned}
$$

These correspond to the nowcast after the first, second, and third month of quarter τ, respectively; the autoregressive term $\widehat{\rho}\, y_{\tau-1}$ enters each update unchanged, while the set of available within-quarter factor realizations expands with m.

3 Matrix Mixed-Frequency Three Pass Regression Filter (Matrix MF–TPRF)

We propose a matrix-valued extension of the mixed-frequency Three-Pass Regression Filter (TPRF) of Hepenstrick and Marcellino (2016). The setting is a high frequency (monthly) predictor panel naturally arranged as matrices, $\mathbf{X}_t \in \mathbb{R}^{p_1 \times p_2}$ (e.g. countries $\times$ indicators), and a low frequency (quarterly) target y_τ (e.g. GDP growth). The goal is to extract a small set of latent factors that are *predictive* for the target by exploiting a low-dimensional proxy vector $\mathbf{z}_\tau \in \mathbb{R}^L$, while preserving the two-way geometry of the predictors instead of collapsing them into a single vector breaking the bilinearity of data. In this setting mixed-frequency alignment and real-time releases in the predictors' panel induce an arbitrary missing-data pattern. Therefore, before estimating the model, we need to complete the high frequency panel by treating missing entries as unobserved and imputing them using the usual EM algorithm. Through factors extracted by PCA from a vectorized dataset missing observations are reconstructed and the completed vectors are then reshaped back into matrices.

The estimation proceeds in three passes as in the case of Hepenstrick and Marcellino (2016). Step 1 works at the quarterly frequency and we target the low frequency proxy span by forming proxy-weighted moment matrices $\mathbf{M}^{(\ell)}$ for every ℓ proxy[5]. This procedure allows to construct whitened row and column moment matrices $(\mathbf{S}_{\text{row}}, \mathbf{S}_{\text{col}})$, whose leading eigenvectors define the loading spaces $(\widehat{\mathbf{R}}, \widehat{\mathbf{C}})$ with possibly different dimensions (r_1, r_2). At the monthly frequency, in the Step 2, we recover factor estimates $\widehat{\mathbf{F}}_t$ by bilinear projection. Finally Step 3, working on a vectorized data as in the vector MF-TPRF case, links monthly factors (embadding the bilinearity) to the quarterly target through a U–MIDAS regression that stacks within-quarter blocks (month 1, 2, 3) and a finite number of quarterly lags. This yields a natural pseudo real-time nowcasting scheme: as quarter $\tau + 1$ unfolds, the nowcast of $y_{\tau+1}$ is updated by progressively augmenting the regressor vector with newly released monthly factor blocks.

3.1 The Model

Let $y_\tau \in \mathbb{R}$ and $\mathbf{z}_\tau \in \mathbb{R}^L$ be the target and a low-dimensional proxy vector observed at the quarterly frequency for each $\tau = 1, \ldots, \mathscr{T}$.[6] Let $\mathbf{X}_t \in \mathbb{R}^{p_1 \times p_2}$ denote the high-frequency predictor matrix observed

[5] TO REVISE: How to select the number of proxies for rows and columns

[6] TO REVISE: do we need to restate the target/proxy equations in matrix form, or is it sufficient to modify only the predictor

at time $t = 1, \ldots, T$ (excluding the target), and let $\mathbf{x}_t = \text{vec}(\mathbf{X}_t) \in \mathbb{R}^{p_1 p_2}$ denote its vectorization. The predictor panel is driven by latent forces that can be partitioned into *predictive* and *non-predictive* components.In the matrix setting, we introduce two matrix-valued factors: a predictive factor $\mathbf{F}_t \in \mathbb{R}^{r_1 \times r_2}$ and a non-predictive factor $\mathbf{G}_t \in \mathbb{R}^{v_1 \times v_2}$. Define the corresponding vector factors $\mathbf{f}_t := \text{vec}(\mathbf{F}_t) \in \mathbb{R}^{r_f}$ and $\mathbf{g}_t := \text{vec}(\mathbf{G}_t) \in \mathbb{R}^{r_g}$, where

$$r_f = r_1 r_2, \qquad r_g = v_1 v_2, \qquad r = r_f + r_g.$$

Hence, r_1 and r_2 (respectively v_1 and v_2) are the row/column ranks of the predictive (respectively non-predictive) common component, while r_f (respectively r_g) is the dimension of the corresponding vectorized factor.

The target and proxy equations are as in (1)–(2) (written in terms of the vectorized factors $\mathbf{h}_t := \text{vec}(\mathbf{H}_t) \in \mathbb{R}^{r_f r_g}$). The high-frequency predictors follow the matrix decomposition

$$\mathbf{X}_t = \mathbf{\Phi}_0 + \mathbf{R}\mathbf{F}_t\mathbf{C}' + \mathbf{R}^{(g)}\mathbf{G}_t\mathbf{C}^{(g)\prime} + \mathbf{\Xi}_t, \tag{13}$$

where $\mathbf{\Phi}_0 \in \mathbb{R}^{p_1 \times p_2}$ is an intercept matrix [7] and $\mathbf{\Xi}_t \in \mathbb{R}^{p_1 \times p_2}$ is an idiosyncratic error matrix. The matrices $\mathbf{R} \in \mathbb{R}^{p_1 \times r_1}$ and $\mathbf{C} \in \mathbb{R}^{p_2 \times r_2}$ load the *predictive* factor $\mathbf{F}_t$, while $\mathbf{R}^{(g)} \in \mathbb{R}^{p_1 \times v_1}$ and $\mathbf{C}^{(g)} \in \mathbb{R}^{p_2 \times v_2}$ load the *non-predictive* factor $\mathbf{G}_t$. Accordingly, $\mathbf{X}_t$ admits an additive decomposition into a predictive common component $\boldsymbol{\chi}_t^{(f)} = \mathbf{R}\mathbf{F}_t\mathbf{C}'$, a non-predictive common component $\boldsymbol{\chi}_t^{(g)} = \mathbf{R}^{(g)}\mathbf{G}_t\mathbf{C}^{(g)\prime}$, and the idiosyncratic part $\mathbf{\Xi}_t$, so that pervasive but forecast-irrelevant comovements are prevented from contaminating the forecasting factors.[8]

equation?

[7]TO REVISE: Data are centered and standardized so intercept can be omitted?

[8]Equivalently, (13) can be written compactly as

$$\mathbf{X}_t = \mathbf{\Phi}_0 + \mathbf{R}^* \mathbf{H}_t \mathbf{C}^{*\prime} + \mathbf{\Xi}_t,$$

where $\mathbf{R}^* = [\mathbf{R} \;\; \mathbf{R}^{(g)}] \in \mathbb{R}^{p_1 \times (r_1 + v_1)}$ and $\mathbf{C}^* = [\mathbf{C} \;\; \mathbf{C}^{(g)}] \in \mathbb{R}^{p_2 \times (r_2 + v_2)}$, and the block-diagonal factor matrix is

$$\mathbf{H}_t = \begin{pmatrix} \mathbf{F}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_t \end{pmatrix} \in \mathbb{R}^{(r_1+v_1)\times(r_2+v_2)}.$$

3.2 Estimation

Estimating the predictive factors $\mathbf{F}_t$ follows the three-pass regression filter logic of Kelly and Pruitt (2015) and, more specifically, its mixed-frequency implementation in Hepenstrick and Marcellino (2016). The key idea is to use a low-dimensional proxy vector to *target* the forecast-relevant component of the predictor panel, thereby preventing strong but forecast-irrelevant common movements from contaminating the factors used for prediction.

To begin, consider a centerd and standardized low-frequency target and a low-frequency proxy, together with a balanced high-frequency panel of predictors. As in standard TPRF, when predictors are centered and standardized prior to estimation, intercept terms can be omitted in the first two passes, while the forecasting regression in Step 3 retains an intercept.[9]

- **Step 1 (targeting): proxy regressions at low frequency.** Step 1 uses the low-frequency proxy vector $\mathbf{z}_\tau \in \mathbb{R}^L$ to isolate the component of the predictor panel that is most aligned with the proxy span. Stack the proxies as

$$Z = \left(\mathbf{z}_1, \dots, \mathbf{z}_{\mathscr{T}}\right)' \in \mathbb{R}^{\mathscr{T} \times L}.$$

Throughout, predictors and proxies are assumed centered and standardized, so that proxy-weighted sums are proportional to sample cross-covariances.

OLS interpretation. For each entry (i, j) of the predictor matrix $\mathbf{X}_\tau$, consider the multivariate time-series regression

$$X_{\tau,ij} = \mathbf{z}_\tau' \boldsymbol{\theta}_{ij} + u_{\tau,ij}, \qquad \boldsymbol{\theta}_{ij} \in \mathbb{R}^L.$$

This is a standard OLS regression with common regressors $\mathbf{z}_\tau$ and entry-specific coefficients $\boldsymbol{\theta}_{ij}$. The OLS estimator is

$$\widehat{\boldsymbol{\theta}}_{ij} = (Z'Z)^{-1} Z' \mathbf{x}_{ij}, \qquad \mathbf{x}_{ij} = (X_{1,ij}, \dots, X_{\mathscr{T},ij})' \in \mathbb{R}^{\mathscr{T}}.$$

Collecting $\widehat{\boldsymbol{\theta}}_{ij}$ over all (i, j) yields L coefficient matrices $\{\widehat{\mathbf{B}}^{(\ell)}\}_{\ell=1}^L$, where $\left(\widehat{\mathbf{B}}^{(\ell)}\right)_{ij} = \widehat{\theta}_{ij}^{(\ell)}$.

[9] TO REVISE: so it should be a PLS.

Equivalently, the entry-wise regressions can be written compactly as the matrix regression

$$\mathbf{X}_\tau = \sum_{\ell=1}^{L} z_\tau^{(\ell)} \mathbf{B}^{(\ell)} + \mathbf{U}_\tau, \qquad \mathbf{U}_\tau \in \mathbb{R}^{p_1 \times p_2}. \tag{14}$$

Moment form and whitening. Define the proxy-weighted moment matrices

$$\mathbf{M}^{(m)} := \sum_{\tau=1}^{\mathscr{T}} z_\tau^{(m)} \mathbf{X}_\tau \in \mathbb{R}^{p_1 \times p_2}, \qquad m = 1, \ldots, L.$$

Under standardization, $\mathbf{M}^{(m)}$ is proportional (entry-wise) to the sample covariance between proxy $z^{(m)}$ and the corresponding predictor series. When proxies are correlated, OLS corrects for this correlation through $(Z'Z)^{-1}$, which acts as a whitening matrix on the proxy span. In particular, for each ℓ,

$$\widehat{\mathbf{B}}^{(\ell)} = \sum_{m=1}^{L} \left[(Z'Z)^{-1}\right]_{\ell m} \mathbf{M}^{(m)}. \tag{15}$$

From proxy-aligned moments to loading spaces. Rather than working with the L coefficient matrices directly, we summarize the proxy-related information into row- and column-loading spaces. Define the whitened row- and column-moment matrices

$$\mathbf{S}_{\text{row}} := \sum_{m=1}^{L} \sum_{\ell=1}^{L} \left[(Z'Z)^{-1}\right]_{\ell m} \mathbf{M}^{(m)} \mathbf{M}^{(\ell)\prime}, \qquad \mathbf{S}_{\text{row}} \in \mathbb{R}^{p_1 \times p_1}, \tag{16}$$

$$\mathbf{S}_{\text{col}} := \sum_{m=1}^{L} \sum_{\ell=1}^{L} \left[(Z'Z)^{-1}\right]_{\ell m} \mathbf{M}^{(m)\prime} \mathbf{M}^{(\ell)}, \qquad \mathbf{S}_{\text{col}} \in \mathbb{R}^{p_2 \times p_2}. \tag{17}$$

Finally, the Step 1 loading spaces are obtained as the leading eigenvectors

$$\widehat{\mathbf{R}} = \operatorname{eig}_{1:r_1}(\mathbf{S}_{\text{row}}) \in \mathbb{R}^{p_1 \times r_1}, \qquad \widehat{\mathbf{C}} = \operatorname{eig}_{1:r_2}(\mathbf{S}_{\text{col}}) \in \mathbb{R}^{p_2 \times r_2}.$$

When $L = 1$, $(Z'Z)^{-1}$ is scalar and the construction reduces to $\mathbf{S}_{\text{row}} \propto \mathbf{M}\mathbf{M}'$ and $\mathbf{S}_{\text{col}} \propto \mathbf{M}'\mathbf{M}$, with $\mathbf{M} = \sum_{\tau=1}^{\mathscr{T}} z_\tau \mathbf{X}_\tau$.

- **Step 2: Estimation of monthly matrix factors.**

  Given the targeted row and column loading spaces obtained in Step 1 $\widehat{\mathbf{R}}, \widehat{\mathbf{C}}$, Step 2 recovers the monthly factor realizations as least-squares *scores* associated with these spaces by the bilinear projection

$$\mathbf{X}_t = \widehat{\mathbf{R}}\,\mathbf{F}_t\,\widehat{\mathbf{C}}' + \boldsymbol{\Xi}_t, \qquad \widehat{\mathbf{R}} \in \mathbb{R}^{p_1 \times r_1},\ \widehat{\mathbf{C}} \in \mathbb{R}^{p_2 \times r_2},$$

where $\mathbf{F}_t \in \mathbb{R}^{r_1 \times r_2}$ collects the monthly latent factors and $\boldsymbol{\Xi}_t \in \mathbb{R}^{p_1 \times p_2}$ is a residual matrix.

For each t, the factor score is defined as the Frobenius-norm minimizer

$$\widehat{\mathbf{F}}_t = \arg\min_{\mathbf{F} \in \mathbb{R}^{r_1 \times r_2}} \|\mathbf{X}_t - \widehat{\mathbf{R}}\,\mathbf{F}\,\widehat{\mathbf{C}}'\|_F^2,$$

which has the closed-form solution

$$\widehat{\mathbf{F}}_t = (\widehat{\mathbf{R}}'\widehat{\mathbf{R}})^{-1}\widehat{\mathbf{R}}'\mathbf{X}_t\widehat{\mathbf{C}}(\widehat{\mathbf{C}}'\widehat{\mathbf{C}})^{-1}. \tag{18}$$

In particular, since Step 1 delivers orthonormal loading spaces (so that $\widehat{\mathbf{R}}'\widehat{\mathbf{R}} = \mathbf{I}_{r_1}$ and $\widehat{\mathbf{C}}'\widehat{\mathbf{C}} = \mathbf{I}_{r_2}$), then (18) simplifies to $\widehat{\mathbf{F}}_t = \widehat{\mathbf{R}}'\mathbf{X}_t\widehat{\mathbf{C}}$. The estimator in (18) is the least-squares score associated with the loading spaces $(\widehat{\mathbf{R}}, \widehat{\mathbf{C}})$, i.e. the Frobenius-norm projection of $\mathbf{X}_t$ onto the bilinear subspace spanned by $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{C}}'$. Related matrix factor–augmented regression frameworks, as in Cai et al. (2025), study the impact of using such estimated factor scores as generated regressors and show that, under strong-factor and weak-dependence conditions, the second-step regression estimator retains standard asymptotic properties.

Finally, if the non-predictive component is absent and the predictive component is pervasive—so that the signal in $\mathbf{X}_t$ can be viewed as generated by a strong factor structure—the targeting restriction becomes immaterial. In this case, the row and column loading spaces $(\widehat{\mathbf{R}}, \widehat{\mathbf{C}})$ obtained from the proxy-weighted moments are expected to align (up to rotation) with the leading principal subspaces of the unconditional second-moment matrices of $\mathbf{X}_t$. Hence, Step 2 reduces to a standard bilinear PCA-type projection, and the recovered scores $\widehat{\mathbf{F}}_t = \widehat{\mathbf{R}}'\mathbf{X}_t\widehat{\mathbf{C}}$ are closely related to factor estimates obtained by tensor/matrix PCA methods by Babii, Ghysels, and Pan (2022).

- **Step 3: Quarterly U–MIDAS regression.**

In Step 3 we link the within-quarter monthly factors from Step 2 to the quarterly target via a linear U–MIDAS forecasting regression. Let y_τ denote the quarterly target, $\tau = 1, \ldots, \mathscr{T}$. For each quarter τ and month-within-quarter position $m \in \{1,2,3\}$, let $\widehat{\mathbf{F}}_\tau^{(m)} \in \mathbb{R}^{r_1 \times r_2}$ be the estimated factor matrix and define its vectorization

$$\widehat{\mathbf{f}}_\tau^{(m)} \equiv \mathrm{vec}\big(\widehat{\mathbf{F}}_\tau^{(m)}\big) \in \mathbb{R}^{r_f}.$$

Stack the three within-quarter factor realizations as

$$\widehat{\mathbf{f}}_\tau^{Q} = \begin{pmatrix} \widehat{\mathbf{f}}_\tau^{(1)} \\ \widehat{\mathbf{f}}_\tau^{(2)} \\ \widehat{\mathbf{f}}_\tau^{(3)} \end{pmatrix} \in \mathbb{R}^{3r_f}.$$

The forecasting equation is the U–MIDAS regression in (10), allowing for up to P_F quarterly lags of the within-quarter factor blocks:

$$y_\tau = \beta_0 + \rho\, y_{\tau-1} + \sum_{p_f=1}^{P_F} \left(\boldsymbol{\beta}_{p_f}^{(1)\prime} \widehat{\mathbf{f}}_{\tau-p_f}^{(1)} + \boldsymbol{\beta}_{p_f}^{(2)\prime} \widehat{\mathbf{f}}_{\tau-p_f}^{(2)} + \boldsymbol{\beta}_{p_f}^{(3)\prime} \widehat{\mathbf{f}}_{\tau-p_f}^{(3)} \right) + \eta_\tau, \qquad \boldsymbol{\beta}_{p_f}^{(m)} \in \mathbb{R}^{r_f},\ m = 1,2,3,$$

which is estimated by OLS on the quarterly sample. Consistency and asymptotic normality of the resulting OLS estimator are established in B. Chen, Han, and Q. Yu (2025) for the case in which the only non-tensor regressor is the autoregressive term $\rho\, y_{\tau-1}$.

In the Matrix MF–TPRF setting, predictors are multi-dimensional objects and the target of interest may be defined at the level of each cross-sectional unit (e.g. country-level GDP). A natural choice is therefore to select proxies that summarize the *common* forecast-relevant variation across the units under study. In this interpretation, $\mathbf{z}_t$ acts as a low-dimensional benchmark that anchors the three-pass procedure to the shared predictive component of the panel. For instance, when nowcasting country-specific GDP using a panel of macroeconomic indicators observed for multiple countries, say belonging to Euro Area, one can set the proxy equal to an aggregate target such as Euro Area GDP (or an average of the countries' GDP series). The targeted factors extracted from the matrix panel are then used downstream (e.g. in the final quarterly U–MIDAS regression) to produce unit-specific nowcasts, while the proxy only serves to isolate the common predictive subspace.

3.3 EM Algorithm for Missing Values

As discussed in Subsubsection 2.2.2, the predictor panel may feature arbitrary missing entries due to mixed-frequency alignment and asynchronous data releases, generating ragged-edge patterns. While the targeting step of the Matrix MF–TPRF is formulated directly in the matrix domain, the completion of the high-frequency panel is conveniently performed in the vectorized domain via the standard EM algorithm for static factor models (missing-data PCA).

For each $t = 1,\dots,T$, define $\mathbf{x}_t = \mathrm{vec}(\mathbf{X}_t) \in \mathbb{R}^{p_1 p_2}$ and collect observations over time in the stacked panel

$$X_{\mathrm{vec}} = (\mathbf{x}_1,\dots,\mathbf{x}_T)' \in \mathbb{R}^{T\times p_1 p_2}.$$

Vectorization is used *only* to implement the EM completion and does not impose additional structure on the subsequent matrix-factor steps. The EM updates are therefore identical[10] to those in Subsubsection 2.2.2, applied to X_{vec} (after centering and standardizing), and iterated until convergence. After completion, adding back the mean and the standard deviation, each imputed vector is reshaped back to matrix form:

$$\widehat{\mathbf{X}}_t = \mathrm{unvec}(\widehat{\mathbf{x}}_t,\ p_1,\ p_2), \qquad t = 1,\dots,T.$$

3.4 Nowcasting

The pseudo real-time nowcasting exercise follows the MF–TPRF scheme described in Subsection 2.2.3. The only difference is that the within-quarter targeted factors are now estimated from the matrix panel via Steps 1–2 and are therefore obtained as factor matrices $\widehat{\mathbf{F}}_\tau^{(m)} \in \mathbb{R}^{r_1\times r_2}$, for $m \in \{1,2,3\}$. Consistently with the U–MIDAS regression in Step 3, we define the corresponding vector factors $\widehat{\mathbf{f}}_\tau^{(m)} \equiv \mathrm{vec}\big(\widehat{\mathbf{F}}_\tau^{(m)}\big) \in \mathbb{R}^{r_f}$ and apply the same sequential nowcast updates after the first, second, and third month of quarter τ.

[10] TO REVIEW: How many factors could we use? max rows and columns /all purpose estimator

4 Mixed-Frequency Dynamic Factor Model (MF–DFM)

In this section we introduce the mixed-frequency dynamic factor model (MF–DFM). Unlike the TPRF class, DFM summarizes a large macroeconomic panel through a small set of latent *dynamic* factors embedded in a linear–Gaussian state-space system. We adopt the estimation strategy of Bańbura and Modugno (2010), which accommodates mixed frequencies, publication delays, and ragged-edge panels in a unified framework. Missing observations are handled directly in the likelihood and filtering step by conditioning on the information available at each month t. Parameters are estimated by quasi-maximum likelihood (QML) through the EM algorithm: the E-step applies the Kalman filter and smoother under missing data to compute the conditional moments of the latent state, while the M-step updates parameters in closed form via Gaussian least-squares regressions based on these moments. Mixed-frequency restrictions are incorporated in the measurement equation through known linear aggregation schemes (e.g. Mariano–Murasawa or stock/flow mappings), implying that quarterly GDP is a known linear function of the monthly state at quarter end. Pseudo real-time nowcasting is conducted by indexing vintages with v, defining the information set Ω_v implied by the release calendar up to the cutoff date t_v, filtering the corresponding ragged-edge dataset, and mapping the filtered state into a current-quarter GDP nowcast through the same mixed-frequency measurement structure.

