Multiple Base Frequency VARs*

Laurent Florin
KOF, ETH Zurich
and
Samad Sarferaz
KOF, ETH Zurich

April 30, 2025

Abstract

This paper develops a generalized mixed-frequency VAR model that efficiently handles data across multiple frequencies. Unlike traditional approaches that rely on a single base frequency, our model features lag lengths that are independent of the frequency gap and a parameter count that grows only linearly with it. We achieve this by decomposing the joint distribution of a mixed-frequency VAR into conditional distributions of bi-frequency VARs, which substantially reduces the parameter burden. The model is estimated via Bayesian MCMC methods within a state-space framework. In an empirical application using quarterly, monthly, and weekly data, our two base frequency VAR demonstrates improved latent state estimation precision and outperforms the conventional single base frequency VAR in real-time forecasting.

JEL Classification: C11,C32,C53
Keywords: Mixed-Frequency VAR Models, Bayesian State-Space, Latent State Estimation, Real-Time Macroeconomic Forecasting

*Authors' contact:...Python code...

1 Introduction

Time series data in economics and finance are observed at a wide range of frequencies, from low-frequency indicators such as quarterly GDP to high-frequency data such as intraday stock prices. Mixed-frequency vector autoregressive models offer a powerful solution for integrating data from various frequencies within a unified framework, addressing the limitations of traditional VAR approaches that require the aggregation of high-frequency data to a lower frequency. By leveraging data from different frequencies, mixed-frequency VARs can deliver more accurate forecasts, which are particularly relevant in contexts requiring timely responses, such as during the COVID-19 pandemic (e.g. Schorfheide and Song, 2020; Huber et al., 2023). Despite their advantages, addressing various frequencies simultaneously and bridging larger frequency gaps, such as those between daily and quarterly data, remains computationally challenging due to the high number of parameters required, as large frequency gaps directly set the minimum viable lag length and drive the overall parameter count.

This paper introduces a multiple base frequency VAR (MBF-VAR) model that accommodates series of various frequencies, leveraging each variable's highest observed frequency. Unlike traditional approaches that rely on a single base frequency (SBF), our methodology accommodates multiple base frequencies. This innovation has two distinct effects on the parameter count. First, it ensures that the lag length is independent of the overall frequency gap. Second, it results in the number of parameters increasing only linearly with that gap. The MBF-VAR allows for the estimation of unobserved variables at different frequencies and the inclusion of observed indicators across various frequencies within the same framework. Central to our approach is decomposing the joint distribution of a mixed-frequency VAR model with more than two frequencies into a complete set of conditional distributions of bi-frequency VARs by incorporating auxiliary variables. This decomposition significantly reduces the number of parameters required compared to conventional mixed-frequency VAR specifications. The model is formulated within a state-space framework and estimated using a Markov Chain Monte Carlo (MCMC) sampler for drawing from the posterior distribution.

To illustrate our approach, consider an MBF-VAR model that incorporates weekly, monthly, and quarterly series. In this example, we employ two bi-frequency VARs: one combining weekly and monthly data and another combining monthly and quarterly data. In the weekly–monthly configuration, the base frequency is set to weekly—meaning that latent variables are modeled at a weekly frequency—whereas in the monthly–quarterly configuration, the base frequency is monthly. In our estimation procedure, we then draw from the posterior of the weekly latent variables conditional on the monthly latent variables from the monthly-quarterly model. In this part of our sampling, the monthly latent variables in the weekly-monthly configuration are treated as observed. This strategy ultimately yields posterior simulations for both the weekly and monthly latent variables.

Various methods have been employed to estimate mixed-frequency VAR models.[1] These include observation-driven approaches (e.g., Ghysels, 2016),[2] as well as state-space approaches (e.g., Zadrozny, 1988; Mariano and Murasawa, 2010; Kuzin et al., 2011; Eraker et al., 2015; Schorfheide and Song, 2015; Zadrozny, 2016). To address the curse of dimensionality, these models often employ Bayesian shrinkage priors, whether in state-space frameworks (e.g., Schorfheide and Song, 2015; Brave et al., 2019; Chan et al., 2023) or in observation-driven settings (e.g., McCracken et al., 2021; Cimadomo et al., 2022; Daniele et al., 2024). More recently, SBF-VAR models featuring time-varying coefficients and stochastic volatilities have been proposed (e.g., Cimadomo and D'Agostino, 2016; Ankargren and Jonéus, 2019; Götz and Hauzenberger, 2021). A common requirement across all these models is the alignment of all variables to a single base frequency for estimation. In state-space models, this process involves estimating latent variables at the highest frequency, which remains a computationally demanding task even with Bayesian shrinkage priors.

We contribute to the aforementioned literature by allowing for multiple base frequencies in our analysis. Our approach utilizes all available variables at their original frequencies, rather than being confined to the highest frequency. Specifically, we decompose the joint density into several bi-frequency VARs, each with its own base frequency and set of indicators. Notably, this decomposition into bi-frequency VARs reduces the number of parameters

[1]See Foroni et al. (2013) for an overview of the SBF-VAR literature.

[2]These approaches partly build on mixed data sampling (MIDAS) regressions (e.g., Ghysels et al., 2004; Ghysels and Wright, 2009; Andreou et al., 2010)

that need to be estimated. A key distinction between the SBF-VAR and our MBF-VAR lies in how the total number of parameters scales with the number of frequencies κ. In the SBF-VAR, both the lag length and the number of variables increase with κ, leading to a parameter count on the order of κ^3, i.e., $O(\kappa^3)$. By contrast, the MBF-VAR is decomposed into a chain of bi-frequency VARs, each with a fixed number of parameters and lag length, causing its total parameter count to scale linearly with κ, i.e., $O(\kappa)$. This difference – cubic growth for the SBF-VAR versus linear growth for the MBF-VAR – highlights the computational efficiency gains of allowing multiple base frequencies instead of relying on a single highest frequency.

We compare the out-of-sample forecasting performance of our multiple base frequency VAR (MBF-VAR), estimated with quarterly, monthly, and weekly data, against the two-frequency VAR model of Schorfheide and Song (2015), which uses quarterly and monthly data. To ensure both models incorporate the same set of variables, we aggregate the weekly series to a monthly frequency in the Schorfheide and Song (2015) specification. Our real-time dataset comprises 11 U.S. economic and financial variables: three at the quarterly frequency, five at the monthly frequency, and three at the weekly frequency. To compare the forecast performance of the two models, we calculate the relative root mean squared error (RMSE) for various forecast horizons. Our findings indicate that the one step ahead forecast of the quarterly GDP improves by X%, while the highest improvement can be observed in the third month and forth week of the quarter. The improvements in forecasting accuracy are observed the strongest in short term forecasts, while forecasting errors further into the future are mostly equivalent.

Moreover, we use simulated data to evaluate the performance of the MBF-VAR compared to the SBF-VAR in estimating latent states for temporal disaggregation. Assuming a high-frequency VAR(p) process, we simulate data and time-aggregate it for lower frequencies using the measurement equations of both models. Our findings reveal that the MBF-VAR model produces narrower error bands and smaller estimation errors on average compared to the SBF-VAR. In addition, we show how the model can be used to generate a weekly GDP series and compare it to the Weekly Economic Index (WEI) published by the Federal Reserve Bank of Dallas (see Lewis et al., 2022). Our generated series closely tracks the WEI. Moreover, the flexibility of our model enables us to construct a consistent weekly economic index dating back to 1964, thereby extending the historical coverage well beyond the WEI, which starts in 2008.

