A Practitioner's Guide and **MATLAB** Toolbox for Mixed Frequency State Space Models

Scott Brave
Federal Reserve Bank of Chicago

R. Andrew Butters
Indiana University

David Kelley
Federal Reserve Bank of Chicago

Abstract

The use of mixed frequency data is now common in many applications, ranging from the analysis of high frequency financial time series to large cross-sections of macroeconomic time series. In this article, we show how state space methods can easily facilitate both estimation and inference in these settings. After presenting a unified treatment of the state space approach to mixed frequency data modeling, we provide a series of applications to demonstrate how our **M**ixed **F**requency **S**tate **S**pace (**MFSS**) MATLAB toolbox can make the estimation and post-processing of these models straightforward.

Keywords: mixed frequency, state space, Kalman filter, maximum likelihood, MATLAB.

1. Introduction

The use of mixed frequency data is now common in many applications, ranging from high frequency financial time series to large cross-sections of macroeconomic time series. For example, a model of a security might jointly involve hourly spread data with the daily closing price. Or, in a macroeconomic model, quarterly Gross Domestic Product (GDP) might be incorporated jointly with monthly information on employment and industrial production. In this article, we show how state space methods can easily facilitate both estimation and inference in mixed frequency settings. By preserving the frequency of the data generating process, this approach offers unique identification and inference possibilities.

In the past, integrating data observed at differing frequencies was often considered a pre-analysis step of data collection, either by temporally aggregating the higher frequency time series or by narrowing the selection of time series to a single frequency. However, given the resulting potential for temporal aggregation bias, the demand for methods that instead explicitly accommodate the mixed frequency nature of a dataset has grown substantially. For example, mixed frequency extensions to traditional dynamic factor models (Mariano and Murasawa (2003), Aruoba, Diebold, and Scotti (2009), Brave and Butters (2012)), dynamic stochastic general equilibrium models (Kim (2010), Foroni and Marcellino (2014)), and vector auto-regressions (Kuzin, Marcellino, and Schumacher (2011), Schorfheide and Song (2015), Brave, Butters, and Justiniano (2016)) have found their way into both economic and financial applications.

All of the above extensions share in common the use of state space methods in order to accommodate the mixed frequency nature of their applications.[1] Generally speaking, state

[1] Accommodating mixed frequency data has also been the focus of statistical procedures for temporal dis-

space methods are well suited to overcome the challenges posed by mixed frequency data. A critical source of their comparative advantage centers on their flexible accommodation of handling missing observations through the use of latent state variables. Somewhat fortuitously, state space methods can treat missing observations due to the mixed frequency or staggered release of data (another kind of missing observation) similarly – provided the appropriate temporal restrictions have been embedded into the state space system. Paying this upfront cost allows the researcher to model the system in the highest (or an even higher) frequency observed in the dataset and opens up the full array of inference and estimation capabilities of a standard state space model. Thus, it should come as no surprise that these methods are quickly becoming a standard in the applied econometrician's toolkit.

Despite the increased use of state space methods in mixed frequency applications, a unified, self-contained description of these methods and the challenges faced in a mixed frequency environment does not, to the best of our knowledge, exist in the extant literature. The same is true for the existence of a singular software package that can readily handle setting up, estimating, and post-estimation processing of these models with limited user cost. Instead, practitioners are often left having to piece together many different programming languages with far from uniform standards and default settings for key specification choices.

In this paper, we present a unified framework that enables low-frequency data to be integrated into state space models with higher frequency data in a manner that respects the temporal aggregation properties of each time series. To further lower the entry cost of utilizing this approach, we also introduce and illustrate the use of our MATLAB[2] toolbox in applying these methods – what we call the Mixed Frequency State Space (**MFSS**) toolbox.

The toolbox offers similar estimation and post-processing capabilities of popular alternatives, e.g., STAMP (Koopman, Harvey, Doornik, and Shephard (2010b)), **SsfPack** in Ox (Koopman, Doornik, and Shephard (2010a), Doornik (2010)), **KFAS** (Helske (2018)), **statsmodels** (Perktold, Seabold, and Taylor (2017)), and the Econometrics Toolbox (in MATLAB)). For a more complete overview of available alternatives, see Commandeur, Koopman, and Ooms (2011). In addition, **MFSS** integrates leading computational methods (e.g. univariate treatment of multivariate time series (Koopman and Durbin (2000)), exact initialization of the state (Koopman and Durbin (2003)), improved calculation of starting values for maximum likelihood estimation (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, and Riddell (2017)), a compiled implementation of the filter & smoother (providing a roughly 15x speedup compared to native MATLAB code), and the computation of filtering and smoothing weights (Koopman and Harvey (2003))) all within an environment that can accommodate an arbitrary mix of data frequencies–a feature unmatched by existing software packages.

Our goal in this paper is to outline and highlight the foundational elements of setting up and estimating a mixed frequency state space model through a series of applications (e.g. a small monetary policy MF-VAR) that illustrate the utility of the **MFSS** toolbox. Given that

aggregation (Harvey and Pierse (1984), Harvey (1989), Harvey and Chung (2000), Proietti (2006)). In these cases, the object of interest is the higher frequency (latent) estimate of one of the lower frequency variables in the model. Furthermore, within the regression framework, the mixed data sampling (MIDAS) approach provides an alternative methodology to the state space approach in explicitly incorporating the mixed frequency nature of time series (Ghysels, Santa-Clara, and Valkanov (2004)).