4.1 The Model

Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{Nt})' \in \mathbb{R}^N$ denote a stationary N-dimensional vector process, observed at high frequency $t = 1, \ldots, T$ and standardized to mean 0 and unit variance. The panel may feature an arbitrary pattern of missing observations due to mixed frequencies and ragged-edge releases. We assume that $\mathbf{x}_t$ admits the following dynamic factor representation in state space form

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \tag{19}$$

$$\mathbf{f}_t = \mathbf{A}\mathbf{f}_{t-1} + \mathbf{u}_t, \qquad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_u), \tag{20}$$

with $\boldsymbol{\varepsilon}_t$ uncorrelated with $\mathbf{f}_t$ at all leads and lags.

- **Measurement equation (19).** $\mathbf{f}_t \in \mathbb{R}^r$ represents the low-rank structure of the data, accommodating dimensional reduction. The number of factors r is selected according to the criteria in Subsection **??**.

$\mathbf{\Lambda} \in \mathbb{R}^{N \times r}$ collects factor loadings and $\boldsymbol{\varepsilon}_t$ captures idiosyncratic components. Based on the specification of $\boldsymbol{\varepsilon}_t$, we can distinguish two cases. When $\mathbf{\Sigma}_\varepsilon$ is diagonal (cross-sectionally uncorrelated idiosyncratic terms) the model is called an *exact* factor model, while when it exhibits weak cross-sectional dependence it is defined an *approximate* factor model. Moreover, $\boldsymbol{\varepsilon}_t$ can be allowed to exhibit serial correlation, according to an AR(P_ε process) as discussed in Subsection 4.3.

- **Factor dynamics (20).** Factors follow a VAR(1) process. $\mathbf{A} \in \mathbb{R}^{r \times r}$ denotes the transition matrix and $\mathbf{\Sigma}_u \in \mathbb{R}^{r \times r}$ the innovations covariance.

In the baseline formulation (19)–(20), we set without loss of generality, a VAR for factors of order $P_f = 1$. Nevertheless, nothing prevents accommodating additional lags for the factor dynamics. The same holds for lags in the idiosyncratic dynamics, assumed of order $P_\varepsilon = 1$.

4.2 Estimation

Since $\mathbf{f}_t$ is latent, direct maximisation of the observed-data log-likelihood does not yield closed-form estimators for the parameters in (19)–(20). Following Bańbura and Modugno (2010), we estimate the MF–DFM by quasi-maximum likelihood (QML) via the EM algorithm. Starting from an initial parameter guess $\boldsymbol{\theta}^{(0)}$, the EM algorithm iterates between two steps until convergence:

- *Expectation step (E-step).* Given $\boldsymbol{\theta}^{(j)}$, compute the conditional expectations of the latent factors using the Kalman filter and smoother, and form the expected complete-data log-likelihood

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j)}) = \mathbb{E}_{\boldsymbol{\theta}^{(j)}}\Big[l(X_T, F_T; \boldsymbol{\theta}) \,\Big|\, \Omega_T\Big].$$

- *Maximisation step (M-step).* Update the parameters by maximising $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j)})$:

$$\boldsymbol{\theta}^{(j+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j)}).$$

This strategy naturally accommodates arbitrary missing-data patterns in $\mathbf{x}_t$, induced by ragged-edge and mixed-frequency releases, by computing Kalman updates and likelihood contributions using only the observed coordinates in Ω_T.

We first present the baseline specification considered in Bańbura and Modugno (2010), featuring VAR(1) factor dynamics and serially uncorrelated idiosyncratic terms,

$$\boldsymbol{\varepsilon}_t \sim i.i.d.\,\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon).$$

QML remains asymptotically valid under weak cross-sectional dependence in the idiosyncratic component, so we work under the general conditions of Doz, Giannone, and Reichlin (2012).

4.2.1 Quasi Log-Likelihood

Let $X_T = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ and $F_T = [\mathbf{f}_1, \ldots, \mathbf{f}_T]$. Denote by $l(X_T, F_T; \boldsymbol{\theta})$ the complete-data (joint) log-likelihood, where

$$\boldsymbol{\theta} = \Big\{\boldsymbol{\Lambda}, \, \mathbf{A}, \, \boldsymbol{\Sigma}_\varepsilon, \, \boldsymbol{\Sigma}_u\Big\}.$$

Up to an additive constant, the complete-data log-likelihood associated with (19)–(20) is

$$\begin{aligned} l(X_T, F_T; \boldsymbol{\theta}) = &-\frac{1}{2}\log|\boldsymbol{\Pi}_0| - \frac{1}{2}\mathbf{f}_0'\boldsymbol{\Pi}_0^{-1}\mathbf{f}_0 - \frac{T}{2}\log|\boldsymbol{\Sigma}_u| - \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_u^{-1}\sum_{t=1}^{T}(\mathbf{f}_t - \mathbf{A}\mathbf{f}_{t-1})(\mathbf{f}_t - \mathbf{A}\mathbf{f}_{t-1})'\right] \\ &- \frac{T}{2}\log|\boldsymbol{\Sigma}_\varepsilon| - \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_\varepsilon^{-1}\sum_{t=1}^{T}(\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{f}_t)(\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{f}_t)'\right], \end{aligned} \tag{21}$$

where $(\mathbf{f}_0, \boldsymbol{\Pi}_0)$ denotes the initial state mean and covariance.

4.2.2 EM algorithm

The EM algorithm maximises the observed-data (quasi) log-likelihood by iteratively maximising the expectation of (21) conditional on Ω_T. In practice, this requires computing the conditional moments of the latent factors via the Kalman filter and smoother; given these moments, the M-step reduces to multivariate Gaussian least-squares updates.

E-step. Given $\boldsymbol{\theta}^{(j)}$, run the Kalman filter and smoother for the state-space system (19)–(20). The smoother delivers the conditional moments of the latent factors given the full-sample information set Ω_T, namely, for $t = 1, \dots, T$,

$$\widehat{\mathbf{f}}_{t|T} := \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mid \Omega_T], \qquad \widehat{\boldsymbol{\Pi}}_{t|T} := \mathrm{Var}_{\boldsymbol{\theta}^{(j)}}(\mathbf{f}_t \mid \Omega_T), \qquad \widehat{\boldsymbol{\Delta}}_{t|T} := \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mathbf{f}'_{t-1} \mid \Omega_T].$$

These quantities imply the sufficient statistics required in the M-step:

$$\mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mathbf{f}'_t \mid \Omega_T] = \widehat{\boldsymbol{\Pi}}_{t|T} + \widehat{\mathbf{f}}_{t|T} \widehat{\mathbf{f}}'_{t|T}, \qquad \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mathbf{f}'_{t-1} \mid \Omega_T] = \widehat{\boldsymbol{\Delta}}_{t|T}.$$

Details on the Kalman filter and smoother with missing observations are reported in Appendix B.

M-step. Conditional on the E-step moments, maximise $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j)})$ in closed form, yielding

$$\boldsymbol{\theta}^{(j+1)} = \left\{ \boldsymbol{\Lambda}^{(j+1)}, \mathbf{A}^{(j+1)}, \boldsymbol{\Sigma}_u^{(j+1)}, \boldsymbol{\Sigma}_\varepsilon^{(j+1)} \right\}.$$

The resulting expressions coincide with standard Gaussian least-squares estimators computed on the sufficient statistics, replaced here by their conditional expectations given Ω_T.

When $\mathbf{x}_t$ is fully observed for all t, the maximisation delivers

$$\boldsymbol{\Lambda}^{(j+1)} = \left(\sum_{t=1}^{T} \mathbf{x}_t \, \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}'_t \mid \Omega_T] \right) \left(\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mathbf{f}'_t \mid \Omega_T] \right)^{-1}, \tag{22}$$

$$\mathbf{A}^{(j+1)} = \left(\sum_{t=2}^{T} \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mathbf{f}'_{t-1} \mid \Omega_T] \right) \left(\sum_{t=2}^{T} \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_{t-1} \mathbf{f}'_{t-1} \mid \Omega_T] \right)^{-1}. \tag{23}$$

Given $\boldsymbol{\Lambda}^{(j+1)}$ and $\mathbf{A}^{(j+1)}$, the covariance matrices are updated as conditional expectations of sums of squared residuals:

$$\boldsymbol{\Sigma}_\varepsilon^{(j+1)} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta}^{(j)}} \left[(\mathbf{x}_t - \boldsymbol{\Lambda}^{(j+1)} \mathbf{f}_t)(\mathbf{x}_t - \boldsymbol{\Lambda}^{(j+1)} \mathbf{f}_t)' \,\middle|\, \Omega_T \right], \tag{24}$$

$$\boldsymbol{\Sigma}_u^{(j+1)} = \frac{1}{T-1} \sum_{t=2}^{T} \mathbb{E}_{\boldsymbol{\theta}^{(j)}} \left[(\mathbf{f}_t - \mathbf{A}^{(j+1)} \mathbf{f}_{t-1})(\mathbf{f}_t - \mathbf{A}^{(j+1)} \mathbf{f}_{t-1})' \,\middle|\, \Omega_T \right]. \tag{25}$$

In the *exact* factor model, $\boldsymbol{\Sigma}_\varepsilon$ is restricted to be diagonal and the update in (24) is replaced by $\boldsymbol{\Sigma}_\varepsilon^{(j+1)} = \mathrm{dg}(\boldsymbol{\Sigma}_\varepsilon^{(j+1)})$.

4.2.3 Extension to missing data

Modification of the EM algorithm When $\mathbf{x}_t$ contains missing entries, the updates for $\mathbf{A}$ and $\boldsymbol{\Sigma}_u$ remain unchanged because they depend only on the smoothed factor moments. Instead, the loadings and the idiosyncratic covariance must be computed using only observed coordinates. Let $\mathbf{W}_t$ be a diagonal selection matrix with $(\mathbf{W}_t)_{ii} = 1$ if x_{it} is observed and 0 otherwise. Then the loading update can be written in vectorised form as

$$\operatorname{vec}\left(\boldsymbol{\Lambda}^{(j+1)}\right) = \left(\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t\mathbf{f}_t' \mid \Omega_T] \otimes \mathbf{W}_t\right)^{-1} \operatorname{vec}\left(\sum_{t=1}^{T} \mathbf{W}_t\,\mathbf{x}_t\,\mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t' \mid \Omega_T]\right). \tag{26}$$

Analogously, under the exact-factor restriction $\boldsymbol{\Sigma}_\varepsilon$ diagonal, the update becomes

$$\begin{aligned}\boldsymbol{\Sigma}_\varepsilon^{(j+1)} = \operatorname{dg}\Bigg(\frac{1}{T}\sum_{t=1}^{T}\Big(&\mathbf{W}_t\mathbf{x}_t\mathbf{x}_t'\mathbf{W}_t - \mathbf{W}_t\mathbf{x}_t\,\mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t' \mid \Omega_T]\,\boldsymbol{\Lambda}^{(j+1)\prime}\mathbf{W}_t - \mathbf{W}_t\boldsymbol{\Lambda}^{(j+1)}\,\mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mid \Omega_T]\,\mathbf{x}_t'\mathbf{W}_t \\ &+ \mathbf{W}_t\boldsymbol{\Lambda}^{(j+1)}\,\mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t\mathbf{f}_t' \mid \Omega_T]\,\boldsymbol{\Lambda}^{(j+1)\prime}\mathbf{W}_t + (\mathbf{I}-\mathbf{W}_t)\boldsymbol{\Sigma}_\varepsilon^{(j)}(\mathbf{I}-\mathbf{W}_t)\Big)\Bigg).\end{aligned} \tag{27}$$

Intuitively, $\mathbf{W}_t$ acts as a selection matrix so that only available observations enter the M-step calculations.

Linear restrictions on the parameters. A key advantage of the quasi-maximum-likelihood state-space approach is that it accommodates ragged edges and mixed-frequency releases and, at the same time, allows one to impose *linear restrictions* on the parameters in a straightforward way. Following Bańbura and Modugno (2010), consider restrictions on the factor loadings of the form

$$\mathbf{H}_\Lambda \operatorname{vec}(\boldsymbol{\Lambda}) = \boldsymbol{\gamma}_\Lambda,$$

where $\mathbf{H}_\Lambda$ is a known matrix selecting and linearly combining entries of $\boldsymbol{\Lambda}$ and $\boldsymbol{\gamma}_\Lambda$ is a known vector. These restrictions are particularly relevant in mixed-frequency DFMs: under Mariano–Murasawa aggregation (for growth rates) or stock/flow aggregation (for levels), quarterly observables imply fixed linear relationships across the loadings on factor lags, which can be encoded in $\mathbf{H}_\Lambda$.

Let $\operatorname{vec}(\boldsymbol{\Lambda}_u^{(j+1)})$ denote the *unrestricted* EM update in the M-step. Under missing data, define the expected weighted design matrix

$$\mathbf{M}_{\Lambda}^{(j)} = \sum_{t=1}^{T} \left(\mathbb{E}_{\boldsymbol{\theta}^{(j)}}[\mathbf{f}_t \mathbf{f}_t' \mid \Omega_T] \otimes \mathbf{W}_t \boldsymbol{\Sigma}_{\varepsilon}^{(j)-1} \mathbf{W}_t \right),$$

where $\mathbf{W}_t$ is diagonal with $(\mathbf{W}_t)_{ii} = 1$ if x_{it} is observed and 0 otherwise. The *restricted* update is obtained by projecting the unrestricted estimator onto the constraint set:

$$\text{vec}(\boldsymbol{\Lambda}_r^{(j+1)}) = \text{vec}(\boldsymbol{\Lambda}_u^{(j+1)}) + \mathbf{M}_{\Lambda}^{(j)-1} \mathbf{H}_{\Lambda}' \left(\mathbf{H}_{\Lambda} \mathbf{M}_{\Lambda}^{(j)-1} \mathbf{H}_{\Lambda}' \right)^{-1} \left(\boldsymbol{\gamma}_{\Lambda} - \mathbf{H}_{\Lambda} \text{vec}(\boldsymbol{\Lambda}_u^{(j+1)}) \right). \tag{28}$$

In the complete-data case ($\mathbf{W}_t = \mathbf{I}_N$ for all t), (28) reduces to the standard restricted least-squares update with weight $\boldsymbol{\Sigma}_{\varepsilon}^{(j)-1}$. Analogous linear restrictions can also be imposed on transition parameters, e.g. $\mathbf{H}_A \text{vec}(\mathbf{A}) = \boldsymbol{\gamma}_A$.

In practice, the mixed-frequency restriction is chosen according to the measurement nature of the quarterly series: Mariano and Murasawa (2003) weights are appropriate for growth rates constructed from overlapping monthly growth rates, whereas stock/flow weights are appropriate when the quarterly variable represents within-quarter averages (stocks) or sums (flows). In both cases the restriction is implemented by augmenting the state to include the required factor lags and imposing the corresponding linear constraints on the loading blocks as shown[11] in Appendix C.

4.3 Modelling serial correlation in the idiosyncratic component

4.3.1 The Model

Following Bańbura and Modugno (2010), we extend the baseline MF–DFM by allowing (i) VAR(P_f) factor dynamics and (ii) persistent idiosyncratic components. Both features are implemented in the same linear Gaussian state-space framework by augmenting the state, so that the model remains first-order Markov.

- *Factor dynamics.* VAR(P_f) dynamics are accommodated by stacking factor lags in the state (companion form), so that the transition can still be written in first-order form with $\mathbf{A}$ denoting the companion matrix.
- *Idiosyncratic dynamics.* For each $i = 1, \ldots, N$, the idiosyncratic dynamics follows specifically a

[11] REVIEW FLOW/STOCK

AR(P_ε). Without loss of generality we consider an AR(1):

$$\varepsilon_{i,t} = \tilde{\varepsilon}_{i,t} + \xi_{i,t}, \qquad \xi_{i,t} \sim i.i.d.\ \mathscr{N}(0, \sigma_\xi^2), \tag{29}$$

$$\tilde{\varepsilon}_{i,t} = \alpha_i \tilde{\varepsilon}_{i,t-1} + e_{i,t}, \qquad e_{i,t} \sim i.i.d.\ \mathscr{N}(0, \sigma_i^2), \tag{30}$$

where $\tilde{\boldsymbol{\varepsilon}}_t$ and $\boldsymbol{\xi}_t$ are cross-sectionally uncorrelated, and $\{e_{i,t}\}$ are cross-sectionally uncorrelated and independent of factor shocks. The variance σ_ξ^2 is fixed to a very small value to improve numerical stability when $\tilde{\boldsymbol{\varepsilon}}_t$ is moved into the state.

Consequently, the augmented state-space formormulation (denoted by $\widetilde{\cdot}$) can be stated as

$$\mathbf{x}_t = \widetilde{\boldsymbol{\Lambda}}\widetilde{\mathbf{f}}_t + \boldsymbol{\xi}_t, \qquad \boldsymbol{\xi}_t \sim \mathscr{N}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}_\varepsilon}), \tag{31}$$

$$\widetilde{\mathbf{f}}_t = \widetilde{\mathbf{A}}\widetilde{\mathbf{f}}_{t-1} + \widetilde{\mathbf{u}}_t, \qquad \widetilde{\mathbf{u}}_t \sim \mathscr{N}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}_u}), \tag{32}$$

where

$$\widetilde{\mathbf{f}}_t = \begin{bmatrix} \mathbf{f}_t \\ \tilde{\boldsymbol{\varepsilon}}_t \end{bmatrix}, \qquad \widetilde{\mathbf{u}}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{e}_t \end{bmatrix}, \qquad \widetilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{I}_N \end{bmatrix}.$$

$$\widetilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\alpha_1, \dots, \alpha_N) \end{bmatrix}, \qquad \widetilde{\boldsymbol{\Sigma}_u} = \begin{bmatrix} \boldsymbol{\Sigma}_u & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\sigma_1^2, \dots, \sigma_N^2) \end{bmatrix}, \qquad \widetilde{\boldsymbol{\Sigma}_\varepsilon} = \kappa \mathbf{I}_N.$$

- *Measurement equation* (31). The persistent idiosyncratic component is moved into the state via $\tilde{\boldsymbol{\varepsilon}}_t$, while the remaining measurement noise $\boldsymbol{\xi}_t$ (with κ fixed to a very small variance) ensures numerical stability.

- *Transition equation* (32). The augmented transition is block-diagonal: the factor block evolves according to $\mathbf{A}$ (companion form for VAR(P_f)), and the idiosyncratic block evolves according to $\mathrm{diag}(\alpha_1, \dots, \alpha_N)$.

The system (31)–(32) nests the baseline model (19)–(20) as a special case (e.g. when the idiosyncratic persistence is not modelled and the VAR process of factors exhibits a $P_f = 1$ order). Estimation can be carried out under arbitrary missing patterns by computing Kalman updates and likelihood contributions using observed coordinates only.

4.3.2 Estimation

Estimation is carried out by quasi-maximum likelihood (QML) via the EM algorithm, following Bańbura and Modugno (2010). Relative to the standard MF–DFM estimation in Section 4.2, the only substantive modification is that the state is augmented to include factor lags (VAR(P_f) in companion form) and a persistent idiosyncratic component (AR(1)). This enlarges the set of parameters updated in the M-step to include $(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_e)$. The parameter vector is

$$\widetilde{\boldsymbol{\theta}} = \left\{ \boldsymbol{\Lambda},\ \mathbf{A},\ \boldsymbol{\Sigma}_u,\ \boldsymbol{\alpha},\ \boldsymbol{\Sigma}_e \right\}, \qquad \boldsymbol{\alpha} := \mathrm{dg}(\alpha_1, \ldots, \alpha_N), \quad \boldsymbol{\Sigma}_e := \mathrm{dg}(\sigma_1^2, \ldots, \sigma_N^2),$$

while the measurement-noise covariance is fixed to $\widetilde{\boldsymbol{\Sigma}}_\varepsilon = \kappa \mathbf{I}_N$ with a small $\kappa > 0$.

In the augmented state, VAR(P_f) factor dynamics are represented in companion form by stacking P_f lags:

$$\mathbf{f}_t^{(P_f)} := \begin{bmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \vdots \\ \mathbf{f}_{t-P_f+1} \end{bmatrix}, \qquad \mathbf{f}_t^{(P_f)} = \mathbf{A}\,\mathbf{f}_{t-1}^{(P_f)} + \mathbf{u}_t, \qquad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_u),$$

and the persistent idiosyncratic component satisfies

$$\tilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{\alpha}\,\tilde{\boldsymbol{\varepsilon}}_{t-1} + \mathbf{e}_t, \qquad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e).$$

Define the augmented state

$$\widetilde{\mathbf{f}}_t := \begin{bmatrix} \mathbf{f}_t^{(P_f)} \\ \tilde{\boldsymbol{\varepsilon}}_t \end{bmatrix}.$$

The state-space form is then (31)–(32), with $\widetilde{\mathbf{A}} = \mathrm{blkdiag}(\mathbf{A}, \boldsymbol{\alpha})$ and $\widetilde{\boldsymbol{\Sigma}}_u = \mathrm{blkdiag}(\boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_e)$.