The paper is organized as follows. Section 2 present the MBF-VAR model and shows the derivation of a MBF-VAR from a SBF-VAR. Section 3 compares the parameter count of the SBF-VAR and the MBF-VAR. Section **??** discusses the real-time data set used for the comparison of the forecast performance of the MBF-VAR and SBF-VAR. Section 4 presents the empirical results of the forecasting performance exercise and the latent state estimation. Section 5 concludes.

2 Multiple Base Frequency Framework

In this section, we introduce our multiple base frequency VAR (MBF-VAR). To provide context, we first begin with a single base frequency VAR (SBF-VAR) and then extend this framework to our multiple base frequency model.

2.1 Single Base Frequency VAR

Let us start with the mixed frequency VAR model developed by Schorfheide and Song (2015), which serves as our point of departure.[3] This model is a special case of our approach with two frequencies and a single base frequency, where ω_H denotes the high frequency (the base frequency), and ω_L represents the low frequency.[4] Following Schorfheide and Song (2015), we assume that the economy evolves at the high frequency ω_H according to the following VAR(p):

$$\mathbf{x}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{x}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{x}_{t-p} + \mathbf{u}_t, \quad \mathbf{u}_t \sim \text{iid } N(0, \mathbf{\Sigma}), \tag{1}$$

where $\mathbf{x}_t$ is an $n \times 1$ vector of macroeconomic variables, which can be decomposed into $\mathbf{x}_t = [\mathbf{x}'_{\omega_H,t}, \mathbf{x}'_{\omega_L,t}]'$, with $\mathbf{x}_{\omega_H,t}$ being the $n_{\omega_H} \times 1$ vector of variables observed at the higher

[3]Other versions of single frequency VARs featuring time-varying parameters or alternative sampling methods could potentially be used to capture additional dynamics in the data; however, for clarity and comparability with previous literature, we adopt the model as proposed by Schorfheide and Song (2015).

[4]Schorfheide and Song (2015) use quarterly and monthly frequencies.

frequency ω_H, and $\mathbf{x}_{\omega_L,t}$ the $n_{\omega_L} \times 1$ vector of unobserved variables that are only published at lower frequency ω_L, $\mathbf{c}$ is an $n \times 1$ vector of constants, $\mathbf{\Phi}_1, ..., \mathbf{\Phi}_p$ are $n \times n$ matrices of VAR coefficients and $\mathbf{u}_t$ is an $n \times 1$ vector of normally distributed error terms.

Define $\mathbf{z}_t = [\mathbf{x}_t', \mathbf{x}_{t-1}', \mathbf{x}_{t-2}', ..., \mathbf{x}_{t-p+1}']'$, $\mathbf{d} = [\mathbf{c}', \mathbf{0}, \mathbf{0}, ..., \mathbf{0}]'$, $\mathbf{v}_t = [\mathbf{u}_t', \mathbf{0}, \mathbf{0}, ..., \mathbf{0}]'$ and

$$\mathbf{F} \equiv \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \mathbf{\Phi}_3 & \dots & \mathbf{\Phi}_{p-1} & \mathbf{\Phi}_p \\ \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}.$$

The VAR can then be written in companion form as:

$$\mathbf{z}_t = \mathbf{d} + \mathbf{F}\mathbf{z}_{t-1} + \mathbf{v}_t \tag{2}$$

where $E(\mathbf{v}_t\mathbf{v}_t') = \mathbf{\Omega}$ with

$$\mathbf{\Omega} \equiv \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Let T represent the last available observation and let $T_b \leq T$ denote the final period at higher frequency ω_H for which all variables at lower frequency ω_L are available. Up to period T_b, the vector of higher-frequency series $\mathbf{x}_{\omega_H,t}$ is observed at each higher frequency ω_H. We denote these observations by $\mathbf{y}_{\omega_H,t}$ and express this relationship as:

$$\mathbf{y}_{\omega_H,t} = \mathbf{x}_{\omega_H,t}, \quad t = 1, \dots, T_b. \tag{3}$$

Assuming the underlying VAR has at least p lags, where $p \geq \tau$ corresponds to the number of periods that match the lower frequency of the data, we express the τ-period average of $\mathbf{x}_{\omega_L,t}$ as:

$$\tilde{\mathbf{y}}_{\omega_L,t} = \frac{1}{\tau}\sum_{i=0}^{\tau-1} \mathbf{x}_{\omega_L,t-i} = \mathbf{\Lambda}_{\omega_L}\mathbf{z}_t. \tag{4}$$

As in Schorfheide and Song (2015), for logarithmic variables, we use a log-linear approximation of the arithmetic average. For flow variables, we use annualized high-frequency flows, with lower-frequency flows represented as averages of the higher-frequency ones. The lower-frequency variables can then be expressed as:

$$\mathbf{y}_{\omega_L,t} = \mathbf{M}_{\omega_L,t}\mathbf{\Lambda}_{\omega_L}\mathbf{z}_t, \quad t = 1, \ldots, T_b, \tag{5}$$

where the dimension of $\mathbf{y}_{\omega_L,t}$ is n_{ω_L} in periods where averages at the lower frequency ω_L are observed, and zero otherwise. $\mathbf{M}_{\omega_L,t}$ is a selection matrix that equals the identity matrix when t corresponds to the last period of the lower-frequency observation, and is empty otherwise. Thus, for periods $t = T_b + 1, \ldots, T$, where no additional lower-frequency observations are available, $\mathbf{y}_{\omega_L,t}$ has dimension zero and the selection matrix $\mathbf{M}_{\omega_L,t}$ is empty. The observed higher frequency variables that become available after T_b are:

$$\mathbf{y}_{\omega_H,t} = \mathbf{M}_{\omega_H,t}\mathbf{x}_{\omega_H,t}, \quad t = T_b + 1, \ldots, T, \tag{6}$$

where $\mathbf{M}_{\omega_H,t}$ is a sequence of selection matrices. The observed variables at the two frequencies can be more compactly written as:

$$y_t = \mathbf{M}_t\mathbf{\Lambda}\mathbf{z}_t \tag{7}$$

where $\mathbf{y}_t = [\mathbf{y}'_{\omega_H,t}, \mathbf{y}'_{\omega_L,t}]'$, $\mathbf{M}_t$ is a sequence of matrices that identifies the variables that are observed by time T and $\mathbf{\Lambda}$ is a matrix that maps $\mathbf{z}_t$ to the observed data. The model can be formulated within a state-space framework, where Equation (2) serves as the state equation, and Equation (7) defines the observation equation.

We employ Bayesian methods to estimate the model, following the approach of Schorfheide and Song (2015). Specifically, we implement a Gibbs sampler for the single base frequency model, identical to their methodology. This sampler iteratively draws from the conditional posterior distribution of the parameters $(\mathbf{\Phi}, \mathbf{\Sigma}) \mid (\mathbf{z}, \mathbf{y})$, where $\mathbf{y} = [\mathbf{y}'_{-p+1}, \mathbf{y}'_1, ..., \mathbf{y}'_T]'$ contains the data and $\mathbf{z} = [\mathbf{z}'_0, \mathbf{z}'_1, ..., \mathbf{z}'_T]'$ the latent states, and the conditional posterior distribution of the latent states $\mathbf{z} \mid (\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{y})$. The Gibbs sampler alternates between two blocks, following the approach suggested by Carter and Kohn (1994): In the first block, conditional on the data and the most recent draw of the latent states, we draw the parameters $(\mathbf{\Phi}, \mathbf{\Sigma})$ from their conditional posterior distribution. In the second block, conditional on the data and the most recent draw of the parameters, we draw the latent states from their conditional posterior distribution.

