Hierarchical Regularizers for Mixed-Frequency Vector Autoregressions

Alain Hecq, Marie Ternes, and Ines Wilms

Department of Quantitative Economics, Maastricht University, Maastricht, Netherlands

ABSTRACT
Mixed-frequency Vector Autoregressions (MF-VAR) model the dynamics between variables recorded at different frequencies. However, as the number of series and high-frequency observations per low-frequency period grow, MF-VARs suffer from the “curse of dimensionality.” We curb this curse through a regularizer that permits hierarchical sparsity patterns by prioritizing the inclusion of coefficients according to the recency of the information they contain. Additionally, we investigate the presence of nowcasting relations by sparsely estimating the MF-VAR error covariance matrix. We study predictive Granger causality relations in a MF-VAR for the U.S. economy and construct a coincident indicator of GDP growth. Supplementary materials for this article are available online.

1. Introduction
Vector AutoRegressive (VAR) models are a cornerstone for modeling multivariate time series; studying their dynamics and forecasting. However, standard VARs require all component series to enter the model at the same frequency, while in practice macro and financial series are typically recorded at different frequencies; quarterly, monthly, or weekly for instance. One could aggregate high-frequency variables to one common low frequency and continue the analysis with a standard VAR, but could aggregate high-frequency variables to one common low frequency and continue the analysis with a standard VAR, but such a practice wastes valuable information contained in high-frequency data due to two main reasons. First, high-frequency data are inherently more timely; they closely track the state of the economy in real time. Second, they can help unmask dynamics that would be hidden under temporal aggregation (see, e.g., recent discussions in Cimadomo et al. 2021; Paccagnini and Parla 2021).

Mixed-frequency (MF) models, instead, exploit the information available in series recorded at different frequencies. One commonly used MF models is the MIXed DATA Sampling (MIDAS) regression (Ghysels, Santa-Clara, and Valkanov 2004). While the literature first focused on a single-equation framework for modeling the low-frequency response, the multivariate extension by Ghysels (2016) enabled one to model the relations between high- and low-frequency series in a mixed-frequency VAR (MF-VAR) system. This is known as the stacked MF-VAR approach since the MF-VAR is estimated at the lowest frequency and all higher-frequency variables are treated as separate components series which are stacked in the MF-VAR.¹

A complication with MF-VARs is that they are severely affected by the “curse of dimensionality.” This curse arises due to two sources. First, the number of parameters grows quadratically with the number of component series, just like for standard VARs. Second, specific to MF-VARs, we have many high-frequency observations per low-frequency observation, which each enter as different component series in the model, thereby adding to the dimensionality. Without further adjustments, one would be limited to MF-VARs with few series and/or a small number of high-frequency observations per low-frequency observation.

This curse of dimensionality has mostly been addressed through mixed-frequency factor models (e.g., Marcellino and Schumacher 2010; Foroni and Marcellino 2014; Andreou et al. 2019) or Bayesian estimation (e.g., Schorfheide and Song 2015; Ghysels 2016; Götz, Hecq, and Smeekes 2016; McCracken, Owyang, and Sekhpoyan 2021; Cimadomo et al. 2021; Paccagnini and Parla 2021). Sparsity-inducing regularizers form an appealing alternative (see, Hastie, Tibshirani, and Wainwright 2015 for an introduction), but despite their popularity in regression and standard VAR settings (e.g., Hsu, Hung, and Chang 2008; Basu and Michailidis 2015; Basu, Shojaie, and Michailidis 2015; Davis, Zang, and Zheng 2016; Callot, Kock, and Meideiros 2017; Derimer et al. 2018; Smeekes and Wijler 2018; Barigozzi and Brownlees 2019; Hecq, Margaritella, and Smeekes 2021a), they have only been rarely explored as a tool for...
dimension reduction in mixed-frequency models. An exception is Babii, Ghysels, and Striaukas (2021) who recently used the sparse-group lasso to accommodate the dynamic nature of high-dimensional, mixed-frequency data, thereby providing a complementary structured machine learning perspective to the penalized Bayesian MIDAS approach of Mogliani and Simoni (2021). Nonetheless, they address univariate MIDAS regressions, leaving regularization of MF-VARs unexplored.

Our article’s first contribution concerns the introduction of a novel convex regularizer that extends the univariate MIDAS approach of Babii, Ghysels, and Striaukas (2021) to the multivariate MF-VAR setting. To this end, we propose a mixed-frequency extension of the single-frequency hierarchical regularizer by Nicholson et al. (2020) used for standard VARs that accounts for covariates at different (high-frequency) lags being temporally ordered. We build upon the group lasso with nested groups and encourage a hierarchical sparsity pattern that prioritizes the inclusion of coefficients according to the recency of the information the corresponding series contains about the state of the economy.

In addition to the development of a new MF-VAR with hierarchical lag structure, our article investigates the presence of nowcasting restrictions in a high-dimensional mixed-frequency setting. According to Eurostat’s glossary, a nowcast is a rapid [estimate] produced during the current reference period [say T* a particular quarter.] for a hard economic variable of interest observed for the same reference period [T*] (Eurostat: Statistics Explained 2014). In this narrow sense, and contrarily to forecasting, nowcasting makes use of all available information becoming available between (and strictly speaking not including) T* − 1 and T*. Götz and Hecq (2014) show that nowcasting in (low-dimensional) MF-VARs can be studied through contemporaneous Granger causality tests; by testing the null of a block diagonal error covariance matrix of the MF-VAR. We build on Götz and Hecq (2014) to study nowcasting relations in high-dimensional MF-VARs by sparsely estimating the covariance matrix of the MF-VAR errors. Its sparsity pattern then provides evidence on those high-frequency variables (i.e., the series and their particular time period) one can use to build coincident indicators for the low-frequency main economic indicators.

In a simulation study (Section 4), we find that the hierarchical regularizer performs well in terms of estimation accuracy and variable selection when compared to alternative methods. Furthermore, we accurately retrieve nowcasting relations between the low- and high-frequency variables by sparsely estimating the error covariance matrix. In the application (Section 5), we study a high-dimensional MF-VAR for the U.S. economy. We apply the hierarchical regularizer to characterize the predictive Granger causality relations through a network analysis. Moreover, we investigate which high-frequency series nowcast quarterly U.S. real gross domestic product (GDP) growth, use those to construct a reliable coincident indicator of GDP growth and evaluate its performance pre and post Covid-19.

The remainder of this article is structured as follows. Section 2 introduces the MF-VAR with hierarchically structured parameters and defines the nowcasting causality relations, Section 3 describes the regularized estimation procedure for MF-VARs and nowcasting causality. Section 4 shows the results on the simulation study, Section 5 on the empirical application of the U.S. economy. Section 6 concludes. Additional results are available in the Appendix, supplementary materials.

2 Alternatively, nowcasts can be obtained as forecasts conditional on the real-time data flow via state-space techniques such as the Kalman filter which is needed to estimate the latent processes. We discuss the method of Cimadomo et al. (2021) as state-of-the-art benchmark in this literature in Section 5.4.
with rich and versatile dynamics, it is plausible to assume that the data generating process of a high-dimensional MF-VAR has a small \( t \). But even a MF-VAR\(_K\)(1) model requires one to estimate \( K^2 \) parameters which quickly becomes large since for a fixed \( T \) the parameter vector grows (quadratically) with the number of time series included but also due to the high-per-low frequency observations \( m_1, \ldots, m_d \). In the remainder, we use matrix notation to compactly express the mixed-frequency VAR. To this end, define

\[
\begin{align*}
\bar{y}_t &= \begin{bmatrix} y(t), x^{m_1}(t), \ldots, x^{m_d}(t) \end{bmatrix}' \quad (K \times 1) \\
Y &= \begin{bmatrix} \bar{y}_1, \ldots, \bar{y}_N \end{bmatrix}' \quad (N \times K) \\
Z &= \begin{bmatrix} y_0, \ldots, y_{N-1} \end{bmatrix}' \quad (N \times K) \\
X &= I_K \otimes Z \quad (NK \times K^2) \\
u_t &= \begin{bmatrix} u(t), u^{m_1}(t), \ldots, u^{m_d}(t) \end{bmatrix}' \quad (K \times 1) \\
U &= \begin{bmatrix} u_1, \ldots, u_N \end{bmatrix}' \quad (N \times K),
\end{align*}
\]

where \( N = T - 1 \) are the number of time points available given the MF-VAR\(_K\)(1), \( I_K \) denotes the identity matrix of dimension \( K \) and \( \otimes \) the Kronecker product. Then the MF-VAR\(_K\)(1) can be written as \( Y = X\beta + u \), with \( y = \text{vec}(Y) \), \( \beta = \text{vec}(B) \), and \( u = \text{vec}(U) \).