EM algorithm Starting from $\widetilde{\boldsymbol{\theta}}^{(0)}$, iterate

$$\widetilde{\boldsymbol{\theta}}^{(j+1)} = \arg\max_{\widetilde{\boldsymbol{\theta}}} \mathbb{E}_{\widetilde{\boldsymbol{\theta}}^{(j)}}\left[l(\mathbf{X}, \widetilde{\mathbf{F}}; \widetilde{\boldsymbol{\theta}}) \,\middle|\, \Omega_T \right], \qquad \widetilde{\mathbf{F}} := [\widetilde{\mathbf{f}}_1, \ldots, \widetilde{\mathbf{f}}_T].$$

All steps involving only $(\mathbf{\Lambda}, \mathbf{A}, \mathbf{\Sigma}_u)$ are identical to those described in Section 5.2.3. The additional AR(1) block contributes only the updates for $(\boldsymbol{\alpha}, \mathbf{\Sigma}_e)$ and replaces moments of $\mathbf{f}_t$ with moments of the augmented state $\widetilde{\mathbf{f}}_t$ in the E-step.

E-step Given $\widetilde{\boldsymbol{\theta}}^{(j)}$, run the Kalman filter and smoother for (31)–(32) under missing observations (via $\mathbf{W}_t$). Store the smoothed first and second moments

$$\widehat{\widetilde{\mathbf{f}}}_{t|T} := \mathbb{E}(\widetilde{\mathbf{f}}_t \mid \Omega_T), \qquad \widehat{\mathbf{P}}_{t|T} := \mathrm{Var}(\widetilde{\mathbf{f}}_t \mid \Omega_T),$$

and the lag-one smoothed cross-covariance

$$\widehat{\mathbf{\Delta}}_{t,t-1|T} := \mathrm{Cov}(\widetilde{\mathbf{f}}_t, \widetilde{\mathbf{f}}_{t-1} \mid \Omega_T).$$

Equivalently, the M-step can be written in terms of $\mathbb{E}(\widetilde{\mathbf{f}}_t \widetilde{\mathbf{f}}_t' \mid \Omega_T)$ and $\mathbb{E}(\widetilde{\mathbf{f}}_t \widetilde{\mathbf{f}}_{t-1}' \mid \Omega_T)$, reconstructed from $\widehat{\widetilde{\mathbf{f}}}_{t|T}$, $\widehat{\mathbf{P}}_{t|T}$, and $\widehat{\mathbf{\Delta}}_{t,t-1|T}$. From these objects we extract the blocks associated with $\mathbf{f}_t^{(P_f)}$ and $\tilde{\boldsymbol{\varepsilon}}_t$.

M-step Updates for $(\mathbf{\Lambda}, \mathbf{A}, \mathbf{\Sigma}_u)$ coincide with the baseline ones (Section 4.2), applied to the factor-companion block $\mathbf{f}_t^{(P_f)}$. The additional AR(1) idiosyncratic updates are:

Idiosyncratic AR(1) block. For each $i = 1, \ldots, N$,

$$\alpha_i^{(j+1)} = \frac{\sum_{t=2}^T \mathbb{E}_{\widetilde{\boldsymbol{\theta}}^{(j)}}[\tilde{\varepsilon}_{i,t}\tilde{\varepsilon}_{i,t-1} \mid \Omega_T]}{\sum_{t=2}^T \mathbb{E}_{\widetilde{\boldsymbol{\theta}}^{(j)}}\left[\tilde{\varepsilon}_{i,t-1}^2 \mid \Omega_T\right]}, \qquad \sigma_i^{2\,(j+1)} = \frac{1}{T-1}\sum_{t=2}^T \mathbb{E}_{\widetilde{\boldsymbol{\theta}}^{(j)}}\left[\left(\tilde{\varepsilon}_{i,t} - \alpha_i^{(j+1)}\tilde{\varepsilon}_{i,t-1}\right)^2 \middle| \Omega_T\right].$$

Collect them as $\boldsymbol{\alpha}^{(j+1)} = \mathrm{dg}(\alpha_1^{(j+1)}, \ldots, \alpha_N^{(j+1)})$ and $\mathbf{\Sigma}_e^{(j+1)} = \mathrm{dg}(\sigma_1^{2\,(j+1)}, \ldots, \sigma_N^{2\,(j+1)})$.

Loadings under missing observations.

$$\mathrm{vec}\left(\mathbf{\Lambda}^{(j+1)}\right) = \left(\sum_{t=1}^T \mathbb{E}_{\widetilde{\boldsymbol{\theta}}^{(j)}}\left[\mathbf{f}_t\mathbf{f}_t' \mid \Omega_T\right] \otimes \mathbf{W}_t\right)^{-1} \mathrm{vec}\left(\sum_{t=1}^T \mathbf{W}_t\left(\mathbf{x}_t\,\mathbb{E}_{\widetilde{\boldsymbol{\theta}}^{(j)}}\left[\mathbf{f}_t' \mid \Omega_T\right] - \mathbb{E}_{\widetilde{\boldsymbol{\theta}}^{(j)}}\left[\tilde{\boldsymbol{\varepsilon}}_t\mathbf{f}_t' \mid \Omega_T\right]\right)\right).$$

4.4 Nowcasting

The nowcasting exercise is implemented in a pseudo real-time fashion within the same (possibly augmented) linear–Gaussian state-space representation used for estimation. At each vintage v, the available dataset features a ragged edge due to asynchronous releases and publication delays.

Let v index pseudo real-time vintages. Each vintage corresponds to a cutoff date on the monthly grid, denoted by t_v. Let $\tau(v)$ be the quarter containing t_v and let $m(v) \in \{1,2,3\}$ denote the month within quarter $\tau(v)$ (M1, M2, M3), so that $t_v = t(\tau(v), m(v))$. Define the information set available at vintage v as $\Omega_v := \Omega_{\tau(v),m(v)}$. Operationally, Ω_v is constructed from the full monthly panel by setting to missing all observations that would not yet be released by date t_v, including all months strictly after t_v and series-specific publication delays. Denote the resulting ragged-edge dataset by $\{\mathbf{x}_t^{(v)}\}_{t \leq t_v}$.

Nowcasts are produced by running the Kalman filter under missing observations, treating the model parameters as fixed at their estimated values. Let $\mathbf{s}_t$ denote the (possibly augmented) latent state collecting the common factors (and, if included, persistent idiosyncratic components), and consider the state-space system in (31)–(32). Running the Kalman filter on $\{\mathbf{x}_t^{(v)}\}_{t \leq t_v}$ yields the filtered state $\widehat{\mathbf{s}}_{t_v|t_v}^{(v)} = \mathbb{E}(\mathbf{s}_{t_v} \mid \Omega_v)$.

Let y_τ denote the quarterly target (GDP) for quarter τ. Under mixed-frequency restrictions, y_τ is a known linear function of the monthly state at quarter-end; denote by $\boldsymbol{\ell}$ the corresponding selector/aggregation vector, consistent with the imposed Mariano–Murasawa or stock/flow scheme. The vintage-v nowcast of the current quarter is

$$\widehat{y}_{\tau(v)}^{(v)} := \mathbb{E}\big[y_{\tau(v)} \,\big|\, \Omega_v\big] = \boldsymbol{\ell}' \widehat{\mathbf{s}}_{t_v|t_v}^{(v)}. \tag{33}$$

Let $\mathscr{V}$ be the set of evaluation vintages for which the corresponding realization $y_{\tau(v)}$ is eventually observed. For each $v \in \mathscr{V}$, define the forecast error

$$e_v := y_{\tau(v)} - \widehat{y}_{\tau(v)}^{(v)}. \tag{34}$$

The root mean squared forecast error (RMSFE) is

$$\text{RMSFE} := \sqrt{\frac{1}{|\mathscr{V}|} \sum_{v \in \mathscr{V}} e_v^2} = \sqrt{\frac{1}{|\mathscr{V}|} \sum_{v \in \mathscr{V}} \Big(y_{\tau(v)} - \widehat{y}_{\tau(v)}^{(v)}\Big)^2}. \tag{35}$$

5 Matrix Mixed-Frequency Dynamic Factor Model (MF–DMFM)

In this section, we extend the mixed-frequency DFM of Section 4 to matrix-valued time series, introducing the Dynamic Matrix Factor Model (DMFM). The key idea is to model a panel observed as a sequence of matrices (e.g. countries $\times$ variables) over time, so that comovements are captured through a low-dimensional common component while preserving the two-way structure. This strategy delivers a more parsimonious and interpretable parameterization with respect to its vector counterpart.

The DMFM is particularly suited to the empirical nowcasting exercise. Its state-space representation naturally handles mixed-frequency releases and ragged edges by conditioning on the entries available at each date, and it produces real-time updates as new information arrives. Moreover, besides dynamic factor persistence, it allows one to augment the state with additional components, notably a persistent idiosyncratic block (e.g. an AR-type term), to capture series-specific inertia.

5.1 The Model

Let $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ be a $p_1 \times p_2$ matrix-valued, zero-mean, second-order stationary process. Assume that the observed series are centered and standardized. The DMFM decomposes $\mathbf{X}_t$ into a low-rank common component and an idiosyncratic component, extending the Matrix Factor Model (MFM) of D. Wang, Liu, and R. Chen (2019) by endowing the latent factors with Matrix Autoregressive (MAR) dynamics as in R. Chen, Xiao, and Yang (2021). Without loss of generality, we consider an MAR process of order $P_F = 1$.

$$\mathbf{X}_t = \mathbf{R}\,\mathbf{F}_t\,\mathbf{C}' + \mathbf{E}_t, \tag{36}$$

$$\mathbf{F}_t = \mathbf{A}\,\mathbf{F}_{t-1}\,\mathbf{B}' + \mathbf{U}_t, \tag{37}$$

Throughout, we assume that $\{\mathbf{F}_t\}$ and $\{\mathbf{E}_t\}$ are uncorrelated at all leads and lags.

- *MFM Equation* (36) The observed matrix $\mathbf{X}_t \in \mathbb{R}^{p_1\times p_2}$ is decomposed as $\mathbf{X}_t = \boldsymbol{\chi}_t + \mathbf{E}_t$, where $\boldsymbol{\chi}_t = \mathbf{R}\mathbf{F}_t\mathbf{C}'$ is a low-rank common component and $\mathbf{E}_t$ collects the idiosyncratic variation. $\mathbf{F}_t \in \mathbb{R}^{r_1\times r_2}$ is the matrix of latent factors with $r_1, r_2 < \min(p_1, p_2)$, and $\mathbf{R} \in \mathbb{R}^{p_1\times r_1}$ and $\mathbf{C} \in \mathbb{R}^{p_2\times r_2}$ are the row and column loading matrices. Since the idiosyncratic term $\mathbf{E}_t$ is left unrestricted, allowing for both cross-sectional and serial dependence, we refer to this DMFM as *approximate*. Cross-sectional dependence is summarized by row and column covariance matrices $\mathbf{H} \in \mathbb{R}^{p_1 \times p_1}$ and $\mathbf{K} \in \mathbb{R}^{p_2 \times p_2}$.

- *MAR Equation* (37) The factors are persistent, hence the term *dynamic*, and follow a MAR(1) process: $\mathbf{A} \in \mathbb{R}^{r_1 \times r_1}$ and $\mathbf{B} \in \mathbb{R}^{r_2 \times r_2}$ govern the temporal dependence, while $\mathbf{U}_t \in \mathbb{R}^{r_1 \times r_2}$ collects the innovations. The latter have separable covariance, with row and column covariance matrices $\mathbf{P} \in \mathbb{R}^{r_1 \times r_1}$ and $\mathbf{Q} \in \mathbb{R}^{r_2 \times r_2}$.

This specification reduces to a standard vector DFM when p_1 or p_2, collapses to one.

Importantly, preserving the matrix structure in which the data are collected yields a parsimonious and interpretable parametrization of the model. Consider a fully vectorized version of the DMFM, where the measurement equation involves the loading matrix $(\mathbf{C} \otimes \mathbf{R})$, so that the dimension becomes $(p_1 p_2) \times (r_1 r_2)$ and the two-way structure is obscured. Working directly with $\mathbf{R}$ and $\mathbf{C}$ keeps the bilinear form explicit and reduces the number of loading parameters to $p_1 r_1 + p_2 r_2$. Nevertheless, parsimony in the loadings does not fully resolve the computational burden, since the idiosyncratic component $\mathbf{E}_t$ can still exhibit both serial and cross-sectional correlation. This typically makes maximum likelihood estimation (MLE) infeasible.

DMFM identification The parametrization of the common component in (36) is not unique: for any invertible matrices $\mathbf{G}_1 \in \mathbb{R}^{r_1 \times r_1}$ and $\mathbf{G}_2 \in \mathbb{R}^{r_2 \times r_2}$, one can reparametrize the triple $(\mathbf{R}, \mathbf{F}_t, \mathbf{C})$ without changing $\mathbf{R}\mathbf{F}_t\mathbf{C}'$. Hence, identification is naturally understood in terms of the spaces spanned by the loadings (and the resulting $\boldsymbol{\chi}_t$), which is sufficient for forecasting/nowcasting.

To select a representative element within the equivalence class, we adopt the usual normalization (D. Wang, Liu, and R. Chen 2019; L. Yu et al. 2022) by imposing (asymptotic) orthonormality of the loading columns:

$$\left\| \frac{1}{p_1} \mathbf{R}^\top \mathbf{R} - \mathbf{I}_{r_1} \right\| \to 0, \qquad \left\| \frac{1}{p_2} \mathbf{C}^\top \mathbf{C} - \mathbf{I}_{r_2} \right\| \to 0. \tag{38}$$

Under (38), the remaining indeterminacy is limited to orthogonal rotations: for any orthogonal $\mathbf{O}_1 \in \mathbb{R}^{r_1 \times r_1}$ and $\mathbf{O}_2 \in \mathbb{R}^{r_2 \times r_2}$, the rotated representation yields the same common component and still satisfies (38). In particular, sign switches are a special case of these rotations.

5.2 Estimation

The approximate DMFM in (36)–(37) is estimated by quasi maximum likelihood (QML) via an EM algorithm combined with Kalman smoother, following Matteo Barigozzi and Trapin (2025). The procedure yields QML estimates of the structural parameters and smoothed estimates of the latent factors, and it naturally handles arbitrary missing-data patterns.

A direct Gaussian maximum likelihood estimation of (36)–(37) is not feasible for two main reasons:

- *large number of parameters to estimate.* In an *approximate* DMFM the idiosyncratic component $\mathbf{E}_t$ may be correlated across rows and columns and potentially over time. In the most general case, the stacked vector $E_T = (\mathbf{e}_1', \ldots, \mathbf{e}_T')'$ has covariance $\boldsymbol{\Omega}_T^E = \mathbb{E}(E_T E_T')$, a full $(p_1 p_2 T) \times (p_1 p_2 T)$ matrix with $\frac{p_1 p_2 T (p_1 p_2 T+1)}{2}$ distinct parameters. Even under a matrix parametrization, cross-sectional dependence summarized by $\mathbf{H}$ and $\mathbf{K}$ already requires $\mathscr{O}(p_1^2 + p_2^2)$ parameters, and allowing serial dependence adds autocovariances across lags, leading to $\mathscr{O}((p_1^2 + p_2^2)T)$ parameters overall.

- *No closed-form maximizer (latent factors).* Because the factors are unobserved, the likelihood typically has no closed-form maximizer, and direct numerical optimization becomes impractical in large systems.

These issues are addressed by:

- *Quasi-log-likelihood (covariance mis-specification).* A tractable Gaussian quasi log-likelihood is obtained by treating $\mathbf{E}_t$ *as if* it were serially and cross-sectionally uncorrelated (diagonal row/column covariances), which avoids estimating the high-dimensional nuisance covariance $\boldsymbol{\Omega}_T^E$. This reduces the number of parameters to be estimated to $\mathscr{O}((p_1 + p_2)T)$, affordable with $p_1 p_2 T$ observations.

- *EM algorithm with Kalman smoothing (closed-form updates).* Conditional on the quasi log-likelihood, EM replaces direct optimization with closed-form updates: the E-step computes the conditional moments of the latent states (also with missing observations), and the M-step updates MFM and MAR parameters while preserving the bilinear structure for loadings and idiosyncratic variances.

Overall, QMLE via the EM algorithm makes estimation feasible for high-dimensional DMFMs while retaining an *approximate* idiosyncratic component and handling ragged-edge and mixed-frequency missingness. In the following, we describe the main steps of the procedure, starting from the quasi log-likelihood specification and then detailing the EM process.

5.2.1 Quasi Log-Likelihood

Let us consider the DMFM as defined in (36)–(37) and define $\mathbf{x}_t = \text{vec}(\mathbf{X_t})$, $\mathbf{f}_t = \text{vec}(\mathbf{F_t})$, and $\mathbf{e}_t = \text{vec}(\mathbf{E_t})$ as the vectorized versions of the observed matrix, the latent factor matrix and the idiosyncratic component, respectively. Stack the sample over T observations and define

$$X_T = (\mathbf{x}_1', \ldots, \mathbf{x}_T')', \qquad F_T = (\mathbf{f}_1', \ldots, \mathbf{f}_T')', \qquad E_T = (\mathbf{e}_1', \ldots, \mathbf{e}_T')',$$

where $X_T, E_T \in \mathbb{R}^{p_1 p_2 T}$ and $F_T \in \mathbb{R}^{r_1 r_2 T}$. Let

$$\mathbf{\Omega}_T^X \equiv \mathbb{E}_{\boldsymbol{\theta}}\left[X_T X_T'\right], \qquad \mathbf{\Omega}_T^E \equiv \mathbb{E}_{\boldsymbol{\theta}}\left[E_T E_T'\right], \qquad \mathbf{\Omega}_T^F \equiv \mathbb{E}_{\boldsymbol{\theta}}\left[F_T F_T'\right]$$

denote the corresponding covariance matrices. By construction, $\mathbf{\Omega}_T^E$ collects all cross-sectional covariances and all temporal autocovariances up to lag $(T-1)$. Since $\mathbf{\Omega}_T^F$ is fully characterized by $(\mathbf{A}, \mathbf{B}, \mathbf{P}, \mathbf{Q})$, we write $\mathbf{\Omega}_T^F = \mathbf{\Omega}_{T(\mathbf{A},\mathbf{B},\mathbf{P},\mathbf{Q})}^F$. Hence, the DMFM implies the following covariance decomposition:

$$\mathbf{\Omega}_T^X = (\mathbf{I}_T \otimes \mathbf{C} \otimes \mathbf{R})\, \mathbf{\Omega}_{T(\mathbf{A},\mathbf{B},\mathbf{P},\mathbf{Q})}^F\, (\mathbf{I}_T \otimes \mathbf{C} \otimes \mathbf{R})' \,+\, \mathbf{\Omega}_T^E. \tag{39}$$

In an *approximate* DMFM, $\mathbf{\Omega}_T^E$ is not restricted a priori. In the most general case, this means treating $\mathbf{\Omega}_T^E$ as a full covariance matrix for $E_T \in \mathbb{R}^{p_1 p_2 T}$. Hence, one would need to estimate a $(p_1 p_2 T) \times (p_1 p_2 T)$ matrix with $\frac{p_1 p_2 T\,(p_1 p_2 T + 1)}{2} = \mathscr{O}\big((p_1 p_2 T)^2\big)$ distinct parameters, which is clearly infeasible given only $p_1 p_2 T$ observations.

One may try to reduce the dimension by exploiting matrix-structured second moments, summarizing cross-sectional dependence through row/column covariances (e.g. $\mathbf{H}$ and $\mathbf{K}$). This lowers the contemporaneous nuisance dimension to $\mathscr{O}(p_1^2 + p_2^2)$. However, once serial dependence in $\{\mathbf{E}_t\}$ is allowed, one must specify such row/column covariance objects for each lag $h = 0, \ldots, T-1$, so the overall number of nuisance parameters still grows with the sample size, on the order of $\mathscr{O}\big((p_1^2 + p_2^2)T\big)$.

Following Matteo Barigozzi and Trapin (2025), we adopt a misspecified Gaussian likelihood treating idiosyncratic components as if they were serially and cross-sectionally uncorrelated, i.e. we replace

$$\mathbf{\Omega}_T^E \approx \mathbf{I}_T \otimes \mathrm{dg}(\mathbf{K}) \otimes \mathrm{dg}(\mathbf{H}), \tag{40}$$

where $\mathrm{dg}(\mathbf{H})$ and $\mathrm{dg}(\mathbf{K})$ collect the diagonal elements of the row and column idiosyncratic covariance matrices $\mathbf{H}$ and $\mathbf{K}$. Under (40), the parameter vector can be written as

$$\boldsymbol{\theta} = \Big(\mathrm{vec}(\mathbf{R})', \mathrm{vec}(\mathbf{C})', \mathrm{vec}(\mathrm{dg}(\mathbf{H}))', \mathrm{vec}(\mathrm{dg}(\mathbf{K}))', \mathrm{vec}(\mathbf{A})', \mathrm{vec}(\mathbf{B})', \mathrm{vec}(\mathbf{P})', \mathrm{vec}(\mathbf{Q})'\Big)'.$$

The resulting Gaussian quasi log-likelihood is

$$\begin{aligned}\ell(X_T;\boldsymbol{\theta}) = & -\frac{p_1 p_2 T}{2}\log(2\pi) - \frac{1}{2}\log\Big|(\mathbf{I}_T\otimes\mathbf{C}\otimes\mathbf{R})\,\mathbf{\Omega}^F_{T(\mathbf{A},\mathbf{B},\mathbf{P},\mathbf{Q})}\,(\mathbf{I}_T\otimes\mathbf{C}\otimes\mathbf{R})' + \mathbf{I}_T\otimes\mathrm{dg}(\mathbf{K})\otimes\mathrm{dg}(\mathbf{H})\Big| \\ & -\frac{1}{2}X_T'\Big[(\mathbf{I}_T\otimes\mathbf{C}\otimes\mathbf{R})\,\mathbf{\Omega}^F_{T(\mathbf{A},\mathbf{B},\mathbf{P},\mathbf{Q})}\,(\mathbf{I}_T\otimes\mathbf{C}\otimes\mathbf{R})' + \mathbf{I}_T\otimes\mathrm{dg}(\mathbf{K})\otimes\mathrm{dg}(\mathbf{H})\Big]^{-1}X_T.\end{aligned} \tag{41}$$

Because (41) is based on the misspecification (40), its maximizer is a Quasi-Maximum Likelihood (QML) estimator. However, missing and latent observations make the likelihood intractable. Dempster, Laird, and Rubin (1977) prospose The EM algorithm an an iterative procedure able to address this issue.