We adopt the same prior distribution as in Schorfheide and Song (2015), utilizing a Minnesota prior, which centers the distribution of the autoregressive coefficients $\mathbf{\Phi}$ at values implying random-walk behavior for each component of the state vector $\mathbf{x}_t$. This prior helps impose a structure that reflects persistent dynamics, while allowing the data to inform deviations.

2.2 Two Base Frequency Example

To illustrate the concept, we extend the single base frequency framework to a model with three frequencies, two of which serve as base frequencies. We decompose the joint density of the model into two base frequency components, with each base frequency component treated separately in the Gibbs sampling framework. The procedure used for the single base frequency model is adapted here to iteratively sample from the two base frequency components.

Consider, for example, a scenario in which we have data observed at quarterly, monthly, and weekly frequencies. To handle these different frequencies, we construct two separate SBF-VAR models: a quarterly-monthly SBF-VAR, denoted as VAR^{QM}, which models the relationship between quarterly and monthly data at a monthly frequency using the latent variable $x_{q,t}$; and a monthly-weekly SBF-VAR, denoted as VAR^{MW}, which models the relationship between monthly and weekly data at a weekly frequency, using the latent variable $x_{m,t}$ and incorporating $x_{q,t}$ from the previous model as an observable.

The VAR^{QM} is structured as follows: The observed variables include monthly observations, $y_{m,t}$, and quarterly observations, $y_{q,t}$, which are only available every third month. In addition to these observed variables, there are unobserved variables $x_{q,t}$, which represent latent quarterly variables at a monthly frequency. In terms of definitions, $\mathbf{y}^{QM}$ is at monthly frequency and refers to the set of both the observed monthly and quarterly data, $\mathbf{\Theta}^{QM}$ are the parameters of the quarterly-monthly VAR model, and α^{QM} is the set of latent quarterly variables at a monthly frequency.

In the VAR^{MW} model the observed variables include weekly observations, $y_{w,t}$, monthly observations, $y_{m,t}$, and the monthly sampled quarterly variables from the previous step, $x_{q,t}$, where the monthly observations are available every 4th week. The unobserved variables in this model include $x_{m,t}$, which are latent monthly variables at a weekly frequency. In terms of definitions, $\mathbf{y}^{MW}$ is at weekly frequency and refers to the set of observed weekly data, monthly data, and the now observed quarterly variables sampled monthly from the VAR^{QM}, $\mathbf{\Theta}^{MW}$ are the parameters of the monthly-weekly VAR model, and α^{MW} is the set of latent monthly variables at a weekly frequency.

The notation can be further simplified by defining $\mathbf{\Xi}^{QM} = [\mathbf{\Theta}^{QM}, \alpha^{QM}]$ and $\mathbf{\Xi}^{MW} = [\mathbf{\Theta}^{MW}, \alpha^{MW}]$. Thus, the overall joint density of the parameters, latent states, and data can be expressed as:

$$P(\mathbf{\Xi}^{MW}, \mathbf{y}^{MW}, \mathbf{\Xi}^{QM}, \mathbf{y}^{QM}) \propto P(\mathbf{\Xi}^{MW} \mid \mathbf{y}^{MW}) P(\mathbf{y}^{MW} \mid \mathbf{\Xi}^{QM}) P(\mathbf{\Xi}^{QM} \mid \mathbf{y}^{QM}), \tag{8}$$

where $P(\mathbf{\Xi}^{MW} \mid \mathbf{y}^{MW})$ is the posterior distribution of the parameters and latent states of the VAR^{MW} model, conditional on the observed weekly variables and the latent quarterly variables $x_{q,t}$ from the VAR^{QM} model. The term $P(\mathbf{y}^{MW} \mid \mathbf{\Xi}^{QM})$ expresses the distribution of the weekly data conditional on the monthly latent states. Given that the transformation from monthly states in $\mathbf{\Xi}^{QM}$ to the weekly data in $\mathbf{y}^{MW}$ is deterministic, $P(\mathbf{y}^{MW} \mid \mathbf{\Xi}^{QM})$ is degenerate and takes the form of a point mass. Finally, $P(\mathbf{\Xi}^{QM} \mid \mathbf{y}^{QM})$ represents the posterior distribution of the parameters and latent states of the VAR^{QM} model, conditional on the observed monthly and quarterly variables.

To sample from the joint posterior distribution in (8), we proceed with the following steps:

1. Sample the parameters and latent states $\mathbf{\Xi}^{QM}$ from the posterior distribution $P(\mathbf{\Xi}^{QM} \mid \mathbf{y}^{QM})$, which corresponds to a pass of the Gibbs sampling procedure for the single base frequency model.

2. Using the draw of $\mathbf{\Xi}^{QM}$, transform the monthly latent states into the weekly frequency to ensure the information from the monthly latent variables is appropriately

incorporated into the weekly data.

3. Sample the parameters and latent states Ξ^{MW} from the conditional posterior distribution $P(\Xi^{MW} \mid \mathbf{y}^{MW})$. This step completes another iteration of the Gibbs sampler, now applied to the extended model that captures dependencies between the two base frequencies.

2.3 Multiple Base Frequency VAR

In this section, we extend the framework to more than two base frequencies, generalizing the two base frequency example from the previous section. Consider $\omega_1, ..., \omega_{\kappa-1}, \omega_\kappa$ frequencies, where ω_1 denotes the lowest frequency and ω_κ represents the highest frequency. The first SBF-VAR then is at frequencies ω_1 and ω_2 denoted as $\text{VAR}^{\omega_{1,2}}$ and the last SBF-VAR is at frequencies $\omega_{\kappa-1}$ and ω_κ denoted as $\text{VAR}^{\omega_{\kappa-1,\kappa}}$. The overall joint density for κ frequencies is given by:

$$
\begin{aligned}
&P(\Xi^{1,2}, \mathbf{y}^{1,2}, \Xi^{2,3}, \mathbf{y}^{2,3}, \ldots, \Xi^{\kappa-1,\kappa}, \mathbf{y}^{\kappa-1,\kappa}) \\
&\propto \prod_{i=3}^{\kappa} P(\Xi^{i-1,i} \mid \mathbf{y}^{i-1,i}) P(\mathbf{y}^{i-1,i} \mid \Xi^{i-2,i-1}) P(\Xi^{i-2,i-1} \mid \mathbf{y}^{i-2,i-1}).
\end{aligned} \tag{9}
$$

where similar to the three frequency example in (8), $P(\Xi^{i-1,i} \mid \mathbf{y}^{i-1,i})$ is the posterior distribution of the parameters and latent states of the $\text{VAR}^{\omega_{i-1,i}}$ model, conditional on the observed data at frequencies ω_{i-1} and ω_i, as well as the latent variables from the previous lower-frequency pair $\Xi^{i-2,i-1}$. The term $P(\mathbf{y}^{i-1,i} \mid \Xi^{i-2,i-1})$ represents the distribution of the data at frequencies ω_{i-1} and ω_i, conditional on the latent states from ω_{i-2} and ω_{i-1}. Since the transformation is deterministic, $P(\mathbf{y}^{i-1,i} \mid \Xi^{i-2,i-1})$ is degenerate and takes the form of a point mass. Finally, $P(\Xi^{i-2,i-1} \mid \mathbf{y}^{i-2,i-1})$ represents the posterior distribution of the parameters and latent states of the $\text{VAR}^{\omega_{i-2,i-1}}$ model, conditional on the observed data at frequencies ω_{i-2} and ω_{i-1}.

