KU LEUVEN

DEPARTMENT OF ECONOMICS

Factor-augmented sparse MIDAS regressions with an application to nowcasting

Jad Beyhum and Jonas Striaukas

FACULTY OF ECONOMICS AND BUSINESS



DISCUSSION PAPER SERIES DPS 24.14

November 2024

Factor-augmented sparse MIDAS regressions with an application to nowcasting*

Jad Beyhum
Department of Economics, KU Leuven, Belgium

Jonas Striaukas
Department of Finance, Copenhagen Business School, Denmark

November 11, 2024

Abstract

This article investigates factor-augmented sparse MIDAS (Mixed Data Sampling) regressions for high-dimensional time series data, which may be observed at different frequencies. Our novel approach integrates sparse and dense dimensionality reduction techniques. We derive the convergence rate of our estimator under misspecification, τ-mixing dependence, and polynomial tails. Our method's finite sample performance is assessed via Monte Carlo simulations. We apply the methodology to nowcasting U.S. GDP growth and demonstrate that it outperforms both sparse regression and standard factor-augmented regression during the COVID-19 pandemic. To ensure the robustness of these results, we also implement factor-augmented sparse logistic regression, which further confirms the superior accuracy of our nowcast probabilities during recessions. These findings indicate that recessions are influenced by both idiosyncratic (sparse) and common (dense) shocks.

Keywords: factor models, high-dimensional data, mixed-frequency data, nowcasting, recessions

*The authors are particularly grateful to Eric Ghysels for his advice and guidance regarding the present paper. In addition, Jonas Striaukas is grateful to his advisor Eric Ghysels for his guidance and friendship during Jonas' PhD studies. The authors also thank Andrii Babii, Luca Barbaglia, Ferre De Graeve, Catherine Doz, Geert Dhaene, Domenico Giannone, Sílvia Gonçalves, Laurent Ferrara, Julia Koh, Massimiliano Marcellino, Peter Reusens, Boriss Siliverstovs, Marie Ternes, Wouter Van der Veken, Raf Wouters, seminar and conference participants at CBS, COMPSTAT 2023, CREST, Paris School of Economics, EcoSta 2023, the MacroFor seminar of the International Institute of Forecasters and Erasmus University Rotterdam for helpful comments. Jad Beyhum undertook part of this work while employed by CREST, ENSAI (Rennes).

1 Introduction

In the current economic climate, where accurate nowcasting of macroeconomic variables such as U.S. GDP is critical for policymakers, addressing the challenges posed by high-dimensional mixed-frequency data has become a key econometric concern. Nowcasting involves predicting the current or near-future values of low-frequency outcome variables using high-frequency data — a task complicated by the need for effective variable selection and control over parameter proliferation. Motivated by the challenges of the data used in nowcasting, this paper introduces a novel estimation method specifically designed for high-dimensional mixed-frequency data.

High dimensionality often arises when the number of regressors is large relative to the number of observations, rendering traditional estimation techniques inadequate. Two primary strategies for addressing this issue are factor-augmented regressions and sparse methods. Factor-augmented regressions, introduced by Stock & Watson (1999) and Stock & Watson (2002), extract information from a large number of variables into a small set of latent factors by assuming a low-dimensional factor structure within the high-dimensional covariate matrix. Factors are first estimated via principal components analysis (PCA) before being used into a linear regression. Factor-augmented approaches are widely used due to their efficiency and interpretability in large datasets, typically following a dense forecasting approach that assumes most predictors meaningfully contribute to the target variable. In contrast, sparse methods, such as LASSO regression (Tibshirani 1996, Bickel et al. 2009), focus on identifying a small subset of predictors most relevant for forecasting, adhering to the principle of parsimony.

Effectively managing mixed-frequency data, where variables are sampled at different time intervals (e.g., weekly financial data alongside monthly macroeconomic data), introduces additional complexities. Standard dimension reduction techniques must be adapted to account for these structural differences. A generic approach is mixed-frequency data sampling (MIDAS), discussed by Andreou et al. (2010). MIDAS regressions apply a weighting scheme to directly link high-frequency data with low-frequency series, mitigating the parameter proliferation problem inherent in high-dimensional lag structures. Recently, Babii et al. (2022) proposed a MIDAS adaptation of sparse regression. A sparse-group LASSO

estimator is used on MIDAS-weighted variables to take advantage of the group structure naturally arising from the MIDAS aggregation scheme. Marcellino & Schumacher (2010), Andreou et al. (2013) and Koh (2023) study factor MIDAS regressions, where factors are first extracted from the high-frequency covariates by PCA before being aggregated by a nonlinear MIDAS scheme.

In this paper, we propose the novel approach of factor-augmented sparse MIDAS regression for forecasting with mixed-frequency high-dimensional data, which harnesses the advantages of both LASSO and factor-augmented regression. Our method works as follows. First, we construct MIDAS-weighted variables using pre-set polynomials, enabling the flexible handling of mixed-frequency data. Second, we estimate factors through PCA applied on the MIDAS-weighted variables. Third, we estimate a high-dimensional linear regression model using the sparse-group LASSO estimator in which we include the MIDAS-weighted covariates and the extracted factors as predictors. This integrated approach captures both dense signals (through the extracted factors) and the sparse signals distributed across different frequencies. By applying the sparse-group LASSO estimator, we effectively model hierarchical relationships among predictors, as it accommodates both grouped variables and sparsity within groups (Babii et al. 2022).

We establish the theoretical properties of our estimator, deriving rates of convergence that account for potential misspecification due to the MIDAS approximation or approximate sparsity. Our theoretical framework accommodates τ-mixing processes with polynomial tails, which are often encountered in financial and macroeconomic time series. This setting allows our method to remain robust even when facing heavy-tailed distributions or complex dependencies, as commonly observed in nowcasting applications. Note that it is important to use τ-mixing — which was introduced by Dedecker & Prieur (2004) — as opposed to more traditionally used β- or α-mixing in our setting. The reason is that both β and α mixing coefficients are too strong for the process we consider, namely, an autoregressive distributed lag (ARDL) model, see Babii et al. (2022) for a formal analysis of τ-mixing processes within the context of the ARDL model. In the present paper, we extend the modeling approach Babii et al. (2022) to a factor-augmented model and show its validity under the τ-mixing assumption. Relative to Babii et al. (2022), the challenge is to obtain results when factors are estimated by PCA.

Our numerical results highlight the practical effectiveness of the proposed method. First, through a series of Monte Carlo simulations, we demonstrate its superior finite sample performance compared to existing methods. Second, we apply the methodology to nowcast U.S. GDP growth, using panels of weekly financial and monthly macroeconomic data. By modeling the macroeconomic predictors with a factor structure and integrating the extracted factors into the regression, that is, imposing a sparse plus dense signal of macro data, our approach outperforms both sparse regression and standard factor-augmented regression models, especially during the COVID-19 pandemic. To validate the robustness of our findings, we also implement a factor-augmented sparse MIDAS logistic regression to predict recessions, which further confirms the improved accuracy of our nowcasted recession probabilities during periods of economic turbulence. These results suggest that economic recessions are driven by a combination of idiosyncratic (sparse) and common (dense) shocks.

The empirical contribution of our research is particularly relevant given the difficulties of nowcasting during the COVID-19 period, widely acknowledged as a challenging time for economic forecasting. Our findings reveal a distinct pattern: during periods of economic stability, factor augmentation does not significantly improve nowcasting accuracy compared to sparse-group regression. This observation suggests that, in stable economic conditions, sparse techniques alone are sufficient for effective nowcasting. However, during times of substantial macroeconomic instability, such as the COVID-19 pandemic, incorporating factors into the model significantly enhances predictive performance. By delineating these differences, our research provides valuable insights into the practical benefits of factor augmentation in nowcasting, especially in the face of heightened economic uncertainty.

Literature review. Our theoretical results contribute to the literature on high-dimensional econometrics. Babii et al. (2022) study the properties of the sparse-group LASSO with mixed-frequency data (without estimated factors). We extend their work by proving a finite sample bound and deriving convergence rates while allowing for estimated factors and τ-mixing dependence. To obtain these results, we establish convergence rates for factor estimation using PCA under τ-mixing, thereby providing novel insights into the statistical properties of PCA in time series settings. Considering τ-mixing processes is important, as highlighted by Dedecker & Prieur (2004), Babii et al. (2022) among others, as other com-

monly used dependence assumptions appear to be too strong for the models we consider. We also want to stress the differences between the factor MIDAS method of Marcellino & Schumacher (2010), Andreou et al. (2013), Koh (2023) and our approach. Their approach assumes a factor model on a high-dimensional, high-frequency set of variables. They implement a two-step procedure: first, the latent factors are estimated via PCA, and then these factors are projected onto the target variable using a low-dimensional nonlinear MIDAS regression. Instead, our approach uses a high-dimensional linear MIDAS regression, which also allows the many original predictors to enter as long as they exhibit a sparse coefficient pattern.

Several recent studies have explored high-dimensional factor-augmented sparse regression models outside a mixed-frequency setting. Fan et al. (2023) propose a general framework that models both sparse and dense signals simultaneously. Building on this research, Beyhum & Striaukas (2024) introduces a fully data-driven testing procedure to test for the presence of the sparse component in factor-augmented sparse regressions. Their analysis of the FRED-MD dataset reveals that the sparse component is statistically significant for most series in the dataset. Barigozzi et al. (2024) introduce the FNETS model, which integrates factor models and sparse vector autoregressions (VARs) to effectively capture network dependencies. Similarly, Krampe & Margaritella (2021) proposes a factor-augmented sparse VAR model and shows its superior forecasting performance over traditional methods. The sparse plus dense data structure has also been considered in the context of panel data, as seen in Hansen & Liao (2019) and Vogt et al. (2022), who develop methods that handle high-dimensional panels with factor structures and sparse idiosyncratic components. These studies focus on estimation and inference problems with sparse plus dense structures. Our work is the first to demonstrate the relevance of sparse plus dense patterns within a mixed-frequency framework. The previously mentioned papers do not consider MIDAS weighting, their theoretical results do not allow for approximation errors and they rely on stronger dependence measures than τ-mixing.

Finally our work relates to the literature on nowcasting. Nowcasting GDP during the COVID-19 pandemic has been particularly challenging due to unprecedented economic shocks. For instance, Foroni et al. (2022) suggests adjusting nowcasts by leveraging prediction errors from the financial crisis, while Huber et al. (2023) develops non-

parametric Bayesian regression trees to achieve robust GDP predictions. Carriero et al. (2024) addresses the impact of COVID-19 outliers by proposing outlier-augmented stochastic volatility Bayesian vector autoregressions (BVARs). Hauzenberger et al. (2024) introduce a Bayesian to MIDAS regressions using Gaussian Processes which improves nowcasting results due nonlinear structure of the model, showing good predictive accuracy for key economic indicators. Additionally, Diebold (2020) analyze the effectiveness of the ADS index in tracking economic activity during the pandemic. Our findings suggest that factor-augmented sparse methods, grounded in a solid statistical framework, provide a promising alternative for nowcasting during periods of significant economic instability. Empirically, our results indicate that dense macroeconomic information is especially beneficial during crisis periods, resulting in improved predictions during the COVID-19 pandemic.

Outline. The paper is organized as follows. First, Section 2 describes our estimation strategy. The theoretical properties of our estimator are studied in Section 3. Finite sample properties of our approach are evaluated thanks to Monte Carlo simulations in Section 4. Our application to nowcasting GDP growth is presented in Section 5. In Section 6, we present our extension to factor-augmented sparse MIDAS logistic regression and apply it to nowcast recessions. Section 7 concludes.

Notation. For an integer $N \in \mathbb{N}$, let $[N] = \{1, \ldots, N\}$. The transpose of a $n_1 \times n_2$ matrix H is written $H^\top$. Its kth singular value is $\sigma_k(H)$. Let us also define the Euclidean norm $\|H\|_2^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} H_{i,j}^2$ and the sup-norm $\|H\|_\infty = \max_{i \in [n_1], j \in [n_2]} |H_{i,j}|$. The quantity $n_1 \vee n_2$ is the maximum of n_1 and n_2. For a real-valued random variable ζ and $g > 0$, we let $\|\zeta\|_g = \mathbb{E}[|\zeta|^g]^{\frac{1}{g}}$. For a d-dimensional random vector ζ, we define $\|\zeta\|_g = \sup_{u \in \mathbb{R}^d:\ \|u\|_2 \leq 1} \|u^\top \zeta\|_g$. For a vector $v \in \mathbb{R}^N$, and $G \subset [N]$, we let v_G be the vector in $\mathbb{R}^N$ such that $(v_G)_i = v_i$ if $i \in G$ and $(v_G)_i = 0$ otherwise.

2 Factor-augmented sparse MIDAS regression

In this section, we describe our novel factor-augmented sparse MIDAS regression. Let $\{y_t,\ t \in [T]\}$ be the low-frequency outcome variable we want to predict where T denotes the sample size. For example, in our application, we consider U.S. real GDP growth, which

is measured quarterly. We use information in two sets of high-frequency regressors denoted

$$\widetilde{z} = \left\{\widetilde{z}_{t-(j-1)/m_z,k},\ t \in [T], k \in [K_z], j \in [m_z]\right\}$$

and

$$\widetilde{x} = \left\{\widetilde{x}_{t-(j-1)/m_x,k},\ t \in [T], k \in [K_x], j \in [m_x]\right\}.$$

Note that, throughout the paper, we use a "$\sim$" to denote high-frequency variables. In total, there are $K = K_z + K_x$ regressors which we observe more frequently than the target variable y_t, e.g., monthly and weekly. The quantities m_z and m_x determine the number of high-frequency lags we use of variables in $\widetilde{z}$ or $\widetilde{x}$, respectively. In our application, $\widetilde{z}$ is a high-dimensional weekly panel of financial variables, while $\widetilde{x}$ is a high-dimensional monthly panel of macroeconomic predictors.

The nowcasting task of y_t by $\widetilde{z}$ and $\widetilde{x}$ is a high-dimensional problem because (i) $\widetilde{z}$ and $\widetilde{x}$ contain many variables and (ii) the higher sampling frequency of the variables $\widetilde{z}$ and $\widetilde{x}$ creates parameter proliferation because many lags can be introduced in the model. This calls for the use of several complementary dimension reduction approaches.

The model is

$$y_t = m_t + \varepsilon_t,\ t \in [T], \tag{1}$$

where

$$\begin{aligned} m_t = \rho_0 + \sum_{j=1}^{J} \rho_j y_{t-j} + \sum_{k=1}^{K_z}\sum_{j=1}^{m_z} \omega_{z,k}\left((j-1)/m_z\right)\widetilde{z}_{t-(j-1)/m_z,k} \\ + \sum_{k=1}^{K_x}\sum_{j=1}^{m_x} \omega_{x,k}\left((j-1)/m_x\right)\widetilde{x}_{t-(j-1)/m_x,k} + \sum_{r=1}^{K_f}\sum_{j=1}^{m_x} \omega_{f,k}\left((j-1)/m_f\right)\widetilde{f}_{t-(j-1)/m_x,r}, \end{aligned} \tag{2}$$

is the true regression function.[1] Here, $\rho = (\rho_0, \ldots, \rho_J)^\top \in \mathbb{R}^{J+1}$ are the autoregressive coefficients, J is the number of lags, $\omega_{z,k},\ k \in [K_z]$, $\omega_{x,k},\ k \in [K_x]$ and $\omega_{f,r},\ r \in [K_f]$ are some functions from $[0,1]$ to $\mathbb{R}$ determining the linear regression coefficients of the different high-frequency lags, and

$$\widetilde{f} = \left\{\widetilde{f}_{t-(j-1)/m_x,r},\ t \in [T], r \in [K_f], j \in [m_x]\right\}$$

[1]Note that our theory does not rely on the definition of m_t in (2) but only requires that the true m_t is well approximated by our predictors. Hence, our theory implicitly allows the true model to contain nonlinearities. The definition of m_t in (2) serves to justify our approach heuristically.

are factors linked to $\widetilde{x}$ through the factor model:

$$\widetilde{x}_{t-(j-1)/m_x,k} = \sum_{r=1}^{K_f} \widetilde{b}_{k,r} \widetilde{f}_{t-(j-1)/m_x,r} + \widetilde{u}_{t-(j-1)/m_x,k}, \tag{3}$$

for $t \in [T], k \in [K_x], j \in [m_x]$, where $\left\{\widetilde{b}_{k,r},\ k \in [K_x], r \in [K_f]\right\}$ are nonrandom loadings and $\left\{\widetilde{u}_{t-(j-1)/m_x,k},\ t \in [T], k \in [K_x], j \in [m_x]\right\}$ are error terms. Note that we only assume that $\widetilde{x}$ (and not $\widetilde{z}$) follows a factor model. This is because, following the literature, it is uncommon to assume a factor structure among financial variables.

The model in (1)-(2) contains both the original regressors and the factors. Estimating all the coefficients of the high-frequency lags in (2) is a complex econometric task. To overcome this difficulty, we approximate the high-frequency lag polynomials through a MIDAS approach. Specifically, we consider a dictionary $(w_d)_{d\geq 0}$ of functions from $[0,1]$ to $\mathbb{R}$ for which there exist linear approximation coefficients $\{\widetilde{\alpha}_{k,d},\ k \in [K_z], d \in [D]\}$, $\left\{\widetilde{\beta}_{k,d},\ k \in [K_x], d \in [D]\right\}$ and $\{\widetilde{\gamma}_{r,d},\ r \in [K_f], d \in [D]\}$ such that, for all $u \in [0,1]$,

$$\begin{aligned}
\sum_{d=1}^{D} \widetilde{\alpha}_{k,d} w_d(u) &\approx \omega_{z,k}(u),\ k \in [K_z], \\
\sum_{d=1}^{D} \widetilde{\beta}_{k,d} w_d(u) &\approx \omega_{x,k}(u),\ k \in [K_x], \\
\sum_{d=1}^{D} \widetilde{\gamma}_{r,d} w_d(u) &\approx \omega_{f,r}(u),\ r \in [K_f],
\end{aligned} \tag{4}$$

where D is the number of polynomials from the dictionary used to approximate high-frequency lag coefficients linearly. In practice, we use Legendre polynomials up to degree 3 as approximating functions, which implies that $D = 4$.