5.2.2 EM Algorithm

Because the quasi log-likelihood in (41) involves latent factors, direct maximization is not available. We therefore adopt the Expectation–Maximization (EM) algorithm of Dempster, Laird, and Rubin (1977), treating $\{\mathbf{F}_t\}$ as missing data and recovering them through Kalman filtering and smoothing.

Starting from an initial value $\boldsymbol{\theta}^{(0)}$, for any iteration j the EM algorithm alternates between an Expectation and a Maximization step.

- *E-step.* Given $\boldsymbol{\theta}^{(j)}$, we compute the conditional moments of the latent state using the Kalman smoother, thus extracting a expected complete-data quasi log-likelihood

- *M-step.* Conditional on these moments, we maximize the expected complete-data quasi log-likelihood with respect to $\boldsymbol{\theta}$, obtaining $\boldsymbol{\theta}^{(j+1)}$. In the DMFM, this yields closed-form updates for the loadings $(\mathbf{R}, \mathbf{C})$, the idiosyncratic variances $\mathrm{dg}(\mathbf{H})$ and $\mathrm{dg}(\mathbf{K})$ under the pseudo-likelihood, and the factor-dynamics parameters $(\mathbf{A}, \mathbf{B}, \mathbf{P}, \mathbf{Q})$, while preserving the bilinear structure of the model.

The procedure generates the sequence $\{\boldsymbol{\theta}^{(j)}\}$ until convergence, which is typically assessed by the stabilization of the quasi log-likelihood.

Kalman filter and smoother To evaluate the quasi log-likelihood and implement the EM algorithm, we run the Kalman filter and smoother on the vectorized DMFM:

$$\mathbf{x}_t = \mathbf{Z}\,\mathbf{f}_t + \mathbf{e}_t, \tag{42}$$

$$\mathbf{f}_t = \mathbf{T}\,\mathbf{f}_{t-1} + \mathbf{u}_t. \tag{43}$$

where $\mathbf{Z} = \mathbf{C} \otimes \mathbf{R}$ and $\mathbf{T} = \mathbf{B} \otimes \mathbf{A}$. Under the pseudo log-likelihood, this linear–Gaussian state-space exhibits innovation covariances:

$$\mathrm{Var}(\mathbf{e}_t) = \mathrm{dg}(\mathbf{K}) \otimes \mathrm{dg}(\mathbf{H}), \qquad \mathrm{Var}(\mathbf{u}_t) = \mathbf{Q} \otimes \mathbf{P}.$$

Missing entries in $\mathbf{x}_t$ are handled by applying the filter to the observed subvector $\mathbf{x}_{t,\mathscr{O}_t}$ and the corresponding rows of the measurement matrix, $\mathbf{Z}_{\mathscr{O}_t}$, as detailed in Appendix B.

The smoother delivers the conditional moments

$$\mathbf{f}_{t|T} = \mathbb{E}(\mathbf{f}_t \mid X_T), \qquad \boldsymbol{\Pi}_{t|T} = \mathrm{Var}(\mathbf{f}_t \mid X_T), \qquad \boldsymbol{\Delta}_{t|T} = \mathrm{Cov}(\mathbf{f}_t, \mathbf{f}_{t-1} \mid X_T),$$

which imply the sufficient second-order statistics used in the EM updates:

$$\mathbb{E}\left[\mathbf{f}_t \mid X_T, \boldsymbol{\theta}^{(j)}\right], \qquad \mathbb{E}\left[\mathbf{f}_t\mathbf{f}_t' \mid X_T, \boldsymbol{\theta}^{(j)}\right], \qquad \mathbb{E}\left[\mathbf{f}_t\mathbf{f}_{t-1}' \mid X_T, \boldsymbol{\theta}^{(j)}\right].$$

E-step The smoothed estimates of the conditional first and second moments of the latent factors are used in the E-step of the EM algorithm to form the expected Gaussian quasi log-likelihood of the approximate DMFM. Given the current parameter estimates $\hat{\boldsymbol{\theta}}^{(j)}$, the observed-data log-likelihood admits the standard

EM decomposition

$$\ell(X_T;\boldsymbol{\theta}) = \underbrace{\mathbb{E}_{\hat{\boldsymbol{\theta}}^{(j)}}\left[\ell(X_T \mid F_T;\boldsymbol{\theta})\,\middle|\,X_T\right] + \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(j)}}\left[\ell(F_T;\boldsymbol{\theta})\,\middle|\,X_T\right]}_{Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^{(j)})} - \underbrace{\mathbb{E}_{\hat{\boldsymbol{\theta}}^{(j)}}\left[\ell(F_T \mid X_T;\boldsymbol{\theta})\,\middle|\,X_T\right]}_{\text{conditional entropy}}, \tag{44}$$

where the last term does not depend on $\boldsymbol{\theta}$ and can therefore be ignored in the M-step. Hence, maximizing $\ell(X_T;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is equivalent to maximizing the *Q-function* $Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^{(j)})$.

The expected log-likelihood of the observed data takes the following form:

$$\begin{aligned}\mathbb{E}_{\hat{\boldsymbol{\theta}}^{(j)}}\left[\ell(X_T \mid F_T;\boldsymbol{\theta})\,\middle|\,X_T\right] &= -\frac{T}{2}\Big(p_1 \log|\mathbf{K}| + p_2 \log|\mathbf{H}|\Big) \\ &\quad - \frac{1}{2}\sum_{t=1}^{T} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(j)}}\left[\operatorname{tr}\Big(\mathbf{H}^{-1}(X_t - \mathbf{R}F_t\mathbf{C}')\mathbf{K}^{-1}(X_t - \mathbf{R}F_t\mathbf{C}')'\Big)\,\middle|\,X_T\right],\end{aligned} \tag{45}$$

while the expected log-likelihood of the latent state is given by:

$$\begin{aligned}\mathbb{E}_{\hat{\boldsymbol{\theta}}^{(j)}}\left[\ell(F_T;\boldsymbol{\theta})\,\middle|\,X_T\right] &= -\frac{T-1}{2}\Big(r_1 \log|\mathbf{Q}| + r_2 \log|\mathbf{P}|\Big) \\ &\quad - \frac{1}{2}\sum_{t=2}^{T} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(j)}}\left[\operatorname{tr}\Big(\mathbf{P}^{-1}(F_t - \mathbf{A}F_{t-1}\mathbf{B}')\mathbf{Q}^{-1}(F_t - \mathbf{A}F_{t-1}\mathbf{B}')'\Big)\,\middle|\,X_T\right].\end{aligned} \tag{46}$$

It is worth noting that both expressions above are directly formulated in terms of the original matrix-valued observations and factor processes.

M-step In the M-step, Equations (45) and (46), derived using the current parameter estimates $\hat{\boldsymbol{\theta}}^{(j)}$, are maximized to obtain an updated set of parameters $\hat{\boldsymbol{\theta}}^{(j+1)}$. At each iteration $j \geq 0$, new estimates of the DMFM parameters are computed, beginning with the row and column loading matrices.

The row loadings are updated as follows:

$$\hat{\mathbf{R}}^{(j+1)} = \left(\sum_{t=1}^{T} \mathbf{X}_t \hat{\mathbf{K}}^{(j)-1}\hat{\mathbf{C}}^{(j)}\hat{\mathbf{f}}_{t|T}^{(j)\prime}\right)\left(\sum_{t=1}^{T}\left(\hat{\mathbf{C}}^{(j)\prime}\hat{\mathbf{K}}^{(j)-1}\hat{\mathbf{C}}^{(j)}\right)\star\left(\hat{\mathbf{f}}_{t|T}^{(j)}\hat{\mathbf{f}}_{t|T}^{(j)\prime} + \mathbf{\Pi}_{t|T}^{(j)}\right)\right)^{-1}, \tag{47}$$

while the column loadings are updated via:

$$\hat{\mathbf{C}}^{(j+1)} = \left(\sum_{t=1}^{T} \mathbf{X}_t' \hat{\mathbf{H}}^{(j)-1} \hat{\mathbf{R}}^{(j+1)} \hat{\mathbf{f}}_{t|T}^{(j)}\right) \left(\sum_{t=1}^{T} \left(\hat{\mathbf{R}}^{(j+1)\prime} \hat{\mathbf{H}}^{(j)-1} \hat{\mathbf{R}}^{(j+1)}\right) \star \left(\mathbf{K}_{r_1 r_2} \left(\hat{\mathbf{f}}_{t|T}^{(j)} \hat{\mathbf{f}}_{t|T}^{(j)\prime} + \mathbf{\Pi}_{t|T}^{(j)}\right) \mathbf{K}_{r_1 r_2}'\right)\right)^{-1}. \tag{48}$$

Since the estimation of $\mathbf{R}$ and $\mathbf{C}$ depends on each other, the empirical strategy adopted is to first compute $\hat{\mathbf{R}}^{(j+1)}$ conditionally on $\hat{\mathbf{C}}^{(j)}$, and then update $\hat{\mathbf{C}}^{(j+1)}$ based on the newly estimated $\hat{\mathbf{R}}^{(j+1)}$. However, the reverse order would also be valid.

Given $\hat{\mathbf{R}}^{(j+1)}$ and $\hat{\mathbf{C}}^{(j+1)}$, we can now estimate the idiosyncratic covariance matrices $\hat{\mathbf{H}}^{(j+1)}$ and $\hat{\mathbf{K}}^{(j+1)}$. These matrices are assumed to be diagonal, consistent with the quasi log-likelihood mis-specification in Equation 53. For $i = 1, \ldots, p_1$, with $[\hat{\mathbf{H}}^{(j+1)}]_{i\ell} = 0$ for $i \neq \ell$, $[\hat{\mathbf{H}}^{(j+1)}]_{ii}$ is computed as:

$$\begin{aligned}
[\hat{\mathbf{H}}^{(j+1)}]_{ii} = \frac{1}{T p_2} \sum_{t=1}^{T} \Big[& \mathbf{X}_t \hat{\mathbf{K}}^{(j)-1} \mathbf{X}_t' - \mathbf{X}_t \hat{\mathbf{K}}^{(j)-1} \hat{\mathbf{C}}^{(j+1)} \hat{\mathbf{F}}_{t|T}^{(j)\prime} \hat{\mathbf{R}}^{(j+1)\prime} - \hat{\mathbf{R}}^{(j+1)} \hat{\mathbf{F}}_{t|T}^{(j)} \hat{\mathbf{C}}^{(j+1)\prime} \hat{\mathbf{K}}^{(j)-1} \mathbf{X}_t' \\
& + \left(\hat{\mathbf{C}}^{(j+1)\prime} \hat{\mathbf{K}}^{(j)-1} \hat{\mathbf{C}}^{(j+1)}\right) \star \left((\mathbf{I}_{r_2} \otimes \hat{\mathbf{R}}^{(j+1)})(\hat{\mathbf{f}}_{t|T}^{(j)} \hat{\mathbf{f}}_{t|T}^{(j)\prime} + \mathbf{\Pi}_{t|T}^{(j)})(\mathbf{I}_{r_2} \otimes \hat{\mathbf{R}}^{(j+1)})'\right) \Big]_{ii}.
\end{aligned} \tag{49}$$

Similarly, for $i = 1, \ldots, p_2$, with $[\hat{\mathbf{K}}^{(j+1)}]_{i\ell} = 0$ for $i \neq \ell$, $[\hat{\mathbf{K}}^{(j+1)}]_{ii}$ takes the form of:

$$\begin{aligned}
[\hat{\mathbf{K}}^{(j+1)}]_{ii} = \frac{1}{T p_1} \sum_{t=1}^{T} \Big[& \mathbf{X}_t' \hat{\mathbf{H}}^{(j+1)-1} \mathbf{X}_t - \mathbf{X}_t' \hat{\mathbf{H}}^{(j+1)-1} \hat{\mathbf{R}}^{(j+1)} \hat{\mathbf{F}}_{t|T}^{(j)} \hat{\mathbf{C}}^{(j+1)\prime} - \hat{\mathbf{C}}^{(j+1)} \hat{\mathbf{F}}_{t|T}^{(j)\prime} \hat{\mathbf{R}}^{(j+1)\prime} \hat{\mathbf{H}}^{(j+1)-1} \mathbf{X}_t \\
& + \left(\hat{\mathbf{R}}^{(j+1)\prime} \hat{\mathbf{H}}^{(j+1)-1} \hat{\mathbf{R}}^{(j+1)}\right) \star \left((\mathbf{I}_{r_1} \otimes \hat{\mathbf{C}}^{(j+1)}) \mathbf{K}_{r_1 r_2} (\hat{\mathbf{f}}_{t|T}^{(j)} \hat{\mathbf{f}}_{t|T}^{(j)\prime} + \mathbf{\Pi}_{t|T}^{(j)}) \mathbf{K}_{r_1 r_2}' (\mathbf{I}_{r_1} \otimes \hat{\mathbf{C}}^{(j+1)})'\right) \Big]_{ii}.
\end{aligned} \tag{50}$$

As discussed for the loadings, the estimation of $\hat{\mathbf{H}}$ and $\hat{\mathbf{K}}$ depends on each other. To maintain the bilinear structure of the model, $\hat{\mathbf{H}}^{(j+1)}$ is computed first, conditional on $\hat{\mathbf{K}}^{(j)}$.

Autoregressive matrices $\mathbf{A}$ and $\mathbf{B}$ and innovation covariance matrices $\mathbf{P}$ and $\mathbf{Q}$ are solely used for the implementation of the Kalman smoother on vectorized data. Therefore, the transition and innovation matrices are estimated via the vectorized MAR model as:

$$\widehat{\mathbf{B} \otimes \mathbf{A}}^{(j+1)} = \left(\sum_{t=2}^{T} \hat{\mathbf{f}}_{t|T}^{(j+1)} \hat{\mathbf{f}}_{t-1|T}^{(j)\prime} + \mathbf{\Delta}_{t|T}^{(j)}\right) \left(\sum_{t=2}^{T} \hat{\mathbf{f}}_{t-1|T}^{(j)} \hat{\mathbf{f}}_{t-1|T}^{(j)\prime} + \mathbf{\Pi}_{t-1|T}^{(j)}\right)^{-1}, \tag{51}$$

$$\widehat{\mathbf{Q}\otimes\mathbf{P}}^{(j+1)} = \frac{1}{T}\sum_{t=2}^{T}\left(\hat{\mathbf{f}}_{t|T}^{(j)}\hat{\mathbf{f}}_{t|T}^{(j)\prime} + \mathbf{\Pi}_{t|T}^{(j)} - \left(\hat{\mathbf{f}}_{t|T}^{(j)}\hat{\mathbf{f}}_{t-1|T}^{(j)\prime} + \mathbf{\Delta}_{t|T}^{(j)}\right)\left(\widehat{\mathbf{B}\otimes\mathbf{A}}^{(j+1)}\right)^{\prime}\right). \tag{52}$$

In this case, the estimators do not enforce the bilinear structure of the MAR model directly. However, singular estimates of the dynamic parameters are not required, and the estimated loadings and factors are asymptotically unaffected by this simplification.

EM initialization Since the EM algorithm maximizes a generally non-concave quasi log-likelihood, its convergence depends on the starting values. We initialize the static MFM block via the projected estimation method of L. Yu et al. (2022), and then obtain pre-estimators for the MAR dynamics by OLS on the estimated factor sequence. The projected approach exploits the simultaneous dimension reduction and denoising induced by row/column projections.

- *Step 1: static matrix factor structure (projected estimation).* Assume a static MFM as in (36). If the (normalized) column loading space is known, projecting onto it yields

$$\mathbf{X}_t^{\mathrm{col}} := \frac{1}{p_2}\mathbf{X}_t\mathbf{C} = \mathbf{R}\mathbf{F}_t + \widetilde{\mathbf{E}}_t^{\mathrm{col}}, \qquad \widetilde{\mathbf{E}}_t^{\mathrm{col}} := \frac{1}{p_2}\mathbf{E}_t\mathbf{C},$$

where the idiosyncratic component is attenuated as p_2 grows. Define the projected covariance

$$\widetilde{\mathbf{M}}_1 = \frac{1}{Tp_1}\sum_{t=1}^{T}\mathbf{X}_t^{\mathrm{col}}\mathbf{X}_t^{\mathrm{col}\prime},$$

and estimate $\mathbf{R}$ as the leading r_1 eigenvectors of $\widetilde{\mathbf{M}}_1$, scaled by $\sqrt{p_1}$. The row-analogue is obtained by projecting

$$\mathbf{X}_t^{\mathrm{row}} := \frac{1}{p_1}\mathbf{X}_t^{\prime}\mathbf{R} = \mathbf{C}\mathbf{F}_t^{\prime} + \widetilde{\mathbf{E}}_t^{\mathrm{row}}, \qquad \widetilde{\mathbf{E}}_t^{\mathrm{row}} := \frac{1}{p_1}\mathbf{E}_t^{\prime}\mathbf{R},$$

leading to an analogous covariance $\widetilde{\mathbf{M}}_2$ and an estimator of $\mathbf{C}$ (scaled by $\sqrt{p_2}$).

- *Step 2: initial projection matrices.* In practice, $\mathbf{R}$ and $\mathbf{C}$ are unknown. We build crude initial estimators via PCA on

$$\widehat{\mathbf{M}}_1 = \frac{1}{Tp_1p_2}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t^{\prime}, \qquad \widehat{\mathbf{M}}_2 = \frac{1}{Tp_1p_2}\sum_{t=1}^{T}\mathbf{X}_t^{\prime}\mathbf{X}_t,$$

and set $\widehat{\mathbf{R}}^{(0)} = \sqrt{p_1}\widehat{\mathbf{Q}}_1$, $\widehat{\mathbf{C}}^{(0)} = \sqrt{p_2}\widehat{\mathbf{Q}}_2$, where $\widehat{\mathbf{Q}}_1, \widehat{\mathbf{Q}}_2$ collect the leading eigenvectors. One projected update is typically sufficient:

$$\widehat{\mathbf{X}}_t^{\mathrm{col}} = \frac{1}{p_2}\mathbf{X}_t\widehat{\mathbf{C}}^{(0)}, \qquad \widehat{\mathbf{X}}_t^{\mathrm{row}} = \frac{1}{p_1}\mathbf{X}_t'\widehat{\mathbf{R}}^{(0)},$$

$$\widetilde{\mathbf{R}} = \sqrt{p_1}\widetilde{\mathbf{Q}}_1, \qquad \widetilde{\mathbf{C}} = \sqrt{p_2}\widetilde{\mathbf{Q}}_2,$$

with $\widetilde{\mathbf{Q}}_1, \widetilde{\mathbf{Q}}_2$ obtained as the leading eigenvectors of the projected covariances.

- *Step 3: factors and idiosyncratic covariance pre-estimators.* Given $\widetilde{\mathbf{R}}, \widetilde{\mathbf{C}}$, estimate factors by least-squares projection,

$$\widehat{\mathbf{F}}_t = \frac{\widetilde{\mathbf{R}}'\mathbf{X}_t\widetilde{\mathbf{C}}}{p_1 p_2}, \qquad \widehat{\mathbf{E}}_t = \mathbf{X}_t - \widetilde{\mathbf{R}}\widehat{\mathbf{F}}_t\widetilde{\mathbf{C}}'.$$

Then obtain diagonal pre-estimators for the separable idiosyncratic covariance via

$$\widehat{\mathbf{K}}^{(0)} = \frac{1}{Tp_1}\sum_{t=1}^{T}\widehat{\mathbf{E}}_t'\widehat{\mathbf{E}}_t, \qquad \widehat{\mathbf{H}}^{(0)} = \frac{1}{Tp_2}\sum_{t=1}^{T}\widehat{\mathbf{E}}_t\widehat{\mathbf{K}}^{(0)-1}\widehat{\mathbf{E}}_t',$$

retaining only diagonal entries in implementation.

- *Step 4: number of row/column factors* (r_1, r_2). All the forecasting procedure assumes that the rank of F_t is known. However, we need to estimate it in practice. We select r_1 and r_2 through an eigenvalue-ratio (ER) rule on the projected covariances. Nevertheless, the contemporary covariance-based unfolded eigenvalue ratio estimator considered in B. Chen, Han, and Q. Yu (2024) can be still considered valid.

$$\widehat{r}_1 = \arg\max_{j\le r_{\max}} \frac{\lambda_j(\widetilde{\mathbf{M}}_1)}{\lambda_{j+1}(\widetilde{\mathbf{M}}_1) + c\delta}, \qquad \delta = \max\left\{\frac{1}{\sqrt{Tp_2}}, \frac{1}{\sqrt{Tp_1}}, \frac{1}{p_1}\right\},$$

and analogously $\widehat{r}_2$ using $\widetilde{\mathbf{M}}_2$. Because $\widetilde{\mathbf{M}}_1$ depends on $\mathbf{C}$ (and vice-versa), we iterate ER updates until $(\widehat{r}_1, \widehat{r}_2)$ stabilizes.