To sample from the joint posterior distribution in (9), we proceed with the following steps:

1. Sample the parameters and latent states $\Xi^{1,2}$ from the posterior distribution $P(\Xi^{1,2} \mid$

$\mathbf{y}^{1,2}$), corresponding to a pass of the Gibbs sampler for the model at frequencies ω_1 and ω_2.

2. Using the draw of $\Xi^{1,2}$, transform the latent states from the lower frequency pair into the next frequency pair $\mathbf{y}^{2,3}$, to ensure the information from the lower frequencies is appropriately incorporated into the higher frequencies.

3. Sample the parameters and latent states $\Xi^{2,3}$ from the posterior distribution $P(\Xi^{2,3} \mid \mathbf{y}^{2,3})$, conditional on the latent states $\Xi^{1,2}$.

4. Repeat the previous steps for all subsequent frequency pairs, i.e., for each $i = 4, \ldots, \kappa$, transform the latent states from $\Xi^{i-2,i-1}$ into the higher-frequency pair $\mathbf{y}^{i-1,i}$, and sample $\Xi^{i-1,i}$ from $P(\Xi^{i-1,i} \mid \mathbf{y}^{i-1,i})$, until all frequency pairs are sampled.

Although the MBF-VAR can accommodate numerous low-frequency variables, only those deemed most relevant should be carried up to the highest frequency to maintain efficiency. Other low-frequency series can still be included, but without being propagated to the highest frequency.

3 Parameter Count: SBF-VAR vs. MBF-VAR

This section compares the parameter counts of the SBF-VAR and MBF-VAR, deriving their dependence on the number of frequencies (κ) based on their structural differences.

3.1 Number of Parameters in SBF-VAR

To analyze the parameter count N_{SBF} for the SBF-VAR, we assume that the number of macroeconomic variables n_s increases with the number of frequencies κ and that at least one new variable is added per frequency. Moreover, given the structure of the SBF-VAR the lag length p_s depends on the frequency ratio τ, which is a function of κ. Specifically, we assume a linear relationship $p_s = c_p\kappa$, where c_p is a constant. The total number of parameters N_{SBF} in the SBF-VAR model thus is

$$N_{\text{SBF}} = n_s + p_s \times n_s^2 + \frac{n_s(n_s+1)}{2}.$$

Substituting $n_s = c_n\kappa$ and $p_s = c_p\kappa$ results into

$$N_{\mathrm{SBF}} = c_n\kappa + c_p\kappa \times (c_n\kappa)^2 + \frac{c_n\kappa(c_n\kappa + 1)}{2}.$$

For large κ, the dominant term is $c_p c_n^2 \kappa^3$ and we thus obtain

$$N_{\mathrm{SBF}} \approx c_p c_n^2 \kappa^3.$$

To express the growth rate of N_{SBF} as κ approaches infinity, we analyze the scaling behavior under the assumption that the constants c_p and c_n remain fixed. Given that the leading term in N_{SBF} is $c_p c_n^2 \kappa^3$, the total parameter count scales cubically with κ such that

$$N_{\mathrm{SBF}} = O(\kappa^3).$$

This result indicates that, with the SBF-VAR lag length p_s increasing linearly with the number of frequencies κ and the number of variables n_s also increasing linearly with κ, the total number of parameters in the SBF-VAR model grows proportionally to κ^3.

3.2 Number of Parameters in MBF-VAR

To calculate the number of parameters in the MBF-VAR, we first determine the number of parameters in each bi-frequency VAR. For each bi-frequency VAR, we define

$$n_{m_i} = n_r + n_{\mathrm{new}_i},$$

where n_r is the number of relevant variables from the lowest frequency that are passed through each frequency level, and n_{new_i} is the number of additional observed indicators introduced locally at each frequency pair. Moreover, p_{m_i} denotes the lag length of the ith bi-frequency VAR, which is determined based on the frequency ratio τ_i of each bi-frequency VAR and does not depend on the total number of frequencies κ. To simplify matters and without loss of generality, we set $p_{m_i} = p_m$ and $n_{\mathrm{new}_i} = n_{\mathrm{new}}$, given that both do not depend on κ. Consequently, the total number of variables in each frequency pair is

$$n_m = n_r + n_{\text{new}}.$$

Hence, the total number of parameters N_{MBF_i} in each bi-frequency VAR is

$$N_{\text{MBF}_i} = n_m + p_m \times n_m^2 + \frac{n_m(n_m + 1)}{2}.$$

The total number of parameters N_{MBF} in the overall MBF-VAR thus is

$$N_{\text{MBF}} = \sum_{i=1}^{\kappa-1} N_{\text{MBF}_i} = (\kappa - 1)\left(n_m + p_m \times n_m^2 + \frac{n_m(n_m + 1)}{2}\right).$$

Since the expression inside the parentheses is constant with respect to κ, the total parameter count scales linearly with κ:

$$N_{\text{MBF}} = O(\kappa).$$

This result indicates that, the total number of parameters grows linearly with the number of base frequencies κ.

4 Empirical Results

In this section, we present our empirical results. To estimate the models, we set the number of lags in the SBF-VAR to $p_m = 6$ and in the MBF-VAR to $p_m = 6$ and $p_w = 4$. For both models, we use 60k draws from the Gibbs sampler, discarding the first 30k as burn-in. For the MBF-VAR, rather than conducting a hyperparameter optimization, we use the hyperparameters suggested by Schorfheide and Song (2015) for all bi-frequency VAR models. Additionally, we discard any explosive draws that violate the stability condition of the VAR models.

The remainder of this section is structured as follows. Section 4.1 describes the real-time dataset used for the forecasting exercise. Section 4.2 compares the latent state accuracy of the SBF-VAR and MBF-VAR models using simulated data. Section 4.2 presents a comparison of the two models' performance in forecasting GDP. Finally, Section 4.4 illustrates how our approach can be used to generate a weekly GDP estimate.

4.1 Real-Time Data

In our forecasting exercise, we analyze the relative out-of-sample performance gains of the MBF-VAR framework, which incorporates quarterly, monthly, and weekly data, versus an SBF-VAR framework that uses quarterly and monthly data. For this purpose, we use a real-time dataset of 11 U.S. macroeconomic and financial variables. The dataset includes three quarterly variables (real GDP, real private fixed investments, and real government expenditure), five monthly variables (hours worked, unemployment rate, consumer prices, industrial production, and personal consumption expenditure), and three variables available in either monthly or weekly frequencies (federal fund rates, 10-Year treasury bond yields, and the S&P 500 Index). Most variables are log-transformed, except for the rate variables (i.e., unemployment rate, federal funds rate, and bond yields), and the S&P 500 data, originally daily, is aggregated as needed. All series begin on January 1, 1964.

The dataset is constructed using ALFRED from the St. Louis Fed and Yahoo Finance for the S&P 500. It has been updated every Friday since January 4, 2002, with two versions generated each week: one in which Federal Fund Rates, Treasury Yields, and the S&P 500 are in monthly frequency (used in the SBF-VAR), and another in which these variables are in weekly frequency (used in the MBF-VAR). This approach ensures that forecasters always use the latest data, enabling a balanced evaluation of the forecasting models. Table 2 in Appendix A.2 illustrates the typically available dataset for each Friday in a given quarter.