For all $t \in [T]$, let us define the MIDAS-weighted variables

$$\begin{aligned}
z_{t,D(k-1)+d} &= \sum_{j=1}^{m_z} w_d((j-1)/m_z)\, \widetilde{z}_{t-(j-1)/m_z,k},\ k \in [K_z], d \in [D]; \\
x_{t,D(k-1)+d} &= \sum_{j=1}^{m_x} w_d((j-1)/m_x)\, \widetilde{x}_{t-(j-1)/m_x,k},\ k \in [K_x], d \in [D]; \\
f_{t,D(r-1)+d} &= \sum_{j=1}^{m_x} w_d((j-1)/m_x)\, \widetilde{f}_{t-(j-1)/m_x,r},\ r \in [K_f], d \in [D]; \\
u_{t,D(k-1)+d} &= \sum_{j=1}^{m_x} w_d((j-1)/m_x)\, \widetilde{u}_{t-(j-1)/m_x,k},\ k \in [K_x], d \in [D].
\end{aligned} \tag{5}$$

Let $p_z = DK_z$, $p_x = DK_x$, $R = DK_f$, $z_t = (z_{t,1}, \ldots, z_{t,p_z})^\top$, $x_t = (x_{t,1}, \ldots, x_{t,p_x})^\top$, $u_t = (u_{t,1}, \ldots, u_{t,p_x})^\top$, $f_t = (f_{t,1}, \ldots, f_{t,R})^\top$. Due to (4), we have

$$y_t = \rho_0 + \sum_{j=1}^{\mathrm{J}} \rho_j y_{t-j} + z_t^\top \alpha + x_t^\top \beta + f_t^\top \gamma + a_t + \varepsilon_t, \tag{6}$$

where

$$a_t = m_t - \left(\rho_0 + \sum_{j=1}^{\mathrm{J}} \rho_j y_{t-j} + z_t^\top \alpha + x_t^\top \beta + f_t^\top \gamma \right) \approx 0,$$

is the approximation error and

$$\alpha_{D(k-1)+d} = \widetilde{\alpha}_{k,d}, k \in [K_z], d \in [D],$$
$$\beta_{D(k-1)+d} = \widetilde{\beta}_{k,d}, k \in [K_x], d \in [D],$$
$$\gamma_{D(r-1)+d} = \widetilde{\gamma}_{r,d}, r \in [K_f], d \in [D].$$

are the linear regression coefficients.[2] Remark also that that x_t follows the factor model

$$x_t = Bf_t + u_t, \ t \in [T], \tag{7}$$

where $B = (b_1, \ldots, b_{p_x})^\top$ is a $p_x \times R$ matrix, with $b_{D(k-1)+d} = (\widetilde{b}_{k,1}, \ldots, \widetilde{b}_{k,R})^\top$, $k \in [K_z], d \in [D]$.

To present the estimation procedure, we rewrite (6) in a matrix form. We introduce further notation. Let $w_t = (1, y_{t-1}, \ldots, y_{t-J}, z_t^\top, x_t^\top)^\top$, $W = (w_1, \ldots, w_T)^\top$, $F = (f_1, \ldots, f_T)^\top$, $A = (a_1, \ldots, a_T)^\top$ and $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_T)^\top$. W is a $T \times p$ matrix where $p = 1 + J + p_z + p_x$, F is a $T \times R$ matrix, while A and $\mathcal{E}$ are $T \times 1$ vectors. The regression model (6), therefore, can be written in a matrix form

$$Y = W\delta + F\gamma + A + \mathcal{E}, \tag{8}$$

where $\delta = (\rho_0, \rho_1, \ldots, \rho_J, \alpha^\top, \beta^\top)^\top$. To proceed with the estimation of (8), we estimate the factors F by PCA. We let the columns of $\widehat{F}/\sqrt{T}$ be the eigenvectors corresponding to the leading R eigenvalues of $XX^\top$. In practice, the number of factors R is unknown and we replace it by an estimator $\widehat{R}$. In our empirical application to nowcasting, we use the

[2]Note that no multicollinearity issue between x_t and f_t arises in model (6). This is because by (7), equation (6) can be rewritten as $y_t = \rho_0 + \sum_{j=1}^{\mathrm{J}} \rho_j y_{t-j} + z_t^\top \alpha + u_t^\top \beta + f_t^\top (\gamma + B^\top \beta) + a_t + \varepsilon_t$, and u_t and f_t are not multicollinear.

growth ratio estimator of Ahn & Horenstein (2013), but there exist various alternatives in the literature (see Bai & Ng 2002, Onatski 2010, Bai & Ng 2019, Fan et al. 2022, among others). Finally, we use the estimator

$$(\widehat{\delta}, \widehat{\gamma}) \in \underset{d \in \mathbb{R}^p, c \in \mathbb{R}^R}{\arg\min} \left\| Y - Wd - \widehat{F}c \right\|_2^2 + \lambda \Omega(d), \tag{9}$$

where $\lambda \in \mathbb{R}$ is the tuning parameter, and the sparse-group LASSO (sg-LASSO) penalty function is

$$\Omega(d) = \mu \|d\|_1 + (1-\mu)\|d\|_{2,1},$$

which interpolates between the ℓ_1 LASSO and group LASSO norms. Here, $\mu \in [0,1]$ is the interpolation parameter between the two norms. The latter is defined as $\|d\|_{2,1} = \sum_{G \in \mathcal{G}} \|d_G\|_2$. Such a penalty is beneficial in MIDAS regression settings because it encourages within-group sparsity, which allows for accurate weight function estimation, as well as across-group sparsity, which controls the dimensionality of regressors (Babii et al. 2022). In practice, λ and μ are selected via cross-validation. Specifically, we use 5-fold cross-validation, defining folds as adjacent blocks over the time dimension to take into account the time series dependence. For the nowcasting application, following Babii et al. (2022), we choose a group structure corresponding to the original high-frequency covariates, that is $\mathcal{G} = \{G_1, \ldots, G_{1+K_z+K_x}\}$, where $G_1 = [J+1]$ and $G_\ell = \{J+1+(\ell-1)d : \ d \in [D]\}$ for $\ell \in \{2, \ldots, 1+K_z+K_x\}$. This structure promotes sparsity across the original predictors.

This is a sparse plus dense dimension reduction approach. Indeed, sparsity is induced through the penalty $\Omega(\cdot)$ and the presence of the factors, which are linear combinations of the variables in X, introduces a dense pattern in the estimation procedure. By combining three different dimension reduction approaches (MIDAS, factor models/PCA and LASSO), we propose a method appropriate for the challenges at hand.

Lastly, we note that in our empirical application, we also explore factor-augmented and sparse MIDAS regressions as alternatives to our approach. Specifically, the factor-augmented MIDAS regression assumes that coefficients related to the sparse component are zero and employs OLS for estimation. Formally, this approach solves

$$\min_{c \in \mathbb{R}^R} \left\| Y - \widehat{F}c \right\|_2^2, \tag{10}$$

On the other hand, the sparse MIDAS regression imposes zero restrictions on γ coefficients, for which we apply the sg-LASSO estimator:

$$\min_{d\in\mathbb{R}^p} \|Y - Wd\|_2^2 + \lambda\Omega(d), \tag{11}$$

These methods are referred to as FAMIDAS (equation 10) and sg-LASSO-MIDAS (equation 11), respectively. Our main approach in equation (9), which uses both sparse and dense signals, is denoted as sg-LASSO-FAMIDAS.

3 Theory

Let us now investigate the theoretical properties of our approach. Our theoretical analysis allows for the approximation error a_t, which can be due to the MIDAS approximation, approximate sparsity or nonlinearites. Notably, we do not need to assume that the true model m_t is linear or sparse (or takes the form (2)), but only that it is well approximated by a sparse linear combination of the predictors that we use. We consider an asymptotic regime where $T \to \infty$, and p_x and p_z go to infinity as a function of T. The number of factors is fixed with T. It would be possible to let it grow with T, but this would greatly complicate the theoretical analysis and, more importantly, the exposition of results.[3] The group structure $\mathcal{G}$ and the parameters δ and γ are not random but can vary with T. For this theoretical analysis, μ is fixed with T. We make the following assumptions.

Assumption 1 *It holds that*

(*i*) *For all* $t \in [T]$, $\mathbb{E}[f_t f_t^\top] = I_R$ *and* $B^\top B$ *is diagonal;*

(*ii*) *All the eigenvalues of the* $R \times R$ *matrix* $p_x^{-1} B^\top B$ *are bounded away from* 0 *and* ∞ *as* $p_x \to \infty$*;*

(*iii*) $\|B\|_\infty = O(1)$.

Assumption 2 *The following holds:*

[3] See, e.g., Beyhum & Gautier (2023), Freeman & Weidner (2023) for an analysis with a growing number of factors.

(*i*) *For all* $t \in [T], k \in [p_x], \ell \in [p], r \in [R]$, *it holds that*

$$\mathbb{E}[u_{t,k}] = \mathbb{E}[u_{t,k} f_{t,r}] = \mathbb{E}[u_{t,k}\varepsilon_t] = \mathbb{E}[f_{t,r}\varepsilon_t] = \mathbb{E}[w_{t,\ell}\varepsilon_t] = 0;$$

(*ii*) *There exist* $q > 2$ *and* $C_1 > 0$, *such that, for all* $t \in [T]$, *we have*

$$\|u_t\|_{2q} + \|\varepsilon_t\|_{2q} + \|f_t\|_{2q} + \|w_t\|_{2q} + \left\| p_x^{-1/2} \sum_{k\in[p_x]} u_{t,k} b_k \right\|_{2q} \leq C_1;$$

(*iii*) *There exists* $C_2 > 0$ *such that, for all* $t \in [T]$, $\max_{k\in[p]} \|E[w_{t,k}u_t]\|_2 \leq C_2$ *and*

$$\max_{s,t\in[T]} \mathbb{E}\left[\left(\frac{u_s^\top u_t}{p_x} - \frac{\mathbb{E}[u_s^\top u_t]}{p_x}\right)^2\right] \leq C_2.$$

Assumption 1 is the same as Assumption 3 in Fan et al. (2023). Its conditions (i) and (ii) form a strong factor assumption as in Bai (2003). Assumption 2 contains restrictions on the moments of random variables. Its condition (i) is a no-correlation condition. It would hold if (a) u_t is mean zero and uncorrelated with f_t, ε_t and (b) ε_t is mean zero and uncorrelated with f_t, w_t. Condition (ii) assumes that variables have strictly more than 4 finite moments, having, therefore, polynomial tails. Finally, condition (iii) bounds some moments. The bound on $\max_{s,t\in[T]} \mathbb{E}\left[\left(\frac{u_s^\top u_t}{p_x} - \frac{\mathbb{E}[u_s^\top u_t]}{p_x}\right)^2\right]$ is standard in the literature on factor models, see Fan et al. (2013, 2023). The bound on $\max_{k\in[p]} \|E[w_{t,k}u_t]\|_2$ is new to the literature. It essentially assumes that $w_{t,k}$ can only be correlated with the entries of u_t.

Next, to control the time series dependence of the variables, we use the concept of τ-mixing. The definition of τ-mixing coefficients is as follows. For a σ-algebra $\mathcal{M}$ and a random vector $\xi \in \mathbb{R}^l$, let

$$\tau(\mathcal{M}, \xi) = \left\| \sup_{f\in \mathrm{Lip}_1} |\mathbb{E}(f(\xi)|\mathcal{M}) - \mathbb{E}(f(\xi))| \right\|_1,$$

where $\mathrm{Lip}_1 = \left\{ f : \mathbb{R}^l \to \mathbb{R} : \ |f(x) - f(y)| \leq |x - y|_1 \right\}$ is a set of 1-Lipschitz functions. Let $\{\xi_t\}_t$ be a stochastic process and let $\mathcal{M}_t = \sigma(\xi_t, \xi_{t-1}, \dots)$ be its filtration. The τ-mixing coefficients of $\{\xi_t\}_t$ are defined as

$$\tau_s = \sup_{j\geq 1} \frac{1}{j} \sup_{t+s\leq t_1<\cdots<t_j} \tau(\mathcal{M}_t, (\xi_{t_1}, \dots, \xi_{t_j})), \ s \in \mathbb{N}.$$

The process $\{\xi_t\}_t$ is called τ-*mixing* when $\tau_s \downarrow 0$ as $s \to \infty$. As noted in the introduction, using τ-mixing conditions is less restrictive than α or β-mixing. The latter are too strong

for the process we consider, namely, an autoregressive distributed lag (ARDL) model, see Dedecker & Prieur (2004) and Babii et al. (2022).

Assumption 3 *The process*

$$\left\{\left(f_t^\top, u_t^\top, \varepsilon_t, w_t^\top, \left(p_x^{-1/2}\sum_{k\in[p_x]} u_{t,k}b_k\right)^\top\right)^\top\right\}_t$$

is stationary, and, for some constants $C_3 > 0$ *and* $a > (q-1)/(q-2)$*, its* τ*-mixing coefficients, denoted by* τ_s *satisfy* $\tau_s \leq C_3 s^{-a}$ *for all* $s \in \mathbb{N}$.

We are ready to state some results on the estimation of the factors. To do so, we introduce further notation. Let V be the matrix with the R largest eigenvalues of $T^{-1}XX^\top$. Next, we define $H = T^{-1}V^{-1}\widehat{F}^\top FB^\top B$, which is a rotation matrix such that the factors rotated by this matrix are well estimated.[4] This rotation matrix is the same as the one used in the literature to show consistency of estimated factors, see Bai (2003) and Bai & Ng (2006). Moreover, we let

$$h_T = \left(\frac{p}{T^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(2p)}{T}},$$

where $\kappa = \frac{(a+1)q-1}{a+q-1}$. We have the following lemma.

Lemma 3.1 *Under Assumptions 1, 2 and 3, we have*

$$(i)\ \left\|\widehat{F} - FH^\top\right\|_2 = O_P\left(\sqrt{\tfrac{T}{p_x}} + 1\right);$$

$$(ii)\ \left\|\left(\widehat{F} - FH^\top\right)^\top \mathcal{E}\right\|_2 = O_P\left(\sqrt{\tfrac{T}{p_x}} + 1\right);$$

$$(iii)\ \left\|\left(\widehat{F} - FH^\top\right)^\top W\right\|_\infty = O_P\left(\left(\tfrac{T}{p_x} + \sqrt{\tfrac{T}{p_x}}\right)\left(\sqrt{p_x}h_T + 1\right)\right).$$

Lemma 3.1 shows consistency of factor estimates. To the best of our knowledge, this is the first result of this type under τ-mixing. The rates in statements (i) and (ii) are the same as in the standard literature, see Bai (2003), Bai & Ng (2006) and Fan et al. (2023). Hence the presence of τ-mixing does not change the rate of convergence in ℓ_2-norm. It only appears in the rate of convergence in ℓ_∞-norm such as statement (iii).

[4] In the Online Appendix, we prove that V is invertible with probability going to 1, see Lemma A.2.

Now, let $\mathcal{S}_0 = \{k \in [p] : \ \delta_k \neq 0\}$ and $\mathcal{G}_0 = \{G \in \mathcal{G} : \ \delta_G \neq 0\}$ be respectively be the support and the group support of δ. We define the effective sparsity $\sqrt{s_\mu} = \mu\sqrt{|\mathcal{S}_0|} + (1-\mu)\sqrt{|\mathcal{G}_0|}$ and the maximum group size $G^* = \max_{G\in\mathcal{G}} |G|$. To avoid complexities arising when $s_\mu = 0$, we assume that $s_\mu \geq 1$. We also define

$$\Sigma = \mathbb{E}[w_t w_t^\top] - \mathbb{E}[w_t f_t^\top]\mathbb{E}[w_t f_t^\top]^\top,$$

which is the population Gram matrix of the w_t once they have been projected on the orthogonal of the factors f_t. We make the following assumption.

Assumption 4 *The following holds:*

(*i*) $\|\gamma\|_2 = O(1)$, $p = o(T^{(\kappa-1)/2})$ *and*

$$G^* s_\mu \left[\left(\left(\frac{p^2}{T^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(2p)}{T}}\right) + \frac{1}{\sqrt{p_x}}\right] = o(1);$$

(*ii*) *There exists* $\nu > 0$ *independent of* T *such that* $\sigma_p(\Sigma) \geq \nu$.

Assumption 4 (i) is a rate condition. It assumes that p cannot grow too quickly with T, this is needed because of the presence of time series dependence and polynomial tails. It also restricts the rate at which the sparsity level s_μ can grow. In the literature on the LASSO estimator, it is typically assumed that $s_*\sqrt{\log(2p)/T} = o(1)$, where s_* is the sparsity level (see Bickel et al. 2009). Our condition is stronger because of time series dependence, polynomial tails and the factors are estimated. Condition (ii) of Assumption 4 assumes that the smallest eigenvalue of Σ is bounded from below. This allows to show that the classical restricted eigenvalue condition of the LASSO literature holds (Bickel et al. 2009).

Then, we state a bound on the prediction error of our estimator, that is the ℓ_2-norm of the prediction $W\widehat{\delta} + \widehat{F}\widehat{\gamma}$ minus the target prediction $W\delta + F\gamma$. To state the bound, we need further notation. Let $P_{\widehat{F}} = \frac{1}{T}\widehat{F}\widehat{F}^\top$ be the orthogonal projector on the vector space spanned by the columns of $\widehat{F}$ and $M_{\widehat{F}} = I_T - P_{\widehat{F}}$. We also introduce $\widetilde{W} = M_{\widehat{F}}W$ which corresponds to W projected on the orthogonal of the vector space spanned by the columns of $\widehat{F}$. Finally, let $\Omega^*(\cdot)$ be the dual norm of $\Omega(\cdot)$, that is, for any $d \in \mathbb{R}^p$, $\Omega^*(d) = \sup_{v\in\mathbb{R}^p} \frac{v^\top d}{\Omega(d)}$.

Theorem 3.1 *Under Assumptions 1, 2, 3 and 4, and letting* $\lambda \geq 2\Omega^*\left(\frac{1}{T}\widetilde{W}^\top \mathcal{E}\right)$, *with probability going to* 1, *we have*

$$\frac{1}{\sqrt{T}}\left\|W\widehat{\delta} + \widehat{F}\widehat{\gamma} - W\delta - F\gamma\right\|_2$$
$$\leq \left(\frac{288 s_\mu \lambda^2}{\nu} + \frac{4}{T}\left(\|A\|_2 + \left\|M_{\widehat{F}} F\gamma\right\|_2\right)^2\right)^{1/2} + \frac{1}{\sqrt{T}}\left(\left\|M_{\widehat{F}} F\gamma\right\|_2 + \left\|P_{\widehat{F}}\mathcal{E}\right\|_2\right).$$

As standard with finite sample bounds on the LASSO (Bickel et al. 2009), we need that the penalty term is large enough, that is $\lambda \geq 2\Omega^*\left(\frac{1}{T}\widetilde{W}^\top \mathcal{E}\right)$. The quantity $\Omega^*\left(\frac{1}{T}\widetilde{W}^\top \mathcal{E}\right)$ is the effective noise of the problem (Lederer & Vogt 2021). In practice, λ is chosen by cross-validation. As in the literature, the finite sample bound here depends on the approximation error $\|A\|_2$, see Bickel et al. (2009) and Babii et al. (2022). Compared to the traditional LASSO literature, our bound contains two additional terms: $\left\|M_{\widehat{F}} F\gamma\right\|_2$ and $\left\|P_{\widehat{F}}\mathcal{E}\right\|_2$. They are present because we do not use the true F for estimation but rather its estimate $\widehat{F}$. The term $\left\|M_{\widehat{F}} F\gamma\right\|_2$ is the ℓ_2-norm of the projection of $F\gamma$ on the orthogonal of the vector space generated by the columns of $\widehat{F}$, while the quantity $\left\|P_{\widehat{F}}\mathcal{E}\right\|_2$ is the ℓ_2-norm of the projection of $\mathcal{E}$ on the the vector space generated by the columns of $\widehat{F}$. Using in particular Lemma 3.1, we can bound $\left\|M_{\widehat{F}} F\gamma\right\|_2$, $\left\|P_{\widehat{F}}\mathcal{E}\right\|_2$ and $\Omega^*\left(\frac{1}{T}\widetilde{W}^\top \mathcal{E}\right)$ in probability. This allows to obtain the following corollary, which states the rate of convergence of the prediction error of our estimator.