In the classical low-dimensional setting \( K < N \), the MF-VAR can be estimated by least squares. However, as the number of parameters grows relative to the time series length \( T \), least squares becomes unreliable as it results in high variance, overfitting and poor out-of-sample forecasting. We therefore resort to penalization methods. Many authors have used the lasso (Tibshirani 1996), which attains so called “patternless” sparsity in the parameter matrices of the VAR (e.g., Hsu, Hung, and Chang 2008; Derimer et al. 2018). We, instead, use the dynamic structure of the MF-VAR as a guide in our sparse estimation procedure. Therefore, we propose regularization estimation of MF-VARs that translates information about the hierarchical structure of low- and high-frequency variables into a convex regularizer that delivers structured sparsity patterns appropriate to the context of MF-VARs.

### 2.1. Hierarchical Structures

We describe the hierarchical sparsity patterns that arise in the autoregressive coefficient matrix \( B \) of a MF-VAR. We demonstrate the intuition, for ease of notation, with \( k_1 = k_1 = \cdots = k_d = 1 \), see Remark 2.1 for an extension to multiple series. The parameters of the matrix \( B \) can be divided in \((d + 1)^2\) different groups as depicted by the submatrices in the left panel of Figure 1. We distinguish three types of submatrices capturing, respectively Own-on-Own, Higher-on-Lower and Lower-on-Higher effects, to extend standard VAR estimation with single-frequencies to mixed-frequencies. The Own-on-Own sub-matrices are \( m \times m \) square matrices that lie on the main diagonal and describe the effects of a series’ own lags on itself (with \( m_i = 1 \) for the low-frequency variable). The Higher-on-Lower sub-matrices are short, wide \( m_1 \times m_H \) matrices, where \( L \) and \( H \) refer to the corresponding lower- and higher-frequency variable. They lie above the main diagonal and contain the lagged effect of a higher-frequency series onto a series with respective lower frequency. Note that \( L \) (and \( H \)) just identify which of the two variables is of lower (and higher) frequency. Thus, this group does not only contain the effects of the higher-frequency variables onto the variable with the lowest frequency but also describes the interactions between the higher-frequency variables. For instance, in a quarter/month/week example, these incorporate the effects of the monthly and weekly variable onto the quarterly variable but also the effects of the weekly onto the monthly variable. The Lower-on-Higher submatrices are long, thin \( m_l \times m_H \) matrices. They lie below the main diagonal and contain the lagged effect depicting the effect of a lower-frequency series onto a higher-frequency series.

For each submatrix, we impose a hierarchical priority structure for parameter inclusion. Parameters with higher priority within one group should be included in the model before parameters with lower priority within the same group, where the priority value of each parameter depends on how informative a practitioner finds the associated regressor. A priority value of one indicates highest priority for inclusion. More precisely, we introduce a priority value \( p_{g}^{ij} \in \{1, \ldots, P_{g}\} \) for each element \( ij \) belonging to parameter group \( g = 1, \ldots, G \) in the matrix \( B \) to denote its inclusion priority. Here, we order the groups in Figure 1 from left to right and top to bottom. For instance, the top left square is group \( g = 1 \), the rectangle to the right is group \( g = 2 \). If \( p_{g}^{ij} < p_{g}^{i'j'} \) for \( i \neq i' \) and \( j \neq j' \), we prioritize parameter \( p_{g}^{ij} \) over \( p_{g}^{i'j'} \) in the model. The hierarchical structure is imposed for each group \( g \) individually and hence in each group the priority values start with value 1 (highest priority) and go up to \( P_{g} \) (lowest priority). We do not encode structures where certain parameters of one group should enter the model before certain parameters of another group as our aim is to encode priority of parameter inclusion with respect to the effects of one frequency component (Own/Lower/Lower) on another (Own/Lower/Lower).

These hierarchical priority structures are highly general and could potentially accommodate various structured sparsity patterns in the AR parameter matrix that researchers or practitioners want to encourage. We give special attention to a recency-based priority structure, where the priority value \( p_{g}^{ij} \) is set according to the recency of the information the \( j \)th time series of \( y_{t-1} \) contains relative to the \( i \)th series of \( y_{t} \). The more recent the information contained in the lagged predictor, the more informative, and thus the higher its inclusion priority. Figure 2 visualizes the recency-based structure for an example with only one quarterly and one monthly variable having a \( 4 \times 4 \) coefficient matrix \( B \) and \( G = 4 \) groups. For instance, consider the Higher-on-Lower parameter block, where month three of the previous quarter contains the most recent information, followed by month two and then month one. The priority values are thus increasing from left to right.

The recency-based priority structure is conceptually similar to other often used restriction schemes in the (MF)-VAR literature. For the Higher-on-Lower group, it is similar to a MIDAS weighting function with decaying shape, for instance an exponential Almon lag polynomial. Both approaches assume a decaying memory pattern in the economic processes, however, in our setting, we do not restrict the parameters to a specific nonlinear function. Besides, our recency-based
priorities are similar in spirit to the Minnesota prior in the Bayesian literature (e.g., Cimadomo et al. 2021) in the sense that longer lags are treated differently than shorter lags as they are expected to contain less information about a variable’s current value.

Remark 2.1. If one were to extend to a MF-VAR$_K$(1) with multiple time series per frequency component, the total number of groups $G$ becomes $(k_L + k_1 + \cdots + k_d)^2$. Then the priority values describing the dependence of the dependent variable having frequency $m_i$ on the independent variable having frequency $m_j$ would be replicated $k_i \times k_j$ times. For example, in a setting with two quarterly and three monthly variables, the priority values describing the effect of monthly series on quarterly series are replicated six times.

Remark 2.2. If one were to increase the lag length, the total number of groups $G$ would remain unchanged, but each group would additionally incorporate all its higher-order lagged coefficients. Since higher-order lags contain older information, we decrease their priority of inclusion. For example, the Higher-on-Lower group for a MF-VAR$_4$(2) would have priority values one up to three for lag 1; four up to six for lag 2. We discuss the sensitivity of our empirical results to the maximum lag length in Section 5.3.

2.2. Nowcasting Restrictions

The hidden contemporaneous links between high- and low-frequency variables can be investigated in the covariance matrix $\Sigma_u$ of the error terms $u_t$ in the MF-VAR model (1). Götz and Hecq (2014) test for block diagonality of $\Sigma_u$ to investigate the null of no contemporaneous Granger causality or, to put it differently, the absence of nowcasting relationships between high- and low-frequency indicators. In the remainder, we refer to contemporaneous Granger causality as "nowcasting causality." The authors show that the conditional single equation model (e.g., MIDAS) with contemporaneous regressors can be misleading as it can change the dynamics observed in a MF-VAR system (e.g., Granger causality relations). We do not face this problem since we work with the reduced form MF-VAR and not with a single equation conditional model derived from the MF-VAR.

In the context of our article, a detailed inspection of the residual covariance matrix $\hat{\Sigma}_u := \frac{1}{T} \sum_{t=\ell+1}^T \mathbf{u}_t \mathbf{u}_t'$ of the high-dimensional MF-VAR is interesting for (at least) two reasons. First, from an economic perspective, we aim to investigate whether there exist high-frequency months, weeks or days of some series that nowcast low-frequency variables. Since this is a correlation measure observed in the symmetric blocks of the residual covariance matrix, we obviously cannot point toward a direction of this contemporaneous link. Lütkepohl (2005, pp.
stresses that the direction of (nowcasting) causation must be obtained from further knowledge (e.g., economic theory) on the relationship between the variables. We can only agree on that. Second, from a statistical perspective, we aim to compare the performance of the MF-VAR from Section 2.1 when additional restrictions on the error covariance matrix are considered. Intuitively, this is a Generalized Least Squares (GLS) type improvement over the “unrestricted” hierarchical MF-VAR.

In our economic application (to be discussed in Section 5), we consider quarterly real GDP growth as one of the main low-frequency indicators, and are interested in investigating whether some high-frequency monthly (e.g., retail sales) and/or weekly (e.g., money stock, federal fund rate) series can deliver a coincident indicator of GDP growth with the advantage being that they are released earlier. We do not claim that they directly impact GDP but simply that the business cycle movements that they are released earlier. We do not claim that they directly impact GDP but simply that the business cycle movements that they are released earlier.

We can non-linearly aggregate that.

Finally, to build our coincident indicator, we use a simple procedure which consists of first selecting the variable-period combinations corresponding to nonzero entries in $\sigma_{K-1}^{2}$, standardizing the series and subsequently constructing the first principal component. We rescale the coincident indicator to GDP growth as in Lewis et al. (2021).