To assess the usefulness of including weekly data in forecasting quarterly GDP, we generate forecasts eight quarters ahead (including the current one) each week (on Friday) using both the MBF-VAR and the SBF-VAR models. For each forecast, the RMSE is calculated. We group the forecasts into 12 groups corresponding to the month and week of the quarter in which they are generated—for example, all forecasts from the first week of the first month of a quarter are grouped together, as are those from the second week of the first month, and so on. We then calculate the within-group relative RMSE between the two models. We use the latest data release as the actual values.

4.2 Latent States Accuracy

In this section, we present a comparison of the latent state accuracy between the SBF-VAR and MBF-VAR models using simulated data. To do this, we simulate data under the assumption that at the highest frequency, all data is generated according to a VAR(p). To simulate the lower frequency data, the series are time-aggregated according to the measurement equation assumed in the SBF-VAR and MBF-VAR. The comparison is based on the root mean square error (RMSE) and the size and standard deviation of 90% and 68% credible intervals (CIs) over time.

Figure 1 (see also Figure 4 in Appendix A.1) illustrates the wider credible intervals of the SBF-VAR model compared to those of the MBF-VAR. The figure also depicts that the posterior mean of the latent state estimates for the SBF-VAR is more variable and further from the true latent states than the posterior mean estimates of the MBF-VAR, as indicated by the RMSE values in Table 1. This again suggests that the MBF-VAR provides narrower credible intervals and potentially more accurate latent state estimates.

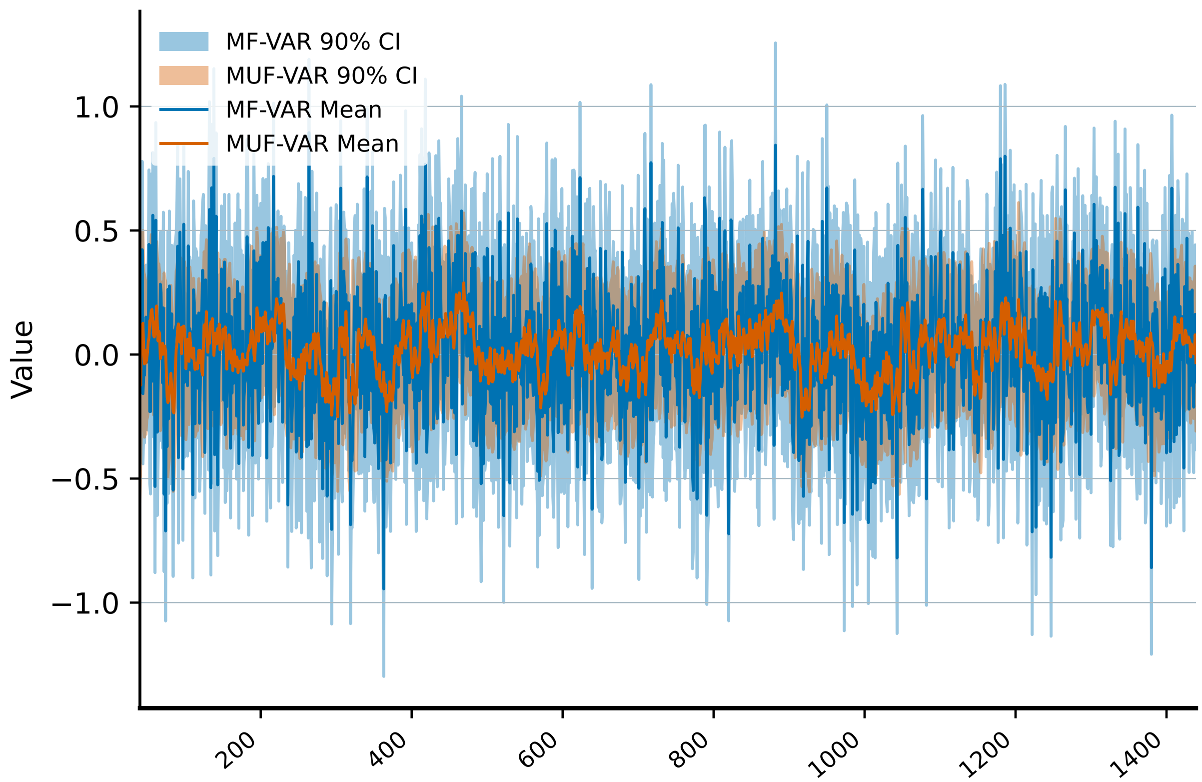

Figure 1: Latent State Estimation of SBF-VAR and MBF-VAR, Mean and 90% Credible Interval

In addition, Table 1 reports an average RMSE for the latent states that is notably lower for the MBF-VAR model, with a value of 0.32 compared to 0.38 for the SBF-VAR model. This represents a relative improvement of 16.30%, indicating that the MBF-VAR model provides more accurate estimates of the latent states. In terms of credible intervals, the table demonstrates that the MBF-VAR model also improves the size of both the 90% and 68% CIs. The average size of the 90% CI is slightly reduced from 0.60 in the SBF-VAR model to 0.57 in the MBF-VAR model, reflecting a 5.39% improvement. Similarly, the size of the 68% CI decreases from 0.36 to 0.31, leading to a more substantial improvement of 14.10%. In addition to the reduced CI sizes, the MBF-VAR model significantly reduces the variability of the credible intervals. The standard deviation of the 90% CI decreases from 0.10 to 0.04, representing a 56.19% improvement. Also, the standard deviation of the 68% CI decreases by 58.22%, from 0.06 to 0.02. These reductions in variability suggest that the MBF-VAR model provides more stable and consistent estimates of latent state uncertainty.

	SBF-VAR	MBF-VAR	Relative Improvement
Avg. RMSE	0.38	0.32	16.30 %
Avg. 90% CI size	0.60	0.57	5.39 %
Avg. 90% CI std.	0.10	0.04	56.19 %
Avg. 68% CI size	0.36	0.31	14.10 %
Avg. 68% CI std.	0.06	0.02	58.22 %

Table 1: Comparison of SBF-VAR and MBF-VAR performance in latent state estimation. For each run, the average 90% and 68% credible intervals (CI) were computed by calculating the size of the CI for each time point and then taking the average over all time points. Similarly, for the standard deviation of the CI size, the standard deviation over all time points was computed instead of the average. The table shows the mean values for both the average CI size and the standard deviation of CI size across all 10 runs.

Overall, the results suggest that the MBF-VAR model performs better than the SBF-VAR model across most metrics, including both accuracy (as reflected by RMSE) and the precision of the latent state estimates (as indicated by CI size and variability). The MBF-VAR model's ability to provide smaller and more stable credible intervals indicates that it may offer more reliable quantification of latent state uncertainty.

4.3 Forecast Performance Comparison

Figure 2 illustrates the overall relative performance of the MBF-VAR compared to the SBF-VAR in forecasting GDP growth. The horizontal axis denotes the forecast horizon, while the vertical axis reports the relative differences in RMSE in percentage terms, with negative values indicating that the MBF-VAR achieves lower RMSEs than the SBF-VAR. The results reveal that the MBF-VAR outperforms the SBF-VAR during the early forecast periods, particularly in the first and second quarters, with the performance gap narrowing from the third quarter onward.