Corollary 1 *Under Assumptions 1, 2, 3 and 4, and letting* $\lambda = 2\Omega^*\left(\frac{1}{T}\widetilde{W}^\top \mathcal{E}\right)$, *we have*

$$\frac{1}{T}\left\|W\widehat{\delta} + \widehat{F}\widehat{\gamma} - W\delta - F\gamma\right\|_2^2 = O_P\left(s_\mu h_T^2 + \|A\|_2^2 + \frac{1}{p_x}\right).$$

When $s_\mu h_T^2 p_x \to \infty$, the term $\frac{1}{p_x}$ in the rate of convergence of Corollary 1 becomes negligible and our rate of convergence becomes the same as that of Babii et al. (2022). Hence, the condition $s_\mu h_T^2 p_x \to \infty$ guarantees that the error in estimating the factors does not influence the rate of convergence of the prediction errors. This condition requires that p_x grows sufficiently quickly with respect to T. The standard rate of convergence of the LASSO estimator in prediction error would replace h_T^2 by $\log(p)/T$. We have h_T^2 in our rate because of the presence of τ-mixing and polynomial tails.

4 Monte Carlo simulations

In this section, we present a Monte Carlo study to examine the finite sample performance of our proposed method. We simulate data using the following data generating process. We use model (1)-(2), where $\rho_0 = 0.5$, $J = 2$, $\rho_1 = 0.2$, $\rho_2 = 0.1$ are fixed across all scenarios. We consider the sample sizes $T \in \{50, 100, 200\}$, which are representative of typical settings in nowcasting applications. We study several scenarios, with different degrees of cross-sectional dependence and fatness of the tails of the variables.

The error term ε_t follows a mean-zero AR(1) process with an autocorrelation coefficient of 0.4 and variance 0.1. The error in this AR(1) process follows either a Gaussian or a student-$t(5)$ distribution.

The panel $\widetilde{z}$ represents high-frequency regressors that enter the model only through a sparse pattern. These variables are simulated at a weekly frequency assuming $m_z = 13$. We set $p_z = 30$ and the weekly variables are generated as AR(1) processes with an autocorrelation coefficient of $\rho_z = 0.4$ and mean-zero error terms that have either a Gaussian or a student-$t(5)$ distribution. The error terms have a cross-sectional covariance matrix specified as $(1-\rho_z^2)\Sigma$, where $\Sigma_{ij}^z = c_z^{|i-j|}$ for $i, j \in [p_z]$, where $c_z \in \{0.1, 0.4\}$ controls the degree of cross-sectional dependence among the regressors. We set $\omega_{z,1}$ and $\omega_{z,2}$ equal to the Beta$(1, 2)$ and Beta$(2, 2)$ densities, respectively. The other variables in $\widetilde{z}$ do not enter the model, that is $\omega_{z,k} \equiv 0$ for all $k \geq 3$.

Next, the factors in $\widetilde{f}$ are monthly variables, that is $m_x = 3$. We fix the number of factors to $K_f = 1$ and simulate the factor as a zero-mean AR(1) process with an autoregressive coefficient of 0.4 and a mean-zero error with unit variance which follows either a Gaussian or a student-$t(5)$ distribution. Given these factors, we generate the $p_x = 130$ monthly regressors in $\widetilde{x}$ according to the factor model (3), where the loadings $\left\{\widetilde{b}_{k,r},\ k \in [K_x], r \in [K_f]\right\}$ are i.i.d. uniform random variables on the interval $[-1, 1]$. The idiosyncratic components $\left\{\widetilde{u}_{t-(j-1)/m_x,k},\ t \in [T], k \in [K_x], j \in [m_x]\right\}$ are modeled as zero-mean autoregressive processes with an autoregressive coefficient of $\rho_u = 0.4$, with error terms following a Gaussian or a student-$t(5)$ distribution with cross-sectional covariance matrix $\Sigma\,(1-\rho_u^2)$. The matrix Σ is defined as $\Sigma_{ij} = c_u^{|i-j|}$ for $i, j \in [p_x]$, where $c_u \in \{0.1, 0.4\}$ controls the degree of cross-sectional dependence among the monthly regressors. For $\omega_{x,1}$ and $\omega_{x,2}$ we use Beta$(1, 2)$

	M1	M2	M3	M1	M2	M3
	Gaussian			student-$t(5)$		
Panel A. *Baseline* ($c_z = c_u = 0.1$)						
T = 50	1.5023	1.3753	1.1760	3.2512	2.2274	1.9531
100	1.4983	1.2031	0.8992	2.5382	1.9978	1.5269
200	1.3992	0.9425	0.6133	2.4736	1.6820	1.1065
Panel B. *Stronger cross-sectional dependence* ($c_z = c_u = 0.4$)						
T = 50	1.5686	1.4210	1.2540	3.0348	2.1126	2.0059
100	1.4382	1.2322	0.8883	2.3524	1.9178	1.4572
200	1.3819	0.8924	0.5658	2.4625	1.6933	1.1337

Table 1: Monte Carlo MSE comparisons — M1 - FAMIDAS, M2 - sg-LASSO-MIDAS, M3 - sg-LASSO-FAMIDAS.

and Beta$(2, 2)$ densities, respectively. For all $k \geq 3$, $\omega_{x,k} \equiv 0$, that is only the first two variables in $\widetilde{x}$ enter the model.

For the model estimation, we use Legendre polynomials of degree 3 to approximate the Beta density functions used in the weighting process ($D = 4$). This choice ensures a flexible yet accurate representation of the underlying relationships. It is important to note that the model is simulated based on Beta density functions which we aim to approximate, hence the regressors that enter the sg-LASSO estimator are estimated based on Legendre polynomials. Lastly, the the number of factors is estimated using the growth ratio estimator of Ahn & Horenstein (2013). The methods evaluated in our study correspond to those described in Section 2.

Table 1 presents the mean squared error (MSE) results of three different methods—M1 (FAMIDAS), M2 (sg-LASSO-MIDAS), and M3 (sg-LASSO-FAMIDAS)—under various simulation settings. Panel A reports results for the baseline scenario with weak cross-sectional dependence ($c_z = c_u = 0.1$), while Panel B shows results for a scenario with stronger cross-sectional dependence ($c_z = c_u = 0.4$). In both panels, the performance is evaluated under two types of error distributions: Gaussian and student-$t(5)$.

Across all sample sizes $T \in \{50, 100, 200\}$, M3 (sg-LASSO-FAMIDAS) consistently

achieves the lowest MSE compared to M1 and M2, indicating superior predictive accuracy. As the sample size increases, all methods demonstrate improved performance (i.e., lower MSE), but the gains are most pronounced for M3. For example, in the baseline scenario (Panel A), M3's MSE decreases from 1.1760 to 0.6133 under the Gaussian errors when T increases from 50 to 200. In the stronger cross-sectional dependence scenario (Panel B), the relative performance of M3 remains robust, maintaining the lowest MSE across all settings. Notably, under the student-$t(5)$ distribution, M3 shows a significant relative improvement over M1 and M2, particularly for smaller sample sizes ($T = 50$), suggesting that the approach is more resilient to heavy-tailed errors when the true data generating process is sparse plus dense.

Overall, the results indicate that sg-LASSO-FAMIDAS (M3) outperforms the other methods in terms of MSE, especially as the sample size increases or when errors exhibit heavy tails, highlighting its effectiveness in handling different data-generating scenarios.

5 Nowcasting GDP growth

5.1 Description of the problem

For our empirical application, we study nowcasting for the current quarter's GDP growth and compute performance metrics across various monthly horizons. It is important to note that the release of the U.S. — or any other country — GDP data typically occurs one month after the quarter of interest or even later. Figure 1 provides a visual representation of the timeline involved in the nowcasting process. Consider the following scenario: we find ourselves at the end of the first month of the year, aiming to nowcast the GDP for the first quarter. In pursuit of this goal, the econometrician trains the model using data spanning up to (and including) the first month of the fourth quarter of the preceding year. The high-frequency data from the first month of the current quarter (Q1) is used to generate the nowcast. The parameter h represents the time remaining until the end of the target quarter. We will consider three nowcasting exercises: forecasting the GDP of the current quarter two months before its conclusion ($h = 2$, as in Figure 1), one month ahead ($h = 1$), and nowcasting at the end of the quarter ($h = 0$).

In our model, we use m_z and m_x to denote the number of past values of high-frequency predictors used. We always use all the data from the previous quarter and the data of the quarter of interest that is available at the time of prediction. For instance, in the context of the monthly macro panel of predictor variables, when $h = 0$, we use $m_x = 6$ lags (the three months of the quarter of interest and the three months of the previous quarter). Similarly, for $h = 1$ (respectively, $h = 2$), we consider the last $m_x = 5$ (respectively, $m_z = 4$) values at the time of prediction. In the weekly financial panel, we set $m_z = 26$ when $h = 0$, $m_z = 21$ when $h = 1$ and $m_z = 17$ when $h = 2$, which approximately corresponds to the number of weeks available in the quarter of interest plus the number of weeks in the previous quarter.

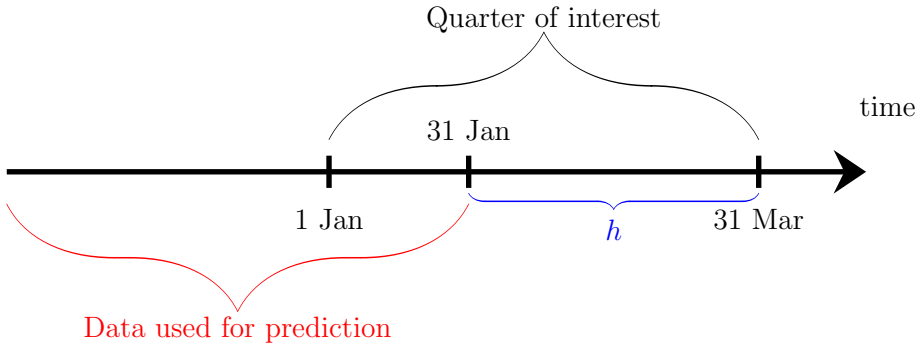


Figure 1: Timeline illustration of nowcasting.

5.2 Data

We apply our methods to nowcast the initial release of real U.S. GDP growth, often referred to as the "advance" release. Typically, this release becomes available towards the end of the first month of the subsequent quarter, except for the 2018 Q4 release. In this case, the first release was made available on February 28th, 2019 (second month) due to a partial government shutdown from December 22, 2018, to January 25, 2019, lasting 35 days. Additional details regarding this event can be found at the Bureau of Economic Analysis. In our analysis, we make use of real-time GDP vintages sourced from the Archival Federal Reserve Economic Data St. Louis Fed (ALFRED) database. We have set the in-sample data period from 1984 Q1 to 2007 Q4. Our first nowcast is for 2008 Q1, and we use an expanding window approach to compute nowcasts for subsequent quarters.

Monthly macroeconomic regressors $\tilde{\boldsymbol{x}}$. The monthly covariates and factors are based on the well-known FRED-MD real-time dataset, see McCracken & Ng (2016) for further details. The variables are transformed using the transformations suggested in McCracken & Ng (2016). We restrict the series so that all of them are available in all vintages used for out-of-sample forecasting. We also omit all financial series from the monthly panel. In total, we use 76 monthly series in our final monthly panel.

Weekly financial regressors $\tilde{\boldsymbol{z}}$. We use the weekly financial variables used to construct the National Financial Conditions Index (NFCI) by the Chicago Fed. The original NFCI is a latent factor computed at a weekly frequency using weekly, monthly, and quarterly series. Instead of using the full list of series, we opt to use the majority of weekly series in the NFCI index without using monthly or quarterly series. In total, we collected 36 weekly financial series. Since financial data is accessible in real time, there is no need to account for publication delays or irregularities for this set of predictor variables. However, some financial series may have shorter samples compared to macroeconomic and other longer financial series. To incorporate all series effectively, we apply matrix completion methods based on nuclear-norm regularization to impute missing data. This technique helps balance the financial series and is applied at the original frequency of the variables, see Section D of the Online Appendix for details.

We provide the full list of series for monthly macro and weekly financial data with additional details in Section B of the Online Appendix.

5.3 Empirical results

Our candidate model set includes the AR(4) model, which we regard as the simplest and designate as our benchmark model. Consequently, our reported root mean squared forecast errors are presented relative to the AR(4). Notably, our conclusions remain consistent even when employing alternative benchmarks such as the random walk or AR(1). Our main results are based on the assumption that monthly macro covariates follow a factor model, hence entering the nowcasting equation with sparse plus dense signal. The weekly financial series are additional regressors that influence the target variables in a sparse way. We have four autoregressive terms ($J = 4$). As in the simulations, we use Legendre polynomials of

degree 3 as the dictionary ($D = 4$). In practice, we choose all tuning parameters by cross-validation adjusting for time series dependence, see, e.g., Section 4 in Babii et al. (2022). We estimate the number of factors using the growth ratio estimator of Ahn & Horenstein (2013).

Table 2 presents results for two sub-samples: Panel A covers the entire dataset, spanning the out-of-sample evaluation period from 2008 Q1 to 2022 Q2 (Full sample), while Panel B focuses on the period up to 2019 Q4, excluding the COVID-19 period (Up to COVID). The results for the autoregressive benchmark model are provided in absolute terms, with the remaining RMSEs reported relative to this benchmark. The benchmark model shows a substantial increase in prediction errors during the COVID-19 period, which is expected given the autoregressive model's inherent lack of high-frequency information.

The results demonstrate that our factor-augmented sparse MIDAS regression greatly outperforms both the FAMIDAS (dense) and sg-LASSO-MIDAS (sparse) methods for the full sample. The most pronounced improvement occurs during the COVID-19 period, highlighting the advantage of incorporating factor-augmented sparse methods in periods of high uncertainty. In contrast, during the pre-pandemic period, the quality of nowcasts of both sparse and factor-augmented sparse MIDAS regressions is similar, indicating that the added value of the factor augmentation is particularly evident under the more volatile economic environments.

To assess statistical significance, we apply the average superior predictive ability (aSPA) test proposed by Quaedvlieg (2021), which allows for a comprehensive evaluation across three nowcasting horizons. However, it is important to interpret these results with caution, as the test has not been specifically validated for the nowcasting methods used in our empirical application. The outcomes of the aSPA tests are presented in Table OA.4 in the Online Appendix. Results confirm that for the full sample, factor-augmented sparse MIDAS regression significantly outperforms the two alternative approaches at 5% significance level.

In Figure 2, we present the square-root cumulative sum of squared forecast errors (CUMSUM) for three competing methods across two sub-samples. These graphs are designed to visualize the performance differences between the models throughout the out-of-sample period. The forecast errors are calculated for the end-of-quarter horizon, and the CUMSUM

	2-month	1-month	EoQ
	Panel A. *Full sample*		
AR(4)	9.553	9.553	9.553
FAMIDAS	0.590	0.495	0.580
sg-LASSO-MIDAS	0.572	0.435	0.559
sg-LASSO-FAMIDAS	0.580	0.340	0.251
	Panel B. *Up to COVID*		
AR(4)	1.934	1.934	1.934
FAMIDAS	0.962	0.911	0.851
sg-LASSO-MIDAS	0.827	0.837	0.733
sg-LASSO-FAMIDAS	0.842	0.827	0.756

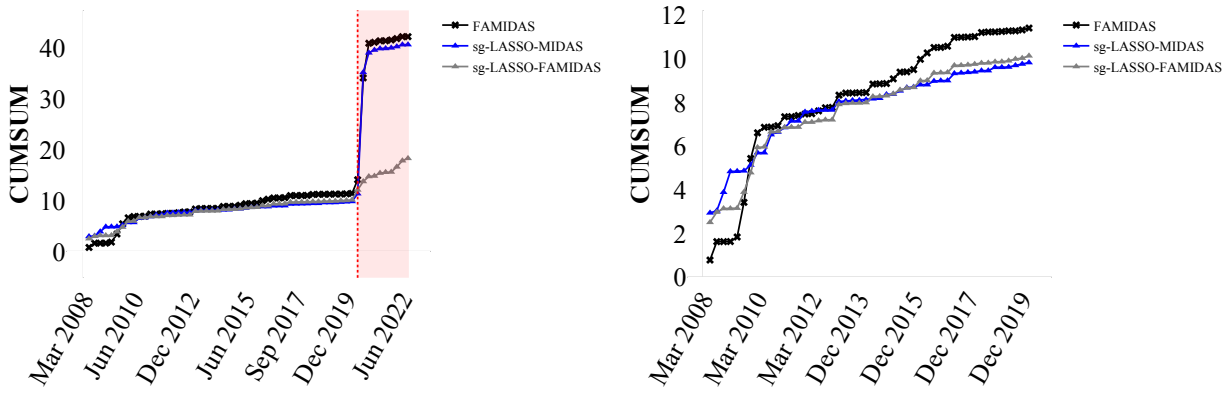
Table 2: Nowcast comparisons — horizons are 2- and 1-month ahead, as well as the end of the quarter (EoQ). We report results for the full sample in Panel (A), and Panel (B) results excluding the COVID pandemic period, while Panel (C) reports results for the COVID pandemic period and beyond. The out-of-sample period starts from 2008 Q1 to 2022 Q2 (Panel A) and from 2008 Q1 to 2019 Q4 (Panel B). The RMSEs are reported relative to the AR model.

is computed as follows:

$$\text{CUMSUM}_{t,t+k} = \sqrt{\sum_{q=t}^{t+k} \hat{\epsilon}_{q,j}^2},$$

where $\hat{\epsilon}_{q,j}$, $j \in$ {FAMIDAS, sg-LASSO-MIDAS, sg-LASSO-FAMIDAS} are the out-of-sample nowcast errors. We plot the CUMSUM for the full sample period, corresponding to Panel A in Table 2, and up to the COVID pandemic which corresponds to Panel B in the same Table.

The plots highlight the differences between the methods, particularly during the onset of the COVID-19 outbreak. Prior to the pandemic, the performance of the sg-LASSO-FAMIDAS and sg-LASSO-MIDAS models is comparable, with the former providing more accurate nowcasts during the financial crisis and the latter performing better during stable periods between crises. This suggests that factor augmentation may not significantly impact performance during stable periods, despite the need to estimate additional unpenalized regression coefficients and the factors themselves. As the COVID-19 period unfolds, both the FAMIDAS and sg-LASSO-MIDAS methods show a marked decline in prediction accuracy, whereas the sg-LASSO-FAMIDAS method demonstrates notable resilience in handling the substantial shock introduced by the pandemic.



(a) Full sample.

(b) Up to COVID.