[2] The toolbox is currently limited to use in MATLAB. Introspection of anonymous functions in Octave is limited, preventing necessary steps in estimating parameters.

goal, we leave the reader to find most of the theoretical derivations and evidence behind the methodological advancements of time series state space modeling to the traditional treatments (e.g. Harvey (1989), Durbin and Koopman (2012)). Each of the applications are selected for their comparative advantage in illustrating the different types of requirements for estimating and analyzing a mixed frequency model. In some cases, such as the trend-cycle decomposition of GDP (section 4.1) and the estimation of the natural rate of interest (section 4.4), they are the first mixed frequency estimates of their kind.

To ensure consistency throughout the article, and to avoid any confusion that could be caused by the minor variations used in the literature, we outline the standard normal-linear state space model in section (2). In section (3), we outline the key departures and extensions from the standard state space model required to accommodate mixed frequency data, leaving some of the underlying details not likely found in usual references to an online appendix. With the foundational elements in place, in section (4) we provide a series of topical applications that benefit from the use of the mixed frequency approach (e.g. a dynamic factor model of the business cycle). For each application, we provide code-examples and visualizations of key results that showcase the approachable syntax of our **MFSS** toolbox in achieving many of the standard tasks associated with setting up, estimating, and conducting post-estimation analysis of a mixed frequency state space model. In section (5), we discuss other likely uses for the toolbox and how it could be extended in the future.

2. A Normal-Linear State Space Model

Our approach to mixed-frequency data modeling centers around normal-linear state space models of the form shown below, where data are noisy observations of an unobserved state.[3]

$$y_t = Z_t \alpha_t + d_t + \beta_t x_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, H_t) \tag{1}$$

$$\alpha_t = T_t \alpha_{t-1} + c_t + \gamma_t w_t + R_t \eta_t \qquad \eta_t \sim \mathcal{N}(0, Q_t) \tag{2}$$

The observation equation (1) specifies the vector of data y_t as a linear combination of the unobserved (latent) states α_t, a deterministic component d_t, possible regressors $\beta_t x_t$, and a measurement error ε_t. The state, or transition, equation (2) specifies that the unobserved state evolves according to a univariate or vector autoregressive process with a deterministic component c_t, possible regressors $\gamma_t w_t$, and shocks η_t. For an extended discussion and examples that build intuition around these models, see Sargent and Stachurski (2018).

A few additional remarks on our specification of these models:

- Exogenous variables ($\beta_t x_t$ and $\gamma_t w_t$) could be subsumed into d_t and c_t, but are included to simplify specification of models with exogenous series.

- The elements of ε_t may be correlated with each other and the elements of η_t can be correlated with each other, but we assume independence across the two sets of shocks and across all time periods.

[3]For our purposes, we describe everything in the normal-linear state space framework typically used in most empirical applications. However, much of the utility from this article applies more generally under alternative distributional and functional forms with only modest modifications.

- The time indexing of system matrices, T_t, R_t, c_t, γ_t and state equation shocks η_t are a slight deviation from the treatment in Durbin and Koopman (2012), and aligns more closely with the treatment in Harvey (1989).
- To allow for potentially fewer shocks than states, the R_t matrix selects which shocks are applied to which states. However, it may have elements other than zero or one. The R_t matrix also allows for lagged variables to be included in the state. For example, it allows a VAR(p) model to be written in companion form in the state equation.
- Due to the recursive nature of the state equation, we must specify an initial condition, $\alpha_0 \sim \mathcal{N}(a_0, P_0)$. These parameters may be set arbitrarily based on outside information, implied from the steady state of the transition equation parameters, or the diagonal elements of P_0 can be arbitrarily large, in which case the values of a_0 are irrelevant.

Given a set of model parameters and observations, we are interested in recovering the implied latent state α_t. The Kalman filter is a set of recursive equations that provides the minimum mean-squared error estimate of α_t at each point in time conditional on all observed data prior to that time, i.e., $a_t = \mathbb{E}(\alpha_t | y_1, \ldots y_{t-1})$, while the Kalman smoother provides the corresponding estimate given the full sample of data, $\hat{\alpha}_t = \mathbb{E}(\alpha_t | y_1, \ldots, y_T)$ (see Appendix A). Additionally, the Kalman filter recursion provides the one-step ahead forecast errors, which can be used to compute the likelihood of a set of model parameters and facilitates maximum likelihood estimation (see Appendix B). Furthermore, because the Kalman filter and smoother recursions are linear, it is possible to decompose estimates of the state according to a set of weights based on the observed data and parameters (see Appendix C).

3. Mixed-Frequency State Space Modeling

Our approach to modeling data observed at differing frequencies can be viewed as interpolating any lower frequency time series according to a set of dynamics that are consistent across all observed time series. In other words, we appropriately aggregate the unobserved dynamic process captured in the transition equation of the state space model for each of these time series before it is related to its lower frequency observations. Doing so separates the observation frequency of the data from the frequency of the model and allows the specification of the model to be time-consistent.