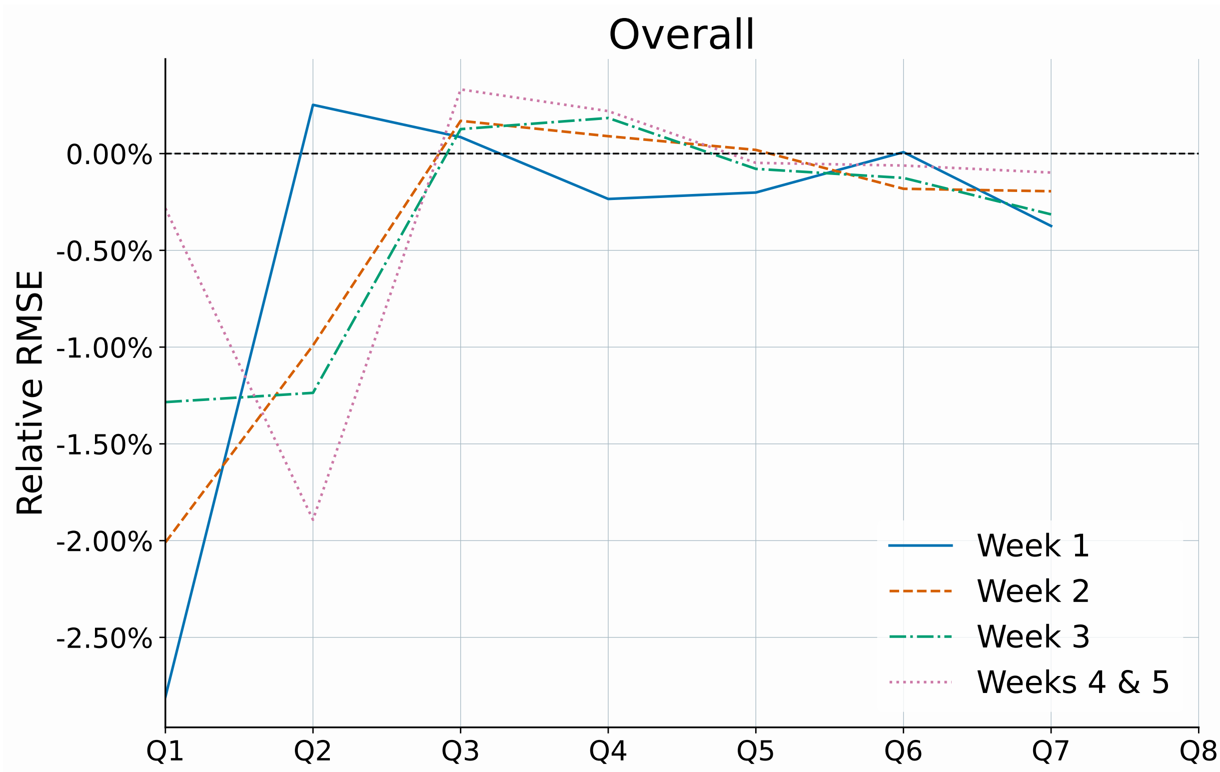

Figure 2: This figure compares the forecast performance of the MBF-VAR model to the SBF-VAR. The horizontal axis represents the forecast horizon, while the vertical axis shows the RMSE differences expressed as a percentage relative to the SBF-VAR. The real-time data vintages used for these forecasts span the period 2002–2014, and the separate lines correspond to forecasts based on data released at weeks 1−3 and weeks 4 & 5.

4.4 Temporal-Disaggregation

We now use our MBF-VAR model to derive estimates for GDP at a weekly frequency, providing a high-resolution view of economic dynamics. Figure 3 illustrates the year-over-year (YoY) GDP growth estimates from three sources: the MBF-VAR model, the Weekly Economic Indicator (WEI) developed by Lewis et al. (2022) and published by the Federal Reserve Bank of Dallas, and the official quarterly GDP growth data. For the period prior to 2008, when the WEI is not available, the MBF-VAR estimates closely match the official quarterly GDP growth rates. In more recent years, the MBF-VAR estimates not only continue to track the quarterly data closely, but also align well with the WEI estimates. See Figure 5 in Appendix A.1 for unsmoothed, high-frequency estimates of weekly GDP growth, which directly compares the MBF-VAR and WEI estimates.

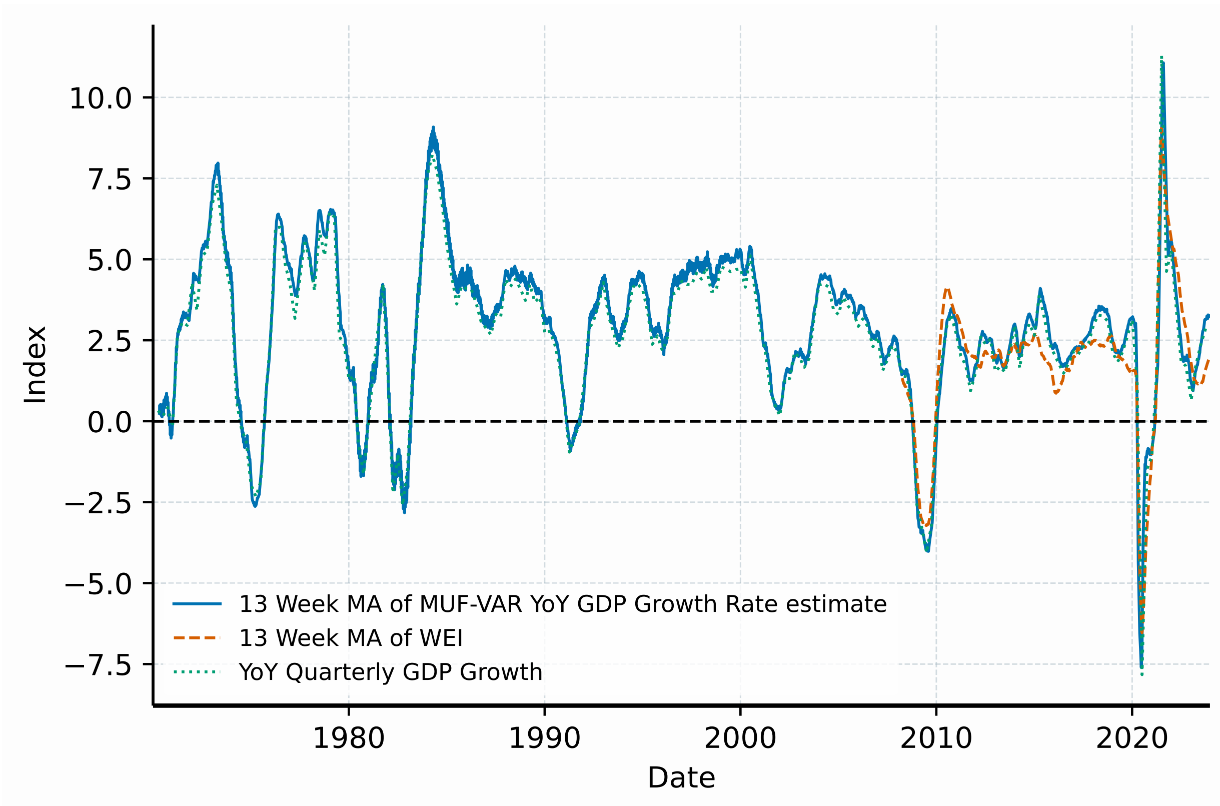

Figure 3: Comparison of 13-week moving average estimates of YoY GDP growth derived from the MBF-VAR model, the Weekly Economic Indicator (WEI), and official quarterly data.