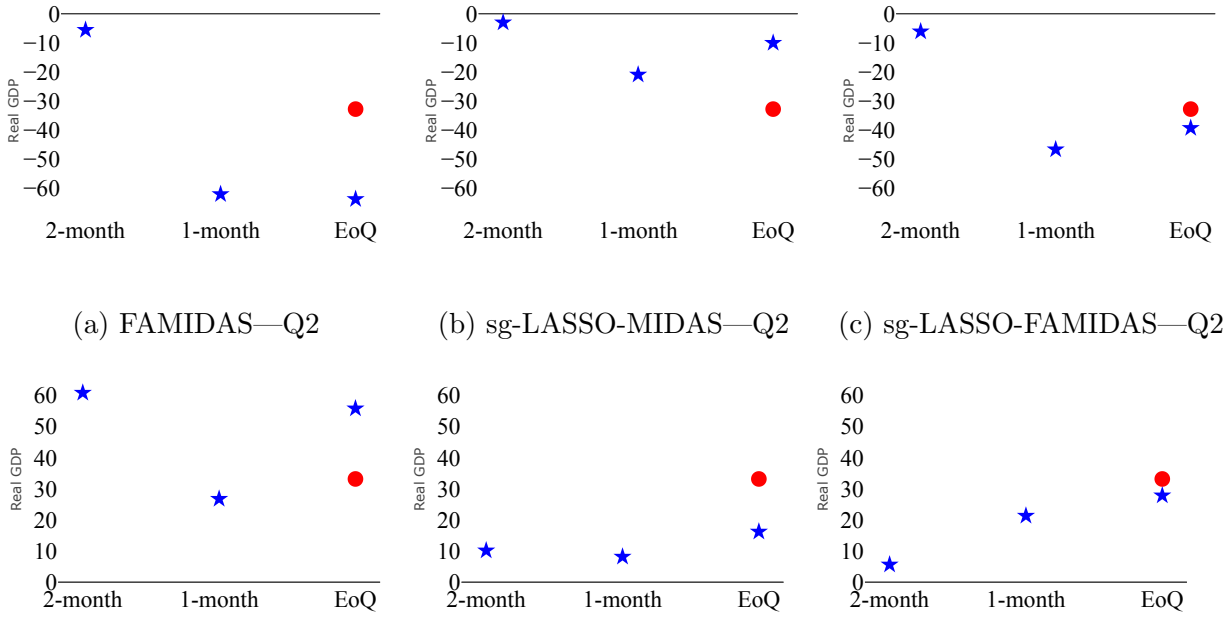
Figure 2: The figure illustrates the CUMSUM for the FAMIDAS, sg-LASSO-MIDAS and sg-LASSO-FAMIDAS using the pre-COVID pandemic period in Figure 2a) while Figure 2b plot the whole out-of-sample period.

In Figure 3, we display the nowcasts for the three forecasting horizons and three meth-

ods under consideration, corresponding to the RMSEs reported in Table 2, alongside the advance release of real GDP, which serves as the target for the nowcasts. These figures visually demonstrate the proximity of the nowcasts from each model to the advance estimate, as well as how the flow of information improves the accuracy of the nowcasts. We focus on the quarters most impacted by the COVID-19 pandemic, specifically Q2 and Q3 of 2020. Notably, for both quarters, the factor-augmented sparse regression model produces highly accurate nowcasts by the end of the quarter. It is important to note that the advance estimate is only available one month into the subsequent quarter. Additionally, the results show that the availability of information throughout the quarter consistently improves the nowcasts. However, the sg-LASSO-MIDAS method, which relies solely on sparsity, tends to under-predict the absolute value of GDP growth. In contrast, the FAMIDAS method, which uses dense signals, generates more erratic predictions. Overall, the figure illustrates that factor-augmented sparse MIDAS regression produces balanced nowcasts across the two consecutive quarters affected by the COVID-19 pandemic.

We also perform a comparative assessment of the sg-LASSO-FAMIDAS method by replacing the macro principal component factors with widely recognized indicators from the literature. Specifically, we consider the Aruoba-Diebold-Scotti index (ADS) as described in Aruoba et al. (2009), the Chicago Fed National Activity Index (CFNAI) as presented in Brave et al. (2019), and the National Financial Conditions Index (NFCI) as discussed in Brave & Kelley (2017) and Amburgey & McCracken (2023). Vintages for all three factors were obtained from the Federal Reserve Banks of Philadelphia, Chicago, and St. Louis, respectively. We maintain the same MIDAS scheme and number of lags for all factors as used for the predictors, treating ADS and NFCI as weekly variables and CFNAI as a monthly variable. Notably, for NFCI, we use vintages provided by Amburgey & McCracken (2023) due to its longer time span, which aligns more with our empirical analysis. Detailed information on the data sources, including web links, can be found in the Online Appendix.

The results, presented in Table 3, indicate that ADS and CFNAI outperform NFCI, yielding more accurate nowcasts, with ADS providing the most consistent improvements. This outcome is likely due to ADS's weekly frequency and its reliance on fewer, more stable predictors, which result in smoother MIDAS weight estimates and a clearer indication of economic conditions, supporting the findings of Diebold (2020). More importantly, both



(a) FAMIDAS—Q2 (b) sg-LASSO-MIDAS—Q2 (c) sg-LASSO-FAMIDAS—Q2

(d) FAMIDAS—Q3 (e) sg-LASSO-MIDAS—Q3 (f) sg-LASSO-FAMIDAS—Q3

Figure 3: The figure plots the nowcasts of U.S. GDP growth for the FAMIDAS, sg-LASSO-MIDAS and sg-LASSO-FAMIDAS during the pandemic period for the three horizons we consider: 2-month, 1-month and end of quarter (EoQ). The blue star in Figure 3a-3c plots the nowcasts for the three methods for 2020 Q2 while Figure 3d-3f plots nowcasts for the same methods but for 2020 Q3. The red circle plots the *advanced release* of the respective quarter, i.e., the target nowcast.

ADS and CFNAI are macroeconomic factors constructed to track real economic activity, while NFCI is financial conditions factor. Results thus support our primary findings, where macroeconomic predictors seem to generate sparse and dense signals, while financial variables signals are sparse. Additionally, when comparing these results with those from our approach, where factors are estimated via PCA, our method demonstrates superior performance.

	2-month	1-month	EoQ
	Panel A. *Full sample*		
ADS	0.553	0.379	0.258
CFNAI	0.325	0.446	0.334
NFCI	0.699	0.505	0.549
	Panel B. *Up to COVID*		
ADS	0.810	0.782	0.742
CFNAI	0.800	0.814	0.732
NFCI	0.895	0.932	0.924

Table 3: Nowcast comparisons — horizons are 2- and 1-month ahead, as well as the end of the quarter (EoQ). We report results for the full sample in Panel (A), and Panel (B) results excluding the COVID pandemic period, while Panel (C) reports results for the COVID pandemic period and beyond. The out-of-sample period starts from 2008 Q1 to 2022 Q2 (Panel A) and from 2008 Q1 to 2019 Q4 (Panel B). The RMSEs are reported relative to the AR(4) model.

6 Extension to factor-augmented sparse MIDAS logistic regression and nowcasting recessions

Our main empirical results reveal that during periods of economic crises, factor-augmented sparse regression yields the most accurate nowcasts. This method integrates both factor-based and sparse approaches, enhancing its predictive accuracy in volatile conditions. To further evaluate its robustness in predicting economic downturns, we modify our target variable from real GDP growth to the NBER recession indicator, as the National Bureau

of Economic Research (NBER) is the authoritative body responsible for officially calling recessions in the U.S. This adjustment shifts our focus to predicting the onset of recessions rather than merely forecasting GDP growth. Additionally, we estimate a logit model instead of a linear one to better handle the binary nature of recession events and improve the model's performance in identifying recessions. This methodological shift aims to refine our assessment of the model's effectiveness in accurately forecasting economic contractions as determined by the NBER.

Specifically, we aim to nowcast the recession indicator r_t using the same regressors as in the nowcasting of real GDP growth y_t discussed in the previous section. We apply the same estimation approach as in the previous section, but in this case, the ℓ_2 loss function is replaced by the logistic loss function:

$$\sum_{t=1}^{T} \left(\log\left(1 + \exp(g_t)\right) - r_t g_t \right),$$

where g_t denotes the linear combination of predictors at time t. Lastly, we lag the data by 8 quarters as suggested by Furno & Giannone (2024), to account for the delay in the reporting of recession periods. That is, we fit the models with the data through time $t-8$ and use the estimated model to nowcast the recession probabilities based on the data available at time t. The rest of the procedure is the same as in the previous section.

Results are reported in Table 4. Notably, factor-augmented sparse regression model, sg-LASSO-FAMIDAS, leads to the most accurate probability nowcasts in terms of Area Under the Curve (AUC). In Figure 4, we plot the nowcasted probabilities for sg-LASSO-FAMIDAS across the three horizons. The results show that as more data becomes available throughout the quarter, the method's predictions improve and become more stable. This superior performance shows the effectiveness of integrating factor models with sparse regression techniques, particularly in capturing the complex dynamics of economic downturns.

7 Conclusion

In this paper, we proposed a novel factor-augmented sparse MIDAS regression model tailored for high-dimensional mixed-frequency data. Our methodology leverages both sparse and dense structures to effectively capture the complex dependencies inherent in data

	2-month	1-month	EoQ
		Full sample	
FAMIDAS	0.710	0.765	0.743
sg-LASSO-MIDAS	0.863	0.840	0.957
sg-LASSO-FAMIDAS	0.908	0.950	0.980

Table 4: Nowcast of recession comparisons — horizons are 2- and 1-month ahead, as well as the end of the quarter (EoQ). We report results for the full sample. The out-of-sample period starts from 2008 Q1 to 2022 Q2. The reported values are AUC.

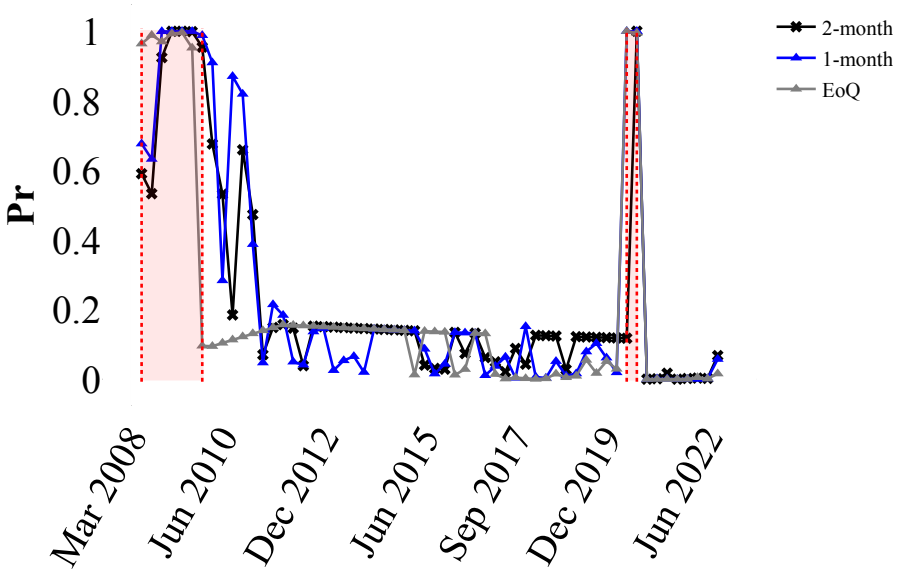


Figure 4: The figure illustrates the nowcasted probabilities of a recession at different horizons for the logistic sg-LASSO-FAMIDAS model. The red shaded areas are the *NBER* reported recession periods.

sampled at different frequencies. We derived the theoretical properties of our estimator, demonstrating its robustness under τ-mixing processes and its adaptability to misspecification due to approximating MIDAS lags polynomials and/or approximate sparsity. This theoretical framework extends the existing literature by establishing convergence rates for factor-augmented sparse regression under a broader set of assumptions, thereby filling a gap in the understanding of high-dimensional mixed-frequency estimation.

In a series of Monte Carlo simulations, the proposed approach consistently outperformed standard sparse and dense-only models, especially during periods of heightened economic volatility. Empirically, our method exhibited strong performance in nowcasting U.S. GDP growth using a combination of weekly financial and monthly macroeconomic data. The gains from using our approach are particularly pronounced during the COVID-19 pandemic. Notably, the findings indicate that during stable economic periods, sparse techniques alone can suffice, whereas the integration of dense factors becomes crucial during times of macroeconomic instability.

Moreover, we extended our approach to a logistic regression framework for nowcasting U.S. recessions, capturing the binary nature of recessionary periods more effectively. This extension confirmed that the factor-augmented sparse MIDAS model is well-suited for predicting the onset of recessions, highlighting the robustness and versatility of our methodology in different forecasting contexts.

Overall, our research provides a comprehensive framework for modeling high-dimensional mixed-frequency data, offering significant improvements in both theory and practice. Our findings have practical implications for policymakers and practitioners involved in economic forecasting and nowcasting, particularly in scenarios characterized by significant economic disruptions. Future research could explore extensions to more complex nonlinear models or the inclusion of alternative high-frequency indicators, thereby broadening the applicability of factor-augmented sparse regression models to a wider array of economic forecasting problems.

Funding sources

Jad Beyhum gratefully acknowledges financial support from the Research Fund KU Leu-

ven through the grant STG/23/014. Jonas Striaukas gratefully acknowledges the financial support of F.R.S.-FNRS PDR project Nr. PDR T.0044.22 and from the European Commission, MSCA-2022-PF Individual Fellowship, Project 101103508, Project Acronym: MACROML.

Supplementary material

The Online Appendix contains the proofs of our theoretical results, details on the data and additional empirical results.

References

Ahn, S. C. & Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**(3), 1203–1227.

Amburgey, A. J. & McCracken, M. W. (2023), 'On the real-time predictive content of financial condition indices for growth', *Journal of Applied Econometrics* **38**(2), 137–163.

Andreou, E., Ghysels, E. & Kourtellos, A. (2010), 'Regression models with mixed sampling frequencies', *Journal of Econometrics* **158**(2), 246–261.

Andreou, E., Ghysels, E. & Kourtellos, A. (2013), 'Should macroeconomic forecasters use daily financial data and how?', *Journal of Business & Economic Statistics* **31**(2), 240–251.

Aruoba, S. B., Diebold, F. X. & Scotti, C. (2009), 'Real-time measurement of business conditions', *Journal of Business & Economic Statistics* **27**(4), 417–427.

Babii, A., Ghysels, E. & Striaukas, J. (2022), 'Machine learning time series regressions with an application to nowcasting', *Journal of Business & Economic Statistics* **40**(3), 1094–1106.

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**(1), 135–171.

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. & Ng, S. (2006), 'Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions', *Econometrica* **74**(4), 1133–1150.

Bai, J. & Ng, S. (2019), 'Rank regularized estimation of approximate factor models', *Journal of Econometrics* **212**(1), 78–96.

Barigozzi, M., Cho, H. & Owens, D. (2024), 'FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series', *Journal of Business & Economic Statistics* **42**(3), 890–902.

Beyhum, J. & Gautier, E. (2023), 'Factor and factor loading augmented estimators for panel regression with possibly nonstrong factors', *Journal of Business & Economic Statistics* **41**(1), 270–281.

Beyhum, J. & Striaukas, J. (2024), 'Testing for sparse idiosyncratic components in factor-augmented regression models', *Journal of Econometrics* **244**(1), 105845.

Bickel, P. J., Ritov, Y. & Tsybakov, A. B. (2009), 'Simultaneous analysis of Lasso and Dantzig selector', *Annals of Statistics* **37**(4), 1705 – 1732.

Brave, S. A., Butters, R. A., Kelley, D. et al. (2019), 'A new "big data" index of us economic activity', *Economic Perspectives, Federal Reserve Bank of Chicago* **1**.

Brave, S. A. & Kelley, D. (2017), 'Introducing the Chicago fed's new adjusted national financial conditions index', *Chicago Fed Letter* **386**, 2017.

Carriero, A., Clark, T. E., Marcellino, M. & Mertens, E. (2024), 'Addressing covid-19 outliers in bvars with stochastic volatility', *Review of Economics and Statistics* **106**(5), 1403–1417.

Dedecker, J. & Prieur, C. (2004), 'Coupling for τ-dependent sequences and applications', *Journal of Theoretical Probability* **17**(4), 861–885.

Diebold, F. X. (2020), Real-time real economic activity: Exiting the great recession and entering the pandemic recession, Technical report, National Bureau of Economic Research.

Fan, J., Guo, J. & Zheng, S. (2022), 'Estimating number of factors by adjusted eigenvalues thresholding', *Journal of the American Statistical Association* **117**(538), 852–861.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **75**(4), 603–680.

Fan, J., Masini, R. P. & Medeiros, M. C. (2023), 'Bridging factor and sparse models', *Annals of Statistics* **51**(4), 1692–1717.

Foroni, C., Marcellino, M. & Stevanovic, D. (2022), 'Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis', *International Journal of Forecasting* **38**(2), 596–612.

Freeman, H. & Weidner, M. (2023), 'Linear panel regressions with two-way unobserved heterogeneity', *Journal of Econometrics* **237**(1), 105498.

Furno, F. & Giannone, D. (2024), 'Nowcasting recession risk', *Research Methods and Applications on Macroeconomic Forecasting* .

Hansen, C. & Liao, Y. (2019), 'The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications', *Econometric Theory* **35**(3), 465–509.

Hauzenberger, N., Marcellino, M., Pfarrhofer, M. & Stelzer, A. (2024), 'Nowcasting with mixed frequency data using gaussian processes', *arXiv preprint arXiv:2402.10574* .

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M. & Schreiner, J. (2023), 'Nowcasting in a pandemic using non-parametric mixed frequency VARs', *Journal of Econometrics* **232**(1), 52–69.

Koh, Y. J. (2023), Inference for factor-MIDAS regression models, Technical report.
URL: *https://www.yookyungjuliakoh.com/assets/img/draft_July2024.pdf*

Krampe, J. & Margaritella, L. (2021), 'Factor models with sparse VAR idiosyncratic components', *arXiv preprint arXiv:2112.07149* .

Lederer, J. & Vogt, M. (2021), 'Estimating the Lasso's effective noise.', *Journal of Machine Learning Research* **22**, 276–1.

Marcellino, M. & Schumacher, C. (2010), 'Factor midas for nowcasting and forecasting with ragged-edge data: A model comparison for german GDP', *Oxford Bulletin of Economics and Statistics* **72**(4), 518–550.

McCracken, M. W. & Ng, S. (2016), 'FRED-MD: a monthly database for macroeconomic research', *Journal of Business & Economic Statistics* **34**(4), 574–589.

Onatski, A. (2010), 'Determining the number of factors from empirical distribution of eigenvalues', *Review of Economics and Statistics* **92**(4), 1004–1016.

Quaedvlieg, R. (2021), 'Multi-horizon forecast comparison', *Journal of Business & Economic Statistics* **39**(1), 40–53.

Stock, J. H. & Watson, M. W. (1999), 'Forecasting inflation', *Journal of monetary economics* **44**(2), 293–335.

Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American statistical association* **97**(460), 1167–1179.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **58**(1), 267–288.

Vogt, M., Walsh, C. & Linton, O. (2022), 'CCE estimation of high-dimensional panel data models with interactive fixed effects', *arXiv preprint arXiv:2206.12152* .

Online appendix of "Factor-augmented sparse MIDAS regressions with an application to nowcasting"

Jad Beyhum
Department of Economics, KU Leuven, Belgium

Jonas Striaukas
Department of Finance, Copenhagen Business School, Denmark

November 8, 2024

Contents

A Proofs **OA. - 1**

A.1 Proof of Theorem 3.1 . . . . . OA. - 1

A.2 Proof of Corollary 1 . . . . . OA. - 5

A.3 Proof of Lemma 3.1 . . . . . OA. - 5

A.4 Auxiliary lemmas . . . . . OA. - 8

A.5 Pre-existing results . . . . . OA. - 22

B Additional details on the data **OA. - 23**

C Additional empirical results **OA. - 28**

D Details on matrix completion **OA. - 28**

A Proofs

This section is organized as follows. First, Section A.1 contains the proof of Theorem 3.1. Next, Corollary 1 is proved in Section A.2. Then, the proof of Lemma 3.1 can be found in Section A.3. The proofs of Theorem 3.1, Corollary 1 and Lemma 3.1 rely on some auxiliary lemmas presented and proved in Section A.4 and some pre-existing results restated in Section A.5.