3.1. Defining Accumulators and Augmenting State Spaces

In handling mixed-frequency data, we follow the direction of Harvey (1989) and Mariano and Murasawa (2003). The latent state process α_t is determined at a base frequency that matches the frequency of the highest-frequency observation.[4] To handle data observed at differing frequencies, we partition y_t into two components. Series observed at the base frequency are denoted y_t^b. Series at a lower frequency are denoted y_t^l. To move from the base frequency to a lower frequency, define an aggregation as a linear function $\mathcal{A}$ that takes the most recent n base-frequency values and creates a low-frequency version, $y_t^l = \mathcal{A}(y_t^h, y_{t-1}^h, \ldots, y_{t-n+1}^h)$, where y_t^h is an unobserved high-frequency version of y_t^l.

[4] It is possible to specify the transition equation at a higher frequency than the data. For example, Aruoba *et al.* (2009) specify a daily frequency despite the highest frequency data in their model being weekly; and in section 4.3 we estimate a trend-cycle decomposition of *quarterly* GDP at a *monthly* frequency.

For the time being, we restrict the measurement equation loadings (Z and β) on time series requiring aggregation to be time-invariant. Given the state α_t and some set of parameters $\{Z^h, d_t^h, \beta^h x_t^h, H_t^h\}$, we could create an interpolated version of the observations with $y_t^h = Z^h\alpha_t + d_t^h + \beta^h x_t^h + \varepsilon_t^h$. Because $\mathcal{A}$ is linear, we can factor this expression as

$$\begin{aligned} y_t^l &= \mathcal{A}(Z^h\alpha_t + d_t^h + \beta^h x_t^h + \varepsilon_t^h, \ldots, Z^h\alpha_{t-n+1} + d_{t-n+1}^h + \beta^h x_{t-n+1}^h + \varepsilon_{t-n+1}^h) \\ &= Z^h\mathcal{A}(\{\alpha_{t-i}\}_{i=0}^{n-1}) + \mathcal{A}(\{d_{t-i}^h\}_{i=0}^{n-1}) + \beta^h\mathcal{A}(\{x_{t-i}^h\}_{i=0}^{n-1}) + \mathcal{A}(\{\varepsilon_{t-i}^h\}_{i=0}^{n-1}) \\ &= Z^l\zeta_t + d_t^l + \beta^l x_t^l + \varepsilon_t^l. \end{aligned}$$

We next define an *accumulator*, ζ_t, that can be included in the state to relate the base frequency state α_t to the low-frequency observations y_t^l, implying that the high-frequency loading matches the lower frequency loading, $Z^l = Z^h$. In order to incorporate a low-frequency observation in a state space model, we need to be able to integrate it into the Kalman filter recursions, and so require a recursive formula for an accumulator,

$$\begin{aligned} \zeta_t &= f_t(\zeta_{t-1}, \alpha_t) && (3) \\ &= A_t^\zeta\zeta_{t-1} + A_t^T\alpha_{t-1} + A_t^c + A_t^\gamma w_t + A_t^R\eta_t, && (4) \end{aligned}$$

where $\{A_t^\zeta, A_t^T, A_t^c, A_t^\gamma, A_t^R\}$ are functions of the system matrices. Formulas are given below for each accumulator type that we consider. Combined with a transition equation for α_t and an observation equation for the base-frequency data, a state space model that accommodates mixed frequency observations is given by[5]

$$\begin{bmatrix} y_t^b \\ y_t^l \end{bmatrix} = \begin{bmatrix} Z_t^b & 0 \\ 0 & Z^l \end{bmatrix} \begin{bmatrix} \alpha_t \\ \zeta_t \end{bmatrix} + \begin{bmatrix} d_t^b \\ d_t^l \end{bmatrix} + \begin{bmatrix} \beta_t^b & 0 \\ 0 & \beta^l \end{bmatrix} \begin{bmatrix} x_t^b \\ x_t^l \end{bmatrix} + \begin{bmatrix} \varepsilon_t^b \\ \varepsilon_t^l \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} \alpha_t \\ \zeta_t \end{bmatrix} = \begin{bmatrix} T_t & 0 \\ A_t^T & A_t^\zeta \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ \zeta_{t-1} \end{bmatrix} + \begin{bmatrix} c_t \\ A_t^c \end{bmatrix} + \begin{bmatrix} \gamma_t \\ A_t^\gamma \end{bmatrix} w_t + \begin{bmatrix} R_t \\ A_t^R \end{bmatrix} \eta_t. \tag{6}$$

This makes clear that in the mixed frequency context, examining the effect of y_t^b and y_t^l on α_t becomes more difficult. Instead of comparing the estimated elements of Z to each other, as in a single-frequency context, we instead must examine how the data affects the estimated states using the Kalman filter and smoother weights to find the effect of y_t^l on α_t.

3.2. Aggregation Types and Accumulator Construction

In general, the matrices in the recursion for ζ_t are time-varying even if T_t, c_t, γ_t, and R_t are not. They depend on the *calendar* that the low-frequency data follows, a scalar function of time t specific to each time series. The calendar definitions will also be specific to each type of accumulator, but some common notation is helpful. Define $\mathcal{P}_1$ as the set of time indexes that start a new low-frequency period. For example, with a monthly base frequency and an observed series at a quarterly frequency, these would be the indexes of the first month of each quarter (January, April, July, and September). Accordingly, for the observations to line up with the calculations for ζ_t, the data must be placed in the last base-frequency period per low-frequency period (March, June, September, and December).

[5] To keep the state dimension as small as possible, we only need to augment the states that will be aggregated for the time series under consideration. Specifically, for any observation requiring an accumulator, we can create augmented states only for those where Z contains non-zero elements.