3. Regularized Estimation Procedure of MF-VARs

3.1. Hierarchical Group Lasso for Structured MF-VAR

To attain the hierarchical structure presented in Section 2.1, we use a nested group lasso (Zhao, Rocha, and Yu 2009) which has been successfully employed for various statistical problems, among which, time series models (e.g., Nicholson et al. 2020), generalized additive models (e.g., Lou et al. 2016), regression models with interactions (e.g., Haris, Witten, and Simon 2016) or banded covariance estimation (e.g., Bien, Bunea, and Xiao 2016). The group lasso uses the sum of (unsquared) Euclidean norms as penalty term to encourage sparsity on the group-level. Then, either all parameters of a group are set to zero or none. By using nested groups, hierarchical sparsity constraints are imposed where one set of parameters being zero implies that another set is also set to zero. We encourage hierarchical sparsity within each group of parameters by prioritizing parameters with a lower priority value over parameters with a higher one. The proposed hierarchical group estimator for the MF-VAR is given by

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2} \| y - X\beta \|^2_2 + \lambda_{\beta} P_{\text{Hier}}(\beta) \right\},$$

(2)

where $P_{\text{Hier}}(\beta)$ denotes the hierarchical group penalty and $\lambda_{\beta} \geq 0$ is a tuning parameter.

Before introducing the hierarchical group penalty, recall that we distinguish $g = 1, \ldots, G$ parameter groups in the autoregressive parameter matrix and that $P_g$ is the maximum priority value of parameter group $g$. To impose the hierarchical structure within each parameter group $g$, we consider $P_g$ nested subgroups $s_{g}^{(1)}, \ldots, s_{g}^{(P_g)}$. Group $s_{g}^{(1)}$ contains all parameters of group $g$, $s_{g}^{(2)}$ omits those parameters having priority value one, and finally the last subgroup $s_{g}^{(P_g)}$ only contains those parameters having the highest priority value $P_g$. Clearly, a nested structure arises with $s_{g}^{(1)} \supset \cdots \supset s_{g}^{(P_g)}$. Now, denote $\beta_g^{(p:P_g)} = [\beta_{g_{1}}^{(p, P)}, \ldots, \beta_{g_{P_g}}^{(p, P)}]$, for $1 \leq p \leq P_g$, where $\beta_{g_{p}}^{(p, P)}$ collects the parameters of group $g$ having priority value $p$. We are now ready to define the hierarchical penalty function as

$$P_{\text{Hier}}(\beta) = \sum_{g=1}^{G} \sum_{p=1}^{P_g} w_{s_{g}^{(p)}} \| \beta_{g_{p}}^{(p, P)} \|_2,$$

(3)

The hierarchical structure within each group $g$ is built in through the condition that if $\beta_{g_{p}}^{(p, P_{g_{p}})} = 0$, then $\beta_{g_{p}}^{(p, P_{g_{p}})} = 0$ where $p < p'$. 

Remark 2.3. The case $\sigma_{K-1}^{2} = 0_{(K-1) \times 1}$ (with one low-frequency variable) corresponds to the block diagonality of $\Sigma$ and hence the null of no nowcasting causality.

Remark 2.4. Unlike for the autoregressive parameter matrix, we do not impose a hierarchical sparsity structure on $\sigma_{K-1}^{2}$ because the middle month could be a better coincident indicator for a quarterly variable than the last month, for instance. Since this is an empirical issue, we prefer to stick to a regularized estimator of the residual covariance matrix that encourages “patternless” sparsity in $\sigma_{K-1}^{2}$.
Remark 3.1. A related sparse-group lasso penalty has been proposed by Babii, Ghysels, and Striaukas (2021). Similar to our penalty, theirs is tailored to the dynamic nature of the mixed-frequency model to account for different high-frequency lags being temporally related. As such, both penalties induce sparsity at two levels: whether a variable is important for explaining another or not; and if important, both steer the effect’s duration via an additional sparsity layer. Via our nested group lasso, we do this by building a hierarchy to cut off the effect of one variable on another after a certain (high-frequency) lag. Via the sparse-group lasso, Babii, Ghysels, and Striaukas (2021) regulate the shape of the MIDAS weight function for each group which consists of all high-frequency lags of a single variable.

The weights \( w^{(p)}_g \) balance unequally sized nested subgroups. We take \( w^{(p)}_g = \text{card}(s^{(1)}_g) - \text{card}(s^{(p)}_g) + 1 \), where \( \text{card}(\cdot) \) corresponds to the cardinality. As the cardinality of the subgroups \( s^{(p)}_g \) is decreasing with \( p \), the weights of the nested subgroups are increasing with \( p \). Subgroups containing parameters with lower priority, that is, with older information about the state of the economy, are thus, penalized more and hence more likely to be zeroed out.

Remark 3.2. A simple alternative to our proposed weights are equal weights as used in Zhao, Rocha, and Yu (2009) or Nicholson et al. (2020), which would lead to less aggressive shrinkage. Simulations in Section 4.1 reveal that the proposed more aggressive weights are preferable for sparsity recognition. In Section 4.4, we include recommendations to guide forecasters in choosing the weights.

Finally, we propose a proximal gradient algorithm (see, e.g., Tseng 2008) to efficiently solve the optimization problem in Equation (2), as detailed in Appendix A supplementary materials.

3.2. Regularized Estimation of Nowcasting Causality Relations

To detect nowcasting relations between the low- and high-frequency variables, we use a lasso-penalty to impose “patternless” sparsity on the covariances between low- and high-frequency errors (see Section 2.2). The proposed sparse covariance estimator is then given by

\[
\hat{\Sigma}^*_n = \arg \min_{\Sigma \succ 0} \left\{ \frac{1}{2} \| \hat{\Sigma}_u - \Sigma \|_F^2 + \lambda_\Sigma \| \Sigma^{-} \|_1 \right\}
\]  

(4)

where \( \lambda_\Sigma \geq 0 \) is a tuning parameter, \( \Sigma^\top \) are the elements of the off-diagonal blocks of \( \Sigma \). Furthermore, for a matrix \( A \), \( \| A \|_F = \| \text{vec}(A) \|_F = \left( \sum_{ij} A_{ij}^2 \right)^{1/2} \) denotes the Frobenius norm and \( \| A \|_1 = \| \text{vec}(A) \|_1 = \sum_{ij} | A_{ij} | \) the \( l_1 \)-norm.

Remark 3.3. The mere addition of the \( l_1 \)-penalty in problem (4) does not guarantee the estimator \( \hat{\Sigma}^*_n \) to be positive definite (see Rothman, Levina, and Zhu 2009). To ensure its positive definiteness, the constraint \( \Sigma \succ 0 \) implies that we only consider solutions with strictly positive eigenvalues. Rothman, Levina, and Zhu (2009) show that (4) without the constraint \( \Sigma \succ 0 \) essentially boils down to element-wise soft-thresholding of \( \Sigma^\top \): the sparse estimate \( \hat{\Sigma}^*_n \) is given by \( \text{sign}(\Sigma^{-}) \max(\| \Sigma^{-} \|_1 - \lambda_\Sigma, 0) \).

If the minimum eigenvalue of the unconstrained solution is greater than 0, then the soft-thresholded matrix is the correct solution to (4). However, if the minimum eigenvalue of the soft-thresholded matrix is below 0, we follow Bien and Tibshirani (2011) and perform the optimization using the alternating direction method of multipliers (Boyd et al. 2011), which is implemented in the R function \( \text{proxADMM} \) of the package \( \text{spcov} \) (Bien and Tibshirani 2012). Note that similarly to the estimation of the MF-VAR, we solve (4) for a decrementing log-scaled grid of \( \lambda_\Sigma \)-values.

If one wishes to incorporate the estimated nowcasting relations, the autoregressive parameters can be re-estimated by taking the error covariance matrix into account. This results in a type of generalized least squares estimator of \( \beta \) as given by

\[
\hat{\beta}^* = \arg \min_{\beta^*} \left\{ \frac{1}{2} \| y^* - X^* \beta^* \|_2^2 + \lambda_{\beta^*} P_{\text{Hier}}(\beta^*) \right\}
\]  

(5)

where \( y^* = \hat{\Sigma}^{-1/2} y \), \( X^* = \hat{\Sigma}^{-1/2} X \) and \( \hat{\Sigma} = \hat{\Sigma}^*_n \otimes I_N \).

4. Simulation Study

We assess the performance of the proposed hierarchical group estimator through a simulation study where we compare its performance to three alternatives, namely the ordinary least squares (OLS), the ridge, and the lasso.

The set-up of the simulation study is driven by our empirical application (Section 5), namely the small MF-VAR with \( K = 22 \) (one quarterly and seven monthly variables) and \( T = 125 \). The parameter matrix is set to reflect the obtained coefficients which result in a stable MF-VAR. To make a clear distinction between zero and nonzero coefficients, we set all coefficients smaller than 0.01 to zero. As a result, the coefficient matrix does not strictly follow the recency-based hierarchical structure anymore, thereby favoring the hierarchical estimator less compared to its benchmarks. Throughout the article, we standardize each series to have sample mean zero and variance one as commonly done in the regularization literature before applying a sparse method such that all coefficients have comparable sizes after standardization. To reduce the influence of initial conditions on the data-generation process (DGP), we burn in the first 300 observations for each simulation run.