We will use the following notation. First, let $Y = (y_1, \dots, y_T)^\top$ and $\widehat{\Sigma} = \widetilde{W}^\top \widetilde{W}/T$. Next, for a matrix H, $\|H\|_{op} = \sigma_1(H)$ is its operator norm. For two matrices $H^{(1)}, H^{(2)}$ of size $n_1 \times n_2$, their scalar product is $\langle H^{(1)}, H^{(2)} \rangle = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} H^{(1)}_{i,j} H^{(2)}_{i,j}$. For $N \in \mathbb{N}$, I_N is the identity matrix of size $N \times N$.

A.1 Proof of Theorem 3.1

Throughout the proof, we work on the event

$$\left\{ \left\| \widehat{\Sigma} - \Sigma \right\|_\infty \leq \frac{\nu}{32 s_\mu G^*} \right\},$$

which has probability going to 1 by Lemma A.5 (i) and Assumption 4 (i). First, remark that, by definition of $M_{\widehat{F}}$ and the Frisch-Waugh-Lovell theorem, we have

$$\widehat{\delta} \in \arg\min_{d \in \mathbb{R}^p} \frac{1}{T} \left\| \widetilde{Y} - \widetilde{W} d \right\|_2^2 + 2\lambda \Omega(d),$$

where $\widetilde{Y} = M_{\widehat{F}} Y$. By Fermat's rule, $\widehat{\delta}$ therefore satisfies

$$\frac{1}{T} \widetilde{W}^\top (\widetilde{W}\widehat{\delta} - \widetilde{Y}) + \lambda z^* = 0, \text{ for some } z^* \in \partial\Omega(\widehat{\delta}),$$

where $\partial\Omega(\widehat{\delta})$ is the sub-differential of $d \mapsto \Omega(d)$ at $\widehat{\delta}$. Taking the inner product with $\delta - \widehat{\delta}$, we get

$$\frac{1}{T} \left\langle \widetilde{W}^\top (\widetilde{Y} - \widetilde{W}\widehat{\delta}), \delta - \widehat{\delta} \right\rangle = \lambda \left\langle z^*, \delta - \widehat{\delta} \right\rangle \leq \lambda \left\{ \Omega(\delta) - \Omega(\widehat{\delta}) \right\}, \tag{OA.1}$$

where the last inequality follows from the definition of the sub-differential. Notice that,

$$\begin{aligned} &\left\langle \widetilde{W}^\top (\widetilde{Y} - \widetilde{W}\widehat{\delta}), \delta - \widehat{\delta} \right\rangle \\ &= \left\langle \widetilde{W}^\top (\widetilde{Y} - \widetilde{W}\delta), \delta - \widehat{\delta} \right\rangle + \left\langle \widetilde{W}^\top \widetilde{W} (\delta - \widehat{\delta}), \delta - \widehat{\delta} \right\rangle \\ &= \left\langle \widetilde{W}^\top \mathcal{E}, \delta - \widehat{\delta} \right\rangle + \left\langle \widetilde{W}^\top (F\gamma + A), \delta - \widehat{\delta} \right\rangle + \left\| \widetilde{W}(\widehat{\delta} - \delta) \right\|_2^2, \end{aligned} \tag{OA.2}$$

where we used $Y = W\delta + F\gamma + A + \mathcal{E}$ in the last equality. By definition of the dual norm and the inequality of Cauchy-Schwarz, (OA.1) and (OA.2) yield

$$\begin{aligned}
&\frac{1}{T}\left\|\widetilde{W}(\hat{\delta}-\delta)\right\|_2^2 - \lambda\left\{\Omega(\delta)-\Omega(\hat{\delta})\right\} \\
&\leq \frac{1}{T}\left\langle \widetilde{W}^\top\left(\widetilde{Y}-\widetilde{W}\delta\right), \hat{\delta}-\delta\right\rangle \\
&\leq \Omega^*\left(\frac{1}{T}\widetilde{W}^\top\mathcal{E}\right)\Omega\left(\hat{\delta}-\delta\right) + \frac{1}{T}\left\|\widetilde{W}(\hat{\delta}-\delta)\right\|_2\left(\|A\|_2 + \left\|M_{\hat{F}}F\gamma\right\|_2\right).
\end{aligned} \tag{OA.3}$$

Using that $\lambda \geq 2\Omega^*\left(\frac{1}{T}\widetilde{W}^\top\mathcal{E}\right)$, we get

$$\begin{aligned}
&\frac{1}{T}\left\|\widetilde{W}\left(\hat{\delta}-\delta\right)\right\|_2^2 - \lambda\left\{\Omega(\delta)-\Omega(\hat{\delta})\right\} \\
&\leq \frac{\lambda}{2}\Omega\left(\hat{\delta}-\delta\right) + \frac{1}{T}\left\|\widetilde{W}(\hat{\delta}-\delta)\right\|_2\left(\|A\|_2 + \left\|M_{\hat{F}}F\gamma\right\|_2\right).
\end{aligned} \tag{OA.4}$$

Now, for $d \in \mathbb{R}^p$, define

$$\begin{aligned}
\Omega_0(d) &= \mu\|d_{\mathcal{S}_0}\|_1 + (1-\mu)\sum_{G\in\mathcal{G}_0}\|d_G\|_2; \\
\Omega_1(d) &= \mu\|d_{\mathcal{S}_0^c}\|_1 + (1-\mu)\sum_{G\in\mathcal{G}_0^c}\|d_G\|_2.
\end{aligned}$$

Let also $\Delta = \hat{\delta}-\delta$. Next, remark that $\Omega(d) = \Omega_0(d)+\Omega_1(d)$, $\Omega_1(\delta) = 0$ and $\Omega_1(\hat{\delta}) = \Omega_1(\Delta)$. This leads to

$$\begin{aligned}
\Omega(\delta) - \Omega(\hat{\delta}) &= \Omega_0(\delta) + \Omega_1(\delta) - \Omega_0(\hat{\delta}) - \Omega_1(\hat{\delta}) \\
&= \Omega_0(\delta+\Delta-\Delta) - \Omega_0(\hat{\delta}) - \Omega_1(\Delta) \\
&\leq \Omega_0(\hat{\delta}) + \Omega_0(\Delta) - \Omega_0(\hat{\delta}) - \Omega_1(\Delta) \quad \text{(By the triangle inequality)} \\
&= \Omega_0(\Delta) - \Omega_1(\Delta).
\end{aligned}$$

Combining this and (OA.4), we get

$$\begin{aligned}
\frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 &\leq \lambda\left\{\Omega_0(\Delta) - \Omega_1(\Delta)\right\} \\
&\quad + \frac{\lambda}{2}\Omega(\Delta) + \frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2\left(\|A\|_2 + \left\|M_{\hat{F}}F\gamma\right\|_2\right).
\end{aligned} \tag{OA.5}$$

We consider two cases.

Case 1: $\|A\|_2 + \left\|M_{\hat{F}}F\gamma\right\|_2 \leq \frac{1}{2}\left\|\widetilde{W}\Delta\right\|_2$. In this case, (OA.5) implies

$$\frac{1}{2T}\left\|\widetilde{W}\Delta\right\|_2^2 \leq \lambda\left\{\Omega_0(\Delta) - \Omega_1(\Delta)\right\} + \frac{\lambda}{2}\Omega(\Delta). \tag{OA.6}$$

Since the left-hand-side of (OA.6) is positive this shows that Δ belongs to the cone $\mathcal{C} = \{d \in \mathbb{R}^p : \ \Omega_1(d) \leq 3\Omega_0(d)\}$. This yields

$$\begin{aligned}
\Omega(\Delta) &\leq 4\Omega_0(\Delta) \\
&\leq \mu\|\Delta_{\mathcal{S}_0}\|_1 + (1-\mu)\sum_{G\in\mathcal{G}_0}\|\Delta_G\|_2 \\
&\leq 4\left(\mu\sqrt{|S_0|}\|\Delta_{S_0}\|_2 + (1-\mu)\sqrt{|\mathcal{G}_0|}\sqrt{\sum_{G\in\mathcal{G}_0}\|\Delta_G\|_2^2}\right) \\
&\leq 4\sqrt{s_\mu}\|\Delta\|_2 \\
&\leq 4\sqrt{s_\mu\Delta^\top\Sigma\Delta/\nu},
\end{aligned} \tag{OA.7}$$

where we used inequality of Cauchy-Schwarz, Assumption 4 (ii), and the definition of $\sqrt{s_\mu}$. Notice also that

$$\begin{aligned}
\Delta^\top\Sigma\Delta &= \Delta^\top\hat{\Sigma}\Delta + \Delta^\top(\Sigma-\hat{\Sigma})\Delta \\
&\leq \frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 + \Omega(\Delta)\Omega^*\left((\hat{\Sigma}-\Sigma)\Delta\right) \\
&\leq \frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 + \Omega(\Delta)^2 G^*\left\|\hat{\Sigma}-\Sigma\right\|_\infty \\
&\leq \frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 + \Omega(\Delta)^2\frac{\nu}{32s_\mu} \quad \left(\text{Because } \left\|\hat{\Sigma}-\Sigma\right\|_\infty \leq \frac{\nu}{32s_\mu G^*}\right),
\end{aligned} \tag{OA.8}$$

where the second inequality is a consequence of the definition of the dual norm, and the third inequality follows from the fact that, by Lemma A.8, we have

$$\begin{aligned}
\Omega^*\left((\hat{\Sigma}-\Sigma)\Delta\right) &= \mu\left\|(\hat{\Sigma}-\Sigma)\Delta\right\|_\infty + (1-\mu)\max_{G\in\mathcal{G}}\left\|\left(\left(\hat{\Sigma}-\Sigma\right)\Delta\right)_G\right\|_2 \\
&\leq \mu\left\|(\hat{\Sigma}-\Sigma)\Delta\right\|_\infty + (1-\mu)\sqrt{G^*}\left\|\left(\hat{\Sigma}-\Sigma\right)\Delta\right\|_\infty \\
&\leq \mu\left\|\hat{\Sigma}-\Sigma\right\|_\infty\|\Delta\|_1 + (1-\mu)\sqrt{G^*}\left\|\hat{\Sigma}-\Sigma\right\|_\infty\|\Delta\|_1 \\
&\leq \mu\left\|\hat{\Sigma}-\Sigma\right\|_\infty\|\Delta\|_1 + (1-\mu)G^*\left\|\hat{\Sigma}-\Sigma\right\|_\infty\|\Delta\|_{2,1} \\
&\leq G^*\left\|\hat{\Sigma}-\Sigma\right\|_\infty\|\Delta\|_1,
\end{aligned}$$

where in the first and third inequalities, we used the inequality of Cauchy-Schwarz and, in the second inequality, we used Hölder's inequality. Combining (OA.7) and (OA.8), we obtain

$$\Omega(\Delta)^2 \leq \frac{16s_\mu}{\nu}\left(\frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 + \Omega(\Delta)^2\frac{\nu}{32s_\mu}\right). \tag{OA.9}$$

Using (OA.6), we get

$$\frac{1}{2T}\left\|\widetilde{W}\Delta\right\|_2^2 \leq \lambda\left\{\Omega_0(\Delta)-\Omega_1(\Delta)\right\}+\frac{\lambda}{2}\Omega\left(\Delta\right) \leq \frac{3}{2}\lambda\Omega(\Delta), \tag{OA.10}$$

so that (OA.9) implies

$$\Omega(\Delta)^2 \leq \frac{16 s_\mu}{\nu}\left(3\lambda\Omega(\Delta)+\Omega(\Delta)^2\frac{\nu}{32 s_\mu}\right).$$

This yields

$$\Omega(\Delta) \leq \frac{96 s_\mu \lambda}{\nu}. \tag{OA.11}$$

Moreover, by (OA.10), we obtain

$$\frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 \leq 3\lambda\Omega(\Delta) \leq \frac{288 s_\mu \lambda^2}{\nu}. \tag{OA.12}$$

Case 2: $\|A\|_2+\frac{1}{\sqrt{T}}\left\|M_{\hat{F}}F\gamma\right\|_2 > \frac{1}{2}\left\|\widetilde{W}\Delta\right\|_2$. This direcly yields

$$\frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 \leq \frac{4}{T}\left(\|A\|_2+\left\|M_{\hat{F}}F\gamma\right\|_2\right)^2. \tag{OA.13}$$

Conclusion of the proof of Theorem 3.1. Combining (OA.12) and (OA.13), we get

$$\frac{1}{T}\left\|\widetilde{W}\Delta\right\|_2^2 \leq \frac{288 s_\mu \lambda^2}{\nu}+\frac{4}{T}\left(\|A\|_2+\left\|M_{\hat{F}}F\gamma\right\|_2\right)^2. \tag{OA.14}$$

Next, remark that

$$\hat{F}\hat{\gamma} = P_{\hat{F}}\left(Y-W\hat{\delta}\right) = P_{\hat{F}}W\left(\delta-\hat{\delta}\right)+P_{\hat{F}}F\gamma+P_{\hat{F}}\mathcal{E}. \tag{OA.15}$$

Hence, we have

$$\begin{aligned}
W\hat{\delta}+\hat{F}\hat{\gamma}-W\delta-F\gamma &= M_{\hat{F}}\left(W\hat{\delta}+\hat{F}\hat{\gamma}-W\delta-F\gamma\right)+P_{\hat{F}}\left(W\hat{\delta}+\hat{F}\hat{\gamma}-W\delta-F\gamma\right)\\
&= \widetilde{W}\left(\hat{\delta}-\delta\right)-M_{\hat{F}}F\gamma+P_{\hat{F}}W\left(\hat{\delta}-\delta\right)+\hat{F}\hat{\gamma}-P_{\hat{F}}F\gamma\\
&= \widetilde{W}\left(\hat{\delta}-\delta\right)-M_{\hat{F}}F\gamma+P_{\hat{F}}\mathcal{E},
\end{aligned}$$

where we used (OA.15) in the last line. Therefore, we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|W\hat{\delta}+\hat{F}\hat{\gamma}-W\delta-F\gamma\right\|_2\\
&\leq \frac{1}{\sqrt{T}}\left(\left\|\widetilde{W}\left(\hat{\delta}-\delta\right)\right\|_2+\left\|M_{\hat{F}}F\gamma\right\|_2+\left\|P_{\hat{F}}\mathcal{E}\right\|_2\right)\\
&= \left(\frac{288 s_\mu \lambda^2}{\nu}+\frac{4}{T}\left(\|A\|_2+\left\|M_{\hat{F}}F\gamma\right\|_2\right)^2\right)^{1/2}+\frac{1}{\sqrt{T}}\left(\left\|M_{\hat{F}}F\gamma\right\|_2+\left\|P_{\hat{F}}\mathcal{E}\right\|_2\right).
\end{aligned}$$

A.2 Proof of Corollary 1

By Lemmas A.5 (ii) and A.8, we have

$$\lambda = 2\Omega^* \left(\frac{1}{T} \widetilde{W}^\top \mathcal{E} \right) = O_P \left(h_T + \frac{1}{p_x} \right).$$

Plugging in this and the results of Lemma A.5 (ii), (iii) and (iv) in the finite sample bound of Theorem 3.1, we obtain

$$\begin{aligned} \frac{1}{T} \left\| W\hat{\delta} + \hat{F}\hat{\gamma} - W\delta - F\gamma \right\|_2^2 &= O_P \left(s_\mu \left(h_T + \frac{1}{p_x} \right)^2 + \|A\|_2^2 + \frac{1}{T} + \frac{1}{p_x} \right) \\ &= O_P \left(s_\mu h_T^2 + s_\mu \frac{1}{p_x^2} + \|A\|_2^2 + \frac{1}{T} + \frac{1}{p_x} \right) \\ &= O_P \left(s_\mu h_T^2 + \|A\|_2^2 + \frac{1}{p_x} \right), \end{aligned}$$

where, in the second equality, we used Assumption 4 to simplify the rate.

A.3 Proof of Lemma 3.1

Proof of (i). By Lemma A.3, we have $\left\| \hat{F} - FH^\top \right\|_2 \leq J_1 + J_2 + J_3$, where

$$\begin{aligned} J_1 &= \frac{1}{T} \left\| FB^\top U^\top \hat{F} V^{-1} \right\|_2; \\ J_2 &= \frac{1}{T} \left\| UBF^\top \hat{F} V^{-1} \right\|_2; \\ J_3 &= \frac{1}{T} \left\| UU^\top \hat{F} V^{-1} \right\|_2. \end{aligned}$$

First, we bound J_1. It holds that

$$\begin{aligned} J_1 &\leq \frac{1}{T} \|F\|_2 \|UB\|_2 \left\| \hat{F} \right\|_2 \|V^{-1}\|_2 \\ &= O_P \left(\frac{1}{T} \sqrt{T} \sqrt{T p_x} \sqrt{T} \frac{1}{p_x} \right) = O_P \left(\sqrt{\frac{T}{p_x}} \right), \end{aligned}$$

where we used Lemmas A.1 (iii), (vi) and A.2 (iii) and the fact that $\left\| \hat{F} \right\|_2 = \sqrt{RT}$. Similarly, we have

$$\begin{aligned} J_2 &\leq \frac{1}{T} \|UB\|_2 \|F\|_2 \left\| \hat{F} \right\|_2 \|V^{-1}\|_2 \\ &= O_P \left(\frac{1}{T} \sqrt{T p_x} \sqrt{T} \sqrt{T} \frac{1}{p_x} \right) = O_P \left(\sqrt{\frac{T}{p_x}} \right). \end{aligned}$$

Finally, it also holds that

$$J_3 \leq \frac{1}{T} \left\|UU^\top\right\|_2 \left\|\hat{F}\right\|_2 \|V^{-1}\|_2$$
$$= O_P\left(\frac{1}{T}\left(T\sqrt{p_x} + p_x\sqrt{T}\right)\sqrt{T}\frac{1}{p_x}\right) = O_P\left(\sqrt{\frac{T}{p_x}} + 1\right),$$

where we used A.1 (xvi) and A.2 (iii) and the fact that $\left\|\hat{F}\right\|_2 = \sqrt{RT}$. Combining the bounds on J_1, J_2 and J_3, we obtain the result.