Construction of the accumulators differs based on the nature of the aggregation of each time series. For the purposes of this article (and the **MFSS** toolbox), there will exist three types. The appropriate accumulator type for any low-frequency time series is a function of the way in which it is sampled as well as any transformation applied to it. These dimensions are intricately linked to a time series' interpretation of being either a stock or flow variable.

As far as sampling is concerned, *point-in-time* sampling (common for "stock" variables) applies to data series which are measured at a specific sub-set of time periods constituting the underlying base frequency (e.g. the exchange rate on business days in a daily base frequency model). Extending the typology of Marcellino (1999), the other two types of sampling processes include *sums* (e.g. total monthly job separations from weekly separations) and *averages* (e.g. annual GDP from annualized quarterly GDP).

A series' transformation, or lack thereof, also affects its temporal aggregation properties. For example, while the (untransformed) level of payroll employment constitutes a *point-in-time* sampling process, its one-period change (Δ_1) reflects a *sum* sampling process. Finally, if the series' transformation constitutes a "long-difference" of length H greater than the base frequency ($\Delta_{H\geq 1}$), a *triangle average* accumulator is required. As the accumulator types used in the **MFSS** toolbox are introduced below, the manner in which they fit data sampling and transformation types (summarized in Table 1) are discussed.[6]

Sampling/Transformation	Level	Δ_1	$\Delta_{H\geq 1}$
Point-in-time	None	None	Sum
Sum	Sum	Sum	Sum
Average	Average	Average	Triangle Average

Table 1: Sampling/Transformation Types and Accumulator Specifications

Sum Accumulators

Sum accumulators are those where the lower frequency is the sum of a time series of higher frequency observations. For a low-frequency time series that is the sum of the previous n observations, we would have

$$\begin{aligned} y_t^l &= y_t^h + y_{t-1}^h + \cdots + y_{t-n+1}^h \\ &= Z^l(\alpha_t + \alpha_{t-1} + \cdots + \alpha_{t-n+1}) + d_t^l + \beta_t^l x_t^l + \varepsilon_t^l, \end{aligned} \tag{7}$$

where y_t^l is the observed low-frequency time series and y_t^f is an unobserved, high-frequency version of that series. We define a sum accumulator, S_t, recursively as follows.

$$S_t = \begin{cases} \alpha_t & t \in \mathcal{P}_1 \\ S_{t-1} + \alpha_t & t \notin \mathcal{P}_1 \end{cases} \tag{8}$$

To set this in the state space context, we define a calendar parameter s_t that is equal to 0 in the first period of a low-frequency period and equal to 1 in all other periods. We can then

[6] Given that the triangle average collapses to the simple average when $H = 1$, we adopt the convention (with a slight abuse of notation) of setting $H = 1$ for any series that is a level in the **MFSS** toolbox.

rewrite equation (8) as $S_t = s_t S_{t-1} + \alpha_t$. The augmented transition equation would then be

$$\begin{pmatrix} \alpha_t \\ S_t \end{pmatrix} = \begin{bmatrix} T_t & 0 \\ T_t & s_t \end{bmatrix} \begin{pmatrix} \alpha_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{bmatrix} c_t \\ c_t \end{bmatrix} + \begin{bmatrix} \gamma_t \\ \gamma_t \end{bmatrix} w_t + \begin{bmatrix} R_t \\ R_t \end{bmatrix} \eta_t. \tag{9}$$

Simple Average Accumulators

Simple average accumulators are those where the lower frequency is the average of a time series of high-frequency observations. For a low-frequency time series that is the average of the previous n observations, we would have

$$y_t^l = \frac{1}{n} \sum_{i=0}^{n} y_{t-i}^h. \tag{10}$$

To arrive at a recursive formulation for the average accumulator, we write out the averages for a 3- and 4-period case below.

$$\frac{1}{3}(y_t^h + y_{t-1}^h + y_{t-2}^h) = \frac{1}{3}y_t^h + \frac{2}{3}(\frac{1}{2}y_{t-1}^h + \frac{1}{2}y_{t-2}^h) \tag{11}$$

$$\frac{1}{4}(y_t^h + y_{t-1}^h + y_{t-2}^h + y_{t-3}^h) = \frac{1}{4}y_t^h + \frac{3}{4}(\frac{1}{3}y_{t-1}^h + \frac{2}{3}(\frac{1}{2}y_{t-2}^h + \frac{1}{2}y_{t-3}^h)) \tag{12}$$

From these examples, we can see that the $(m-1)$-period average, appropriately weighted, can be combined with the next observation to produce an m-period average. For the general case, we define an averaging calendar m_t equal to the number of high-frequency periods within the low-frequency period.[7] For example, with a monthly base frequency, a quarterly average accumulator's calender would be the repetition of (1, 2, 3). Differing numbers of high-frequency periods within a low-frequency period are allowed by changing how high the m_t counts before it resets to 1, as is required for instance with monthly observations of a weekly process. We then recursively define the average accumulator M_t as

$$M_t = \frac{1}{m_t}\alpha_t + \frac{m_t - 1}{m_t} M_{t-1}. \tag{13}$$

To augment the state space with this accumulator, we can write

$$\begin{pmatrix} \alpha_t \\ M_t \end{pmatrix} = \begin{bmatrix} T_t & 0 \\ T_t/m_t & (m_t-1)/m_t \end{bmatrix} \begin{pmatrix} \alpha_{t-1} \\ M_{t-1} \end{pmatrix} + \begin{bmatrix} c_t \\ c_t/m_t \end{bmatrix} + \begin{bmatrix} \gamma_t \\ \gamma_t/m_t \end{bmatrix} w_t + \begin{bmatrix} R_t \\ R_t/m_t \end{bmatrix} \eta_t. \tag{14}$$