We consider four simulation designs and run \( R = 500 \) simulations in each. The first design compares the estimators in terms of their estimation accuracy and variable selection performance of the autoregressive parameter vector. In the second design, we analyze how well the proposed regularization method can detect the nowcasting relations between the low- and high-frequency variables in \( \hat{\Sigma}^*_n \). The third design compares the point forecasts between the hierarchical estimator and its GLS-type version. The fourth design investigates the performance of the proposed estimator for DGPs with varying degree of sparsity.
4.1. Autoregressive Estimation Accuracy and Variable Selection

We take the error covariance matrix to be the identity matrix and compare estimation accuracy of the autoregressive parameter vector by calculating the mean squared error

\[ \text{MSE} = \frac{1}{R} \sum_{r=1}^{R} \frac{1}{K^2} \sum_{k=1}^{K^2} (\hat{\beta}_k - \hat{\beta}_k^{(r)})^2, \]

where \( \hat{\beta}_k^{(r)} \) refers to the \( k \)th element of the estimated parameter vector in simulation run \( r \). To investigate variable selection performance, we use the false positive rate (FPR), the false negative rate (FNR) and Matthews correlation coefficient (MCC):

\[ \text{FPR} = \frac{1}{R} \sum_{r=1}^{R} \frac{FP}{\#(k: \beta_k \neq 0)} \]

\[ \text{FNR} = \frac{1}{R} \sum_{r=1}^{R} \frac{FN}{\#(k: \beta_k = 0)} \]

\[ \text{MCC} = \frac{1}{R} \sum_{r=1}^{R} \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \]

where \( TP \) (and \( TN \)) are the number of regression coefficients that are estimated as nonzero (0) and are also truly nonzero (zero) in the model and \( FP \) (and \( FN \)) are the number of regression coefficients that are estimated as zero (nonzero), but are truly nonzero (0) in the model. Both FPR and FNR should be as small as possible. The MCC balances the two measures and is in essence a correlation coefficient between the true and estimated binary classifications. It returns a value between -1 and +1 with +1 representing a perfect prediction, 0 no better than random prediction and -1 complete discrepancy between prediction and observation. For the regularization methods (i.e., the proposed hierarchical estimator, lasso and ridge), we each time use a grid of 10 tuning parameters and select the one that minimizes the MSE between the estimated and true parameter vector.

Results. We first focus on estimation accuracy, see the left panel in Figure 3. The hierarchical estimator generates the lowest estimation errors. It significantly outperforms all others in terms of MSE as confirmed by paired sample \( t \)-tests at 5% significance level. OLS suffers as it is an unregularized estimator and thus cannot impose the necessary sparsity on the parameter vector; similarly for ridge which can only perform shrinkage but not variable selection. The hierarchical estimator performs slightly better than lasso in terms of MSE, but the difference is less profound since lasso can also handle sparsity.

Second, we compare the variable selection performance of the lasso and hierarchical estimator, see the middle and right panel of Figure 3, respectively. We plot the average FPR, FNR and MCC for different values of the tuning parameter \( \lambda_\beta \). The variable selection performance of the hierarchical estimator is in line with its good performance in terms of estimation accuracy. The maximum MCC lies at roughly 0.5, in comparison to the maximum MCC of lasso which only reaches 0.23. The larger FPR of lasso indicates that its estimate is overly sparse, thereby missing important variables in the model. The larger FNR of the hierarchical estimator in comparison to lasso can be explained by the fact that the DGP does not favor the recency-based structure of the hierarchical estimator. Recall that we have set several small coefficients to zero. Thus, it is possible that within one parameter group we have large coefficients with low priority and zero coefficients with higher priority. Alternatively, we can have hierarchical groups (all coefficients having the same priority) in which some coefficients are zero and some are large. To estimate and capture those important coefficients, our hierarchical estimator estimates some true zero coefficients as nonzero which automatically increases its FNR. Lastly, the gray area in the figures indicates the 2.5% and 97.5% quantiles of the selected position in the tuning parameter grid across the simulation runs. It illustrates that the maximum MCC lies within the gray area of the hierarchical estimator but does not for lasso.

Finally, we repeat the simulation study using the hierarchical estimator with equal weights; see Figure 8 in Appendix B.1 supplementary materials. While the equal weights version generates lower estimation errors, the original weights version performs better in terms of sparsity recognition (MCC). The hierarchical estimator with equal weights shrinks less aggressively, thereby having a higher FNR as many zero coefficients are estimated as small nonzeros.

4.2. Nowcasting Relations

We evaluate the detection of the nowcasting relations. To this end, we set the error covariance matrix to the regularized error covariance matrix estimated in the application Section 5.2. Coefficients smaller than 0.03 are set to zero to again ensure a clear distinction between the zeros and nonzeros. We first estimate the model using the four different estimators, calculate the resulting residual covariance matrix and then compute its regularized version through optimization problem (4). In line with our empirical application, we are mainly concerned with the sparsity pattern of the first row/column, namely the one corresponding to the low-frequency variable. Precisely, we investigate its variable selection performance using MCC, FPR, and FNR. We estimate the MF-VAR and covariance matrix of the corresponding residuals for a two-dimensional \((10 \times 10 \lambda_\Sigma)\) grid of tuning parameters. We select the tuning parameter couple that maximizes the MCC of the regularized covariance matrix in each simulation run. Table 1 contains the results.

Results. The performance across estimators is very similar, but the hierarchical estimator and lasso do perform best. Their MCCs lie at roughly 0.78. Their FPRs are slightly higher than their FNR which implies that the nowcasting relations tend to be estimated too sparsely. On the other hand, the low FNR suggests that in general we do not select variables which do not nowcast the low-frequency variable. Lastly, all estimators perform comparably across the tuning parameter grid for \( \lambda_\Sigma \), but the variability around the selected \( \lambda_\Sigma \) is higher for OLS and ridge than for lasso and the hierarchical estimator.
4.3. Forecast Comparison

We assess whether the best possible forecasting performance of the hierarchical estimator can be improved with its GLS version that incorporates the best nowcasting relations.

The error covariance matrix is set to the same matrix as in Section 4.2. We generate time series of length $T = 105$ (as in the forecast of Section 5.3), fit the models to the first $T - 1$ observations and use the last observation to compute the one-step-ahead mean squared forecast error for each series. In line with the empirical application, we focus on forecast accuracy of the first series, which represents the low-frequency variable, and select the tuning parameter $\lambda_\beta$ that minimizes its squared forecast error. For the selected model, we then estimate its regularized covariance matrix using $10 \lambda_\Sigma$ values and choose the one that maximizes the MCC. Finally, with the selected $\hat{\Sigma}_u$, we re-estimate the model according to Equation (5) to compare the forecast performance of the MF-VAR when the additional restrictions on the error covariance matrix are accounted for.

Results. The one-step-ahead MSFE for the first series of the hierarchical unrestricted estimator and its GLS version are 0.5508 and 0.5810, respectively. The former significantly outperforms the latter, as confirmed with a paired sample $t$-test at 1% significance level. The addition of the nowcasting relation may not improve the forecast because the values of the covariances in the covariance matrix of the DGP, particularly of the first row/column, are relatively small in absolute terms. Moreover, it is important to point out that the running time for the estimation of the restricted version is substantially larger than for the unrestricted version. Thus, even if the covariance matrix would be denser, there is a clear tradeoff between forecast accuracy and computational efficiency one needs to make.

4.4. Varying Degrees of Sparsity

In light of the recent debate on the validity of sparse methods for macroeconomic data (Giannone, Lenza, and Primiceri 2021), we revisit simulation studies 1 and 2 by varying the degree of sparsity of the DGP. Full details on the simulation designs are given in Appendix B.2 supplementary materials.

First, we vary the degree of sparsity of the autoregressive parameters in simulation study 1. Figure 9, supplementary materials demonstrates that the MSE of the hierarchical estimator stays relatively constant when the degree of sparsity decreases. Only for the fully dense DGP, we observe a small increase in MSE. The hierarchical estimator with equal weights does not suffer from this. Hence, if one expects a dense DGP or variable selection is not the main priority, the hierarchical estimator with equal weight can be used. On the other hand, if one expects a sparse DGP or seeks parsimonious models to facilitate interpretation, more aggressive shrinkage can be enforced via the hierarchical estimator with proposed weights.

Next, we vary the degree of sparsity in the first row/column of the error covariance matrix of simulation study 2. Here, denser settings do have an effect on the detection of the nowcasting relations. Table 4, supplementary materials shows that the MCC decreases mainly due to an increase in the FPR, meaning that the nowcasting relations are estimated too sparsely.