Proof of (ii). By Lemma A.3, we have $\left\|(\hat{F} - FH^\top)^\top \mathcal{E}\right\|_2 \leq J_1 + J_2 + J_3$, where

$$J_1 = \frac{1}{T}\left\|\mathcal{E}^\top F B^\top U^\top \hat{F} V^{-1}\right\|_2;$$
$$J_2 = \frac{1}{T}\left\|\mathcal{E}^\top U B F^\top \hat{F} V^{-1}\right\|_2;$$
$$J_3 = \frac{1}{T}\left\|\mathcal{E}^\top U U^\top \hat{F} V^{-1}\right\|_2.$$

We have

$$J_1 \leq \frac{1}{T}\left\|\mathcal{E}^\top F\right\|_2 \left(\|UB\|_2 \left\|\hat{F} - FH^\top\right\|_2 + \|H\|_2 \left\|B^\top U^\top F\right\|_2\right) \|V^{-1}\|_2$$
$$= O_P\left(\frac{1}{T}\sqrt{T}\left(\sqrt{Tp_x}\sqrt{\frac{T}{p_x} + 1} + \sqrt{Tp_x}\right)\frac{1}{p_x}\right) = O_P\left(\frac{\sqrt{T}}{p_x} + \frac{1}{\sqrt{p_x}}\right),$$

by Lemmas A.1 (v), (vi), (ix), A.4 (i) and A.2 (iii), and statement (i). Moreover, it holds that

$$J_2 \leq \frac{1}{T}\|\mathcal{E}^\top UB\|_2 \|F\|_2 \left\|\hat{F}\right\|_2 \|V^{-1}\|_2$$
$$= O_P\left(\frac{1}{T}\sqrt{Tp_x}\sqrt{T}\sqrt{T}\frac{1}{p_x}\right) = O_P\left(\sqrt{\frac{T}{p_x}}\right),$$

by Lemmas A.1 (x), (iii) and A.2 (iii) and the fact that $\left\|\hat{F}\right\|_2 = \sqrt{RT}$. We also have

$$J_3 \leq \frac{1}{T}\left\|\mathcal{E}^\top U\right\|_2 \left(\|U\|_2 \left\|\hat{F} - FH^\top\right\|_2 + \left\|U^\top F\right\|_2\right) \|V^{-1}\|_2$$
$$= O_P\left(\frac{1}{T}\sqrt{Tp_x}\left(\sqrt{Tp_x}\sqrt{\frac{T}{p_x} + 1} + \sqrt{Tp_x}\right)\frac{1}{p_x}\right)$$
$$= O_P\left(\sqrt{\frac{T}{p_x}} + 1\right),$$

by Lemmas A.1 (viii), (ii), (vii), A.4 (i) and A.2 (iii), and statement (i). Combining the bounds on J_1, J_2 and J_3, we obtain (ii).

Proof of (iii). By Lemma A.3, we have $\left\|(\hat{F} - FH^\top)^\top W\right\|_\infty \leq J_1 + J_2 + J_3$, where

$$J_1 = \frac{1}{T}\left\|W^\top F B^\top U^\top \hat{F} V^{-1}\right\|_\infty;$$
$$J_2 = \frac{1}{T}\left\|W^\top U B F^\top \hat{F} V^{-1}\right\|_\infty;$$
$$J_3 = \frac{1}{T}\left\|W^\top U U^\top \hat{F} V^{-1}\right\|_\infty.$$

We have

$$\begin{aligned}
J_1 &= \frac{1}{T}\max_{k\in[p],r\in[R]}\left|\sum_{j\in[R]}\left(W^\top F\right)_{k,j}\left(B^\top U^\top \hat{F} V^{-1}\right)_{j,r}\right| \\
&\leq \frac{R}{T}\left\|W^\top F\right\|_\infty \left\|B^\top U^\top \hat{F} V^{-1}\right\|_\infty \\
&\leq \frac{R}{T}\left\|W^\top F\right\|_\infty \left(\|UB\|_2 \left\|\hat{F} - FH^\top\right\|_2 + \|H\|_2 \left\|B^\top U^\top F\right\|_2\right)\|V^{-1}\|_2 \\
&= O_P\left(\frac{1}{T}(Th_T + T)\left(\sqrt{Tp_x}\sqrt{\frac{T}{p_x}+1} + \sqrt{Tp_x}\right)\frac{1}{p_x}\right) \\
&= O_P\left(\frac{T}{p_x} + \sqrt{\frac{T}{p_x}}\right),
\end{aligned}$$

by Lemmas A.1 (xi), (vi), (ix), A.4 (i) and A.2 (iii), and statement (i). Moreover, it holds that

$$\begin{aligned}
J_2 &= \frac{1}{T}\max_{k\in[p],r\in[R]}\left|\sum_{j\in[R]}\left(W^\top U B\right)_{k,j}\left(F^\top \hat{F} V^{-1}\right)_{j,r}\right| \\
&\leq \frac{R}{T}\left\|W^\top U B\right\|_\infty \left\|F^\top \hat{F} V^{-1}\right\|_\infty \\
&\leq \frac{R}{T}\left\|W^\top U B\right\|_\infty \|F\|_2 \left\|\hat{F}\right\|_2 \|V^{-1}\|_2 \\
&= O_P\left(\frac{1}{T}\left(T\sqrt{p_x}h_T + T\right)\sqrt{T}\sqrt{T}\frac{1}{p_x}\right) \\
&= O_P\left(\frac{T}{p_x}\left(\sqrt{p_x}h_T + 1\right)\right),
\end{aligned}$$

by Lemmas A.1 (xiv), (iii) and A.2 (iii), and the fact that $\left\|\hat{F}\right\|_2 = \sqrt{RT}$. Finally, by the

inequality of Cauchy-Schwarz, we also have

$$\begin{aligned}
J_3 &= \frac{1}{T}\max_{k\in[p],r\in[R]}\left|\sum_{\ell\in[p_x]}\left(W^\top U\right)_{k,\ell}\left(U^\top\hat{F}V^{-1}\right)_{\ell,r}\right| \\
&\leq \frac{1}{T}\max_{k\in[p]}\left\|\sum_{t\in[T]}w_{t,k}u_t\right\|_2\left\|U^\top\hat{F}V^{-1}\right\|_2 \\
&\leq \frac{1}{T}\max_{k\in[p]}\left\|\sum_{t\in[T]}w_{t,k}u_t\right\|_2\left(\|U\|_2\left\|\hat{F}-FH^\top\right\|_2+\left\|U^\top F\right\|_2\right)\|V^{-1}\|_2 \\
&= O_P\left(\frac{1}{T}\left(T\sqrt{p_x}h_T+T\right)\left(\sqrt{Tp_x}\sqrt{\frac{T}{p_x}+1}+\sqrt{Tp_x}\right)\frac{1}{p_x}\right) \\
&= O_P\left(\left(\frac{T}{p_x}+\sqrt{\frac{T}{p_x}}\right)\left(\sqrt{p_x}h_T+1\right)\right),
\end{aligned}$$

by Lemmas A.1 (xv), (ii), (vii) and A.2 (iii), and statement (i). Combining the bounds on J_1, J_2 and J_3, we obtain (iii).

A.4 Auxiliary lemmas

Lemma A.1 *Under the assumptions of Lemma 3.1, the following holds:*

(*i*) $\left\|\frac{1}{T}F^\top F - I_R\right\|_2 = O_P\left(\frac{1}{\sqrt{T}}\right)$;

(*ii*) $\|U\|_2 = O_P\left(\sqrt{Tp_x}\right)$;

(*iii*) $\|F\|_2 = O_P(\sqrt{T})$;

(*iv*) $\|\mathcal{E}\|_2 = O_P\left(\sqrt{T}\right)$;

(*v*) $\left\|F^\top\mathcal{E}\right\|_2 = O_P\left(\sqrt{T}\right)$;

(*vi*) $\|UB\|_2 = O_P\left(\sqrt{Tp_x}\right)$;

(*vii*) $\left\|F^\top U\right\|_2 = O_P\left(\sqrt{Tp_x}\right)$;

(*viii*) $\left\|\mathcal{E}^\top U\right\|_2 = O_P\left(\sqrt{Tp_x}\right)$;

(*ix*) $\left\|F^\top UB\right\|_2 = O_P\left(\sqrt{Tp_x}\right)$;

(*x*) $\left\|\mathcal{E}^\top UB\right\|_2 = O_P\left(\sqrt{Tp_x}\right)$;

(xi) $\left\|W^\top F\right\|_\infty = O_P\left(Th_T + T\right)$;

(xii) $\left\|\frac{1}{T}W^\top W - E[w_t w_t^\top]\right\|_\infty = O_P\left(\left(\frac{p^2}{T^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(2p)}{T}}\right)$;

(xiii) $\left\|\frac{1}{T}W^\top F\left(\frac{1}{T}F^\top W\right) - E[w_t f_t^\top]E[w_t f_t^\top]^\top\right\|_\infty = O_P\left(h_T\right)$;

(xiv) $\left\|W^\top U B\right\|_\infty = O_P\left(T\sqrt{p_x}h_T + T\right)$;

(xv) $\max_{k\in[p]}\left\|\sum_{t\in[T]} w_{t,k}u_t\right\|_2 = O_P\left(T\sqrt{p_x}h_T + T\right)$;

(xvi) $\left\|UU^\top\right\|_2 = O_P\left(T\sqrt{p_x} + \sqrt{T}p_x\right)$;

(xvii) $\left\|W^\top \mathcal{E}\right\|_\infty = O_P\left(h_T\right)$.

Proof. In this proof, we will repeatedly apply Lemma A.6. Its conditions hold by Assumptions 2 and 3.

Proof of (i). This follows directly from Lemma A.6 applied to $\zeta_t = f_t f_t^\top - E[f_t f_t^\top] = f_t f_t^\top - I_R$.

Proof of (ii). We use that $E[\|U\|_2^2] = E[\sum_{t\in[T]}\sum_{k=1}^p u_{t,j}^2] = O(Tp)$ by Assumption 2 (ii) and Markov's inequality.

Proof of (iii). We use that $E[\|F\|_2^2] = E[\sum_{t\in[T]}\sum_{r=1}^R f_{t,r}^2] = O(T)$ by Assumption 2 (ii) and Markov's inequality.

Proof of (iv). We use that $E[\|\mathcal{E}\|_2^2] = E[\sum_{t\in[T]}\varepsilon_t^2] = O(T)$ by Assumption 2 (ii) and Markov's inequality.

Proof of (v). Applying Lemma A.6 to $\zeta_t = f_{t,r}\varepsilon_t$ gives $\sum_{t\in[T]} f_{t,r}\varepsilon_t = O_P(1/\sqrt{T})$, which yields $\left\|F^\top\mathcal{E}\right\|_2^2 = \sum_{r=1}^R\left(\sum_{t\in[T]} f_{t,r}\varepsilon_t\right)^2 = O_P(T)$.

Proof of (vi). We have

$$E[\|UB\|_2^2] = p_x\sum_{t\in[T]}\sum_{r=1}^R E\left[\left(p_x^{-1/2}\sum_{k=1}^{p_x} u_{t,k}b_{k,r}\right)^2\right] = O(Tp_x),$$

by Assumption 2 (ii). We obtain the result by Markov's inequality.

Proof of (vii). We have

$$E\left[\left\|F^\top U\right\|_2^2\right] = T\sum_{k=1}^{p_x}\sum_{r=1}^{R} E\left[\left(T^{-1/2}\sum_{t=1}^{T} u_{t,k}f_{t,r}\right)^2\right] = O(Tp_x).$$

We obtain the result by Markov's inequality.

Proof of (viii). We have

$$E\left[\left\|\mathcal{E}^\top U\right\|_2^2\right] = T\sum_{k=1}^{p_x} E\left[\left(T^{-1/2}\sum_{t=1}^{T} u_{t,k}\varepsilon_t\right)^2\right] = O(Tp_x).$$

We obtain the result by Markov's inequality.

Proof of (ix). This follows directly from Lemma A.6 applied to $\zeta_t = f_t\left(p_x^{-1/2}\sum_{k=1}^{p_x} u_{t,k}b_{k,r}\right)$.

Proof of (x). This follows directly from Lemma A.6 applied to $\zeta_t = \varepsilon_t\left(p_x^{-1/2}\sum_{k=1}^{p_x} u_{t,k}b_{k,r}\right)$.

Proof of (xi). By Lemma A.6 applied to $\zeta_t = w_t f_t^\top - E[w_{t,k}f_{t,r}]$ and the triangle inequality, we have

$$\begin{aligned}
\frac{1}{T}\left\|W^\top F\right\|_\infty &= \max_{k\in[p],r\in[R]}\left|\frac{1}{T}\sum_{t\in[T]} w_{t,k}f_{t,r}\right| \\
&\leq \max_{k\in[p],r\in[R]}\left|\frac{1}{T}\sum_{t\in[T]} w_{t,k}f_{t,r} - E[w_{t,k}f_{t,r}]\right| + \max_{k\in[p],r\in[R]}\left|E[w_{t,k}f_{t,r}]\right| \\
&= O_P\left(h_T\right) + O(1).
\end{aligned}$$

Proof of (xii). This follows directly from Lemma A.6 applied to $\zeta_t = w_t w_t^\top - E[w_t w_t^\top]$.

Proof of (xiii). First, note that

$$\left\|\frac{1}{T}W^\top F\left(\frac{1}{T}F^\top W\right) - E[w_t f_t^\top]E[w_t f_t^\top]^\top\right\|_\infty \leq J_1 + J_2 + J_3, \tag{OA.16}$$

where

$$J_1 = \left\| \left(\frac{1}{T} W^\top F - E[w_t f_t^\top] \right) E[w_t f_t^\top]^\top \right\|_\infty ;$$
$$J_2 = \left\| E[w_t f_t^\top] \left(\frac{1}{T} W^\top F - E[w_t f_t^\top] \right)^\top \right\|_\infty ;$$
$$J_3 = \left\| \left(\frac{1}{T} W^\top F - E[w_t f_t^\top] \right) \left(\frac{1}{T} W^\top F - E[w_t f_t^\top] \right)^\top \right\|_\infty .$$

Moreover, by Lemma A.6 applied to $\zeta_t = w_t f_t^\top - E[w_t f_t^\top]$, we have

$$\left\| \frac{1}{T} W^\top F - E[w_t f_t^\top] \right\|_\infty = O_P\left(h_T\right). \tag{OA.17}$$

This yields

$$\begin{aligned} J_1 &= \max_{k,\ell \in [p_x]} \left| \sum_{r,j \in [R]} \left(\frac{1}{T} W^\top F - E[w_t f_t^\top] \right)_{k,r} \left(E[w_t f_t^\top]^\top \right)_{j,\ell} \right| \\ &\leq R \left\| \frac{1}{T} W^\top F - E[w_t f_t^\top] \right\|_\infty \left\| E[w_t f_t^\top] \right\|_\infty = O_P(h_T). \end{aligned}$$

Similarly, we have $J_2 = O_P(h_T)$. Finally, by (OA.17), it also holds that

$$\begin{aligned} J_3 &= \max_{k,\ell \in [p_x]} \left| \sum_{r,j \in [R]} \left(\frac{1}{T} W^\top F - E[w_t f_t^\top] \right)_{k,r} \left(\left(\frac{1}{T} W^\top F - E[w_t f_t^\top] \right)^\top \right)_{j,\ell} \right| \\ &\leq R \left\| \frac{1}{T} W^\top F - E[w_t f_t^\top] \right\|_\infty^2 = O_P(h_T^2) = o_P(h_T). \end{aligned}$$

Combining (OA.16) and the bounds on J_1, J_2 and J_3.

Proof of (xiv). By Lemma A.6 applied to

$$\zeta_t = w_{t,k} \left(p_x^{-1/2} \sum_{k \in [p_x]} u_{t,k} b_k^\top \right) - E\left[w_{t,k} \left(p_x^{-1/2} \sum_{k \in [p_x]} u_{t,k} b_k^\top \right) \right],$$

we have

$$\max_{k \in [p], r \in [R]} \left| \frac{1}{T} \sum_{t \in [T]} w_{t,k} \left(p_x^{-1/2} \sum_{\ell \in [p_x]} u_{t,\ell} b_{\ell,r} \right) - E\left[w_{t,k} \left(p_x^{-1/2} \sum_{\ell \in [p_x]} u_{t,\ell} b_{\ell,r} \right) \right] \right| = O_P\left(h_T\right).$$

This implies that

$$\left\| W^\top U B \right\|_\infty = T \max_{k \in [p], r \in [R]} \left| \frac{1}{T} \sum_{t \in [T]} w_{t,k} \left(p_x^{-1/2} \sum_{\ell \in [p_x]} u_{t,\ell} b_{\ell,r} \right) \right|$$

$$
\begin{aligned}
&\leq T\sqrt{p_x}\max_{k\in[p]}\left\|\frac{1}{T}\sum_{t\in[T]}w_{t,k}\left(p_x^{-1/2}\sum_{\ell\in[p_x]}u_{t,\ell}b_{\ell,r}\right)-E\left[w_{t,k}\left(p_x^{-1/2}\sum_{\ell\in[p_x]}u_{t,\ell}b_{\ell,r}\right)\right]\right\|_2\\
&+T\max_{k\in[p]}\left\|E\left[w_{t,k}\left(\sum_{k\in[p_x]}u_{t,k}b_k\right)\right]\right\|_2\\
&=O_P\left(T\sqrt{p_x}h_T\right)+O(T).
\end{aligned}
$$

Proof of (xv). By Lemma A.6 applied to

$$
\zeta_t=w_tu_t^\top-E\left[w_tu_t^\top\right],
$$

we have

$$
\left\|\frac{1}{T}\sum_{t\in[T]}w_tu_t^\top-E\left[w_tu_t^\top\right]\right\|_\infty=O_P\left(h_T\right).
$$

We have

$$
\begin{aligned}
&\max_{k\in[p]}\left\|\frac{1}{T}\sum_{t\in[T]}w_{t,k}u_t\right\|_2\\
&\leq\max_{k\in[p]}\left\|\frac{1}{T}\sum_{t\in[T]}w_{t,k}u_t-E\left[w_{t,k}u_t\right]\right\|_2+\max_{k\in[p]}\left\|E\left[w_{t,k}u_t\right]\right\|_2\\
&\leq\sqrt{p_x}\left\|\frac{1}{T}\sum_{t\in[T]}w_tu_t^\top-E\left[w_tu_t^\top\right]\right\|_\infty+\max_{k\in[p]}\left\|E\left[w_{t,k}u_t\right]\right\|_2\\
&=O_P\left(\sqrt{p_x}h_T\right)+O(1).
\end{aligned}
$$

Proof of (xvi). It holds that

$$
\left\|UU^\top\right\|_2\leq J_1+J_2,
$$

where

$$
\begin{aligned}
J_1&=\left\|UU^\top-E\left[UU^\top\right]\right\|_2;\\
J_2&=\left\|E\left[UU^\top\right]\right\|_2.
\end{aligned}
$$

First, we bound J_1. Since $\max_{s,t\in[T]}E\left[\left(\frac{u_s^\top u_t}{p_x}-\frac{E[u_s^\top u_t]}{p_x}\right)^2\right]=O(1)$, it holds that

$$
E\left[\frac{1}{T^2}\sum_{s,t\in[T]}\left(p_x^{-1/2}\left[u_s^\top u_t-E[u_s^\top u_t]\right]\right)^2\right]=O(1).
$$

By Markov's inequality, this yields

$$\left\| \frac{1}{T\sqrt{p_x}} \left(UU^\top - E\left[UU^\top\right] \right) \right\|_2^2 = \frac{1}{T^2} \sum_{s,t\in[T]} \left(p_x^{-1/2} \left[u_s^\top u_t - E[u_s^\top u_t] \right] \right)^2 = O_P(1).$$

This implies $J_1 = O_P\left(T\sqrt{p_x}\right)$. Now, we bound J_2. We have

$$\begin{aligned} \left\| E\left[\frac{1}{\sqrt{T}p_x} UU^\top \right] \right\|_2^2 &= \sum_{s,t\in[T]} \frac{1}{Tp_x^2} E[u_s^\top u_t]^2 \\ &\leq \max_{s,t\in[T]} \left| \frac{E[u_s^\top u_t]}{p_x} \right| \max_{s\in[T]} \left(\sum_{t\in[T]} \frac{1}{p_x} \left| \frac{E[u_s^\top u_t]}{p_x} \right| \right). \end{aligned} \tag{OA.18}$$

We have

$$\max_{s,t\in[T]} \frac{1}{p_x} |E[u_s^\top u_t]| \leq \max_{s,t\in[T]} \max_{k\in[p_x]} E[u_{s,k} u_{t,j}] \leq \max_{t\in[T]} \max_{k\in[p_x]} E[u_{t,j}^2] = O(1). \tag{OA.19}$$

Moreover, by Lemma A.7, for all $t \in [T], k \in [p]$, we have

$$\sum_{s\in[T]} \frac{1}{p_x} |E[u_{s,k} u_{t,k}]| \leq \sum_{s\in\mathbb{N}} \max_{k\in[p_x]} |E[u_{0,k} u_{t,k}]| \leq \sum_{s\in\mathbb{N}} \tau_s^{\frac{q-2}{q-1}} \|u_{s,k}\|_q^{\frac{q}{q-1}} = O(1),$$

where the sum $\sum_{s\in\mathbb{N}} \tau_s^{\frac{q-2}{q-1}} \leq \sum_{s\in\mathbb{N}} c s^{-a\frac{q-2}{q-1}}$ converges since $a > (q-1)/(q-2)$ by Assumption 3. This, (OA.18) and (OA.19) imply that $J_2 = O(\sqrt{T}p_x)$. Combining the bounds on J_1 and J_2, we obtain the result.