Triangle Average Accumulators

A long-differenced transformation with an average sampling process requires an alteration to the simple average accumulator, and thus an additional accumulator type. To illustrate, consider a model where we observe quarterly GDP with a monthly base frequency. Denote monthly levels of GDP as X_t and its *base frequency*-differences as $x_t = X_t - X_{t-1}$. We want to relate the observed year-over-year change in quarterly GDP (i.e., the Q4/Q4 change in

[7] $m_t = t - p_t + 1$ where $p_t = \max \mathcal{P}_{1_{p_t \leq t}}$.

GDP) to the latent monthly first-differences.[8] At the end of the second year of the sample, this is the 12-month change (a long-difference) in the 3-month average of X_t:

$$
\begin{aligned}
x_{24}^{(3,12)} &= \frac{1}{3}(X_{22} + X_{23} + X_{24}) - \frac{1}{3}(X_{10} + X_{11} + X_{12}) && (15)\\
&= \frac{1}{3}[3X_{10} + 3\sum_{t=11}^{22} x_t + 2x_{23} + x_{24} - (3X_{10} + 2x_{11} + x_{12})] && (16)\\
&= \frac{1}{3}x_{11} + \frac{2}{3}x_{12} + \sum_{t=13}^{22} x_t + \frac{2}{3}x_{23} + \frac{1}{3}x_{24} && (17)
\end{aligned}
$$

We can see that the *base frequency*-differences are averaged using a set of weights that forms a trapezoid ending at time t. The name "triangle" average comes from the most common case in which the horizon of the change in the averaged quantities matches the difference in frequencies, in which case the inner portion of the trapezoid collapses and the weights form a triangle. This accumulator, with an averaging period S and a differencing period H (and $D = \min\{S, H\}$), is defined on the differenced time series so that the following relationship to its level holds,

$$
\begin{aligned}
x_t^{(S,H)} &= \frac{1}{S}\sum_{i=0}^{S-1} X_{t-i} - \frac{1}{S}\sum_{i=0}^{S-1} X_{t-i-H} && (18)\\
&= \frac{1}{S}\left(x_t + 2x_{t-1} + \cdots + \sum_{i=t-D+1}^{t-H-S+D+1} Dx_i + \cdots + 2x_{t-H-S+3} + x_{t-H-S+2}\right). && (19)
\end{aligned}
$$

To put this into recursive form, for a given horizon the m-period average can be written in terms of the $(m-1)$-period average, similar to the earlier result, as

$$x_t^{(m,H)} = \frac{1}{m}x_t^{(1,H)} + \frac{m-1}{m}x_{t-1}^{(m-1,H)}. \tag{20}$$

Since the 1-period average of an H-difference series is $x_t^{(1,H)} = X_t - X_{t-H} = \sum_{i=0}^{H-1} x_{t-i}$, we have a slightly modified form of the simple average accumulator where we have to take the sum of the past H base-frequency observations into account,

$$M_t = \frac{1}{m_t}(\alpha_t + \alpha_{t-1} + \cdots + \alpha_{t-H+1}) + \frac{m_t - 1}{m_t}M_{t-1}. \tag{21}$$

This implies that we will need to make sure that there are at least $H-1$ lags of the state we are accumulating in α_t. To augment the state with this accumulator, we can write

$$\begin{pmatrix} \alpha_t \\ \{\alpha_{t-i}\}_{i=1}^{H} \\ M_t \end{pmatrix} = \begin{bmatrix} T_t & 0 & 0 \\ & \mathcal{L}_{H-1,n} & 0 \\ \frac{T_t + I_m}{m_t} & \frac{I_{(H-1)m}}{m_t} & \frac{m_t - 1}{m_t} \end{bmatrix} \begin{pmatrix} \alpha_{t-1} \\ \{\alpha_{t-i}\}_{i=2}^{H+1} \\ M_{t-1} \end{pmatrix} + \begin{bmatrix} c_t \\ 0 \\ \frac{c_t}{m_t} \end{bmatrix} + \begin{bmatrix} \gamma_t \\ 0 \\ \frac{\gamma_t}{m_t} \end{bmatrix} w_t + \begin{bmatrix} R_t \\ 0 \\ \frac{R_t}{m_t} \end{bmatrix} \eta_t \tag{22}$$

[8] If this model is run on log-levels of U.S. GDP, as in Mariano and Murasawa (2010), this treatment involves an approximation given that the standard accounting identity used for quarterly GDP is the arithmetic average and not the geometric average of the time series. The approximation is necessary, however, if the linearity of the state space model is to be preserved. Mitchell, Smith, Weale, Wright, and Salazar (2005) finds this to be a good first-order approximation in the case of GDP.

where $\mathcal{L}_{H-1,n}$ is a lag matrix ensuring that we have a total of $H-1$ lags of the n states.[9]