- *Step 5: MAR initialization (factor dynamics).* Let $\widehat{\mathbf{f}}_t := \mathrm{vec}(\widehat{\mathbf{F}}_t) \in \mathbb{R}^{r_1 r_2}$. For a VAR(1) representation

$$\widehat{\mathbf{f}}_t = (\mathbf{B}\otimes\mathbf{A})\widehat{\mathbf{f}}_{t-1} + \mathbf{u}_t,$$

the OLS pre-estimator is

$$\widehat{\mathbf{B}\otimes\mathbf{A}}^{(0)} = \left(\sum_{t=2}^{T}\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}'_{t-1}\right)\left(\sum_{t=2}^{T}\widehat{\mathbf{f}}_{t-1}\widehat{\mathbf{f}}'_{t-1}\right)^{-1},$$

and the innovation covariance is initialized as

$$\widehat{\mathbf{Q}\otimes\mathbf{P}}^{(0)} = \frac{1}{T-1}\sum_{t=2}^{T}\left(\widehat{\mathbf{f}}_t - \widehat{\mathbf{B}\otimes\mathbf{A}}^{(0)}\widehat{\mathbf{f}}_{t-1}\right)\left(\widehat{\mathbf{f}}_t - \widehat{\mathbf{B}\otimes\mathbf{A}}^{(0)}\widehat{\mathbf{f}}_{t-1}\right)'.$$

These quantities provide the starting values for the state-space EM iterations.

EM Convergence Any iterative procedure, especially within optimization frameworks, requires a well-defined convergence criterion to determine when the algorithm should stop. This also applies to the EM Algorithm. At each iteration j, the algorithm provides an updated set of estimated parameters, denoted by $\hat{\boldsymbol{\theta}}^{(j+1)}$. Based on these values, it is possible to compute the Prediction Error Log-Likelihood (PE-LLK).

The algorithm stops when the improvement in log-likelihood between two successive iterations becomes negligible. Specifically, convergence is reached when the relative difference in PE-LLK falls below an arbitrary and user-specified tolerance level ε. Formally this can be expressed as:

$$\Delta\mathscr{L}_j = \frac{\mathscr{L}(X_T;\hat{\boldsymbol{\theta}}^{(j+1)}) - \mathscr{L}(X_T;\hat{\boldsymbol{\theta}}^{(j)})}{\frac{1}{2}\left(\mathscr{L}(X_T;\hat{\boldsymbol{\theta}}^{(j+1)}) + \mathscr{L}(X_T;\hat{\boldsymbol{\theta}}^{(j)})\right)}. \tag{53}$$

For the seek of clarity, the log-likelihood $L(X_T;\boldsymbol{\theta})$ is computed using the Kalman filter, which produces one-step-ahead predictions $f_{t|t-1}$ and associated prediction error covariances $\Pi_{t|t-1}$, as follows:

$$\begin{aligned}\mathscr{L}(X_T;\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\Bigg[&\log\left|(\mathbf{C}\otimes\mathbf{R})\,\mathbf{P}_{t|t-1}\,(\mathbf{C}\otimes\mathbf{R})' + \mathrm{dg}(\mathbf{K})\otimes\mathrm{dg}(\mathbf{H})\right| \\ &+ \left(\mathbf{x}_t - (\mathbf{C}\otimes\mathbf{R})\,\mathbf{f}_{t|t-1}\right)'\left((\mathbf{C}\otimes\mathbf{R})\,\mathbf{P}_{t|t-1}\,(\mathbf{C}\otimes\mathbf{R})' + \mathrm{dg}(\mathbf{K})\otimes\mathrm{dg}(\mathbf{H})\right)^{-1}\left(\mathbf{x}_t - (\mathbf{C}\otimes\mathbf{R})\,\mathbf{f}_{t|t-1}\right)\Bigg].\end{aligned} \tag{54}$$

The numerator in Equation 53 represents the absolute improvement in the PE-LLK between two consecutive iterations of the algorithm, while the denominator provides a normalization based on the mean value of the PE-LLK at these two iterations. This expression measures the relative percentage improvement of the log-likelihood with respect to the average log-likelihood value across the two iterations.

However, convergence may not occur within a finite number of steps, or the procedure may become computationally intensive in high or ultra-high dimensional settings such as in this case. A possible solution for practical implementation is setting an additional stopping rule based on a maximum number of iterations $j_{\max}$.

Then, the EM algorithm will stop the iterative process as soon as the relative log-likelihood increment ΔL_j falls below ε, or the number of iterations reaches $j_{\max}$.

When the stopping condition is satisfied at iteration j^*, the final EM estimate is defined as $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(j^*+1)}$ and a last run of the Kalman smoother is performed to compute the smoothed estimates of the latent factor matrices.

5.2.3 Extension to missing data

Missing observations are pervasive in macroeconomic panels, especially around exceptional events (e.g. COVID-19) and in mixed-frequency settings, where the release calendar generates ragged edges. Common fixes such as aggregation or ad hoc imputations are problematic in factor models: aggregation may remove informative high-frequency variation, while simple imputations can distort the dependence structure the model aims to capture.

The EM algorithm provides a coherent likelihood-based treatment of missing values. This subsection extends EM estimation of the DMFM to general missingness patterns, building on the matrix-variate framework of Matteo Barigozzi and Trapin (2025) and the vector-state-space treatment of missing data in Bańbura and Modugno (2014). Operationally, missing entries are handled within the Kalman filter/smoother by conditioning on the observed subvector at each date, producing filtered and smoothed factors that are also suitable for nowcasting.

Imputation for the EM initialization The EM algorithm requires initial values for the parameters and, in particular, a preliminary estimate of the common component when the data contain missing observations. This is also necessary here because the PCA-based initialization in paragraph 5.2.2 presumes a balanced panel. Following Matteo Barigozzi and Trapin (2025), we obtain a first completed panel and initial second-moment estimates using the matrix-factor imputation method of Cen and Lam (2025). This approach extends the "all-purpose" covariance idea of Xiong and Pelger (2023) to matrix/tensor time series under general missingness patterns, while preserving the multi-way structure that would be weakened by full vectorization.

Let $\mathbf{X}_t \in \mathbb{R}^{p_1 \times p_2}$ denote the monthly panel at time t, where rows index countries and columns index variables, and let $\mathbf{W}_t \in \{0,1\}^{p_1 \times p_2}$ be the missingness indicator, with $(\mathbf{W}_t)_{ij} = 1$ if $X_{t,ij}$ is observed and 0 otherwise. Since $\mathbf{X}_t$ is a matrix (($K = 2$)) observed over time, exhibiting two modes. the mode-1 unfolding coincides with $\mathrm{mat}_1(\mathbf{X}_t) = \mathbf{X}_t$ (countries $\times$ variables), while the mode-2 unfolding is $\mathrm{mat}_2(\mathbf{X}_t) = \mathbf{X}_t'$ (variables $\times$ countries).

- *Step 1 (mode-wise covariance reconstruction).* For each mode, for indices i, j in that mode (such as countries) and for fibre index h in the remaining mode (such as variables), define the set of time points in which the two entries are jointly observed along fibre h (one particular variable fixed) as:

$$\psi_{k,ij,h} := \Big\{ t \in \{1, \ldots, T\} : \left[\mathrm{mat}_k(\mathbf{W}_t)\right]_{ih} \left[\mathrm{mat}_k(\mathbf{W}_t)\right]_{jh} = 1 \Big\}.$$

  Using only the available observations, the generalized mode-k covariance matrix is estimated elementwise by

$$(\widehat{\mathbf{S}}_k)_{ij} := \sum_{h=1}^{p_{-k}} \left\{ \frac{1}{|\psi_{k,ij,h}|} \sum_{t \in \psi_{k,ij,h}} \left[\mathrm{mat}_k(\mathbf{X}_t)\right]_{ih} \left[\mathrm{mat}_k(\mathbf{X}_t)\right]_{jh} \right\}, \qquad p_{-1} = p_2, \;\; p_{-2} = p_1,$$

  so that each cross-covariance is computed fibre-by-fibre by averaging over the time indices where the relevant pair is jointly observed.

- *Step 2 (loading spaces via PCA).* Let r_1 and r_2 be the chosen ranks in the country- and variable-dimensions, respectively. We apply PCA to $\widehat{\mathbf{S}}_1$ and $\widehat{\mathbf{S}}_2$ and define the preliminary loading spaces as the leading eigenvectors,

$$\widehat{\mathbf{R}} \in \mathbb{R}^{p_1 \times r_1} \;\text{ from } \widehat{\mathbf{S}}_1, \qquad \widehat{\mathbf{C}} \in \mathbb{R}^{p_2 \times r_2} \;\text{ from } \widehat{\mathbf{S}}_2,$$

  where $\widehat{\mathbf{R}}$ captures the common country-space and $\widehat{\mathbf{C}}$ captures the common variable-space in the matrix panel.

- *Step 3 (factor matrices by weighted least squares).* Given $(\widehat{\mathbf{R}}, \widehat{\mathbf{C}})$, we estimate the time-$t$ factor matrix $\mathbf{F}_t \in \mathbb{R}^{r_1 \times r_2}$ from the matrix factor representation

$$\mathbf{X}_t \approx \widehat{\mathbf{R}} \mathbf{F}_t \widehat{\mathbf{C}}'.$$

Equivalently, in vectorized form $\mathbf{x}_t = \mathrm{vec}(\mathbf{X}_t)$,

$$\mathbf{x}_t \approx (\widehat{\mathbf{C}} \otimes \widehat{\mathbf{R}}) \mathbf{f}_t, \qquad \mathbf{f}_t = \mathrm{vec}(\mathbf{F}_t).$$

Let $\mathcal{O}_t := \{(i,j) : (\mathbf{W}_t)_{ij} = 1\}$ be the set of observed entries at time t, and let $\mathbf{x}_{t,\mathcal{O}_t}$ collect the corresponding observed elements of $\mathbf{x}_t$. Denoting by $(\widehat{\mathbf{C}} \otimes \widehat{\mathbf{R}})_{\mathcal{O}_t}$ the submatrix formed by the rows associated with $\mathcal{O}_t$, we obtain $\widehat{\mathbf{f}}_t$ by least squares on the observed block,

$$\widehat{\mathbf{f}}_t = \arg\min_{\mathbf{f} \in \mathbb{R}^{r_1 r_2}} \left\| \mathbf{x}_{t,\mathcal{O}_t} - (\widehat{\mathbf{C}} \otimes \widehat{\mathbf{R}})_{\mathcal{O}_t} \mathbf{f} \right\|_2^2, \qquad \widehat{\mathbf{F}}_t = \mathrm{mat}_{r_1 \times r_2}(\widehat{\mathbf{f}}_t).$$

- *Step 4 (common component reconstruction and completion).* The preliminary common component is then reconstructed as $\widehat{\boldsymbol{\chi}}_t = \widehat{\mathbf{R}} \widehat{\mathbf{F}}_t \widehat{\mathbf{C}}'$ and an initial completed series is obtained entrywise by

$$X_{t,ij}^{(0)} = \begin{cases} X_{t,ij}, & (\mathbf{W}_t)_{ij} = 1, \\ \widehat{\chi}_{t,ij}, & (\mathbf{W}_t)_{ij} = 0. \end{cases}$$

The resulting $\{\mathbf{X}_t^{(0)}\}_{t=1}^T$ is fully observed and can be used as the input panel for the subsequent projection-based initialization of the EM algorithm.

Modification of the EM algorithm Once the initial estimates have been computed, accounting for missing values, the EM algorithm can be initialized. Even in this case, some refinements are needed. While this extension is appealing, it comes at the cost of slightly more demanding theoretical and computational considerations.

Recalling Paragraph 5.2.2, the expected log-likelihood can be decomposed into a component that depends only on the latent factors ((45)), and a component that depends on the observed data ((46)).

Since the first component involves only the latent factors, the presence of missing values does not alter its analytical form. As a result, the MAR parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{P}$, and $\mathbf{Q}$ are unaffected. On the other hand, the second component depends explicitly on the available data. Consequently, the parameters in the measurement equation—the loading matrices $\mathbf{R}$ and $\mathbf{C}$ and the idiosyncratic covariance matrices $\mathbf{H}$ and $\mathbf{K}$—must be adjusted. Missing values are handled through a selection matrix.

The M-step is therefore modified following Bańbura and Modugno (2014) for the vector case and Matteo Barigozzi and Trapin (2025) for the matrix case. Missing values are handled through the selection matrix $\mathbf{W}_t$ defined above.

Consequently, the parameters are updated at each iteration $j \geq 0$. In particular, the update formulas for the loading matrices $\mathbf{R}$ and $\mathbf{C}$, given the current parameter estimates $\widehat{\boldsymbol{\theta}}^{(j)}$, are:

$$
\begin{aligned}
\operatorname{vec}\left(\widehat{\mathbf{R}}^{(j+1)}\right) = & \left(\sum_{t=1}^{T}\sum_{s=1}^{p_1}\sum_{q=1}^{p_1}\left[\left(\widehat{\mathbf{C}}^{(j)\prime}\mathbb{D}_{\mathbf{W}_t}^{[s,q]}\widehat{\mathbf{K}}^{(j)-1}\widehat{\mathbf{C}}^{(j)}\right)\star\left(\widehat{\mathbf{f}}_{t|T}^{(j)}\widehat{\mathbf{f}}_{t|T}^{(j)\prime}+\boldsymbol{\Pi}_{t|T}^{(j)}\right)\right]\otimes\left(\mathbb{E}_{p_1,p_1}^{[s,q]}\widehat{\mathbf{H}}^{(j)-1}\right)\right)^{-1} \\
& \times\left(\sum_{t=1}^{T}\operatorname{vec}\left(\left[\mathbf{W}_t\circ\widehat{\mathbf{H}}^{(j)-1}\mathbf{X}_t\widehat{\mathbf{K}}^{(j)-1}\right]\widehat{\mathbf{C}}^{(j)}\widehat{\mathbf{f}}_{t|T}^{(j)\prime}\right)\right).
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{vec}\left(\widehat{\mathbf{C}}^{(j+1)}\right) = & \left(\sum_{t=1}^{T}\sum_{k=1}^{p_2}\sum_{q=1}^{p_2}\left[\left(\widehat{\mathbf{R}}^{(j+1)\prime}\mathbb{D}_{\mathbf{W}_t'}^{[k,q]}\widehat{\mathbf{H}}^{(j)-1}\widehat{\mathbf{R}}^{(j+1)}\right)\star\mathbf{K}_{r_1,r_2}\left(\widehat{\mathbf{f}}_{t|T}^{(j)}\widehat{\mathbf{f}}_{t|T}^{(j)\prime}+\boldsymbol{\Pi}_{t|T}^{(j)}\right)\mathbf{K}_{r_1,r_2}'\right]\otimes\left(\mathbb{E}_{p_2,p_2}^{[k,q]}\widehat{\mathbf{K}}^{(j)-1}\right)\right)^{-1} \\
& \times\left(\sum_{t=1}^{T}\operatorname{vec}\left(\left[\mathbf{W}_t\circ\widehat{\mathbf{H}}^{(j)-1}\mathbf{X}_t\widehat{\mathbf{K}}^{(j)-1}\right]'\widehat{\mathbf{R}}^{(j+1)}\widehat{\mathbf{f}}_{t|T}^{(j)}\right)\right).
\end{aligned}
$$

For the same iteration $j \geq 0$ and current estimates $\widehat{\boldsymbol{\theta}}^{(j)}$, the idiosyncratic variances $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{K}}$ in the presence of missing data are computed as:

$$
\begin{aligned}
\left[\widehat{\mathbf{H}}^{(j+1)}\right]_{ii} = \frac{1}{Tp_2}\sum_{t=1}^{T}\Bigg\{ & (\mathbf{W}_t\circ\mathbf{X}_t)\widehat{\mathbf{K}}^{(j)-1}(\mathbf{W}_t\circ\mathbf{X}_t)' - (\mathbf{W}_t\circ\mathbf{X}_t)\widehat{\mathbf{K}}^{(j)-1}\left(\mathbf{W}_t\circ\left[\widehat{\mathbf{R}}^{(j+1)}\widehat{\mathbf{f}}_{t|T}^{(j)}\widehat{\mathbf{C}}^{(j+1)\prime}\right]\right)' \\
& -\left(\mathbf{W}_t\circ\left[\widehat{\mathbf{R}}^{(j+1)}\widehat{\mathbf{f}}_{t|T}^{(j)}\widehat{\mathbf{C}}^{(j+1)\prime}\right]\right)\widehat{\mathbf{K}}^{(j)-1}(\mathbf{W}_t\circ\mathbf{X}_t)' \\
& +\widehat{\mathbf{K}}^{(j)-1}\star\left[\mathbb{D}_{\mathbf{W}_t}\left(\widehat{\mathbf{C}}^{(j+1)}\otimes\widehat{\mathbf{R}}^{(j+1)}\right)\left(\widehat{\mathbf{f}}_{t|T}^{(j)}\widehat{\mathbf{f}}_{t|T}^{(j)\prime}+\boldsymbol{\Pi}_{t|T}^{(j)}\right)\left(\widehat{\mathbf{C}}^{(j+1)}\otimes\widehat{\mathbf{R}}^{(j+1)}\right)'\mathbb{D}_{\mathbf{W}_t}'\right] \\
& +\left[\widehat{\mathbf{H}}^{(j)}\mathbf{1}_{p_1,p_2}\widehat{\mathbf{K}}^{(j)}\right]\widehat{\mathbf{K}}^{(j)-1}\left(\mathbf{1}_{p_1,p_2}-\mathbf{W}_t\right)'\Bigg\}_{ii}.
\end{aligned}
$$

$$
\begin{aligned}
\left[\widehat{\mathbf{K}}^{(j+1)}\right]_{ii} = \frac{1}{Tp_1}\sum_{t=1}^{T}\Bigg\{ & (\mathbf{W}_t\circ\mathbf{X}_t)'\widehat{\mathbf{H}}^{(j+1)-1}(\mathbf{W}_t\circ\mathbf{X}_t) - (\mathbf{W}_t\circ\mathbf{X}_t)'\widehat{\mathbf{H}}^{(j+1)-1}\left(\mathbf{W}_t\circ\left[\widehat{\mathbf{R}}^{(j+1)}\widehat{\mathbf{f}}^{(j)}_{t|T}\widehat{\mathbf{C}}^{(j+1)\prime}\right]\right)\\
& - \left(\mathbf{W}_t\circ\left[\widehat{\mathbf{R}}^{(j+1)}\widehat{\mathbf{f}}^{(j)}_{t|T}\widehat{\mathbf{C}}^{(j+1)\prime}\right]\right)'\widehat{\mathbf{H}}^{(j+1)-1}(\mathbf{W}_t\circ\mathbf{X}_t)\\
& + \widehat{\mathbf{H}}^{(j+1)-1}\star\left[\mathbb{D}'_{\mathbf{W}_t}\left(\widehat{\mathbf{R}}^{(j+1)}\otimes\widehat{\mathbf{C}}^{(j+1)}\right)\mathbb{K}_{r_1,r_2}\left(\widehat{\mathbf{f}}^{(j)}_{t|T}\widehat{\mathbf{f}}^{(j)\prime}_{t|T}+\mathbf{\Pi}^{(j)}_{t|T}\right)\mathbb{K}'_{r_1,r_2}\left(\widehat{\mathbf{R}}^{(j+1)}\otimes\widehat{\mathbf{C}}^{(j+1)}\right)'\mathbb{D}_{\mathbf{W}_t}\right]\\
& + (\mathbf{1}_{p_1,p_2}-\mathbf{W}_t)'\widehat{\mathbf{H}}^{(j+1)-1}\left(\widehat{\mathbf{H}}^{(j+1)}\mathbf{1}_{p_1,p_2}\widehat{\mathbf{K}}^{(j)}\right)\Bigg\}_{ii}.
\end{aligned}
$$

5.3 Modelling serial correlation in the idiosyncratic component

MODELLING SERIAL CORRELATION IN THE IDIOSYNCRATIC COMPONENT[12]

In the approximate DMFM in (36)–(37), the idiosyncratic component is treated as a measurement disturbance with a parsimonious diagonal covariance. The diagonal restriction controls cross-sectional dependence, while the measurement-disturbance assumption implies that it is serially uncorrelated (i.i.d. over t) and therefore carries no idiosyncratic persistence. However, in forecasting applications an idiosyncratic persistence can be beneficial since it allows to extrapolate part of the series-specific dynamics and improve the efficiency of the common-factor extraction, especially in real-time settings. Following Bańbura and Modugno (2010), we therefore introduce a persistent idiosyncratic component (e.g. AR(1)) by augmenting the state vector, while preserving the bilinear structure of the DMFM.

This augmented formulation changes the interpretation of the pseudo-likelihood covariance in (40). Rather than accounting for the entire idiosyncratic variability as measurement noise, the persistent portion is moved into the state dynamics and is governed by the innovation variance of the idiosyncratic transition, whereas the measurement equation retains only a small residual noise term used as a nugget for regularization and numerical stability.

5.3.1 The Model

Consider the approximate DMFM in Subsection 5.1. To allow for serial correlation in the idiosyncratic component, decompose the error term in (36) as

$$
\mathbf{E}_t = \widetilde{\mathbf{E}}_t + \mathbf{\Xi}_t, \qquad \mathrm{vec}(\mathbf{\Xi}_t) \sim \mathscr{N}(\mathbf{0}, \kappa\mathbf{I}_{p_1p_2}), \tag{55}
$$

[12]TO REVIEW: Completely new section

where $\widetilde{\mathbf{E}}_t$ is the persistent idiosyncratic component and $\boldsymbol{\Xi}_t$ is a small measurement "nugget" with fixed variance $\kappa > 0$, included to avoid degeneracy of the measurement-error covariance in the augmented state-space form.