Proof of (xvii). This follows directly from Lemma A.6 applied to $\zeta_t = \varepsilon_t w_t$. □

Lemma A.2 *Under the assumptions of Theorem 3.1, the following holds*

(*i*) $\|U\|_{op} = O_P\left(\sqrt{T}p_x^{1/4} + \sqrt{p_x}T^{1/4}\right)$;

(*ii*) *There exists* $c > 0$ *independent of* T *such that*

$$\mathbb{P}\left(\sigma_R\left(\frac{1}{Tp_x} BF^\top FB^\top \right) \geq c \right) \to 1;$$

(*iii*) *With probability going to* 1, V *is invertible. Moreover, we have* $\|V^{-1}\|_2 = O_P\left(p_x^{-1}\right)$.

Proof. *Proof of* (i). We have $\|U\|_{op} \leq \sqrt{\|UU^\top\|_2} = O_P\left(\sqrt{T}p_x^{\frac{1}{4}} + \sqrt{p_x}T^{\frac{1}{4}}\right)$ by Lemma A.1 (xvi).

Proof of (ii). By Weyl's inequality, we have

$$\begin{aligned}
&\left|\sigma_R\left(\frac{1}{Tp_x}BF^\top FB^\top\right) - \sigma_R\left(\frac{1}{p_x}BB^\top\right)\right| \\
&\leq \left\|\frac{1}{p_x}B\left(\frac{1}{T}F^\top F - I_R\right)B^\top\right\|_{op} \\
&\leq \frac{1}{p_x}\|B\|_{op}^2\left\|\frac{1}{T}F^\top F - I_R\right\|_2, \\
&\leq \frac{1}{p_x}\|B^\top B\|_2^2\left\|\frac{1}{T}F^\top F - I_R\right\|_2 = O_P\left(\frac{1}{\sqrt{T}}\right),
\end{aligned} \tag{OA.20}$$

where we used Lemma A.1 (i) and Assumption 1 (iii). Moreover, let $2c$ be the lower bound mentioned in Assumption 1 (ii), then we have $\sigma_R\left(\frac{1}{p_x}BB^\top\right) \geq 2c$, so that

$$\mathbb{P}\left(\sigma_R\left(\frac{1}{Tp_x}BF^\top FB^\top\right) \geq c\right) \to 1,$$

by (OA.20).

Proof of (iii). First, by Weyl's inequality, we have

$$\begin{aligned}
&\left|\sigma_R\left(\frac{1}{T}X^\top X\right) - \sigma_R\left(\frac{1}{T}BF^\top FB^\top\right)\right| \\
&\leq \frac{1}{T}\left\|X^\top X - BF^\top FB^\top\right\|_{op} \\
&\leq \frac{1}{T}\left\|(FB^\top + U)^\top(FB^\top + U) - BF^\top FB^\top\right\|_{op} \\
&\leq \frac{1}{T}\left\|FB^\top U^\top + UBF^\top + U^\top U\right\|_{op} \\
&\leq \frac{2}{T}\left\|FB^\top\right\|_{op}\|U\|_{op} + \frac{1}{T}\left\|U^\top U\right\|_{op} \\
&\leq \frac{2}{T}\|F\|_2\|B\|_2\|U\|_{op} + \frac{1}{T}\|U\|_{op}^2 \\
&= O_P\left(\frac{1}{T}\sqrt{T}\sqrt{p_x}\left(\sqrt{T}p_x^{1/4} + \sqrt{p_x}T^{1/4}\right) + \frac{1}{T}\left(T\sqrt{p_x} + \sqrt{T}p_x\right)\right) \\
&= o_P\left(p_x\right),
\end{aligned}$$

where we used (i), Lemma A.1 (iii) and Assumption 1 (iii). This shows that $\mathbb{P}\left(\sigma_R\left(\frac{1}{T}XX^\top\right) \geq cp_x/2\right) \to 1$, which also yields $\|V^{-1}\|_2 \leq \sqrt{R}\left(\sigma_R\left(\frac{1}{T}XX^\top\right)\right)^{-1} = O_P\left(p_x^{-1}\right)$. □

Lemma A.3 *Under the assumptions of Lemma 3.1, it holds that*

$$\hat{F} - FH^\top = \frac{1}{T}FB^\top U^\top \hat{F}V^{-1} + \frac{1}{T}UBF^\top \hat{F}V^{-1} + \frac{1}{T}UU^\top \hat{F}V^{-1}.$$

Proof. Recall that $H = \frac{1}{T}V^{-1}\hat{F}^\top FB^\top B$ and $\hat{F}V = \frac{1}{T}XX^\top \hat{F}$. As a result, we have

$$\begin{aligned}\hat{F}V &= T^{-1}XX^\top \hat{F} \\ &= \frac{1}{T}(FB^\top + U)(FB^\top + U)^\top \hat{F} \\ &= \frac{1}{T}FB^\top BF^\top \hat{F} + \frac{1}{T}FB^\top U^\top \hat{F} + \frac{1}{T}UBF^\top \hat{F} + T^{-1}UU^\top \hat{F}.\end{aligned}$$

Multiplying both sides by V^{-1}, we get the result. □

Lemma A.4 *Under the assumptions of Lemma 3.1, the following holds:*

(*i*) $\|H\|_2 = O_P(1)$;

(*ii*) $\|H^\top H - I_R\|_2 = O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_x}}\right)$.

Proof. *Proof of* (i). Recall that $H = \frac{1}{T}V^{-1}\hat{F}^\top FB^\top B$. This yields

$$\|H\|_2 \leq \frac{1}{T}\|V^{-1}\|_2 \left\|\hat{F}\right\|_2 \|F\|_2 \|B^\top B\|_2 = O_P(1),$$

by Lemmas A.2 (iii) and A.1 (iii), $\|\hat{F}\|_2 = \sqrt{T}$ and the fact that $\|B^\top B\|_2 = O(p_x)$ by Assumption 1 (ii).

Proof of (ii). We have

$$\begin{aligned}\left\|H^\top H - I_R\right\|_2 &\leq \left\|H^\top H - H^\top \frac{F^\top F}{T}H\right\|_2 + \left\|H^\top \frac{F^\top F}{T}H - I_R\right\|_2 \\ &\leq \left\|H^\top \left(I_R - \frac{F^\top F}{T}\right)H\right\|_2 + \left\|H^\top \frac{F^\top F}{T}H - I_R\right\|_2 \\ &\leq \|H\|_2^2 \left\|\frac{F^\top F}{T} - I_R\right\|_2 + \left\|H^\top \frac{F^\top F}{T}H - I_R\right\|_2 \\ &= O_P\left(\frac{1}{\sqrt{T}}\right) + \left\|H^\top \frac{F^\top F}{T}H - I_R\right\|_2,\end{aligned} \tag{OA.21}$$

by Lemma A.1 (i). Note that

$$H^\top \frac{F^\top F}{T}H - I_R = H^\top \frac{F^\top F}{T}H - \frac{\hat{F}^\top \hat{F}}{T}$$

$$= \frac{1}{T}\left(HF^\top - \hat{F}^\top\right)FH + \frac{1}{T}\hat{F}^\top\left(FH^\top - \hat{F}\right).$$

Hence, we have

$$\begin{aligned}\left\|H^\top\frac{F^\top F}{T}H - I_R\right\|_2 &\leq \frac{1}{T}\left\|\hat{F} - FH^\top\right\|_2\|F\|_2\|H\|_2 + \left\|\hat{F} - FH^\top\right\|_2\|\hat{F}\|_2 \\ &= O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_x}}\right),\end{aligned}$$

where we used Lemmas 3.1 (i) and A.1 (iii) and $\|\hat{F}\|_2 = \sqrt{RT}$. (Note that (ii) is not used to prove Lemma 3.1 (i) so that we can indeed use the latter.) This and (OA.21) show (ii). □

Lemma A.5 *Under the assumptions of Theorem 3.1, the following holds:*

(*i*) $\left\|\hat{\Sigma} - \Sigma\right\|_\infty = O_P\left(\left(\left(\frac{p^2}{T^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(2p)}{T}}\right) + \frac{1}{\sqrt{p_x}}\right)$;

(*ii*) $\frac{1}{T}\left\|\widetilde{W}^\top\mathcal{E}\right\|_\infty = O_P\left(h_T + \frac{1}{p_x}\right)$;

(*iii*) $\left\|M_{\hat{F}}F\gamma\right\|_2 = O_P\left(\sqrt{\frac{T}{p_x}+1}\right)$;

(*iv*) $\left\|P_{\hat{F}}\mathcal{E}\right\|_2 = O_P\left(\sqrt{\frac{T}{p_x}+1}\right)$.

Proof. *Proof of* (i). We have

$$\begin{aligned}\left\|\hat{\Sigma} - \Sigma\right\|_\infty &\leq \left\|\frac{1}{T}\widetilde{W}^\top\widetilde{W} - \Sigma\right\|_\infty \\ &\leq \left\|\frac{1}{T}W^\top\left(I_T - \frac{1}{T}\hat{F}\hat{F}^\top\right)W - \Sigma\right\|_\infty \\ &\leq J_1 + J_2 + J_3 + J_4 + J_5,\end{aligned}$$

where

$$\begin{aligned}J_1 &= \left\|\frac{1}{T}W^\top\frac{1}{T}(\hat{F} - FH^\top)(\hat{F} - FH^\top)^\top W\right\|_\infty; \\ J_2 &= \left\|\frac{1}{T}W^\top\frac{1}{T}(\hat{F} - FH^\top)HF^\top W\right\|_\infty; \\ J_3 &= \left\|\frac{1}{T}W^\top\frac{1}{T}FH^\top(\hat{F} - FH^\top)^\top W\right\|_\infty; \\ J_4 &= \left\|\frac{1}{T}W^\top\frac{1}{T}F\left(I_R - H^\top H\right)F^\top W\right\|_\infty;\end{aligned}$$

$$J_5 = \left\| \frac{1}{T} W^\top \left(I_T - \frac{1}{T} F F^\top \right) W - \Sigma \right\|_\infty .$$

Let us first bound J_1. It holds that

$$\begin{aligned}
J_1 &= \frac{1}{T^2} \max_{k,\ell \in [p_x]} \left| \sum_{r \in [R]} \left(W^\top \left(\widehat{F} - F H^\top \right) \right)_{k,r} \left(\left(\widehat{F} - F H^\top \right)^\top W \right)_{r,\ell} \right| \\
&\leq \frac{R}{T^2} \left\| \left(\widehat{F} - F H^\top \right)^\top W \right\|_\infty^2 \\
&= O_P \left(\frac{1}{T^2} \left(\frac{T}{p_x} + \sqrt{\frac{T}{p_x}} \right)^2 \left(\sqrt{p_x} h_T + 1 \right)^2 \right) \\
&= o_P \left(\frac{1}{p_x} + \frac{1}{T} \right),
\end{aligned}$$

by Lemma 3.1 (iii). Next, we consider J_2. We have

$$\begin{aligned}
J_2 &= \frac{1}{T^2} \max_{k,\ell \in [p_x]} \left| \sum_{r \in [R]} \left(W^\top \left(\widehat{F} - F H^\top \right) \right)_{k,r} \left(H F^\top W \right)_{r,\ell} \right| \\
&\leq \frac{R}{T^2} \left\| \left(\widehat{F} - F H^\top \right)^\top W \right\|_\infty \left\| H F^\top W \right\|_\infty \\
&\leq \frac{R}{T^2} \left\| \left(\widehat{F} - F H^\top \right)^\top W \right\|_\infty \max_{r \in [R], k \in [p_x]} \left| \sum_{j \in [R]} H_{r,j} \left(F^\top W \right)_{j,k} \right| \\
&\leq \frac{R^2}{T^2} \left\| \left(\widehat{F} - F H^\top \right)^\top W \right\|_\infty \|H\|_2 \left\| F^\top W \right\|_\infty \\
&= O_P \left(\frac{1}{T^2} \left(\frac{T}{p_x} + \sqrt{\frac{T}{p_x}} \right) \left(\sqrt{p_x} h_T + 1 \right) \left(T h_T + T \right) \right) \\
&= O_P \left(\frac{1}{\sqrt{p_x}} + \frac{1}{\sqrt{T}} \right),
\end{aligned}$$

by Lemmas 3.1 (iii), A.4 (i) and A.1 (xi). By similar reasoning, it holds that

$$J_3 = o_P \left(\frac{1}{\sqrt{p_x}} + \frac{1}{\sqrt{T}} \right).$$

Then, we focus on J_4. It holds that

$$\begin{aligned}
J_4 &= \frac{1}{T^2} \max_{k,\ell \in [p_x]} \left| \sum_{r \in [R]} \left(W^\top F \right)_{k,r} \left(\left(I_R - H^\top H \right) F^\top W \right)_{r,\ell} \right| \\
&\leq \frac{R}{T^2} \| W^\top F \|_\infty \left\| \left(I_R - H^\top H \right) F^\top W \right\|_\infty
\end{aligned}$$

$$
\begin{aligned}
&= \frac{R}{T^2}\|W^\top F\|_\infty \max_{r\in[R],k\in[p_x]} \left|\sum_{j\in[R]} \left(I_R - H^\top H\right)_{r,j} \left(F^\top W\right)_{j,k}\right| \\
&\leq \frac{R^2}{T^2}\|W^\top F\|_\infty^2 \left\|I_R - H^\top H\right\|_2 \\
&= O_P\left(\frac{1}{T^2}(Th_T + T)^2\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_x}}\right)\right) \\
&= O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_x}}\right),
\end{aligned}
$$

by Lemmas A.4 (ii) and A.1 (xi). We finish by bounding J_5. We have

$$
\begin{aligned}
J_5 &\leq \left\|\frac{1}{T}W^\top W - E[w_t w_t^\top]\right\|_\infty + \left\|\frac{1}{T}W^\top F\left(\frac{1}{T}F^\top W\right) - E[w_t f_t^\top]E[w_t f_t^\top]^\top\right\|_\infty \\
&= O_P\left(\left(\frac{p^2}{T^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(2p)}{T}}\right),
\end{aligned}
$$

by Lemmas A.1 (xii) and (xiii). Combining the bounds on J_1 to J_5, we obtain (i).

Proof of (ii). We have

$$
\begin{aligned}
\frac{1}{T}\left\|\widetilde{W}^\top \mathcal{E}\right\|_\infty &= \frac{1}{T}\left\|T^{-1}W^\top\left(I_T - \frac{1}{T}\hat{F}\hat{F}^\top\right)\mathcal{E}\right\|_\infty \\
&\leq J_1 + J_2 + J_3 + J_4 + J_5 + J_6,
\end{aligned}
$$

where

$$
\begin{aligned}
J_1 &= \left\|\frac{1}{T}W^\top\frac{1}{T}(\hat{F} - FH^\top)(\hat{F} - FH^\top)^\top\mathcal{E}\right\|_\infty; \\
J_2 &= \left\|\frac{1}{T}W^\top\frac{1}{T}(\hat{F} - FH^\top)HF^\top\mathcal{E}\right\|_\infty; \\
J_3 &= \left\|\frac{1}{T}W^\top\frac{1}{T}FH^\top(\hat{F} - FH^\top)^\top\mathcal{E}\right\|_\infty; \\
J_4 &= \left\|\frac{1}{T}W^\top\frac{1}{T}F\left(I_R - H^\top H\right)F^\top\mathcal{E}\right\|_\infty; \\
J_5 &= \left\|\frac{1}{T^2}W^\top FF^\top\mathcal{E}\right\|_\infty; \\
J_6 &= \left\|\frac{1}{T}W^\top\mathcal{E}\right\|_\infty.
\end{aligned}
$$