5. Macroeconomic Application

We investigate a high-dimensional MF-VAR for the U.S. economy. We use data from 1987 Q3 until 2018 Q4 ($T = 126$) on various aspects of the economy; among others output, income, prices and employment, see Table 5 in Appendix C.1, supplementary materials for an overview. The quarterly and monthly series are directly taken from the FRED-QD and FRED-MD datasets which are available at the Federal Reserve Bank of St. Louis FRED database (see McCracken and Ng 2016, 2020 for more details). The weekly time series are additionally retrieved
from the FRED database. The FRED-MD and -QD datasets contain transformation codes to make the data approximately stationary (see column “T-code” in Table 5, supplementary materials) which we apply to all series, thereby facilitating replicability and comparison with related research.

To evaluate the influence of additional variables and higher-frequency components, we estimate three MF-VAR models: The small MF-VAR ($K = 22$) consists of the quarterly variable at interest, real GDP (GDPC1), and seven monthly variables focusing on (industrial) production, employment and inflation. The medium MF-VAR ($K = 55$) contains the small group and 11 additional monthly variables containing further information on different aspects of the economy including financial variables. The large MF-VAR incorporates the variables of the medium group and replaces four monthly variables ($CLAIMSx, M2SL, FEDFUNDS, S&P 500$) with their equivalent weekly series. See columns “$K = 22$,” “$K = 55$,” and “$K = 91$” of Table 5, supplementary materials. To ensure that $n_T = 12$ in the large MF-VAR, we consider all months with more than four weekly observations and disregard the excessive weeks at the beginning of the corresponding month, as in Götz, Hecq, and Smeekes (2016).

We aim to investigate several aspects. First, we focus on the autoregressive parameter estimates of the hierarchical estimator to investigate the predictive Granger causality relations between the series through a network analysis. Second, we concentrate on nowcasting and investigate whether some high-frequency monthly and/or weekly economic series nowcast quarterly U.S. GDP growth and thus can deliver a coincident indicator of GDP growth. Third, we perform a sensitivity analysis of these results. Finally, we perform an out-of-sample expanding window nowcasting exercise including recent data on the Covid-19 pandemic. For ease of the discussion of the results, we follow the variable classification of McCracken and Ng (2016) which can be found in Table 6 in Appendix C.1 supplementary materials.

### 5.1. Autoregressive Effects

We first investigate predictive Granger causality relations. To that end, we estimate the MF-VAR using the hierarchical estimator with proposed weights and recency-based priority structure (seasonally adjusted data are used) for all parameter groups and a grid of 10 tuning parameters $\lambda_P$. The tuning parameter is selected using rolling window time series cross-validation with window size $T_1$. For each rolling window whose in-sample period ends at time $t = T_1, \ldots, T - 1$, we first standardize each time series to have sample mean zero and variance one using the most recent $T_1$ observations, thereby taking possible time variations in the first and second moment of the data into account (see, e.g., the Great Moderation Campbell 2007; Stock and Watson 2007). Given the evaluation period $[T_1, T]$, we use the one-step-ahead mean squared forecast error as cross-validation score. $T_1 = 105$, leaving us with 20 observations for tuning parameter selection. We first discuss the results for the small MF-VAR, then summarize the findings for the medium and large MF-VARs.

**Small MF-VAR.** Figure 4(a) depicts the estimated autoregressive coefficient matrix of the hierarchical estimator; 197 of the 484 coefficients (roughly 41%) are estimated as nonzero as indicated by the colored cells. Figure 4(b) visualizes the same results through a directed network. The vertices represent the variables, the edges the nonzero autoregressive coefficients. The edge’s width is proportional to the absolute value of the estimate. The colors of the vertices and their outgoing edges indicate to which macroeconomic group in McCracken and Ng (2016) the variable and its outgoing effect belong. We summarize the linkages between the macroeconomic categories in Table 2. The columns reflect a macroeconomic category’s out-degree (influence), the rows its in-degree (responsiveness).

We first concentrate on our main variable of interest GDPC1. Eight variables contribute toward its prediction (see first row of Figure 4(a) or incoming edges in panel (b)). Most influential is the macroeconomic group Output and Income as month three and two are included for both variables ($INDPRO$ and $CUMFNS$). The three variables related to employment ($UNRATE, PAYEMS, and USFIRE$) and $HOUST$ are each selected with their third monthly component. Besides, 14 economic variables are influenced by lagged GDPC1 (see first column of Figure 4(a) or outgoing edges of its vertex in panel (b)). The Employment category has the most incoming edges, nevertheless, the most prominent (thicker) edges indicate that GDPC1 contributes the most to month one of $INDPRO$ and $CUMFNS$ (i.e., the Output and Income group).

Next, we focus on the linkages between macroeconomic groups containing the higher-frequency variables in Table 2. The categories Output and Income and Employment are strongly connected with each other and have the most outgoing edges, even when taking into account that these groups contain more than one monthly variable. Output and Income also contributes to the prediction of Prices and Employment toward Prices as well as Housing. While Prices play a substantial role for Output and Income and Employment, Housing is more relevant for Employment than it is for Output and Income.

Finally, we inspect the linkages within each macroeconomic group (diagonal in Table 2). Employment and Housing display the highest within-group interaction. Zooming in on $PAYEMS$ or $USFIRE$ in Figure 4, we see that their own lagged effects are the strongest, despite them belonging to a group containing more than one high-frequency variable. Within macroeconomic groups, a series’ own lags thus, remain the most informative.

When focusing again on the network, we find that the third months of the variables have the most outgoing edges, whereas the first months have the most incoming edges. This is a logical consequence of the recency-based priority structure that we imposed.

**Medium and Large MF-VAR.** To evaluate the influence of dimensionality, we now compare our results to the two larger MF-VARs. Their networks are given in Figures 10 and 11 in Appendix C.2, supplementary materials, whereas Tables 7 and 8 in Appendix C.1, supplementary materials present the linkages between the macroeconomic categories, respectively. The medium MF-VAR has 889 out of 3025 nonzero coefficients (30%), the large MF-VAR 1199 out of 8281 (only 15%). The
Looking at the influencer of GDPC1, we find that the majority of the variables selected for $K = 22$ are also selected for the medium and large MF-VAR. In addition, the monetary (M2SL) and financial variables ($FEDFUNDS$ and $S&P500$) as well as two monthly variables related to sales (CMRMTSPLx and RETAILx) deem to be relevant. Apart from month three of $OILSPRICEx$ and $PCEPI$ in the medium system, no variables related to prices are selected for the prediction of GDP growth. This coincides with the results for the small MF-VAR where CPIAUCSL also only had minor influence. Similarly, the variables that GDPC1 influences in the medium and large system overlap with the selected ones in the small system. Particularly, GDPC1 strongly contributes to the prediction of the variables in the macroeconomic groups Output and Income, illustrated by its thick outgoing edges, and Employment, indicated by the most incoming edges.

Next we focus on the linkages between the macroeconomic groups. For the medium MF-VAR, Table 7, supplementary materials underlines that Output and Income, Sales, Employment and Prices are highly interconnected. Moreover, the group Interest Rates influences the groups Employment and Prices. The latter one is not surprising as $FEDFUNDS$ is usually set to control inflation, hence, one can expect changes in the previous quarter to aid in predicting inflation, measured by changes in prices. A similar argument supports the interconnection between Prices and Money. When looking at the diagonal entries of Table 7, supplementary materials, we notice that similarly to the small MF-VAR, the groups Employment, Housing and Output and Income have a high within-groups interaction. In contrast to the small system, the within-group linkages among Prices highly increases, which is likely due to the addition of four price variables. The introduction of the weekly variables in the large system does not change the relations among the macroeconomic categories (see Table 8, supplementary materials).

### Table 2. Small $(K = 22)$ MF-VAR: Linkages between macroeconomic group.

| To/from | GDP | Output & Income | Housing | Employment | Prices | In-degree |
|---------|-----|----------------|---------|------------|--------|-----------|
| GDP     | 0   | 4              | 1       | 3          | 0      | 8         |
| Output & Income | 4 | 22             | 1       | 16         | 4      | 49        |
| Housing | 1   | 1              | 8       | 6          | 0      | 16        |
| Employment | 7 | 26             | 5       | 52         | 7      | 97        |
| Prices  | 2   | 10             | 1       | 9          | 5      | 27        |
| Out-degree | 14 | 63             | 18      | 86         | 16     | 197       |

**NOTE:** Entry $(i,j)$ indicates the number of edges from group $j$ to group $i$.

A significant increase in dimensionality thus, induces a higher degree of selectiveness.