Accumulated Observations

Restricting Z^l to be time-invariant above allowed us to focus on creating low-frequency versions of the latent states that can be used in the observation equation of low-frequency time series. This can provide substantial performance gains when estimating these models, since the number of observations is commonly far larger than the number of states (e.g. in dynamic factor models) and keeping the size of the state small has computational advantages. It is also possible to define accumulators with a time-varying observation matrix, Z_t^l, but only at the expense of expanding the state by as many observables requiring accumulation. Following Harvey (1989), we instead bring the observation into the state equation as follows:

$$\begin{bmatrix} y_t^b \\ y_t^l \end{bmatrix} = \begin{bmatrix} Z_t^b & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_t \\ y_t^l \end{bmatrix} + \begin{bmatrix} d_t^b \\ 0 \end{bmatrix} + \begin{bmatrix} \beta_t^b & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_t^b \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^b \\ 0 \end{bmatrix} \tag{23}$$

$$\begin{bmatrix} \alpha_t \\ y_t^l \end{bmatrix} = \begin{bmatrix} T_t & 0 \\ Z_t^l A_t^T & Z_t^l A_t^\zeta \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ y_{t-1}^l \end{bmatrix} + \begin{bmatrix} c_t \\ Z_t^l A_t^c + d_t^l + \beta_t^l x_t^l \end{bmatrix} + \begin{bmatrix} \gamma_t \\ Z_t^l A_t^\gamma \end{bmatrix} w_t + \begin{bmatrix} R_t & 0 \\ Z_t^l A_t^R & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \varepsilon_t^l \end{bmatrix}. \tag{24}$$

For models with a time-varying Z^l, this redefinition allows for the accumulators to still be calculated in the state, since it forces the redefined Z^l to be time-invariant.

4. Mixed Frequency State Space Applications

The primary contribution of the MATLAB toolbox **MFSS** is to allow for easy manipulation of the accumulator specifications described above across a variety of mixed frequency datasets and applications. The toolbox also readily handles the computation of the state decompositions, in addition to being able to match the estimation capabilities and computational performance of other packages.

To illustrate how to use the **MFSS** toolbox and appropriately handle a variety of different modeling aspects of a mixed frequency application, we use it in four different empirical applications of topical interest in the rest of this section. The four applications are: (i) a dynamic factor model of the business cycle, (ii) a trend-cycle decomposition of GDP, (iii) a small monetary policy VAR, and (iv) an estimate of the natural rate of interest. While each of these applications vary in their object of interest and empirical strategy, what they share is a tradition of being estimated in a single-frequency set-up. And, while the recent literature has developed mixed frequency estimates of (i) and (iii), the results to follow offer the first mixed frequency estimates of the trend-cycle decomposition of GDP and the natural rate of interest.

4.1. Dynamic Factor Model (DFM) of the Business Cycle

Our first application introduces the **MFSS** toolbox and its core functionality. For this introduction, we estimate perhaps one of the very first successful mixed frequency economic applications; a dynamic factor model of the business cycle. Specifically, we estimate a model similar to the mixed frequency dynamic factor model of Mariano and Murasawa (2003), which

[9]The triangle average collapses to the simple average when $H = 1$.

```
% y is a panel of 5 time series at a monthly base frequency.
% The 1st series in y is quarterly with observations in the 3rd month of the quarter.
Z = [1; nan(4,1)];
d = [nan; zeros(4,1)];
H = diag(nan(5,1));
T = nan;
Q = 1;
ssE = StateSpaceEstimation(Z, H, T, Q, 'd', d);

LB = ssE.ThetaMapping.LowerBound;
UB = ssE.ThetaMapping.UpperBound;
LB.T(1,1) = -1;
UB.T(1,1) = 1;
ssE.ThetaMapping = ssE.ThetaMapping.addRestrictions(LB, UB);

accum = Accumulator.GenerateRegular(y, {'avg','','','',''}, [3 1 1 1 1]);
ssEA = accum.augmentStateSpaceEstimation(ssE);
ssML = ssEA.estimate(y);
alphaHat = ssML.smooth(y);
```

Code Example 1: Estimating a Mixed Frequency Dynamic Factor Model

involves using quarterly GDP together with a selection of monthly indicators of economic activity to extract a common (monthly) factor that is linked to the business cycle.

Code Example 1 outlines the few steps involved in estimating the dynamic factor model using the **MFSS** toolbox. The observed vector of time series in this model have been log-differenced and standardized. We require a triangle average accumulator with a horizon of 3 months given the monthly base frequency of this model. After creating the accumulator object, we augment the state space to align with the mixed-frequency nature the data, estimate the parameters via maximum likelihood, and obtain the smoothed state given the estimated parameters.