Let us first bound J_1. It holds that

$$\begin{aligned}
J_1 &= \frac{1}{T^2}\max_{k\in[p_x]}\left|\sum_{r\in[R]}\left(W^\top\left(\widehat{F}-FH^\top\right)\right)_{k,r}\left(\left(\widehat{F}-FH^\top\right)^\top\mathcal{E}\right)_r\right| \\
&\leq \frac{R}{T^2}\left\|\left(\widehat{F}-FH^\top\right)^\top W\right\|_\infty\left\|\left(\widehat{F}-FH^\top\right)^\top\mathcal{E}\right\|_\infty \\
&= O_P\left(\frac{1}{T^2}\left(\frac{T}{p_x}+\sqrt{\frac{T}{p_x}}\right)(\sqrt{p_x}h_T+1)\sqrt{\frac{T}{p_x}+1}\right) \\
&= o_P\left(\frac{1}{T\sqrt{p_x}}+\frac{1}{T^{3/2}}+\frac{1}{p_x\sqrt{T}}\right),
\end{aligned}$$

by Lemma 3.1 (ii) and (iii). Next, we consider J_2. We have

$$\begin{aligned}
J_2 &= \frac{1}{T^2}\max_{k\in[p_x]}\left|\sum_{r\in[R]}\left(W^\top\left(\widehat{F}-FH^\top\right)\right)_{k,r}\left(HF^\top\mathcal{E}\right)_r\right| \\
&\leq \frac{R}{T^2}\left\|\left(\widehat{F}-FH^\top\right)^\top W\right\|_\infty\left\|HF^\top\mathcal{E}\right\|_\infty \\
&\leq \frac{R}{T^2}\left\|\left(\widehat{F}-FH^\top\right)^\top W\right\|_\infty\|H\|_2\left\|F^\top\mathcal{E}\right\|_2 \\
&= O_P\left(\frac{1}{T^2}\left(\frac{T}{p_x}+\sqrt{\frac{T}{p_x}}\right)(\sqrt{p_x}h_T+1)(Th_T+T)\right) \\
&= O_P\left(\frac{1}{\sqrt{Tp_x}}+\frac{1}{p_x}+\frac{h_T}{\sqrt{p_x}}+\frac{h_T}{\sqrt{T}}\right).
\end{aligned}$$

by Lemmas 3.1 (iii), A.4 (i) and A.1 (xi). Next, we consider J_3. We have

$$\begin{aligned}
J_3 &\leq \frac{1}{T^2}\max_{k\in[p_x]}\left|\sum_{r\in[R]}\left(\mathcal{E}^\top\left(\widehat{F}-FH^\top\right)\right)_{k,r}\left(HF^\top W\right)_r\right| \\
&\leq \frac{R}{T^2}\left\|\left(\widehat{F}-FH^\top\right)^\top\mathcal{E}\right\|_\infty\left\|HF^\top W\right\|_\infty \\
&\leq \frac{R}{T^2}\left\|\left(\widehat{F}-FH^\top\right)^\top\mathcal{E}\right\|_\infty\max_{r\in[R],k\in[p_x]}\left|\sum_{j\in[R]}H_{r,j}\left(F^\top W\right)_{j,k}\right| \\
&\leq \frac{R^2}{T^2}\left\|\left(\widehat{F}-FH^\top\right)^\top\mathcal{E}\right\|_\infty\|H\|_2\left\|F^\top W\right\|_\infty \\
&= O_P\left(\frac{1}{T^2}\left(\frac{T}{p_x}+1\right)^{1/2}(Th_T+T)\right) \\
&= o_P\left(\frac{1}{\sqrt{Tp_x}}+\frac{1}{T}\right),
\end{aligned}$$

by Lemmas 3.1 (ii), A.4 (i) and A.1 (xi). Then, we focus on J_4. It holds that

$$\begin{aligned}
J_4 &\leq \frac{1}{T^2} \max_{k\in[p_x]} \left| \sum_{r\in[R]} \left(W^\top F\right)_{k,r} \left(\left(I_R - H^\top H\right) F^\top \mathcal{E}\right)_r \right| \\
&\leq \frac{1}{T^2} \|W^\top F\|_\infty \left\| \left(I_R - H^\top H\right) F^\top \mathcal{E} \right\|_2 \\
&\leq \frac{1}{T^2} \|W^\top F\|_\infty \|\mathcal{E}^\top F\|_2 \left\| I_R - H^\top H \right\|_2 \\
&= O_P\left(\frac{1}{T^2}(Th_T + T)\sqrt{T} \left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_x}} \right) \right) \\
&= O_P\left(\frac{1}{T} + \frac{1}{\sqrt{Tp_x}} \right),
\end{aligned}$$

by Lemmas A.1 (xi), (v) and A.4 (ii). Next, we bound J_5, We have

$$\begin{aligned}
J_5 &\leq \frac{1}{T^2} \max_{k\in[p_x]} \left| \sum_{r\in[R]} \left(W^\top F\right)_{k,r} \left(F^\top \mathcal{E}\right)_r \right| \\
&\leq \frac{R}{T^2} \left\|W^\top F\right\|_\infty \left\|F^\top \mathcal{E}\right\|_\infty \\
&\leq \frac{R}{T^2} \left\|W^\top F\right\|_\infty \left\|F^\top \mathcal{E}\right\|_2 \\
&= O_P\left(\frac{1}{T^2}(Th_T + T)\sqrt{T} \right) = O_P\left(\frac{1}{\sqrt{T}} \right),
\end{aligned}$$

Finally, we have

$$J_6 = O_P\left(h_T\right)$$

by Lemma A.1 (xvii). Combining the bounds on J_1 to J_6, we obtain (ii).

Proof of (iii). We have

$$\left\|M_{\widehat{F}} F\gamma\right\|_2 = \left\| \left(I_T - \frac{1}{T}\widehat{F}\widehat{F}^\top \right) F\gamma \right\|_2 \leq J_1 + J_2 + J_3 + J_4 + J_5,$$

where

$$\begin{aligned}
J_1 &= \frac{1}{T} \left\| \left(\widehat{F} - FH^\top\right) HF^\top F\gamma \right\|_2; \\
J_2 &= \frac{1}{T} \left\| FH^\top \left(\widehat{F} - FH^\top\right)^\top F\gamma \right\|_2; \\
J_3 &= \frac{1}{T} \left\| \left(\widehat{F} - FH^\top\right) \left(\widehat{F} - FH^\top\right)^\top F\gamma \right\|_2;
\end{aligned}$$

$$J_4 = \frac{1}{T}\left\|F\left(H^\top H - I_R\right)F\gamma\right\|_2;$$
$$J_5 = \left\|\left(I_T - \frac{1}{T}FF^\top\right)F\gamma\right\|_2.$$

We consider first J_1. It holds that

$$\begin{aligned} J_1 &\leq \frac{1}{T}\left\|\widehat{F} - FH^\top\right\|_2 \|F\|_2^2\|H\|_2\|\gamma\|_2 \\ &= O_P\left(\frac{1}{T}\left(\frac{T}{p_x}+1\right)^{1/2}T\right) = O_P\left(\sqrt{\frac{T}{p_x}+1}\right), \end{aligned}$$

by Lemmas 3.1 (i) and A.1 (iii). Similarly, we have

$$J_2 = O_P\left(\sqrt{\frac{T}{p_x}+1}\right).$$

Next, we focus on J_3 and obtain

$$\begin{aligned} J_3 &\leq \frac{1}{T}\left\|\widehat{F} - FH^\top\right\|_2^2 \|F\|_2\|\gamma\|_2 \\ &= O_P\left(\frac{1}{T}\left(\frac{T}{p_x}+1\right)\sqrt{T}\right) = O_P\left(\frac{\sqrt{T}}{p_x} + \frac{1}{\sqrt{T}}\right) = O_P\left(\sqrt{\frac{T}{p_x}+1}\right)., \end{aligned}$$

by Lemmas 3.1 (i) and A.1 (iii). Then, we consider J_4. We have

$$\begin{aligned} J_4 &\leq \frac{1}{T}\|F\|_2^2\|\gamma\|_2\left\|H^\top H - I_R\right\|_2 \\ &= O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_x}}\right), \end{aligned}$$

by Lemmas A.1 (iii) and A.4 (ii). Finally, we have

$$\begin{aligned} J_5 &= \left\|\left(I_T - \frac{1}{T}FF^\top\right)F\left(I_R - \frac{1}{T}F^\top F\right)\gamma\right\|_2 \\ &\leq \left\|I_R - \frac{1}{T}F^\top F\right\|_2 \|F\|_2\|\gamma\|_2 \\ &= O_P\left(\frac{1}{\sqrt{T}}\sqrt{T}\right) = O_P(1), \end{aligned}$$

by Lemma A.1 (i) and (iii). Combining the bounds on J_1 to J_5, we obtain (iii).

Proof of (iv). It holds that

$$\left\|P_{\widehat{F}}\mathcal{E}\right\|_2 = \left\|\frac{1}{T}\widehat{F}\widehat{F}^\top\mathcal{E}\right\|_2$$

$$
\begin{aligned}
&\leq \frac{1}{T}\left\|\hat{F}\left(\hat{F}-FH^{\top}\right)\mathcal{E}\right\|_2+\frac{1}{T}\left\|\hat{F}F^{\top}\mathcal{E}\right\|_2\\
&\leq \frac{1}{T}\left\|\hat{F}\right\|_2\left\|\hat{F}-FH^{\top}\right\|_2\|\mathcal{E}\|_2+\frac{1}{T}\left\|\hat{F}\right\|_2\|F^{\top}\mathcal{E}\|_2\\
&= O_P\left(\frac{1}{T}\sqrt{T}\left(\frac{T}{p_x}+1\right)^{1/2}\sqrt{T}+\frac{1}{T}\sqrt{T}\sqrt{T}\right)\\
&= O_P\left(\sqrt{\frac{T}{p_x}+1}\right),
\end{aligned}
$$

by Lemmas 3.1 (i) and A.1 (iv), (v) and the fact that $\left\|\hat{F}\right\|_2=\sqrt{RT}$. □

A.5 Pre-existing results

The following lemma is a direct consequence of the Fuk-Nagaev inequality for τ-mixing processes of Babii et al. (2022) (see their Theorem A.1).

Lemma A.6 *Let $\{\zeta_t\}_t$ be a p-dimensional mean zero stationary random process such that*

(*i*) *for some $m>2$, $\|\zeta_t\|_m=O(1)$;*

(*ii*) *there exists $c>0$ and $a>(m-1)/(m-2)$ such that, for every $j\in[p]$, the τ-mixing coefficients of $\{\zeta_t\}_t$ satisfy $\tau_s^{(j)}\leq cs^{-a}$.*

Then, we have

$$
\left\|\frac{1}{T}\sum_{t\in[T]}\zeta_t\right\|_{\infty}=O_P\left(\left(\frac{p}{T^{\kappa-1}}\right)^{1/\kappa}\vee\sqrt{\frac{\log(2p)}{T}}\right),
$$

where $\kappa=\frac{(a+1)m-1}{a+m-1}$.

The following lemma is a direct implication of Lemma A.1.2 in Babii et al. (2024).

Lemma A.7 *Let $\{\zeta_t\}_t$ be a 1-dimensional mean zero stationary random process such that, for some $q>2$, $\|\zeta_t\|_q=O(1)$. Then, we have*

$$
|E[\zeta_0\zeta_t]|\leq\tau_s^{\frac{q-2}{q-1}}\|\zeta_t\|_q^{\frac{q}{q-1}},
$$

where $\{\tau_s\}_s$ are the τ-mixing coefficients of $\{\zeta_t\}_t$.

The following Lemma is a direct consequence of Lemma A.2.1 in Babii et al. (2022).

Lemma A.8 *For any $d \in \mathbb{R}^T$, we have*

$$\Omega^*(d) = \mu \|d\|_\infty + (1-\mu) \max_{G \in \mathcal{G}} \|d_G\|_2 .$$

This implies that

$$\Omega^*(d) \leq \left(\mu + (1-\mu)\sqrt{G^*}\right) \|d\|_\infty .$$

B Additional details on the data

Monthly macro data

	Description	Category	T-code
1	Real Personal Income	Output and income	5
2	Real personal income ex transfer receipts	Output and income	5
3	Industrial Production Index	Output and income	5
4	IP: Final Products and Nonindustrial Supplies	Output and income	5
5	IP: Final Products (Market Group)	Output and income	5
6	IP: Consumer Goods	Output and income	5
7	IP: Materials	Output and income	5
8	IP: Manufacturing (SIC)	Output and income	5
9	Capacity Utilization: Manufacturing	Output and income	5
10	Civilian Labor Force	Labour market	5
11	Civilian Employment	Labour market	2
12	Civilian Unemployment Rate	Labour market	5
13	Average Duration of Unemployment (Weeks)	Labour market	5
14	Civilians Unemployed - Less Than 5 Weeks	Labour market	2
15	Civilians Unemployed for 5-14 Weeks	Labour market	2
16	Civilians Unemployed - 15 Weeks & Over	Labour market	5
17	Civilians Unemployed for 15-26 Weeks	Labour market	5
18	Civilians Unemployed for 27 Weeks and Over	Labour market	5
19	Initial Claims	Labour market	5

20	All Employees: Total nonfarm	Labour market	5
21	All Employees: Goods-Producing Industries	Labour market	5
22	All Employees: Mining and Logging: Mining	Labour market	5
23	All Employees: Construction	Labour market	5
24	All Employees: Manufacturing	Labour market	5
25	All Employees: Durable goods	Labour market	5
26	All Employees: Nondurable goods	Labour market	5
27	All Employees: Service-Providing Industries	Labour market	5
28	All Employees: Wholesale Trade	Labour market	5
29	All Employees: Retail Trade	Labour market	5
30	All Employees: Financial Activities	Labour market	5
31	All Employees: Government	Labour market	5
32	Avg Weekly Hours : Goods-Producing	Labour market	5
33	Avg Weekly Overtime Hours : Manufacturing	Labour market	5
34	Avg Weekly Hours: Manufacturing	Labour market	1
35	Avg Hourly Earnings: Goods-Producing	Labour market	2
36	Avg Hourly Earnings: Construction	Labour market	1
37	Avg Hourly Earnings: Manufacturing	Labour market	4
38	Housing Starts: Total New Privately Owned	Housing	4
39	Housing Starts, Northeast	Housing	4
40	Housing Starts, Midwest	Housing	4
41	Housing Starts, South	Housing	4
42	Housing Starts, West	Housing	4
43	New Private Housing Permits, Northeast (SAAR)	Housing	4
44	New Private Housing Permits, Midwest (SAAR)	Housing	4
45	New Private Housing Permits, South (SAAR)	Housing	4
46	New Private Housing Permits, West (SAAR)	Housing	5
47	Real Manu. and Trade Industries Sales	Consumption	5
48	Retail and Food Services Sales	Consumption	5

49	New Orders for Durable Goods	Consumption	5
50	New Orders for Nondefense Capital Goods	Consumption	2
51	Unfilled Orders for Durable Goods	Consumption	6
52	Total Business Inventories	Consumption	6
53	Total Business: Inventories to Sales Ratio	Consumption	5
54	Consumer Sentiment Index	Consumption	6
55	M1 Money Stock	Money and credit	7
56	M2 Money Stock	Money and credit	6
57	Real M2 Money Stock	Money and credit	6
58	Total Reserves of Depository Institutions	Money and credit	6
59	Reserves Of Depository Institutions	Money and credit	2
60	Commercial and Industrial Loans	Money and credit	6
61	Real Estate Loans at All Commercial Banks	Money and credit	6
62	Total Nonrevolving Credit	Money and credit	6
63	Nonrevolving consumer credit to Personal Income	Money and credit	6
64	Consumer Motor Vehicle Loans Outstanding	Money and credit	6
65	Total Consumer Loans and Leases Outstanding	Money and credit	6
66	Securities in Bank Credit at All Commercial Banks	Money and credit	6
67	Crude Oil, spliced WTI and Cushing	Prices	6
68	PPI: Metals and metal products:	Prices	6
69	CPI : All Items	Prices	6
70	CPI : Apparel	Prices	6
71	CPI : Transportation	Prices	6
72	CPI : Medical Care	Prices	6
73	CPI : Commodities	Prices	2
74	CPI : Services	Prices	6
75	CPI : All Items Less Food	Prices	6
76	CPI : All items less medical care	Prices	6

Table OA.1: FRED MD monthly data subset. Definitions of t-codes are available in the primary data source.

Source: https://research.stlouisfed.org/econ/mccracken/fred-databases/

Weekly financial data

	Description	Category	T-code
1	1-mo. Nonfinancial commercial paper A2P2/AA credit spread	Credit	1
2	Moody's Baa corporate bond/10-yr Treasury yield spread	Credit	1
3	BofAML High Yield/Moody's Baa corporate bond yield spread	Credit	1
4	30-yr Jumbo/Conforming fixed rate mortgage spread	Credit	1
5	30-yr Conforming Mortgage/10-yr Treasury yield spread	Credit	1
6	10-yr Constant Maturity Treasury yield	Leverage	2
7	S&P 500 Financials/S&P 500 Price Index (Relative to 2-yr MA)	Leverage	5
8	S&P 500, S&P 500 mini, NASDAQ 100, NASDAQ mini Open Interest	Leverage	4
9	3-mo. Eurodollar, 10-yr/3-mo. swap, 2-yr and 10-yr Treasury Open Interest	Leverage	4
10	1-mo. Asset-backed/Financial commercial paper spread	Risk	1
11	BofAML Home Equity ABS/MBS yield spread	Risk	1
12	3-mo. Financial commercial paper/Treasury bill spread	Risk	1
13	Commercial Paper Outstanding	Risk	3
14	BofAML 3-5 yr AAA CMBS OAS spread	Risk	1
15	3-mo./1-wk AA Financial commercial paper spread	Risk	1
16	Treasury Repo Delivery Fails Rate	Risk	4
17	Agency Repo Delivery Failures Rate	Risk	4
18	Government Securities Repo Delivery Failures Rate	Risk	4
19	Agency MBS Repo Delivery Failures Rate	Risk	4
20	3-mo. Eurodollar spread (LIBID-Treasury)	Risk	1
21	On-the-run vs. Off-the-run 10-yr Treasury liquidity premium	Risk	1
22	Fed Funds/Overnight Treasury Repo rate spread	Risk	1

23	Fed Funds/Overnight Agency Repo rate spread	Risk	1
24	Fed Funds/Overnight MBS Repo rate spread	Risk	1
25	3-mo./1-wk Treasury Repo spread	Risk	1
26	10-yr/2-yr Treasury yield spread	Risk	1
27	2-yr/3-mo. Treasury yield spread	Risk	1
28	10-yr Interest Rate Swap/Treasury yield spread	Risk	1
29	2-yr Interest Rate Swap/Treasury yield spread	Risk	1
30	1-yr Interest Rate Swap/1-Year Treasury spread	Risk	1
31	3-mo. LIBOR/CME Term SOFR-Treasury spread	Risk	1
32	1-yr./1-mo. LIBOR/CME Term SOFR spread	Risk	1
33	Advanced Foreign Economies Trade-weighted US Dollar Value Index	Risk	3
34	CBOE Market Volatility Index VIX	Risk	1
35	1-mo. BofAML Option Volatility Estimate Index	Risk	1
36	3-mo. BofAML Swaption Volatility Estimate Index	Risk	1

Table OA.2: Financial weekly data set. Sources: Bloomberg & Haver Analytics. Definitions of t-codes are available on NFCI Chicago Fed website: https://www.chicagofed.org/research/data/nfci/current-data

Factors

Factor	Frequency	Source website
ADS	Weekly	https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/ads
CFNAI	Monthly	https://www.chicagofed.org/research/data/cfnai/historical-data
NFCI	Weekly	https://research.stlouisfed.org/econ/mccracken/fred-databases/

Table OA.3: Factors.

C Additional empirical results

	sg-LASSO-FAMIDAS	sg-LASSO-MIDAS
	Panel A. *Full sample*	
sg-LASSO-MIDAS	0.016**	
FAMIDAS	0.020**	0.020**
	Panel B. *Up to COVID*	
sg-LASSO-MIDAS	0.560	
FAMIDAS	0.028**	0.070*

Table OA.4: Nowcast comparisons — We report the p-values of the average superior predictive ability bootstrap tests (aSPA) of Quaedvlieg (2021) over all three horizons comparing sg-LASSO-FAMIDAS, sg-LASSO-MIDAS, and FAMIDAS methods. We test the null hypothesis that the average out-of-sample loss over the three horizons is smaller for the methods in the column versus in the row. * and ** indicates 10% and 5% significance, respectively.

D Details on matrix completion

To implement matrix completion, we use the R package `softImpute` version 1.4—1 downloaded from CRAN. The algorithm fits a low-rank matrix approximation to a matrix with missing values via nuclear-norm regularization. We set the maximum number of rank, `max.rank`, to 6, which restricts the rank of the solution. Starting from λ_0, where λ is the regularization parameter for the nuclear norm minimization problem, we find λ so that the solution reached has rank slightly less than `rank.max`, as suggested in the package manual. λ_0 is the initial guess, which we set to a value computed by a function `lambda0` within the package. This function computes the smallest value for λ such that `softImpute` returns the zero solution.

References

Babii, A., Ghysels, E. & Striaukas, J. (2022), 'Machine learning time series regressions with an application to nowcasting', *Journal of Business & Economic Statistics* **40**(3), 1094–1106.

Babii, A., Ghysels, E. & Striaukas, J. (2024), 'High-dimensional granger causality tests with an application to vix and news', *Journal of Financial Econometrics* **22**(3), 605–635.

Quaedvlieg, R. (2021), 'Multi-horizon forecast comparison', *Journal of Business & Economic Statistics* **39**(1), 40–53.

Copyright © 2024 @ the author(s). Discussion papers are in draft form. This discussion paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.