We model $\widetilde{\mathbf{E}}_t$ as a MAR process and, without loss of generality, set order $P_E = 1$:

$$\widetilde{\mathbf{E}}_t = \mathbf{A}_E \widetilde{\mathbf{E}}_{t-1} \mathbf{B}_E' + \boldsymbol{\Psi}_{E,t}, \tag{56}$$

where $\mathbf{A}_E \in \mathbb{R}^{p_1 \times p_1}$ and $\mathbf{B}_E \in \mathbb{R}^{p_2 \times p_2}$ govern row/column propagation and $\boldsymbol{\Psi}_{E,t}$ is the innovation matrix.

- *Diagonal AR propagation (no dynamic spillovers)* We impose diagonal autoregressive matrices,

  $$\mathbf{A}_E = \mathrm{diag}(a_1, \ldots, a_{p_1}), \qquad \mathbf{B}_E = \mathrm{diag}(b_1, \ldots, b_{p_2}), \tag{57}$$

  so that each entry follows a scalar AR(1),

  $$\tilde{e}_{ij,t} = (a_i b_j)\, \tilde{e}_{ij,t-1} + \psi_{ij,t}, \qquad i = 1, \ldots, p_1, \;\; j = 1, \ldots, p_2. \tag{58}$$

  Hence each idiosyncratic series depends only on its own lag, with a two-way coefficient $\phi_{ij} = a_i b_j$. Stationarity requires $|a_i b_j| < 1$ for all i, j. Relative to an unrestricted entrywise AR(1) with $p_1 p_2$ coefficients, the restriction $\phi_{ij} = a_i b_j$ uses $(p_1 + p_2)$ parameters (up to a scale normalization).

- *Innovation covariance (separable entrywise variances and no cross-covariances)* We assume

  $$\mathrm{vec}(\boldsymbol{\Psi}_{E,t}) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_E), \qquad \mathbf{Q}_E = \mathrm{dg}(\mathbf{K}_E) \otimes \mathrm{dg}(\mathbf{H}_E), \tag{59}$$

  where $\mathbf{H}_E \in \mathbb{R}^{p_1 \times p_1}$ and $\mathbf{K}_E \in \mathbb{R}^{p_2 \times p_2}$ are diagonal with positive entries. Entrywise, this implies

  $$\mathrm{Var}(\psi_{ij,t}) = h_i k_j, \qquad \mathrm{Cov}(\psi_{ij,t}, \psi_{i'j',t}) = 0 \text{ for } (i,j) \neq (i',j'). \tag{60}$$

  Relative to an unrestricted diagonal covariance with $p_1 p_2$ free variances $\{\mathrm{Var}(\psi_{ij,t})\}$, the separable parametrization uses $(p_1 + p_2)$ parameters $\{h_i\}_{i=1}^{p_1}$ and $\{k_j\}_{j=1}^{p_2}$ (up to a scale normalization, since only the products $h_i k_j$ are identified).

5.3.2 Estimation

Let $\mathbf{x}_t = \text{vec}(\mathbf{X}_t) \in \mathbb{R}^{p_1 p_2}$, $\mathbf{f}_t = \text{vec}(\mathbf{F}_t) \in \mathbb{R}^{r_1 r_2}$, $\tilde{\mathbf{e}}_t = \text{vec}(\widetilde{\mathbf{E}}_t) \in \mathbb{R}^{p_1 p_2}$, and $\boldsymbol{\xi}_t = \text{vec}(\boldsymbol{\Xi}_t) \in \mathbb{R}^{p_1 p_2}$. Fix a small $\kappa > 0$ and define the augmented state $\mathbf{s}_t := (\mathbf{f}_t', \tilde{\mathbf{e}}_t')' \in \mathbb{R}^{r_1 r_2 + p_1 p_2}$. The vectorized DMFM with a persistent idiosyncratic component can be written as the linear–Gaussian state-space system

$$\mathbf{x}_t = \underbrace{\begin{bmatrix} \mathbf{C} \otimes \mathbf{R} & \mathbf{I}_{p_1 p_2} \end{bmatrix}}_{\mathbf{Z}} \mathbf{s}_t + \boldsymbol{\xi}_t, \qquad \boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \kappa \mathbf{I}_{p_1 p_2}), \tag{61}$$

$$\mathbf{s}_t = \underbrace{\begin{bmatrix} \mathbf{B} \otimes \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_E \otimes \mathbf{A}_E \end{bmatrix}}_{\mathbf{T}} \mathbf{s}_{t-1} + \underbrace{\begin{bmatrix} \mathbf{u}_t \\ \boldsymbol{\upsilon}_t \end{bmatrix}}_{\boldsymbol{\eta}_t}, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_s), \tag{62}$$

$$\mathbf{Q}_s = \begin{bmatrix} \mathbf{Q} \otimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_E \end{bmatrix}, \qquad \mathbf{Q}_E = \text{dg}(\mathbf{K}_E) \otimes \text{dg}(\mathbf{H}_E), \tag{63}$$

where $\mathbf{H}_E$ and $\mathbf{K}_E$ are diagonal with positive entries. When data are missing, inference is carried out by applying the Kalman filter and smoother to the observed subvector at each date. The full recursion (prediction, update, and smoothing) is reported in Appendix B.

Using the Kalman filter (applied to the observed subvector when data are missing), let $\boldsymbol{\nu}_{t,\mathscr{O}_t}$ and $\mathbf{S}_{t,\mathscr{O}_t}$ denote the one-step-ahead innovation and its variance, computed on the observed block $\mathscr{O}_t$. Then, the observed-data likelihood admits the prediction-error (innovation) decomposition

$$\ell(X_T; \boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{T} \Big(m_t \log(2\pi) + \log |\mathbf{S}_{t,\mathscr{O}_t}| + \boldsymbol{\nu}_{t,\mathscr{O}_t}' \mathbf{S}_{t,\mathscr{O}_t}^{-1} \boldsymbol{\nu}_{t,\mathscr{O}_t} \Big), \tag{64}$$

where $m_t := |\mathscr{O}_t|$ is the number of observed entries in $\mathbf{x}_t$ at time t.

Complete-data log-likelihood and EM Algorithm

- **E-step (Kalman smoother).** For any EM iteration $j \geq 0$, given current estimates $\widehat{\boldsymbol{\theta}}^{(j)}$, we run the Kalman smoother on the augmented state-space system (61)–(62) to obtain the conditional moments of $\{\mathbf{s}_t\}_{t=1}^T$. Given $\widehat{\boldsymbol{\theta}}^{(j)}$, the observed-data log-likelihood admits the EM decomposition

$$\ell(X_T;\boldsymbol{\theta}) = \underbrace{\mathbb{E}_{\widehat{\boldsymbol{\theta}}^{(j)}}\Big[\ell(X_T,F_T,\widetilde{E}_T;\boldsymbol{\theta}) \,\Big|\, X_T\Big]}_{Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)})} - \underbrace{\mathbb{E}_{\widehat{\boldsymbol{\theta}}^{(j)}}\Big[\ell(F_T,\widetilde{E}_T \mid X_T;\boldsymbol{\theta}) \,\Big|\, X_T\Big]}_{\text{conditional entropy (constant in } \boldsymbol{\theta})}, \tag{65}$$

where the second term does not depend on $\boldsymbol{\theta}$ and can be ignored in the M-step. Under Gaussianity, the complete-data log-likelihood decomposes (up to constants) as

$$\ell(X_T,F_T,\widetilde{E}_T;\boldsymbol{\theta}) = \ell_X(X_T \mid F_T,\widetilde{E}_T;\boldsymbol{\theta}) + \ell_F(F_T;\boldsymbol{\theta}) + \ell_E(\widetilde{E}_T;\boldsymbol{\theta}), \tag{66}$$

so that

$$Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)}) = Q_X(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)}) + Q_F(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)}) + Q_E(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)}). \tag{67}$$

- *Measurement equation.* The measurement equation is $\mathbf{X}_t = \mathbf{R}\mathbf{F}_t\mathbf{C}' + \widetilde{\mathbf{E}}_t + \boldsymbol{\Xi}_t$ with $\mathrm{vec}(\boldsymbol{\Xi}_t) \sim \mathcal{N}(\mathbf{0}, \kappa\mathbf{I}_{p_1p_2})$ and fixed $\kappa > 0$. Hence

$$Q_X(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)}) = -\frac{Tp_1p_2}{2}\log\kappa - \frac{1}{2\kappa}\sum_{t=1}^T \mathbb{E}_{\widehat{\boldsymbol{\theta}}^{(j)}}\left[\left\|\mathbf{X}_t - \mathbf{R}\mathbf{F}_t\mathbf{C}' - \widetilde{\mathbf{E}}_t\right\|_F^2 \,\Big|\, X_T\right]. \tag{68}$$

- *Factor dynamics.* The factor dynamics are $\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1}\mathbf{B}' + \mathbf{U}_t$ with $\mathrm{vec}(\mathbf{U}_t) \sim \mathcal{N}(\mathbf{0},\mathbf{Q}\otimes\mathbf{P})$, so that

$$\begin{aligned} Q_F(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)}) = &-\frac{T-1}{2}\Big(r_1\log|\mathbf{Q}| + r_2\log|\mathbf{P}|\Big) \\ &-\frac{1}{2}\sum_{t=2}^T \mathbb{E}_{\widehat{\boldsymbol{\theta}}^{(j)}}\left[\mathrm{tr}\Big(\mathbf{P}^{-1}(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1}\mathbf{B}')\mathbf{Q}^{-1}(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1}\mathbf{B}')'\Big) \,\Big|\, X_T\right]. \end{aligned} \tag{69}$$

- *Persistent idiosyncratic dynamics.* Finally, $\widetilde{\mathbf{E}}_t = \mathbf{A}_E\widetilde{\mathbf{E}}_{t-1}\mathbf{B}_E' + \boldsymbol{\Psi}_{E,t}$ with $\mathrm{vec}(\boldsymbol{\Psi}_{E,t}) \sim \mathcal{N}(\mathbf{0},\mathbf{Q}_E)$ and $\mathbf{Q}_E = \mathrm{dg}(\mathbf{K}_E)\otimes\mathrm{dg}(\mathbf{H}_E)$. Therefore

$$\begin{aligned} Q_E(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}^{(j)}) = &-\frac{T-1}{2}\Big(p_1\log|\mathrm{dg}(\mathbf{K}_E)| + p_2\log|\mathrm{dg}(\mathbf{H}_E)|\Big) \\ &-\frac{1}{2}\sum_{t=2}^T \mathbb{E}_{\widehat{\boldsymbol{\theta}}^{(j)}}\left[\mathrm{tr}\Big(\mathrm{dg}(\mathbf{H}_E)^{-1}(\widetilde{\mathbf{E}}_t - \mathbf{A}_E\widetilde{\mathbf{E}}_{t-1}\mathbf{B}_E')\mathrm{dg}(\mathbf{K}_E)^{-1}(\widetilde{\mathbf{E}}_t - \mathbf{A}_E\widetilde{\mathbf{E}}_{t-1}\mathbf{B}_E')'\Big) \,\Big|\, X_T\right]. \end{aligned} \tag{70}$$

All conditional expectations in (68)–(70) are obtained from the Kalman smoother applied to

(61)–(62).

- **M-step.**

The M-step updates for the factor-dynamics block $(\mathbf{A}, \mathbf{B}, \mathbf{P}, \mathbf{Q})$ retain the same closed forms as in the approximate DMFM with a serially and cross-sectionally uncorrelated idiosyncratic component, since Q_F is unaffected by the introduction of the persistent idiosyncratic state. The loading updates for $(\mathbf{R}, \mathbf{C})$ also preserve the same bilinear least-squares structure, but are computed using the "cleaned" observation $\mathbf{X}_t - \widetilde{\mathbf{E}}_t$ (in conditional expectation) and the smoothed moments produced by the Kalman smoother applied to the augmented system. We therefore focus on the parameters governing the persistent idiosyncratic block $(\mathbf{A}_E, \mathbf{B}_E, \mathbf{H}_E, \mathbf{K}_E)$.

Let $\mathbf{A}_E = \mathrm{diag}(a_1, \dots, a_{p_1})$, $\mathbf{B}_E = \mathrm{diag}(b_1, \dots, b_{p_2})$, $\mathbf{H}_E = \mathrm{diag}(h_1, \dots, h_{p_1})$, and $\mathbf{K}_E = \mathrm{diag}(k_1, \dots, k_{p_2})$. The part of $Q_E(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(j)})$ that depends on (a_i) and (b_j) can be written entrywise as

$$\sum_{i=1}^{p_1} \sum_{j=1}^{p_2} \frac{1}{h_i k_j} \sum_{t=2}^{T} \mathbb{E}_{\widehat{\boldsymbol{\theta}}^{(j)}} \left[\left(\tilde{e}_{ij,t} - a_i b_j \tilde{e}_{ij,t-1} \right)^2 \,\middle|\, X_T \right], \tag{71}$$

where $\tilde{e}_{ij,t}$ denotes the (i, j) entry of $\widetilde{\mathbf{E}}_t$.

Define the sufficient statistics

$$S_{ij}^{(10)} = \sum_{t=2}^{T} \mathbb{E}\left[\tilde{e}_{ij,t} \tilde{e}_{ij,t-1} \mid X_T \right], \qquad S_{ij}^{(00)} = \sum_{t=2}^{T} \mathbb{E}\left[\tilde{e}_{ij,t-1}^2 \mid X_T \right], \tag{72}$$

which are obtained from the smoothed mean and covariance of the augmented state $\mathbf{s}_t = (\mathbf{f}_t', \tilde{\mathbf{e}}_t')'$.

Joint maximization in (a_i) and (b_j) does not admit a closed-form solution. However, conditional on one set of parameters, the other admits a closed-form update. We therefore implement an alternating least-squares (ALS) scheme within the M-step. Specifically, given $(b_j^{(j)})$ and $(k_j^{(j)})$, update (a_i) as

$$a_i^{(j+1)} = \frac{\sum_{j=1}^{p_2} \frac{b_j^{(j)}}{k_j^{(j)}} S_{ij}^{(10)}}{\sum_{j=1}^{p_2} \frac{\left(b_j^{(j)}\right)^2}{k_j^{(j)}} S_{ij}^{(00)}}, \qquad i = 1, \dots, p_1, \tag{73}$$

and, given $(a_i^{(j+1)})$ and $(h_i^{(j)})$, update (b_j) as

$$b_j^{(j+1)} = \frac{\sum_{i=1}^{p_1} \dfrac{a_i^{(j+1)}}{h_i^{(j)}} S_{ij}^{(10)}}{\sum_{i=1}^{p_1} \dfrac{\big(a_i^{(j+1)}\big)^2}{h_i^{(j)}} S_{ij}^{(00)}}, \qquad j = 1, \dots, p_2. \tag{74}$$

Given $(a_i^{(j+1)})$ and $(b_j^{(j+1)})$, define the expected residual sums of squares

$$\mathrm{RSS}_{ij}^{(j+1)} = \sum_{t=2}^{T} \mathbb{E}\Big[\big(\tilde{e}_{ij,t} - a_i^{(j+1)} b_j^{(j+1)} \tilde{e}_{ij,t-1}\big)^2 \;\Big|\; X_T\Big]. \tag{75}$$

Then the innovation-variance parameters can be updated by alternating scaling. Using $(k_j^{(j)})$, update (h_i) as

$$h_i^{(j+1)} = \frac{1}{(T-1)p_2} \sum_{j=1}^{p_2} \frac{R_{ij}^{(j+1)}}{k_j^{(j)}}, \qquad i = 1, \dots, p_1, \tag{76}$$

and, given $(h_i^{(j+1)})$, update (k_j) as

$$k_j^{(j+1)} = \frac{1}{(T-1)p_1} \sum_{i=1}^{p_1} \frac{R_{ij}^{(j+1)}}{h_i^{(j+1)}}, \qquad j = 1, \dots, p_2. \tag{77}$$

The parametrization $\phi_{ij} = a_i b_j$ is invariant under $(a_i, b_j) \mapsto (c a_i, b_j / c)$ for any $c \neq 0$, and the same scale indeterminacy holds for (h_i, k_j) in $h_i k_j$. Therefore, after each ALS cycle we impose a normalization such as $\mathrm{tr}(\mathbf{H}_E^{(j+1)}) = p_1$ (equivalently, $\frac{1}{p_1} \sum_i h_i^{(j+1)} = 1$) and rescale $\mathbf{K}_E^{(j+1)}$ accordingly.

Extension to missing values. With the persistent idiosyncratic component included in the state, the EM algorithm proceeds as in the baseline case, but the Kalman filter/smoother is applied to the *augmented* system and, under missingness, only to the observed subvector selected by $\mathbf{W}_t$. As a consequence, the M-step updates for the factor-dynamics block $(\mathbf{A}, \mathbf{B}, \mathbf{P}, \mathbf{Q})$ keep the same closed forms, since they depend only on the latent factor state once conditioning is taken on the available information.

The loading updates for $(\mathbf{R}, \mathbf{C})$ preserve the bilinear least-squares structure, but are computed on the *cleaned* observation, i.e. $\mathbf{W}_t \circ \left(\mathbf{X}_t - \mathbb{E}[\widetilde{\mathbf{E}}_t \mid X_T]\right)$, where the conditional moments are obtained from the smoother under current parameters.

Finally, the idiosyncratic block $(\mathbf{A}_E, \mathbf{B}_E, \mathbf{H}_E, \mathbf{K}_E)$ is updated entrywise using sufficient statistics built from the smoothed mean and covariance of the augmented state. Their definitions are unchanged, but they are implicitly computed conditional on the observed data only; when many entries are missing, the smoother relies more on the state dynamics, while observed entries contribute through the measurement update. Joint updates of row/column AR coefficients are carried out via an alternating least-squares (ALS) scheme, followed by a normalization to remove the scale indeterminacy of the bilinear parametrization.

5.4 Nowcasting

To produce nowcasts with the DMFM, we adapt the pseudo real-time procedure of Bańbura and Modugno (2014) to the matrix-variate setting. The implementation relies on a linear–Gaussian state-space representation obtained by vectorizing the DMFM as in (42)–(43).

Let v index pseudo real-time vintages, with cutoff date t_v on the monthly grid, and denote by Ω_v the corresponding information set. The ragged-edge dataset at vintage v is obtained by imposing the release calendar and publication delays up to t_v, yielding $\{\mathbf{X}_t^{(v)}\}_{t \leq t_v}$ (and the associated availability mask, if used in the implementation).

At each vintage v, the model is re-estimated using the information available up to t_v and the Kalman filter is run under missing observations. Let $\mathbf{s}_t$ denote the *augmented* latent state in the vectorized system, collecting the dynamic factors and, when included, persistent idiosyncratic components and/or the lags required by the mixed-frequency scheme. Since no forward information is available at the cutoff date t_v, the nowcast is based on the filtered state $\widehat{\mathbf{s}}_{t_v|t_v}^{(v)} = \mathbb{E}(\mathbf{s}_{t_v} \mid \Omega_v)$.

In matrix form, the filtered common component can be reconstructed from the filtered factors as

$$\widehat{\boldsymbol{\chi}}_{t_v|t_v}^{(v)} = \widehat{\mathbf{R}}^{(v)}\, \widehat{\mathbf{F}}_{t_v|t_v}^{(v)}\, \widehat{\mathbf{C}}^{(v)\prime},$$

noting that $\widehat{\mathbf{F}}_{t_v|t_v}^{(v)}$ is itself a sub-block (or a linear projection) of $\widehat{\mathbf{s}}_{t_v|t_v}^{(v)}$.

Let y_τ denote the quarterly target (GDP) for quarter τ. Under mixed-frequency restrictions, $y_{\tau(v)}$ is a known linear function of the augmented state at the cutoff date. Denote by $\boldsymbol{\ell}$ the corresponding selector/aggregation vector (consistent with Mariano–Murasawa or stock/flow, and with the chosen state augmentation). Then the vintage-v nowcast is

$$\widehat{y}_{\tau(v)}^{(v)} \;:=\; \mathbb{E}\big[y_{\tau(v)} \,\big|\, \Omega_v\big] \;=\; \boldsymbol{\ell}'\widehat{\mathbf{s}}_{t_v|t_v}^{(v)}. \tag{78}$$

The forecast error and RMSFE are defined as in (34)–(35).

6 Empirics

6.1 Data Description

6.2 Dataset Preparation

In what follows I describe row data and the initial data manipulation applied to all the models explored in this paper.

- **Dataset and sources:** mixed–frequency *EA–MD–QD* database from M. Barigozzi and Lissona (2024), collecting macroeconomic and financial indicators from institutional providers (e.g. Eurostat, ECB, OECD, national statistical institutes) and already delivered in de-seasonalized form.
- **Countries Considered:** Germany (DE), France (FR), Italy (IT), and Spain (ES).
- **Time span:** April 2000 – October 2025.[13]
- **Variable blocks and economic type:** the panel includes multiple groups of indicators:
  - *Real activity* (`Class = R`): national accounts in volumes and other quantity measures (production, turnover, employment/unemployment, etc.).
  - *Nominal and prices* (`Class = N`): price indices and deflators (HICP, PPI, GDP deflator), nominal aggregates and value measures.
  - *Financial/monetary* (`Class = F`): monetary aggregates, interest rates, credit/financial quantities, exchange rates and related series.
  - *Surveys/confidence* (`Class = C`): business and consumer confidence indicators and other qualitative measures.
- **Frequency handling:** the dataset contains monthly and quarterly series. Quarterly series are aligned by placing each quarterly observation in the third month of the corresponding quarter.
- **Transformations and stationarity (light transformations):** all series are transformed applying the `light transformation` option, i.e. series are assumed to be at most $I(1)$ (hence no $I(2)$ transformations are imposed). Operationally, for a generic series x_t,

[13] The sample starts in April 2000 to avoid missing values in the annual growth rate at the beginning of the time span, when the EA aggregate did not yet exist in the previous year.