5 Conclusion

This paper introduces a novel multiple base frequency VAR model that addresses the computational and methodological challenges of integrating time series data observed at different frequencies. By allowing for multiple base frequencies, the MBF-VAR model significantly reduces the parameter count compared to traditional single base frequency VAR models, which require aligning all variables to a single frequency. Our approach decomposes the joint distribution into a set of conditional bi-frequency VARs, ensuring that the parameter count scales linearly with the frequency gap rather than cubically as in SBF-VAR models. This innovation not only enhances computational efficiency but also improves the accuracy of forecasts and latent state estimates.

Our empirical results demonstrate that the MBF-VAR model outperforms the SBF-

VAR model in out-of-sample real-time forecasting, particularly for short-term horizons. The inclusion of high-frequency data, such as weekly indicators, adds predictive power, as evidenced by the improved forecasting accuracy for quarterly GDP and other key economic variables. Furthermore, our simulated data experiments reveal that the MBF-VAR model produces narrower error bands and smaller estimation errors for latent states, underscoring its superiority in temporal disaggregation tasks. The model's flexibility is further illustrated by its ability to generate a consistent weekly GDP series, extending historical coverage beyond existing high-frequency economic indices.

These findings have important implications for real-time economic forecasting and policy analysis, especially in contexts where timely and accurate data integration is critical, such as during economic crises or pandemics. By leveraging the highest observed frequency of each variable, the MBF-VAR model provides a robust framework for handling mixed-frequency data. Future research should focus on the optimization of hyperparameters, as the number of hyperparameters increases with each additional base frequency. This challenge necessitates the development of more efficient optimization techniques beyond a simple grid search. Additionally, extending the MBF-VAR framework to incorporate features such as time- varying parameters and stochastic volatility, or applying the model to other economic indicators, could further enhance its applicability and robustness.

References

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

Ankargren, S. and P. Jonéus (2019). Estimating large mixed-frequency bayesian var models.

Brave, S. A., R. A. Butters, and A. Justiniano (2019). Forecasting economic activity with mixed frequency bvars. *International Journal of Forecasting 35*(4), 1692–1707.

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika 81*, 541–553.

Chan, J. C., A. Poon, and D. Zhu (2023). High-dimensional conditionally gaussian state space models with missing data. *Journal of Econometrics 236*(1), 105468.

Cimadomo, J. and A. D'Agostino (2016). Combining time variation and mixed frequencies: An analysis of government spending multipliers in italy. *Journal of Applied Econometrics 31*, 1276–1290.

Cimadomo, J., D. Giannone, M. Lenza, F. Monti, and A. Sokol (2022). Nowcasting with large bayesian vector autoregressions. *Journal of Econometrics 231*, 500–519.

Daniele, M., H. Mikosch, and S. Neuwirth (2024). An observation-driven mixed-frequency var model with closed-form solution. *Available at SSRN*.

Eraker, B., C.-W. Chiu, A. T. Foerster, T. B. Kim, and H. D. Seoane (2015). Bayesian mixed frequency vars. *Journal of Financial Econometrics 13*(3), 698–721.

Foroni, C., E. Ghysels, and M. Marcellino (2013). *Mixed-Frequency Vector Autoregressive Models*, Volume 32, Chapter 32, pp. 247–272. Emerald Group Publishing Limited.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics 193*(2), 294–314. The Econometric Analysis of Mixed Frequency Data Sampling.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The midas touch: Mixed data sampling regression models. *Technical Report*.

Ghysels, E. and J. H. Wright (2009). Forecasting professional forecasters. *Journal of Business & Economic Statistics 27*, 504–516.

Götz, T. B. and K. Hauzenberger (2021). Large mixed-frequency vars with a parsimonious time-varying parameter structure. *The Econometrics Journal 24*(3), 442–461.

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

Kuzin, V., M. Marcellino, and C. Schumacher (2011). Midas vs. mixed-frequency var: Nowcasting gdp in the euro area. *International Journal of Forecasting 27*, 529–542.

Lewis, D. J., K. Mertens, J. H. Stock, and M. Trivedi (2022). Measuring real activity using a weekly economic index. *Journal of Applied Econometrics 37*(4), 667–687.

Mariano, R. S. and Y. Murasawa (2010). A coincident index, common factors, and monthly real gdp. *Oxford Bulletin of Economics and Statistics 72*, 27–46.

McCracken, M. W., M. T. Owyang, and T. Sekhposyan (2021). Real-Time Forecasting and Scenario Analysis Using a Large Mixed-Frequency Bayesian VAR. *International Journal of Central Banking 17*(71), 1–41.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency var. *Journal of Business & Economic Statistics 33*(3), 366–380.

Schorfheide, F. and D. Song (2020). Real-time forecasting with a (standard) mixed-frequency var during a pandemic. *Working paper (Federal Reserve Bank of Philadelphia)*.

Zadrozny, P. A. (1988). Gaussian likelihood of continuous-time armax models when data are stocks and flows at different frequencies. *Econometric Theory 4*(1), 108–124.

Zadrozny, P. A. (2016). Extended yule–walker identification of varma models with single- or mixed-frequency data. *Journal of Econometrics 193*, 438–446.

A Appendix

A.1 Additional Figures

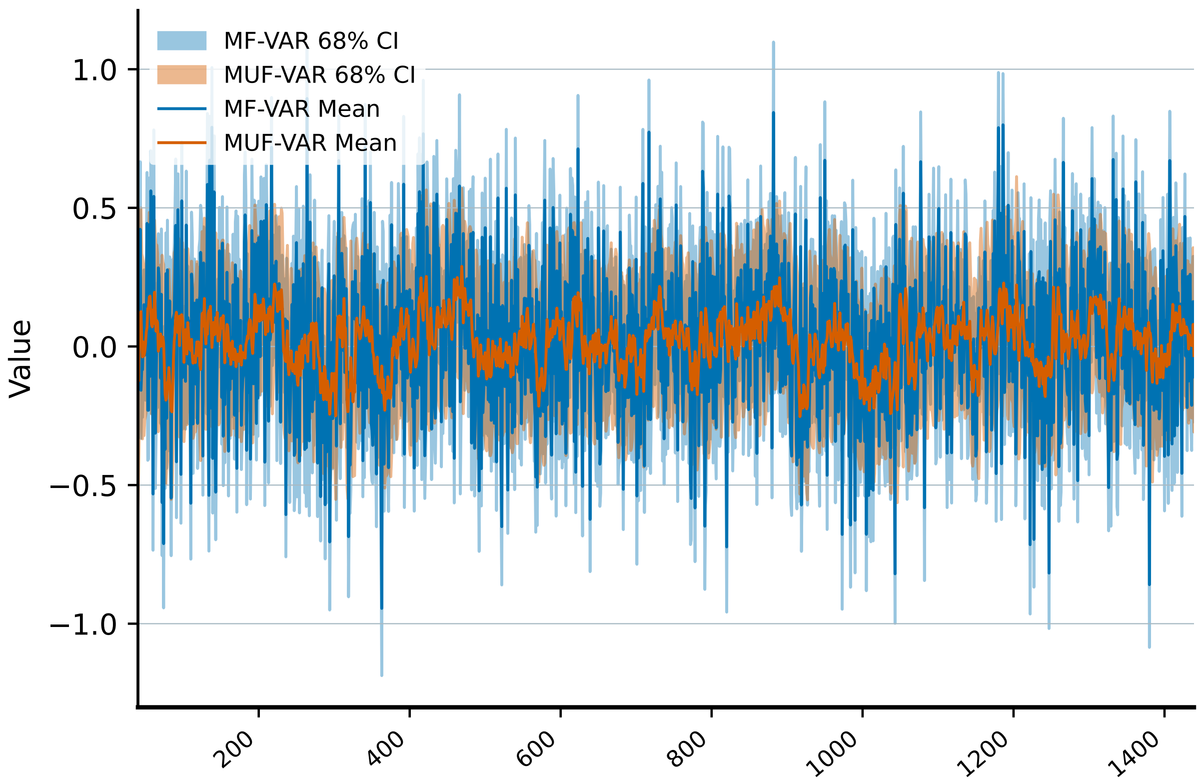

Figure 4: Latent State Estimation of MF-VAR and MUF-VAR, Mean and 68% Credible Interval