5.2. Coincident Indicators

We investigate whether some high-frequency monthly and/or weekly economic series nowcast quarterly GDP growth and thus, can deliver a coincident indicator. More specifically, we analyze how the performance of the indicator is influenced through (i) the formalization of “sparse” nowcasting relations in the MF-VAR in comparison to a naive approach of constructing a coincident indicator based on the first principal component of all high-frequency variables and (ii) the number of high-frequency components included in the model.

To construct the coincident indicator, we first estimate the MF-VAR using the hierarchical estimator as in Section 5.1. Second, we compute the regularized $\hat{\Sigma}_u$ from the MF-VAR residuals. We then select the high-frequency variables having a
nonzero covariance element in the GDP column and construct
the first principal component of the corresponding correlation
matrix.\(^5\) We estimate the MF-VAR and corresponding corre-
lation matrix of the MF-VAR residuals for a two-dimensional
\((10 \times 10)\) grid of tuning parameters \(\lambda_\beta\) and \(\lambda_\Sigma\). We report the
results for the tuning parameter couple that maximizes the cor-
relation between the most parsimonious coincident indicator
and GDP growth.\(^6\)

**Small MF-VAR.** Figure 5(a) plots the U.S. GDP growth
against the coincident indicator for the small MF-VAR that
maximizes the correlation to a value of 0.7429. The indicator
tracks the movements of the GDP growth fairly well. The
selected monthly variables from which the first principal
component has been constructed are listed on the x-axis of
Figure 5(b). Thus, 16 out of the 21 high-frequency variables
are selected, since their corresponding covariance with GDP
growth is estimated as nonzero. Each variable is included with
at least one monthly component. The variables related to
Housing (\textit{HOUST}) and Output and Income (\textit{INDPRO} and
\textit{CUMFNS}) follow the fluctuations of GDP growth particularly
well as all three months are selected. The variables measuring
Employment (\textit{UNRATE}, \textit{PAYEMS}, and \textit{USFIRE}) have varying
levels of contribution. Lastly, the second month of \textit{CPIAUCSL},
which measures price changes, is also chosen.

For comparison, a coincident indicator constructed from
all high-frequency variables would result in a correlation of
0.7216. Capturing the nowcasting relations in a MF-VAR where
the lagged and instantaneous dynamics are separated, thus,
increases the correlation by two percentage points. The advan-
tage of carefully selecting variables before conducting a prin-
cipal component analysis is a result consistent with Bai and
Ng (2008). Furthermore, the correlations with GDP growth are
fairly stable across the two-dimensional grid of tuning param-
eters. In roughly half of the cases, the correlation is at least as high
as the one computed from all high-frequency variables. Our
findings are thus, rather robust to a different choice of selection
criterion for the tuning parameters.\(^7\)

\textit{GDPCI} is typically released with a relatively long delay (usual-
ly one month after the quarter for the first release), whereas
the monthly variables are released in blocks at different dates
throughout the following month.\(^8\) For instance, the previous
month’s \textit{UNRATE}, \textit{PAYEMS}, and \textit{USFIRE} are typically released
on the first Friday of the following month, whereas the remain-
ing variables are only available around the middle of the month.
We follow the release scheme of Giannone, Reichlin, and Small
(2008) to study the marginal impact of these data releases on
the construction of our coincident indicator. We do not take data
revisions into account. Hence, our vintages are “pseudo” real-
time vintages rather than true real-time vintages.

Figure 5(b) illustrates the achieved correlation between GDP
growth and the coincident indicator, updated according to the
release dates of the selected variables. As such, the first coinci-
dent indicator only uses month one for \textit{UNRATE} and \textit{PAYEMS},
whereas the last one is constructed from all selected variables
and its correlation with GDP growth is indicated by the hori-
zontal line. The figure shows that intra-quarter information
matters. The second month releases have a large impact on the
accuracy of the coincident indicator. Particularly, the addition
of the variables \textit{INDPRO} and \textit{CUMFNS} raises the correlation
significantly. In fact, it is possible to construct an almost equally
reliable indicator with the data from month one and two com-
pared to one constructed with the data from the entire quarter.
Hence, one can build a reliable coincident indicator roughly 1.5
months before the first release of GDP, thereby accounting for

\(^5\)Alternatively, we constructed a coincident indicator by Partial Least Squares.
Results are comparable.

\(^6\)We cannot select according to MCC so we use the maximal correlation as
proxy.

\(^7\)Time series cross-validation (with one-step-ahead MSFE as CV-score) results
in a similar correlation.

\(^8\)In our case, the variables belonging to the same macroeconomic category
(McCracken and Ng 2020) are published on the same day of the month.
the publishing lag of approximately half a month for the monthly series.

Medium and Large MF-VAR. Table 3 summarizes the correlations between GDP growth and the coincident indicators constructed from the different MF-VAR systems. The correlation for the large system (0.7928) is slightly higher than for the medium MF-VAR (0.7715) and both outperform the small system (0.7429).

Table 3. Correlation between U.S. GDP growth and the coincident indicators for the small (K = 22), medium (K = 55) and large (K = 91) MF-VAR groups.

| Type of coincident indicator     | K = 22 | K = 55 | K = 91 |
|----------------------------------|--------|--------|--------|
| Nowcasting relations M1          | 0.6150 | 0.5128 | 0.4887 |
| Nowcasting relations M1 + M2     | 0.7288 | 0.7495 | 0.6680 |
| All nowcasting relations         | 0.7429 | 0.7715 | 0.7928 |
| All variables                    | 0.7216 | 0.7564 | 0.7557 |

Full details on the medium and large MF-VAR are available in Appendix C.2: Figures 12(a) and 13(a) show that both indicators behave similarly and can better pick up the drop in GDP growth during the financial crisis than the coincident indicator of the small system. Many of the variables selected for the coincident indicator of the small MF-VAR are also selected for the two larger MF-VARs (see Figures 12(b) and 13(b), supplementary materials). There is a clear focus on variables related to (industrial) output, sales and employment whereas variables measuring price changes have a smaller influence. For the large MF-VAR 28 weekly variables are selected, thereby making it valuable to incorporate higher-frequency variables. But the mere addition of variables does not lead to a larger correlation, emphasizing that selection is important. Table 3 illustrates that the advantage of selecting the nowcasting relations in the MF-VARs persists as the correlation achieved from the first principal component with all variables is lower. Finally, the coincident indicator constructed from the selected variables from month one and two for the medium system performs comparable to the one constructed from all selected variables, in line with our finding for the small MF-VAR.

5.3. Sensitivity Analyses

We investigate the sensitivity of our results regarding (i) the choice of weights for the hierarchical estimator, (ii) the maximal lag length of the MF-VAR, (iii) the inclusion of daily high-frequency data, and (iv) forecast comparisons.

Equal Weights. We use the hierarchical estimator with equal weights and present the autoregressive linkages between the macroeconomic groups for the MF-VARs in Tables 9–11 of Appendix C.3, supplementary materials. Using equal weights corresponds to weaker penalization, hence the networks become denser, but the main conclusions regarding the interdependencies remain unchanged. Next, we revisit the results on the coincident indicators in Table 12, supplementary materials. For the small MF-VAR, the results are identical. For the medium and the large MF-VAR, respectively, more and fewer high-frequency variables are selected for the coincident indicator which only affects its correlation with GDP growth in the first two months.

Lag Length. We re-estimate the small MF-VAR with the hierarchical estimator using a maximal lag ℓ = 1, 2, or 4. Practically all coefficients from the second lag onwards (more than 98%) are zeroed out, and more first-order coefficients are shrunked toward zero. Both the Bayesian and Akaike information criteria indicate one to be the “optimal” lag length.

Daily Data. To investigate the behavior of the hierarchical estimator in presence of daily high-frequency components, we replace the weekly financial variables (FEDFUNDS, S&P500) in the large MF-VAR with 60 daily component series for each. This leads to an “ultra-large” MF-VAR with K = 187 component series instead of the large MF-VAR with K = 91 series. Results are detailed in Appendix C.4, supplementary materials. Only 5% of the autoregressive coefficients are estimated as nonzero, supporting our previous finding that an increase in dimensionality induces a higher degree of selectiveness. The influences of GDPCI stay roughly the same as in large MF-VAR. Interest Rate and Stock Price are selected with only two and three components, respectively. The coincident indicator benefits from the usage of the more timely daily instead of weekly financial data only in the beginning of the quarter (for the first two months), but not when using the daily data from the whole quarter.

Forecast Comparison. We investigate the out-of-sample forecast performance of the hierarchical estimator, as detailed in Appendix C.5, supplementary materials. We compare its performance to the random walk (RW) and AR(1) model as two popular, simple univariate benchmarks; a traditional quarterly VAR(1), estimated by OLS, with all higher-frequency variables aggregated to the quarterly level; and the three alternative estimators from Section 4 (OLS, ridge, and lasso). Table 15, supplementary materials provides the one-step ahead MSFE for GDPCI across all MF-VARs.