- `TR=0`: $\tilde{x}_t = x_t$ (levels, already stationary);
- `TR=1`: $\tilde{x}_t = 100\log(x_t)$ (log-level, stationary in log units);
- `TR=2`: $\tilde{x}_t = 100\Delta\log(x_t)$ (log-difference / growth rate);
- `TR=4`: $\tilde{x}_t = \Delta x_t$ (first difference in levels),

where Δ is the first-difference operator.

- **Crises and missing data treatment:**
  - *Global Financial Crisis (2008–2009):* no special treatment for any class of variables.
  - *COVID (March 2020 – July 2021):* real variables (`Class = R`) are *masked* by setting the corresponding observations to missing values during the Covid window; crucially, these masked values are *not* imputed in our analysis (i.e. we do not run any Covid-specific EM/Kalman/PCA reconstruction of the masked block).
- **Types of missingness:**
  - Regular missingness induced by the mixed-frequency structure (quarterly variables and publication lags).
  - Covid-induced blocks of missing values for real variables during the masked period (no imputation).
  - Ragged-edge missingness from heterogeneous release calendars, handled in the pseudo real–time nowcasting exercise.

6.3 Standardization under missing data

Standardization. Let $\mathbf{X}^{\text{raw}} \in \mathbb{R}^{T_m \times N}$ denote the high-frequency panel, with entries $x_{t,i}^{\text{raw}}$, allowing for missing observations. For each series i, define the set of observed time indices

$$\mathscr{O}_i := \{t \in \{1, \dots, T_m\} : x_{t,i}^{\text{raw}} \text{ is observed}\}, \qquad |\mathscr{O}_i| > 0,$$

and the observation indicator

$$w_{t,i} := \mathbb{1}\{t \in \mathscr{O}_i\} \in \{0, 1\}.$$

Compute the mean and standard deviation using available observations only:

$$\mu_i := \frac{1}{|\mathscr{O}_i|} \sum_{t \in \mathscr{O}_i} x_{t,i}^{\text{raw}}, \qquad \sigma_i := \left(\frac{1}{|\mathscr{O}_i|} \sum_{t \in \mathscr{O}_i} (x_{t,i}^{\text{raw}} - \mu_i)^2 \right)^{1/2}.$$

The standardized panel $\mathbf{X}^{\text{std}} \in \mathbb{R}^{T_m \times N}$ has entries, for $t \in \mathscr{O}_i$,

$$x_{t,i}^{\text{std}} := \frac{x_{t,i}^{\text{raw}} - \mu_i}{\sigma_i},$$

and remains missing for $t \notin \mathscr{O}_i$.

De-standardization. Any standardized estimate/forecast $\widehat{x}_{t,i}^{\text{std}}$ is mapped back to the original scale as

$$\widehat{x}_{t,i}^{\text{raw}} = \mu_i + \sigma_i \widehat{x}_{t,i}^{\text{std}}.$$

References

Ahn, Seung C and Alex R Horenstein (2013). "Eigenvalue ratio test for the number of factors". In: *Econometrica* 81.3, pp. 1203–1227.

Babii, Andrii, Eric Ghysels, and Junsu Pan (2022). "Tensor principal component analysis". In: *arXiv*. arXiv:2212.12981.

Bai, Jushan and Serena Ng (2002). "Determining the number of factors in approximate factor models". In: *Econometrica* 70.1, pp. 191–221.

– (2006). "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions". In: *Econometrica* 74.4, pp. 1133–1150.

– (2008). "Forecasting economic time series using targeted predictors". In: *Journal of Econometrics* 146.2, pp. 304–317.

Bańbura, Marta and Michele Modugno (2010). *Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data*. Tech. rep. ECB Working Paper.

– (2014). "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data". In: *Journal of applied econometrics* 29.1, pp. 133–160.

Barigozzi, M. and C. Lissona (2024). *EA-MD-QD: Large Euro Area and Euro Member Countries Datasets for Macroeconomic Research (Version 12.2023)*. Data set. URL: https://doi.org/10.5281/ZENODO.10514668.

Barigozzi, Matteo, Marco Lippi, and Matteo Luciani (2021). "Large-dimensional dynamic factor models: Estimation of impulse–response functions with I (1) cointegrated factors". In: *Journal of Econometrics* 221.2, pp. 455–482.

Barigozzi, Matteo and Luca Trapin (2025). "Quasi maximum likelihood estimation of high-dimensional approximate dynamic matrix factor models via the EM algorithm". In: *arXiv preprint arXiv:2502.04112*.

Billio, Monica et al. (2023). "Bayesian dynamic tensor regression". In: *Journal of Business & Economic Statistics* 41.2, pp. 429–439.

Cai, Xiong et al. (2025). "Matrix-factor-augmented regression". In: *Journal of Business & Economic Statistics*, pp. 1–13.

Cen, Zetai and Clifford Lam (2025). "Tensor time series imputation through tensor factor modelling". In: *Journal of Econometrics* 249, p. 105974.

Chen, Bin, Elynn Y Chen, et al. (2024). “Time-varying matrix factor models”. In: *arXiv preprint arXiv:2404.01546*.

Chen, Bin, Yuefeng Han, and Qiyang Yu (2024). “Estimation and inference for CP tensor factor models”. In: *arXiv preprint arXiv:2406.17278*.

– (2025). “Diffusion Index Forecast with Tensor Data”. In: *Available at SSRN 5202316*.

Chen, Elynn Y and Rong Chen (2019). “Modeling dynamic transport network with matrix factor models: with an application to international trade flow”. In: *arXiv preprint arXiv:1901.00769*.

Chen, Rong, Han Xiao, and Dan Yang (2021). “Autoregressive models for matrix-valued time series”. In: *Journal of Econometrics* 222.1, pp. 539–560.

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008). “Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?” In: *Journal of Econometrics* 146.2, pp. 318–328.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the royal statistical society: series B (methodological)* 39.1, pp. 1–22.

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin (2012). “A quasi–maximum likelihood approach for large, approximate dynamic factor models”. In: *Review of economics and statistics* 94.4, pp. 1014–1024.

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov (2007). “MIDAS regressions: Further results and new directions”. In: *Econometric reviews* 26.1, pp. 53–90.

Giannone, Domenico, Lucrezia Reichlin, and David Small (2008). “Nowcasting: The real-time informational content of macroeconomic data”. In: *Journal of monetary economics* 55.4, pp. 665–676.

Hepenstrick, Christian and Massimiliano Marcellino (2016). *Forecasting with Large Unbalanced Datasets: The Mixed Frequency Three-Pass Regression Filter*. Tech. rep. Swiss National Bank.

Kapetanios, George, Laura Serlenga, and Yongcheol Shin (2021). “Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure”. In: *Journal of Econometrics* 220.2, pp. 504–531.

Kelly, Bryan and Seth Pruitt (2015). “The three-pass regression filter: A new approach to forecasting using many predictors”. In: *Journal of Econometrics* 186.2, pp. 294–316.

Krämer, Nicole and Masashi Sugiyama (2011). “The degrees of freedom of partial least squares regression”. In: *Journal of the American Statistical Association* 106.494, pp. 697–705.

Mariano, Roberto S and Yasutomo Murasawa (2003). “A new coincident index of business cycles based on monthly and quarterly series”. In: *Journal of applied Econometrics* 18.4, pp. 427–443.

Stock, James H. and Mark W. Watson (2002). “Forecasting Using Principal Components from a Large Number of Predictors”. In: *Journal of the American Statistical Association* 97.460, pp. 1167–79.

Wang, Dong, Xialu Liu, and Rong Chen (2019). “Factor models for matrix-valued high-dimensional time series”. In: *Journal of econometrics* 208.1, pp. 231–248.

Xiong, Ruoxuan and Markus Pelger (2023). “Large dimensional latent factor modeling with missing observations and applications to causal inference”. In: *Journal of Econometrics* 233.1, pp. 271–301.

Yu, Long et al. (2022). “Projected estimation for large-dimensional matrix factor models”. In: *Journal of Econometrics* 229.1, pp. 201–217.

Zhu, Ziwei, Tengyao Wang, and Richard J Samworth (2022). “High-dimensional principal component analysis with heterogeneous missingness”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.5, pp. 2000–2031.

A Notation

Table 3: Notation: Indices, dimensions, calendars, information sets, missing-data, and basic operators

Symbol	Type / dim.	Meaning
Indices, calendars, and information sets		
$t,\ \tau$	indices	month / quarter index
$T,\ \mathscr{T}$	scalars	number of months / quarters in sample
m	index	month-in-quarter, $m \in \{1,2,3\}$
v	index	pseudo real-time vintage index
t_v	scalar	monthly cutoff date for vintage v
Ω_T, Ω_v	set	information set available at T/t_v
$\tau(v), m(v)$	index	quarter containing t_v/ month-in-quarter for t_v
$\mathscr{V}$	set	evaluation vintages (realization eventually observed)
i	index	series index
Target variable and related forecast matrices		
y_t, y_τ	scalars	monthly/quarterly target
$\widehat{y}^{(v)}_{t(v)}, \widehat{y}^{(v)}_{\tau(v)}$	scalar	forecast nowcast $\mathbb{E}(y_{t(v)} \mid \Omega_v)/\mathbb{E}(y_{\tau(v)} \mid \Omega_v)$
e_v	scalar	forecast/nowcast error
RMSFE	scalar	$\sqrt{\lvert\mathscr{V}\rvert^{-1}\sum_{v\in\mathscr{V}} e_v^2}$
Dimensions (vector vs matrix models)		
N	scalar	number of monthly indicators in the vector panel
$p_1,\ p_2$	scalars	row/column dimensions of a matrix panel $\mathbf{X}_t$
r	scalar	vector/vectorized factor dimension
$r_1,\ r_2$	scalars	row/column factor dimensions in matrix models
r_g, r_f	scalars	vectorized row/column (not)relevant factor dim.
L	scalar	proxy dimension (number of proxies)
k	scalar	state dimension in the Kalman filter, $k = \dim(\mathbf{s}_t)$
P_f, P_ε	scalars	VAR/AR order for factor/idiosyncratic dynamics
Missing-data treatment		
$\mathscr{O}_t$	set	indices observed at t in the full observation vector
$\mathscr{O}_i$	set	time observed for i in the full observation vector
m_t	scalar	number of observed entries, $m_t := \lvert\mathscr{O}_t\rvert$
$\mathbf{W}_t$	$\{0,1\}^{d\times d}$	diagonal selection matrix on the observation space

Table 4: Notation: Kalman filter and smoother outputs (generic state-space)

Symbol	**Type / dim.**	**Meaning**
Filter (prediction and update)		
$\widehat{\mathbf{s}}_{t\mid t-1}$	$\mathbb{R}^{k}$	predicted state mean $\mathbb{E}(\mathbf{s}_t \mid \Omega_{t-1})$
$\widehat{\mathbf{P}}_{t\mid t-1}$	$\mathbb{R}^{k\times k}$	predicted state variance $\mathrm{Var}(\mathbf{s}_t \mid \Omega_{t-1})$
$\widehat{\boldsymbol{\nu}}_{t,\mathscr{O}_t}$	$\mathbb{R}^{m_t}$	innovation on observed block
$\widehat{\mathbf{S}}_{t,\mathscr{O}_t}$	$\mathbb{R}^{m_t\times m_t}$	innovation variance on observed block
$\widehat{\mathbf{G}}_{t,\mathscr{O}_t}$	$\mathbb{R}^{k\times m_t}$	Kalman gain (observed block)
$\widehat{\mathbf{s}}_{t\mid t}$	$\mathbb{R}^{k}$	filtered state mean $\mathbb{E}(\mathbf{s}_t \mid \Omega_t)$
$\widehat{\mathbf{P}}_{t\mid t}$	$\mathbb{R}^{k\times k}$	filtered state variance $\mathrm{Var}(\mathbf{s}_t \mid \Omega_t)$
Smoother (smoothed moments and lag-one cross-moments)		
$\widehat{\mathbf{s}}_{t\mid T}$	$\mathbb{R}^{k}$	smoothed state mean $\mathbb{E}(\mathbf{s}_t \mid \Omega_T)$
$\widehat{\mathbf{P}}_{t\mid T}$	$\mathbb{R}^{k\times k}$	smoothed state variance $\mathrm{Var}(\mathbf{s}_t \mid \Omega_T)$
$\widehat{\mathbf{P}}_{t,t-1\mid T}$	$\mathbb{R}^{k\times k}$	smoothed cross-*covariance* $\mathrm{Cov}(\mathbf{s}_t, \mathbf{s}_{t-1} \mid \Omega_T)$
$\widehat{\boldsymbol{\Delta}}_{t,t-1\mid T}$	$\mathbb{R}^{k\times k}$	lag-one smoothed cross-*moment* $\mathbb{E}(\mathbf{s}_t\mathbf{s}_{t-1}' \mid \Omega_T)$
Information-form smoother auxiliaries (optional)		
$\widehat{\mathbf{r}}_t$	$\mathbb{R}^{k}$	information vector recursion
$\widehat{\mathbf{N}}_t$	$\mathbb{R}^{k\times k}$	information matrix recursion
$\widehat{\mathbf{L}}_{t,\mathscr{O}_t}$	$\mathbb{R}^{k\times k}$	smoother transition matrix $\mathbf{T}\big(\mathbf{I}-\mathbf{G}_{t,\mathscr{O}_t}\mathbf{Z}_{t,\mathscr{O}_t}\big)$
State space		
$\mathbf{x}_{t,\mathscr{O}_t}$	$\mathbb{R}^{m_t}$	observed subvector $\mathbf{x}_t[\mathscr{O}_t]$
$\mathbf{Z}, \mathbf{Z}_{t,\mathscr{O}_t}$	$\mathbb{R}^{p_1p_2\times k}$, $\mathbb{R}^{m_t\times k}$	(observed-block) measurement matrix
$\mathbf{T}$	$\mathbb{R}^{k\times k}$	transition matrix
$\mathbf{H}, \mathbf{H}_{t,\mathscr{O}_t}$	$\mathbb{R}^{p_1p_2\times p_1p_2}, \mathbb{R}^{m_t\times m_t}$	(observed-blok) measurement covariance
$\mathbf{Q}$	$\mathbb{R}^{k\times k}$	state-innovation covariance

Table 5: Notation: TPRF (single-frequency three-pass regression filter)

Symbol	Type / dim.	Meaning
Targets, proxies, and predictors		
y_{t+1}	$\mathbb{R}$	one-step-ahead target
$\mathbf{z}_t$	$\mathbb{R}^L$	proxy vector
$\mathbf{x}_t$	$\mathbb{R}^N$	predictor vector
X	$\mathbb{R}^{T\times N}$	stacked predictors, $X = (\mathbf{x}_1,\ldots,\mathbf{x}_T)'$
$Z,\ \widetilde{Z}$	$\mathbb{R}^{T\times L},\ \mathbb{R}^{T\times(L+1)}$	$Z = (\mathbf{z}_1,\ldots,\mathbf{z}_T)',\ \widetilde{Z} = [\mathbf{1}_T\ \ Z]$
Latent structure (DGP) and disturbances		
$\mathbf{F}_t$	$\mathbb{R}^r$	latent factors in the DGP
$\mathbf{\Phi}$	$\mathbb{R}^{N\times r}$	predictor loadings in the DGP
$\mathbf{\Lambda}$	$\mathbb{R}^{L\times r}$	proxy loadings in the DGP
η_{t+1}	scalar	target disturbance
$\boldsymbol{\omega}_t$	$\mathbb{R}^L$	proxy disturbance
$\boldsymbol{\xi}_t$	$\mathbb{R}^N$	predictor disturbance (vector form)
Pass 1–2 coefficients and targeted factors		
$\phi_{0,i},\ \boldsymbol{\phi}_i$	scalar, $\mathbb{R}^L$	intercept and proxy slopes in Pass 1 (series i)
$\boldsymbol{\phi}_0,\ \Phi$	$\mathbb{R}^N,\ \mathbb{R}^{N\times L}$	$\boldsymbol{\phi}_0 = (\phi_{0,1},\ldots,\phi_{0,N})',\ \Phi$ has rows $\boldsymbol{\phi}_i'$
$\widehat{\Phi}$	$\mathbb{R}^{N\times L}$	estimated slopes matrix (Pass 1 output)
$\phi_{0,t}$	scalar	intercept in Pass 2 (time t)
$\widehat{\mathbf{f}}_t,\ \widehat{F}$	$\mathbb{R}^L,\ \mathbb{R}^{T\times L}$	targeted factor and its stacking, $\widehat{F} = (\widehat{\mathbf{f}}_1,\ldots,\widehat{\mathbf{f}}_T)'$
$\widetilde{\Phi}$	$\mathbb{R}^{N\times(L+1)}$	augmented regressor in Pass 2, $\widetilde{\Phi} = [\mathbf{1}_N\ \ \widehat{\Phi}]$
$F,\ \widetilde{F}$	$\mathbb{R}^{T\times L},\ \mathbb{R}^{T\times(L+1)}$	$F = (\mathbf{f}_1,\ldots,\mathbf{f}_T)',\ \widetilde{F} = [\boldsymbol{\phi}_0\ \ F]$
Forecast regression and prediction (Pass 3)		
$\beta_0,\ \rho$	scalars	intercept and AR(1) coefficient
$\boldsymbol{\beta},\ \widetilde{\boldsymbol{\beta}}$	$\mathbb{R}^L,\ \mathbb{R}^{L+2}$	$\widetilde{\boldsymbol{\beta}} = (\beta_0,\rho,\boldsymbol{\beta}')'$
$\mathbf{y},\ \mathbf{y}_{-1}$	$\mathbb{R}^{T-1},\ \mathbb{R}^{T-1}$	$\mathbf{y} = (y_2,\ldots,y_T)',\ \mathbf{y}_{-1} = (y_1,\ldots,y_{T-1})'$
$\widehat{F}_{-1}$	$\mathbb{R}^{(T-1)\times L}$	$(\widehat{\mathbf{f}}_1,\ldots,\widehat{\mathbf{f}}_{T-1})'$
$\widetilde{F}^{(3)}$	$\mathbb{R}^{(T-1)\times(L+2)}$	$[\mathbf{1}_{T-1}\ \ \mathbf{y}_{-1}\ \ \widehat{F}_{-1}]$
$\widehat{y}_{t+1}$	$\mathbb{R}$	$\widehat{\beta}_0 + \widehat{\rho}\, y_t + \widehat{\boldsymbol{\beta}}'\widehat{\mathbf{f}}_t$

Table 6: Notation: MF–TPRF (mixed-frequency three-pass regression filter)

Symbol	Type / dim.	Meaning
Proxies, and alignment (Pass 1)		
$\mathbf{z}_\tau$	$\mathbb{R}^L$	quarterly proxy vector
$x_{i,\tau}^{(m\to q)}$	$\mathbb{R}$	quarterly-aligned predictor for series i
$\alpha_{0,i},\ \boldsymbol{\alpha}_i$	scalar, $\mathbb{R}^L$	Pass 1 intercept and slopes (series i)
$\widehat{A}$	$\mathbb{R}^{N\times L}$	matrix with rows $\widehat{\boldsymbol{\alpha}}_i'$
Within-quarter targeted factors and U–MIDAS (Pass 2–3)		
$\widehat{\mathbf{f}}_\tau^{(m)}$	$\mathbb{R}^L$	within-quarter targeted factors after month m
$\widehat{\mathbf{f}}_\tau^{Q}$	$\mathbb{R}^{3L}$	$(\widehat{\mathbf{f}}_\tau^{(1)\prime}, \widehat{\mathbf{f}}_\tau^{(2)\prime}, \widehat{\mathbf{f}}_\tau^{(3)\prime})'$
$\boldsymbol{\beta}_m$	$\mathbb{R}^L$	coefficient on block m
$\boldsymbol{\beta}_{p_f}^{(m)}$	$\mathbb{R}^L$	coefficient on lag p_f of block m
$\widehat{y}_\tau^{(m)}$	$\mathbb{R}$	nowcast after month m
Ragged edges and completion (EM step)		
A_i	$\{0,1\}^{\lvert\mathcal{O}_i\rvert\times T}$	selection matrix (keeps observed entries of series i)
$\mathbf{x}_{(i)}^{\text{std}}$	$\mathbb{R}^T$	centered and standardized series i
$\widehat{X}^{\text{std}}$	$\mathbb{R}^{T\times N}$	completed standardized panel (PCA/EM completion)
$\widehat{x}_{i,t}$	$\mathbb{R}$	completed value on the original scale

Table 7: Notation: Matrix MF–TPRF (matrix-specific)