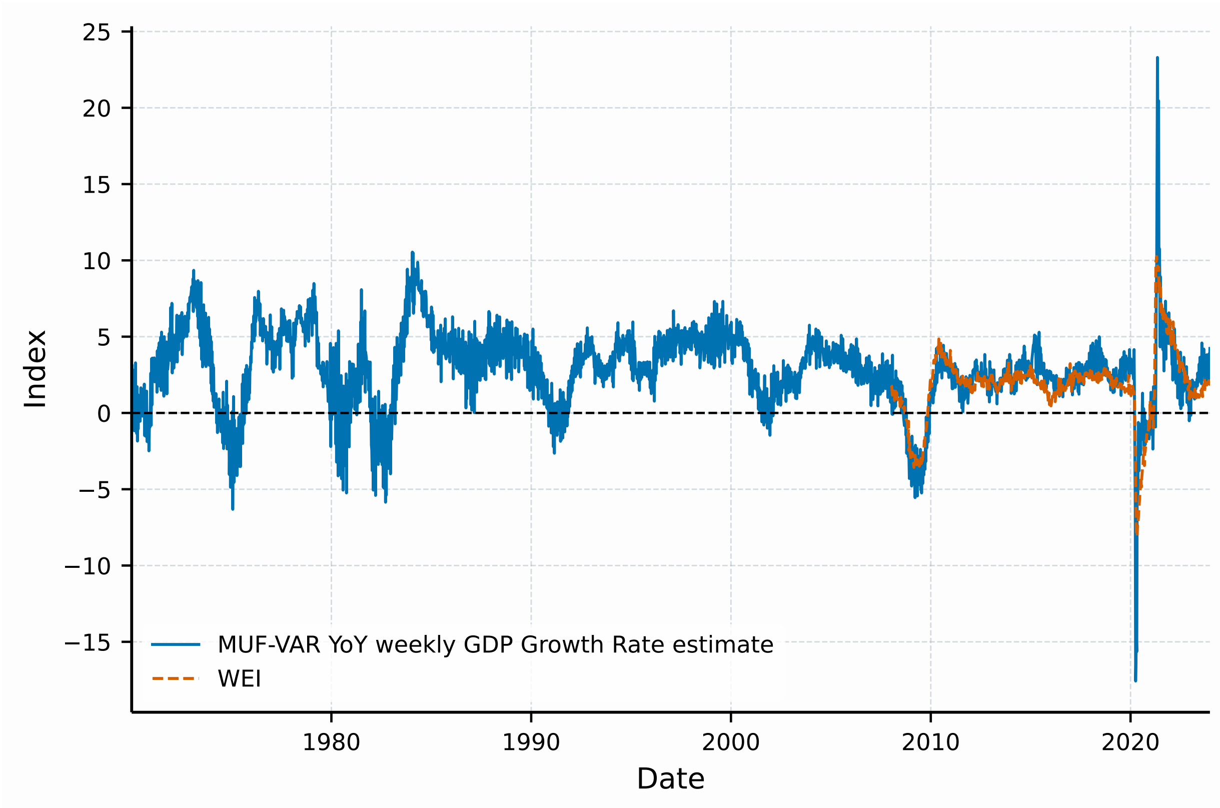

Figure 5: Comparison of unsmoothed weekly YoY GDP growth estimates from the MBF-VAR model and the Weekly Economic Indicator (WEI).

A.2 Data

	Current Quarter											
	Month 1				Month 2				Month 3			
	Week 1	Week 2	Week 3	Week 4/5	Week 1	Week 2	Week 3	Week 4/5	Week 1	Week 2	Week 3	Week 4/5
GDPC1	Q(-2)	Q(-2)	Q(-2)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)
FPIC1	Q(-2)	Q(-2)	Q(-2)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)
GCEC1	Q(-2)	Q(-2)	Q(-2)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)	Q(-1)
AWHI	Q(-1)M2	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q0M1	Q0M1	Q0M1	Q0M1	Q0M2	Q0M2	Q0M2	Q0M2
FEDFUNDS	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q0M1	Q0M1	Q0M1	Q0M1	Q0M2	Q0M2	Q0M2	Q0M2
CPIAUCSL	Q(-1)M2	Q(-1)M2 / Q(-1)M3	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q(-1)M3 / Q0M1	Q0M1	Q0M1	Q0M1	Q0M1 / Q0M2	Q0M2	Q0M2
INDPRO	Q(-1)M2	Q(-1)M2	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q(-1)M3 / Q0M1	Q0M1	Q0M1	Q0M1	Q0M1	Q0M2	Q0M2
PCEC96	Q(-1)M2	Q(-1)M2	Q(-1)M2	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q0M1	Q0M1	Q0M1	Q0M1	Q0M2
GS10	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q0M1	Q0M1	Q0M1	Q0M1	Q0M2	Q0M2	Q0M2	Q0M2
UNRATE	Q(-1)M2 / Q(-1)M3	Q(-1)M3	Q(-1)M3	Q(-1)M3	Q0M1	Q0M1	Q0M1	Q0M1	Q0M2	Q0M2	Q0M2	Q0M2
FF	Q(-1)M3W4/5	Q0M1W1	Q0M1W2	Q0M1W3	Q0M1W4/5	Q0M2W1	Q0M2W2	Q0M2W3	Q0M2W4/5	Q0M3W1	Q0M3W2	Q0M3W3
WGS10	Q(-1)M3W4/5	Q0M1W1	Q0M1W2	Q0M1W3	Q0M1W4/5	Q0M2W1	Q0M2W2	Q0M2W3	Q0M2W4/5	Q0M3W1	Q0M3W2	Q0M3W3
SP500	Q0M1W1	Q0M1W2	Q0M1W3	Q0M1W4/5	Q0M2W1	Q0M2W2	Q0M2W3	Q0M2W4/5	Q0M3W1	Q0M3W2	Q0M3W3	Q0M3W4/5

Table 2: Overview over the data set. The quarterly observed variables are, Real GDP (GDPC1), Real Private Fixed Investments (FPIC1) and Real Governement Expenditure (GCEC1). Hours Worked (AWHI), Unemployment (UNRATE), Consumer Prices (CPIAUCSL), Industrial Production (INDPRO) and Personal Consumption Expenditure (PCEC96) are all observed in a monthly frequency. Federal Fund Rates as well as 10 Year Treasury Bond Yields are available in monthly (FEDFUNDS, GS10) and weekly (FF, WGS10) frequencies. The S&P 500 Index (SP500) is available in a daily frequency. The S&P 500 is time aggregated to a weekly as well as monthly frequency depending on the model. Q0 refers to the current quarter, while Q(-1) and Q(-2) refer to the two previous quarters. M# and W# refer to the month and week inside these quarters (i.e. Q0M1W1 means, that the newest available data is from week one of month one of the current quarter).