The hierarchical estimator generally outperforms the AR, RW, quarterly VAR, and MF-VAR OLS, thereby indicating that higher-frequency variables as well as variable selection help to predict GDP growth. The large MF-VAR with hierarchical estimator attains the lowest MSFE. Both the hierarchical estimator and lasso are included in the Model Confidence Set (Hansen, Lunde, and Nason 2011) for two out of the three MF-VARs. The quarterly VAR is not included; information contained in high-frequency series is thus, beneficial in our application for forecasting GDP growth.

5.4. Nowcasting GDP Growth Pre and Post Covid-19

The Covid-19 pandemic has caused a dramatic drop in economic activity worldwide including the U.S. We end by assessing the performance of our coincident indicators in an out-of-sample expanding window nowcasting exercise with evaluation period 2019 Q1 until 2021 Q2, thereby covering the recent pandemic. For each current quarter in the expanding window approach, we estimate the MF-VAR with the hierarchical estimator (using data until the previous quarter), select the high-frequency variables for the coincident indicator from the residual covariance matrix as in Section 5.2 and obtain their loadings on the first principal component. The nowcasts are then calculated by multiplying these loadings with the out-of-sample high-frequency data of the current quarter.
The mean squared nowcast errors between GDP growth and the coincident indicators constructed from the three MF-VARs are depicted on the last line of the table in Figure 6(a). The coincident indicator of the small MF-VAR attains the best accuracy, the error more than doubles for the larger MF-VARs. In addition, we track the nowcasting accuracy as the quarter progresses, thereby constructing the coincident indicator only from the selected high-frequency variables from the first month (first line), or the first two months (second line in the table). For the best performing small MF-VAR, the nowcasts gradually improve as new data gets released along the quarter, with the largest improvement occurring when adding high-frequency data from the second month to the first.

Next, we discuss the stability through time of the selected variables from which the coincident indicator of the small MF-VAR has been constructed. Panel(a) of Figure 7 shows the selected high-frequency variables for each endpoint of the expanding window. Panel(b) summarizes them according to the macroeconomic group they belong to. Before the pandemic hits the U.S. economy in 2020 Q1, the selection of the high-frequency variables was very stable. Almost all variables are selected with all monthly components, the exceptions are UNRATE, USFIRE, and CPIAUCSL. At the onset of the pandemic, a clear break in variable selection occurs: the coincident indicator is much more sparsely constructed thereby only using data on the Output & Income group from the first month.

Finally, we compare the performance of the coincident indicator of the small MF-VAR with two state-of-the-art benchmarks, one Bayesian and one survey. We use the “blocking” Bayesian MF-VAR (B-BVAR) by Cimadomo et al. (2021) since it also relies on the stacked MF-VAR approach.9 Second, we use the GDP nowcasts of the Survey of Professional Forecasters (SPF). Figure 6(b) plots GDP growth against the three different nowcasts for 2019 Q1 to 2020 Q2.10 Overall, all methods track the same fluctuations in the business cycle. Before the pandemic, all nowcasts are practically the same and very close to actual GDP growth. When the pandemic hits the economy, some differences can be observed. First, the SPF does not pick up the slight drop in GDP growth in Q1 2020. This can be explained by the fact that the SPF nowcasts are released during the second month of a quarter, while most countries only went into lockdown by March. Second, the drop in economic

9We would like to thank Michele Lenza for providing the code for the B-BVAR method which we applied (with default settings for stationary data) to the small and medium MF-VAR. In line with our findings, their small MF-VAR was best performing, hence, the results for their medium MF-VAR are omitted but available upon request.

10Please note that all insights obtained from Figure 6(b) are still subject to GDP being revised.
activity in 2020 Q2 is very pronounced for the SPF, much less so for our coincident indicator and the B-BVAR nowcast. The recovery in Q3 thereafter is best tracked by the B-BVAR. Our coincident indicator and the B-BVAR nowcasts align again from Q4 onwards.

6. Conclusion

We introduce a convex regularization method tailored toward the hierarchically ordered structure of mixed-frequency VARs. To this end, we use a group lasso with nested groups which permits various forms of hierarchical sparsity patterns that allows one to discriminate between recent and obsolete information. Our simulation study shows that the proposed regularizer can improve estimation and variable selection performance. Furthermore, nowcasting relations can be detected from the sparsity pattern of the covariance matrix of the MF-VAR errors. Those high-frequency variables that nowcast the low-frequency variables, as evident from their nonzero contemporaneous link, can deliver a coincident indicator of the low-frequency variable. Constructing coincident indicators from a group of selected variables rather than all permits policy makers to get an earlier grasp of the state of the economy, as can be seen from our economic application on U.S. GDP growth.

The proposed MF-VAR method is quite flexible and can be extended in various ways. First, regularization via restrictions other than sparsity can be explored. Temporal aggregation restrictions, for instance, can be imposed in the MF-VAR by exploiting fusion penalties (e.g., Yan and Bien 2021) that permits various forms of hierarchical sparsity patterns that allows one to discriminate between recent and obsolete information. Our simulation study shows that the proposed regularizer can improve estimation and variable selection performance. Furthermore, nowcasting relations can be detected from the sparsity pattern of the covariance matrix of the MF-VAR errors. Those high-frequency variables that nowcast the low-frequency variables, as evident from their nonzero contemporaneous link, can deliver a coincident indicator of the low-frequency variable. Constructing coincident indicators from a group of selected variables rather than all permits policy makers to get an earlier grasp of the state of the economy, as can be seen from our economic application on U.S. GDP growth.

The proposed MF-VAR method is quite flexible and can be extended in various ways. First, regularization via restrictions other than sparsity can be explored. Temporal aggregation restrictions, for instance, can be imposed in the MF-VAR by exploiting fusion penalties (e.g., Yan and Bien 2021) that encourages similarity across certain coefficients. For monthly stock data it could, for instance, be interesting to encourage the effects of all months on the quarterly variable to be similar, thereby implicitly aggregating the monthly variable to the quarterly level. Second, an interesting path for future research concerns the extension of our method to enable structural analysis as recently done in a Bayesian set-up by, for example, Cimadomo et al. (2021) who use generalized impulse responses to track transmission mechanisms of low-frequency shocks hitting the U.S. economy, or Paccagnini and Parla (2021) who use orthogonalized impulse responses to identify the impact of high-frequency shocks thereby revealing a temporal aggregation bias when adopting single low-frequency models instead of mixed-frequency ones. Lastly, while we consider MF-VAR for stationary data, a natural next step would be to allow for nonstationarity by building on the lag-summation idea of Toda and Yamamoto (1995) as done in Götzh and Hecq (2019) for low-dimensional mixed MF-VARs or Hecq, Margariella, and Smeekes (2021b) for high-dimensional VARs.

Supplemental Materials

R-code: Supplemental files for this article include R-code to reproduce all results. Please consult the file README in the zip file for more details. (Code.zip, zip archive)

Appendix: The Appendix contains implementation details on the algorithm, and additional results on the simulations and empirical application. (Appendix.pdf)

Acknowledgments

We thank the editor, associate editor and reviewers for their thorough review and highly appreciate their constructive comments which substantially improved the quality of the article.

Funding

Wilms gratefully acknowledges funding from the European Union’s Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant no. 832671).

References

Andreu, E., Gagliardini, P., Ghysels, E., and Rubin, M. (2019), “Inference in Group Factor Models with an Application to Mixed-Frequency Data,” Econometrica, 87, 1267–1305. [1076]

Babii, A., Ghysels, E., and Striaukas, J. (2021), “Machine Learning Time Series Regressions with an Application to Nowcasting,” Journal of Business & Economic Statistics, 1–23. doi:10.1080/07350015.2021.1899933. [1077,1081]

Bai, J., and Ng, S. (2008), “Forecasting Economic Time Series Using Targeted Predictors,” Journal of Econometrics, 146, 304–317. [1086]

Barigozzi, M., and Brownlees, C. (2019), “Nets: Network Estimation for Time Series,” Journal of Applied Econometrics, 34, 347–364. [1076]

Basu, S., and Michailidis, G. (2015), “Regularized Estimation in Sparse High-Dimensional Time Series Models,” The Annals of Statistics, 43, 1535–1567. [1076]

Basu, S., Shojaie, A., and Michailidis, G. (2015), “Network Granger Causality with Inherent Grouping Structure,” The Journal of Machine Learning Research, 16, 417–453. [1076]

Bien, J., Bunea, F., and Xiao, L. (2016), “Convex Banding of the Covariance Matrix,” Journal of the American Statistical Association, 11, 834–845. [1080]

Bien, J., and Tibshirani, R. (2012), spcov: Sparse Estimation of a Covariance Matrix. R package version 1.01. [1081]