Symbol	**Type / dim.**	**Meaning**
Latent spaces		
$\mathbf{R},\ \mathbf{C}$	$\mathbb{R}^{p_1 \times r_1},\ \mathbb{R}^{p_2 \times r_2}$	row/column loading spaces (Steps 1–2)
Targeting moments and whitening (Step 1)		
$\mathbf{M}^{(\ell)}$	$\mathbb{R}^{p_1 \times p_2}$	proxy-weighted moment $\sum_{\tau=1}^{\mathscr{T}} z_\tau^{(\ell)} \mathbf{X}_\tau$
$\mathbf{S}_{\text{row}}, \mathbf{S}_{\text{col}}$	$\mathbb{R}^{p_1 \times p_1}, \mathbb{R}^{p_2 \times p_2}$	whitened row/column-moment matrix
Matrix factors (Step 2) and vectorization		
$\mathbf{F}_t$	$\mathbb{R}^{r_1 \times r_2}$	monthly targeted factor matrix
$\widehat{\mathbf{F}}_t$	$\mathbb{R}^{r_1 \times r_2}$	estimated factor matrix (Step 2)
$\mathbf{\Xi}_t$	$\mathbb{R}^{p_1 \times p_2}$	residual matrix in Step 2
$\widehat{\mathbf{f}}_t$	$\mathbb{R}^{r_f}$	$\widehat{\mathbf{f}}_t = \text{vec}(\widehat{\mathbf{F}}_t)$
Within-quarter factor blocks (Step 3)		
$\widehat{\mathbf{F}}_\tau^{(m)}$	$\mathbb{R}^{r_1 \times r_2}$	factor matrix available at $m \in \{1,2,3\}$ in τ
$\widehat{\mathbf{f}}_\tau^{(m)}$	$\mathbb{R}^{r_f}$	$\text{vec}(\widehat{\mathbf{F}}_\tau^{(m)})$
$\widehat{\mathbf{f}}_\tau^{Q}$	$\mathbb{R}^{3r_f}$	$(\widehat{\mathbf{f}}_\tau^{(1)\prime}, \widehat{\mathbf{f}}_\tau^{(2)\prime}, \widehat{\mathbf{f}}_\tau^{(3)\prime})^\prime$

Table 8: Notation: MF–DFM (vector mixed-frequency dynamic factor model)

Symbol	**Type / dim.**	**Meaning**
Core objects		
$\mathbf{x}_t, X_T$	$\mathbb{R}^N, \mathbb{R}^{N\times T}$	(stacked) observed monthly vector (standardized)
$\mathbf{f}_t, F_T$	$\mathbb{R}^r, \mathbb{R}^{r\times T}$	(stacked) latent factor vector
$\boldsymbol{\varepsilon}_t$	$\mathbb{R}^N$	idiosyncratic component (measurement disturbance)
$\mathbf{u}_t$	$\mathbb{R}^r$	factor innovation
State-space parameters (baseline)		
$\boldsymbol{\Lambda}$	$\mathbb{R}^{N\times r}$	loading matrix
$\mathbf{A}$	$\mathbb{R}^{r\times r}$	factor transition matrix
$\boldsymbol{\Sigma}_u$	$\mathbb{R}^{r\times r}$	$\mathrm{Var}(\mathbf{u}_t)$
$\boldsymbol{\Sigma}_\varepsilon$	$\mathbb{R}^{N\times N}$	$\mathrm{Var}(\boldsymbol{\varepsilon}_t)$
Augmentation: VAR(P_f) factors and AR(1) idiosyncratic terms		
$\tilde{\boldsymbol{\varepsilon}}_t$	$\mathbb{R}^N$	persistent idiosyncratic state
$\boldsymbol{\xi}_t$	$\mathbb{R}^N$	nugget/measurement noise with $\mathrm{Var}(\boldsymbol{\xi}_t) = \kappa \mathbf{I}_N$
$\boldsymbol{\alpha}$	$\mathbb{R}^{N\times N}$	diagonal AR(1) coefficients, $\boldsymbol{\alpha} = \mathrm{dg}(\alpha_1, \ldots, \alpha_N)$
$\boldsymbol{\Sigma}_e$	$\mathbb{R}^{N\times N}$	diagonal innovation variances, $\boldsymbol{\Sigma}_e = \mathrm{dg}(\sigma_1^2, \ldots, \sigma_N^2)$
$\widetilde{\mathbf{f}}_t$	$\mathbb{R}^{rP_f+N}$	augmented state (e.g. $(\mathbf{f}', \tilde{\boldsymbol{\varepsilon}}_t')'$)

Table 9: Notation: MF–DMFM (matrix dynamic factor model)

Symbol	Type / dim.	Meaning
Core objects (matrix form)		
$\mathbf{X}_t,\ \mathbf{x}_t$	$\mathbb{R}^{p_1\times p_2},\ \mathbb{R}^{p_1p_2}$	matrix observation at month t / $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$
$\mathbf{E}_t,\ \mathbf{e}_t$	$\mathbb{R}^{p_1\times p_2},\ \mathbb{R}^{p_1p_2}$	idiosyncratic component / $\mathbf{e}_t = \text{vec}(\mathbf{E}_t)$
$\mathbf{H},\mathbf{K}$	$\mathbb{R}^{p_1\times p_1},\ \mathbb{R}^{p_2\times p_2}$	$\text{Var}(\text{vec}(\mathbf{E}_t)) = \mathbf{K}\otimes\mathbf{H}$
$\mathbf{R},\mathbf{C}$	$\mathbb{R}^{p_1\times r_1},\mathbb{R}^{p_2\times r_2}$	row/column loading matrix
$\mathbf{F}_t,\ \mathbf{f}_t$	$\mathbb{R}^{r_1\times r_2},\ \mathbb{R}^{r_1r_2}$	latent factor matrix / $\mathbf{f}_t = \text{vec}(\mathbf{F}_t)$
Factor dynamics (MAR block)		
$\mathbf{A},\mathbf{B}$	$\mathbb{R}^{r_1\times r_1},\mathbb{R}^{r_2\times r_2}$	row/column transition matrix
$\mathbf{U}_t$	$\mathbb{R}^{r_1\times r_2}$	factor innovation matrix
$\mathbf{P},\mathbf{Q}$	$\mathbb{R}^{r_1\times r_1},\ \mathbb{R}^{r_2\times r_2}$	$\text{Var}(\text{vec}(\mathbf{U}_t)) = \mathbf{Q}\otimes\mathbf{P}$
Idiosyncratic block (optional persistence)		
$\widetilde{\mathbf{E}}_t,\ \tilde{\mathbf{e}}_t$	$\mathbb{R}^{p_1\times p_2},\ \mathbb{R}^{p_1p_2}$	persistent idiosyncratic state / $\tilde{\mathbf{e}}_t = \text{vec}(\widetilde{\mathbf{E}}_t)$
$\boldsymbol{\Xi}_t$	$\mathbb{R}^{p_1\times p_2}$	nugget/measurement noise (vectorized variance $\kappa\mathbf{I}$)
$\mathbf{A}_E,\mathbf{B}_E$	$\mathbb{R}^{p_1\times p_1},\ \mathbb{R}^{p_2\times p_2}$	idiosyncratic MAR propagation (diagonal)
$\mathbf{H}_E,\mathbf{K}_E$	$\mathbb{R}^{p_1\times p_1},\ \mathbb{R}^{p_2\times p_2}$	idiosyncratic innovation covariances (diagonal)

B Kalman Filter and Smoother with Missing Observations

We consider a linear–Gaussian state-space model where the observation vector may contain missing entries. Let Ω_t denote the information set generated by the available data up to time t, and let Ω_T denote the full-sample information set. For both the vector DFM and the matrix DMFM (in vectorized form), the state-space representation is

$$\mathbf{x}_t = \mathbf{Z}\mathbf{s}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0},\mathbf{H}), \tag{79}$$

$$\mathbf{s}_t = \mathbf{T}\mathbf{s}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0},\mathbf{Q}). \tag{80}$$

Mapping to the models.

- **Vector DFM.** $\mathbf{x}_t \in \mathbb{R}^N$, $\mathbf{s}_t = \widetilde{\mathbf{f}}_t$, $\mathbf{Z} = \widetilde{\boldsymbol{\Lambda}}$, $\mathbf{T} = \widetilde{\mathbf{A}}$, $\mathbf{H} = \widetilde{\mathbf{R}}$, $\mathbf{Q} = \widetilde{\mathbf{Q}}$.

- **Matrix DFM (vectorized).** $\mathbf{x}_t = \mathrm{vec}(\mathbf{X_t}) \in \mathbb{R}^{p_1 p_2}$, $\mathbf{s}_t = \mathbf{f}_t = \mathrm{vec}(\mathbf{F_t}) \in \mathbb{R}^{r_1 r_2}$,

$$\mathbf{Z} = \mathbf{C} \otimes \mathbf{R}, \qquad \mathbf{T} = \mathbf{B} \otimes \mathbf{A},$$

and, under the pseudo-likelihood,

$$\mathbf{H} = \mathrm{dg}(\mathbf{K}) \otimes \mathrm{dg}(\mathbf{H}), \qquad \mathbf{Q} = \mathbf{Q} \otimes \mathbf{P}.$$

- **Augmented matrix DFM (vectorized) with persistent idiosyncratic component.** $\mathbf{x}_t = \mathrm{vec}(\mathbf{X}_t) \in \mathbb{R}^{p_1 p_2}$ and $\mathbf{s}_t = (\mathbf{f}_t', \tilde{\mathbf{e}}_t')' \in \mathbb{R}^{r_1 r_2 + p_1 p_2}$, where $\mathbf{f}_t = \mathrm{vec}(\mathbf{F}_t) \in \mathbb{R}^{r_1 r_2}$ and $\tilde{\mathbf{e}}_t = \mathrm{vec}(\widetilde{\mathbf{E}}_t) \in \mathbb{R}^{p_1 p_2}$.

  In this formulation the persistent idiosyncratic component is treated as part of the state, so the measurement disturbance in (79) is only the residual "nugget" $\boldsymbol{\varepsilon}_t = \boldsymbol{\xi}_t$ with $\boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \kappa \mathbf{I}_{p_1 p_2})$. Accordingly,

$$\mathbf{Z} = \begin{bmatrix} \mathbf{C} \otimes \mathbf{R} & \mathbf{I}_{p_1 p_2} \end{bmatrix}, \qquad \mathbf{H} = \kappa \mathbf{I}_{p_1 p_2}.$$

  The transition equation in (80) is block-diagonal,

$$\mathbf{T} = \begin{bmatrix} \mathbf{B} \otimes \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_E \otimes \mathbf{A}_E \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} \mathbf{Q} \otimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_E \end{bmatrix},$$

  where $\mathbf{Q}_E = \mathrm{dg}(\mathbf{K}_E) \otimes \mathrm{dg}(\mathbf{H}_E)$ with diagonal $\mathbf{H}_E, \mathbf{K}_E$. The blocks $(\mathbf{B}_E \otimes \mathbf{A}_E, \mathbf{Q}_E)$ encode the chosen persistent idiosyncratic dynamics (e.g. separable MAR(1)).

Observed block under missingness. Let $\mathcal{O}_t$ be the set of observed indices at time t in the full observation vector (dimension N in the DFM, or $p_1 p_2$ in the vectorized DMFM) and let $m_t := |\mathcal{O}_t|$. Define the observed subvector and reduced measurement objects as

$$\mathbf{x}_{t,\mathcal{O}_t} := \mathbf{x}_t[\mathcal{O}_t] \in \mathbb{R}^{m_t}, \qquad \mathbf{Z}_{t,\mathcal{O}_t} := \mathbf{Z}[\mathcal{O}_t, \cdot] \in \mathbb{R}^{m_t \times k}, \qquad \mathbf{H}_{t,\mathcal{O}_t} := \mathbf{H}[\mathcal{O}_t, \mathcal{O}_t] \in \mathbb{R}^{m_t \times m_t},$$

where $k = \dim(\mathbf{s}_t)$. If $m_t = 0$, no measurement update is performed.

Kalman filter. Initialize $\mathbf{s}_{0|0}$ and $\mathbf{P}_{0|0}$. For $t = 1, \ldots, T$:

- **Prediction (state and covariance).**

$$\mathbf{s}_{t|t-1} = \mathbf{T}\,\mathbf{s}_{t-1|t-1}, \qquad \mathbf{P}_{t|t-1} = \mathbf{T}\,\mathbf{P}_{t-1|t-1}\,\mathbf{T}' + \mathbf{Q}. \tag{81}$$

- **Innovation (only on the observed block).**

$$\boldsymbol{\nu}_{t,\mathcal{O}_t} = \mathbf{x}_{t,\mathcal{O}_t} - \mathbf{Z}_{t,\mathcal{O}_t}\,\mathbf{s}_{t|t-1}, \qquad \mathbf{S}_{t,\mathcal{O}_t} = \mathbf{Z}_{t,\mathcal{O}_t}\,\mathbf{P}_{t|t-1}\,\mathbf{Z}'_{t,\mathcal{O}_t} + \mathbf{H}_{t,\mathcal{O}_t}. \tag{82}$$

- **Update (Kalman gain and filtered moments).**

$$\mathbf{G}_{t,\mathcal{O}_t} = \mathbf{P}_{t|t-1}\,\mathbf{Z}'_{t,\mathcal{O}_t}\,\mathbf{S}^{-1}_{t,\mathcal{O}_t} \tag{83}$$

$$\mathbf{s}_{t|t} = \mathbf{s}_{t|t-1} + \mathbf{G}_{t,\mathcal{O}_t}\,\boldsymbol{\nu}_{t,\mathcal{O}_t}, \tag{84}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{G}_{t,\mathcal{O}_t}\,\mathbf{S}_{t,\mathcal{O}_t}\,\mathbf{G}'_{t,\mathcal{O}_t}. \tag{85}$$

- **Missing-at-t special case.** If $m_t = 0$, skip the measurement update and set $\mathbf{s}_{t|t} = \mathbf{s}_{t|t-1}$ and $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1}$.

Gaussian log-likelihood (innovation form).

$$\ell_t(\boldsymbol{\theta}) = -\tfrac{1}{2}\Big(m_t \log(2\pi) + \log|\mathbf{S}_{t,\mathcal{O}_t}| + \boldsymbol{\nu}'_{t,\mathcal{O}_t}\mathbf{S}^{-1}_{t,\mathcal{O}_t}\boldsymbol{\nu}_{t,\mathcal{O}_t}\Big), \qquad \ell(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}). \tag{86}$$

Kalman smoother (information form). Define

$$\mathbf{L}_{t,\mathcal{O}_t} := \mathbf{T}\big(\mathbf{I} - \mathbf{G}_{t,\mathcal{O}_t}\mathbf{Z}_{t,\mathcal{O}_t}\big),$$

initialize $\mathbf{r}_T = \mathbf{0}$ and $\mathbf{N}_T = \mathbf{0}$, and for $t = T, \ldots, 1$:

- **Backward recursion for information vectors/matrices.**

$$\mathbf{r}_{t-1} = \mathbf{Z}'_{t,\mathcal{O}_t}\mathbf{S}^{-1}_{t,\mathcal{O}_t}\boldsymbol{\nu}_{t,\mathcal{O}_t} + \mathbf{L}'_{t,\mathcal{O}_t}\mathbf{r}_t, \qquad \mathbf{N}_{t-1} = \mathbf{Z}'_{t,\mathcal{O}_t}\mathbf{S}^{-1}_{t,\mathcal{O}_t}\mathbf{Z}_{t,\mathcal{O}_t} + \mathbf{L}'_{t,\mathcal{O}_t}\mathbf{N}_t\mathbf{L}_{t,\mathcal{O}_t}. \tag{87}$$

- **Missing-at-t convention.** If $m_t = 0$, the instantaneous terms vanish and only the propagation via $\mathbf{L}'_{t,\mathscr{O}_t}(\cdot)$ remains.

- **Smoothed moments.** For $t = 1, \ldots, T$:

$$\mathbf{s}_{t|T} = \mathbf{s}_{t|t-1} + \mathbf{P}_{t|t-1}\,\mathbf{r}_{t-1}, \qquad \mathbf{P}_{t|T} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\,\mathbf{N}_{t-1}\,\mathbf{P}_{t|t-1}. \tag{88}$$

C Mixed-frequency restrictions

Consider a monthly time grid and partition the observable vector as

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{x}_t^M \\ \bar{\mathbf{x}}_t^Q \end{bmatrix}, \qquad \mathbf{x}_t^M \in \mathbb{R}^{N_M}, \;\; \bar{\mathbf{x}}_t^Q \in \mathbb{R}^{N_Q},$$

where $\bar{\mathbf{x}}_t^Q$ is observed only at quarter-end months (and possibly affected by ragged edges) and is missing otherwise. Let $\mathbf{x}_t^Q$ denote the latent monthly counterpart underlying the quarterly observable and assume the aggregation

$$\bar{\mathbf{x}}_t^Q = \omega(L)\,\mathbf{x}_t^Q = \sum_{\ell=0}^{L-1} \omega_\ell\,\mathbf{x}_{t-\ell}^Q,$$

with weights $\omega = (\omega_0, \ldots, \omega_{L-1})$ determined by the construction of the quarterly series. Two common choices are: (i) Mariano–Murasawa for growth rates with $L = 5$ and $\omega = (1,2,3,2,1)$; (ii) stock/flow for levels with $L = 3$ and $\omega = (1,1,1)$ (flows) or $\omega = (1/3, 1/3, 1/3)$ (stocks).

Assume factor representations on the monthly grid

$$\mathbf{x}_t^M = \mathbf{\Lambda}^M \mathbf{f}_t + \boldsymbol{\varepsilon}_t^M, \qquad \mathbf{x}_t^Q = \mathbf{\Lambda}^Q \mathbf{f}_t + \boldsymbol{\varepsilon}_t^Q.$$

Then the quarterly observable loads on *factor lags* with fixed coefficients:

$$\bar{\mathbf{x}}_t^Q = \bar{\mathbf{\Lambda}}^Q(\omega)\,\mathbf{f}_t^{(L)} + \bar{\boldsymbol{\varepsilon}}_t^Q(\omega), \qquad \bar{\mathbf{\Lambda}}^Q(\omega) = \begin{bmatrix} \omega_0 \mathbf{\Lambda}^Q & \cdots & \omega_{L-1}\mathbf{\Lambda}^Q \end{bmatrix},$$

where $\mathbf{f}_t^{(L)} = (\mathbf{f}_t', \ldots, \mathbf{f}_{t-L+1}')'$ and $\bar{\boldsymbol{\varepsilon}}_t^Q(\omega) = \sum_{\ell=0}^{L-1} \omega_\ell\, \boldsymbol{\varepsilon}_{t-\ell}^Q$.

State-space implementation (iid idiosyncratic terms). If idiosyncratic components are serially uncorrelated, the restriction is implemented by augmenting the state to include the required factor lags:

$$\mathbf{f}_t^{(L)} := \begin{bmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \vdots \\ \mathbf{f}_{t-L+1} \end{bmatrix}, \qquad \mathbf{Z}^{MF}(\omega) := \begin{bmatrix} \mathbf{\Lambda}^M & 0 & \cdots & 0 \\ \omega_0 \mathbf{\Lambda}^Q & \omega_1 \mathbf{\Lambda}^Q & \cdots & \omega_{L-1}\mathbf{\Lambda}^Q \end{bmatrix},$$

so that, on the monthly grid, the measurement equation becomes

$$\begin{bmatrix} \mathbf{x}_t^M \\ \bar{\mathbf{x}}_t^Q \end{bmatrix} = \mathbf{Z}^{MF}(\omega)\,\mathbf{f}_t^{(L)} + \begin{bmatrix} \boldsymbol{\varepsilon}_t^M \\ \bar{\boldsymbol{\varepsilon}}_t^Q(\omega) \end{bmatrix},$$

while the transition equation is obtained by writing the factor VAR dynamics for $\mathbf{f}_t$ and the corresponding companion form for $\mathbf{f}_t^{(L)}$.

Extension (persistent idiosyncratic components). If idiosyncratic components are persistent, e.g. $\boldsymbol{\varepsilon}_t^M$ and/or $\boldsymbol{\varepsilon}_t^Q$ follow diagonal AR(1) dynamics, the aggregation implies that $\bar{\boldsymbol{\varepsilon}}_t^Q(\omega)$ inherits serial dependence through the same weights ω. A convenient state-space implementation is to treat the idiosyncratic components as additional latent states and keep the mixed-frequency restriction in the measurement equation via fixed loading blocks.

Specifically, define the augmented state

$$\tilde{\mathbf{s}}_t^{MF} := \begin{bmatrix} \mathbf{f}_t^{(L)} \\ \boldsymbol{\varepsilon}_t^M \\ \boldsymbol{\varepsilon}_t^{Q,(L)} \end{bmatrix}, \qquad \boldsymbol{\varepsilon}_t^{Q,(L)} := \begin{bmatrix} \boldsymbol{\varepsilon}_t^Q \\ \boldsymbol{\varepsilon}_{t-1}^Q \\ \vdots \\ \boldsymbol{\varepsilon}_{t-L+1}^Q \end{bmatrix}.$$

Then the mixed-frequency measurement equation can be written as

$$\begin{bmatrix} \mathbf{x}_t^M \\ \bar{\mathbf{x}}_t^Q \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}^M & \mathbf{I}_{N_M} & 0 \\ \bar{\boldsymbol{\Lambda}}^Q(\omega) & 0 & \bar{\mathbf{I}}_{N_Q}(\omega) \end{bmatrix}}_{\tilde{\mathbf{Z}}^{MF}(\omega)} \tilde{\mathbf{s}}_t^{MF} + \begin{bmatrix} \boldsymbol{\xi}_t^M \\ \boldsymbol{\xi}_t^Q \end{bmatrix}, \qquad \bar{\mathbf{I}}_{N_Q}(\omega) := \begin{bmatrix} \omega_0 \mathbf{I}_{N_Q} & \cdots & \omega_{L-1}\mathbf{I}_{N_Q} \end{bmatrix}.$$

where $\boldsymbol{\xi}_t^M$ and $\boldsymbol{\xi}_t^Q$ are optional small measurement noises used for numerical stability. The transition equation is block-diagonal between (i) the factor companion form for $\mathbf{f}_t^{(L)}$ and (ii) the idiosyncratic AR(1) blocks, with the additional shift-register structure for $\boldsymbol{\varepsilon}_t^{Q,(L)}$. Importantly, the mixed-frequency restriction continues to enter only through the fixed weights ω in $\tilde{\mathbf{Z}}^{MF}(\omega)$.