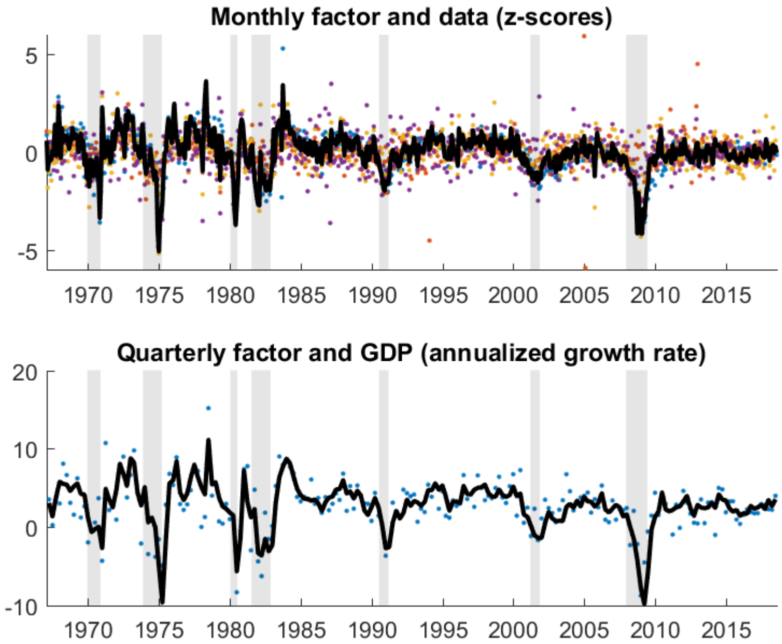
When estimating reduced form models like the one above, we often need to restrict the values the parameters can take to ensure identification. The simplest parameter restrictions occur when elements of the state space matrices need to be restricted to obey explicit bounds. This type of restriction was is in Code Example 1 to require that the AR(1) controlling the dynamics of the latent factor ensures that it remains stationary throughout estimation by restricting the appropriate element of T to be between -1 and 1.

Results for the estimation of the DFM are presented in figure 1. A couple of points are worth highlighting. First, the estimated factor is a clear coincident indicator of the business cycle. Positive realizations of the factor generally coincide with business cycle expansions, while negative realizations are typically associated with contractionary periods (gray shaded periods), with very few false signals in either case. Given the idiosyncratic variation each series exhibits beyond the common factor, it is clear that being able to jointly include both key quarterly (e.g. GDP) and monthly variables (e.g. payroll employment) into the dynamic factor model facilitates the extraction of the business cycle "signal" from the inherent noise present in many macroeconomic time series.

4.2. Small Monetary Policy Mixed Frequency VAR (MF-VAR)

The use of VARs to identify monetary policy shocks has a long tradition. Ideally, these VARs

Figure 1: DFM Estimation Results



This figure displays the time series used in the estimation of the dynamic factor model together with the estimate of the common factor. In the top panel: (i) payroll employment (blue), (ii) real personal income less transfers (orange), (iii) industrial production (yellow), (iv) real manufacturing and trade sales (purple), as well as the estimate of the factor (black) are reported in standardized units. In the bottom panel, GDP (blue) and the factor are reported in annualized growth units.

inform policymakers to what extent unanticipated monetary policy interventions impact the ultimate path of economic activity and inflationary pressures. In its most concise form, a monetary policy VAR is comprised of an economic activity variable (e.g. payroll employment, or GDP), an aggregate price variable (e.g. CPI), and the federal funds rate. Once estimated, policymakers examine impulse response functions, variance decompositions and/or forecasts of economic activity and price levels to historical policy surprises.

Even in this admittedly small system, a critical trade-off exists in the modeling decisions involving frequency and variable selection. While monthly variations in monetary policy and prices are observed, some of the most critical aggregate measures of economic activity are only observed at the quarterly frequency, most notably GDP. Alternatively, other perhaps less desirable measures of aggregate activity are available at the higher monthly frequency; such as monthly payroll employment. So, without the ability of estimating the VAR in a mixed frequency setting, the econometrician is left with a less than ideal trade-off that is likely to distort the ultimate set of inferences trying to be recovered. A perfect example of the ambiguous nature of this trade-off comes from the different choices made by Bańbura, Giannone, and Reichlin (2010) and Koop (2013) in their examination of the role of Bayesian shrinkage techniques on the performance of a small monetary policy VAR.

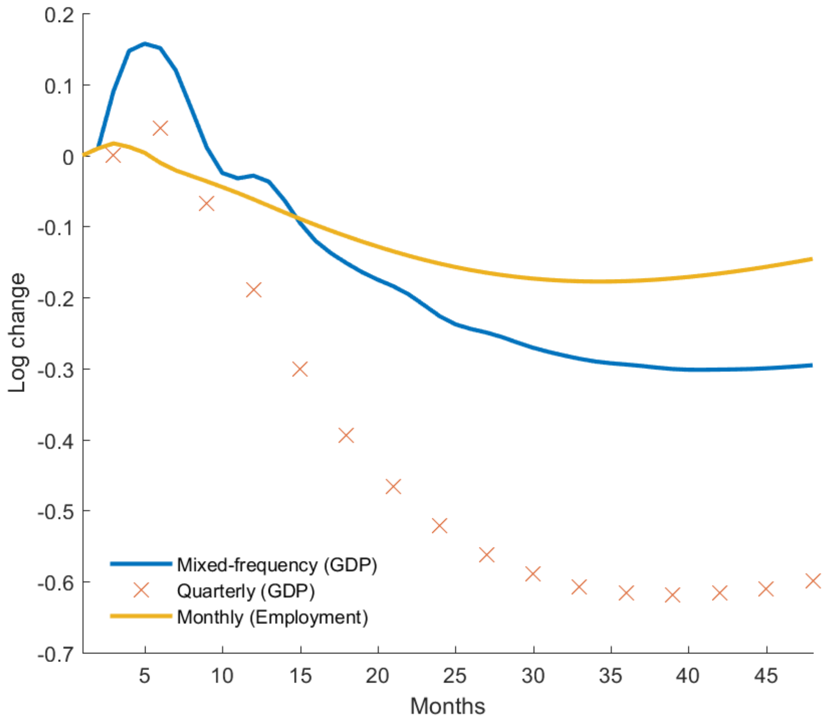
By leveraging a mixed frequency setup and the relative ease in which a model of this size can be estimated using the **MFSS** toolbox, we alleviate the potential drawbacks of their alternative modeling choices. Specifically, we estimate a four variable monthly VAR comprised of GDP, CPI, commodity prices, and the federal funds rate. Here, the mixed frequency aspect of the application stems from the fact that we observe the price indexes and the federal funds rate at a monthly frequency, and GDP at a quarterly frequency. To illustrate the different inferences such an approach can yield relative to the most natural single frequency alternatives, we examine the impulse response functions of GDP to a monetary policy shock for our mixed frequency VAR(6) model in addition to a quarterly version of the model (2 lags), as well as a monthly version that replaces GDP with monthly payroll employment as the measure of economic activity (Bańbura *et al.* (2010)).