Bien, J., and Tibshirani, R. J. (2011), “Sparse Estimation of a Covariance Matrix,” Biometrika, 98, 807–820. [1081]

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011), “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” Foundations and Trends in Machine Learning, 3, 1–122. [1081]

Brave, S. A., Butters, R. A., and Justiniano, A. (2019), “Forecasting Economic Activity with Mixed Frequency BVARs,” International Journal of Forecasting, 35, 1692–1707. [1076]

Callot, L. A., Kock, A. B., and Medeiros, M. C. (2017), “Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice,” Journal of Applied Econometrics, 32, 140–158. [1076]

Campbell, S. D. (2007), “Macroeconomic Volatility, Predictability, and Uncertainty in the Great Moderation: Evidence from the Survey of Professional Forecasters,” Journal of Business & Economic Statistics, 25, 191–200. [1084]

Chevillon, G., Hecq, A., and Laurent, S. (2018), “Generating Univariate Fractional Integration Within a Large VAR(1),” Journal of Econometrics, 204, 54–65. [1077]

Cimadomo, J., Giannone, D., Lenza, M., Monti, F., and Sokol, A. (2021), “Nowcasting with Large Bayesian Vector Autoregressions,” Journal of Econometrics, forthcoming. [1076,1077,1079,1088,1089]

Caballada, G., Hecq, A., and Palm, F. C. (2009), “Studying Co-movements in Large Multivariate Data Prior to Multivariate Modeling,” Journal of Econometrics, 148, 25–35. [1077]

Davis, R. A., Zang, P., and Zheng, T. (2016), “Sparse Vector Autoregressive Modeling,” Journal of Computational and Graphical Statistics, 25, 1077–1096. [1076]

Derimer, M., Diebold, F. X., Liu, L., and Yilmaz, K. (2018), “Estimating Global Bank Network Connectedness,” Journal of Applied Econometrics, 33, 1–15. [1076,1078]
Eurostat: Statistics Explained (2014), Glossary: Nowcasting. Available at https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary: Nowcasting. [1077]

Foroni, C., and Marcellino, M. (2014), “A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates,” International Journal of Forecasting, 30, 554–568. [1076]

Gefang, D., Koop, G., and Poon, A. (2020), “Computationally Efficient Inference in Large Bayesian Mixed Frequency VARs,” Economics Letters, 191, 109120. [1076]

Ghysels, E. (2016), “Macroeconomics and the Reality of Mixed Frequency Data,” Journal of Econometrics, 193, 294–314. [1076,1077]

Ghysels, E., Hill, J. B., and Moteki, K. (2016), “Testing for Granger Causality with Mixed Frequency Data,” Journal of Econometrics, 192, 207–230. [1077]

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004), “The MIDAS Touch: Mixed Data Sampling Regression Models,” CIRANO working papers, CIRANO. [1076]

Giannone, D., Lenza, M., and Primiceri, G. E. (2021), “Economic Predictions with Big Data: The Illusion of Sparsity,” Econometrica, forthcoming. [1083]

Giannone, D., Reichlin, L., and Small, D. (2008), “Nowcasting: The Real-Time Informational Content of Macroeconomic Data,” Journal of Monetary Economics, 55, 665–676. [1086]

Götz, T. B., and Hecq, A. (2014), “Nowcasting Causality in Mixed Frequency Vector Autoregressive Models,” Economics Letters, 122, 74–78. [1077]

Götz, T. B., Hecq, A., and Smeekes, S. (2016), “Testing for Granger Causality in Large Mixed-Frequency VARs,” Journal of Econometrics, 193, 418–432. [1076,1077,1084]

Götz, T. B., and Hecq, A. W. (2019), “Granger Causality Testing in Mixed-Frequency VARs with Possibly (co)integrated Processes,” Journal of Time Series Analysis, 40, 914–935. [1089]

Hansen, P. R., Lunde, A., and Nason, J. M. (2011), “The Model Confidence Set,” Econometrica, 79, 453–497. [1087]

Harris, A., Witten, D., and Simon, N. (2016), “Convex Modeling of Interactions with Strong Heredity,” Journal of Computational and Graphical Statistics, 25, 981–1004. [1080]

Hastie, T., Tibshirani, R., and Wainwright, M. (2015), Statistical Learning with Sparsity: The Lasso and Generalizations, Boca Raton, FL: Chapman and Hall/CRC Press. [1076]

Hecq, A., Margaritella, L., and Smeekes, S. (2021a), “Granger Causality Testing in High-Dimensional VARs: A Post-Double-Selection Procedure,” Journal of Financial Econometrics, forthcoming. nbab023. [1076]

Hecq, A., Margaritella, L., and Smeekes, S. (2021b), “Inference in Nonstationary High-dimensional VARs,” Technical report. [1089]

Hsu, N.-J., Hung, H.-L., and Chang, Y.-M. (2008), “Subset Selection for Vector Autoregressive Processes Using Lasso,” Computational Statistics & Data Analysis, 52, 3645–3657. [1076,1078]

Koebll, L., and Deistler, M. (2020), “A New Approach for Estimating VAR Systems in the Mixed-Frequency Case,” Statistical Papers, 61, 1203–1212. [1076]

Kuzin, V., Marcellino, M., and Schumacher, C. (2011), “MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area,” International Journal of Forecasting, 27, 529–542. [1076]

Lewis, D., Mertens, K., Stock, J. H., and Trivedi, M. (2021), “Measuring Real Activity Using a Weekly Economic Index,” Journal of Applied Econometrics, forthcoming. [1080]

Lou, Y., Bien, J., Caruana, R., and Gehrke, J. (2016), “Sparse Partially Linear Additive Models,” Journal of Computational and Graphical Statistics, 25, 1126–1140. [1080]

Lütkepohl, H. (2005), New Introduction to Multiple Time Series Analysis, Berlin: Springer. [1079]

Marcellino, M., and 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, 518–550. [1076]

McCracken, M., and Ng, S. (2016), “FRED-MD: A Monthly Database for Macroeconomic Research,” Journal of Business & Economic Statistics, 34, 574–589. [1083,1084]

McCracken, M., Owyang, M., and Sekhposyan, T. (2021), “Real-Time Forecasting and Scenario Analysis Using a Large Mixed-Frequency Bayesian VAR,” International Journal of Central Banking, 18, 327–367. [1076]

Mogliani, M., and Simon, A. (2021), “BAYESIAN MIDAS Penalized Regressions: Estimation, Selection, and Prediction,” Journal of Econometrics, 222, 833–860. [1077]

Nicholson, W. B., Wilms, I., Bien, J., and Matteson, D. S. (2020), “High Dimensional Forecasting via Interpretable Vector Autoregression,” Journal of Machine Learning Research, 21, 1–52. [1077,1080,1081]

Paccagnini, A., and Parla, F. (2021), “Identifying High-Frequency Shocks with Bayesian Mixed-Frequency VARs,” working paper No. 26/2021, CAMA. [1076,1089]

Rothman, A. J., Levina, E., and Zhu, J. (2009), “Generalized Thresholding of Large Covariance Matrices,” Journal of the American Statistical Association, 104, 177–186. [1081]

Schorfeide, F., and Song, D. (2015), “Real-Time Forecasting with a Mixed-Frequency VAR,” Journal of Business & Economic Statistics, 33, 366–380. [1076]

Smeekes, S., and Wijler, E. (2018), “Macroeconomic Forecasting Using Penalized Regression Methods,” International Journal of Forecasting, 34, 408–430. [1076]

Stock, J., and Watson, M. (2007), “Why has U.S. Inflation Become Harder to Forecast?,” Journal of Money, Credit and Banking, 39, 3–33. [1084]

Tibshirani, R. (1996), “Regression Shrinkage and Selection via the Lasso,” Journal of the Royal Statistical Society, Series B, 58, 267–288. [1078]

Toda, H. Y., and Yamamoto, T. (1995), “Statistical Inference in Vector Autoregressions with Possibly Integrated Processes,” Journal of Econometrics, 66, 225–250. [1089]

Tseng, P. (2008), “On Accelerated Proximal Gradient Methods for Convex-Concave Optimization,” submitted to SIAM Journal on Optimization, 1. [1081]

Yan, X., and Bien, J. (2021), “Rare Feature Selection in High Dimensions,” Journal of the American Statistical Association, 116, 887–900. [1089]

Zellner, A., and Palm, F. (1974), “Time Series Analysis and Simultaneous Equation Econometric Models,” Journal of Econometrics, 2, 17–54. [1077]

Zhao, P., Rocha, G., and Yu, B. (2009), “The Composite Absolute Penalties Family for Grouped and Hierarchical Variable Selection,” The Annals of Statistics, 37, 3468–3497. [1080,1081]