Figure 2 reports the resulting impulse response functions for economic activity from a 100 basis point increase to the federal funds rate using data from 1965-2017.[10] Each of the three models result in similar interpretations regarding an increase in the federal funds rate: a contraction in economic activity. The differences across the single frequency quarterly and monthly models highlight the potential ambiguity facing policymakers faced with having to make assessments in real time given the publication delays associated with many headline, but lower frequency, releases like GDP. Here, the quarterly VAR that used GDP as its measure of economic activity predicts a much steeper descent as a result of a federal funds rate increase than the monthly VAR which used payroll employment.

Interestingly, the mixed frequency model seems to strike a balance between the two models in terms of its interpretation of the depth of the contraction in GDP following a federal funds rate increase. Taking a closer look at the mixed frequency model's impulse response, it is clear that the more immediate effects (within two quarters after the shock) resemble the short run impacts of the quarterly model (which also used GDP as its measure), while the longer run dynamics (beyond two years) more closely resemble the monthly frequency

[10]The code supplement reports the estimated VAR coefficients for each model. Alone, these coefficient estimates can often be difficult to interpret, and is why we choose to focus on the impulse response functions.

Figure 2: Monetary Policy VAR Estimation Results



This figure plots the estimated impulse response functions to a 100 basis point increase to the federal funds rate for GDP and payroll employment for the three different small monetary policy VARs. The impulse for payroll employment (yellow) and GDP from the mixed frequency VAR (blue) are plotted over 48 months after the shock, while the impulse for GDP from the quarterly VAR (red crosses) are shown for the parallel 12 quarters.

model. In practice, the mixed-frequency model would also facilitate a richer set of possible identification strategies to the extent that higher frequency movements in monetary policy are able to identify critical responses of interest for policymakers.

4.3. Trend-cycle Decomposition of GDP

The next application takes a closer look at separating the trend and business cycle components of GDP. Due to the inherent non-stationary behavior of time series like GDP, establishing empirical regularities as they relate to business cycle phenomenon versus the longer run movements in GDP often require such a decomposition. The trend-cycle decomposition, in particular, serves both as an object of interest in its own right (e.g. Antolin-Diaz, Drechsel, and Petrella (2017)), and also as a critical input to any multivariate analysis that attempts to explain the cyclical co-movements in the economy (e.g. Hodrick and Prescott (1997)).

Given these demands, it seems costly to require that such decomposition only be available at the frequency of observation – especially for a series as important as GDP. To estimate such a decomposition at a higher frequency than the frequency of observation naturally lends itself to a mixed frequency set up. In particular, we attempt to decompose GDP into its trend and cycle components as in Harvey and Jaeger (1993), but at the monthly frequency. This application has the added benefit of illustrating the **MFSS** toolbox's flexibility in being able to handle more complicated restrictions in the estimation process.

Evaluating the likelihood of a vector of time series given a proposed parameter vector θ^U is, in this instance, a 3-step process – an unrestricted "structural" parameter vector θ^U is domain restricted according to element-wise restriction functions $\mathcal{R}$ to θ, transformed to reduced-form parameters ψ, then element-wise transformed into the state space parameters

$$\theta_i = \mathcal{R}_i(\theta_i^U) \quad \rightarrow \quad \psi_i = \Psi_i(\theta) \quad \rightarrow \quad X_{j,i} = \mathcal{T}_{X,j}(\psi_i) \tag{25}$$

for $X \in \{Z_t, d_t, \beta_t, H_t, T_t, c_t, \gamma_t, R_t, Q_t\}$. In cases where multiple parameters are related, we define the state space matrices as functions of the underlying structural parameters.

The stochastic trend-cycle decomposition of Harvey and Jaeger (1993) is exactly a setting where such functionality is required. For instance, a version of the stochastic trend-cycle decomposition model is as follows:

$$y_t = \mu_t + \psi_t \tag{26}$$
$$\mu_t = \mu_{t-1} + \phi_{t-1} \tag{27}$$
$$\phi_t = \phi_{t-1} + \xi_t \tag{28}$$
$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \tag{29}$$

where ξ_t is normally distributed and κ_t and κ_t^* are independently normally distributed with a common variance. The structural parameters ρ and λ are restricted so that the cyclical component remains stationary with an expected period between 1.5 and 12 years. Casting

```
% y is a time series with quarterly observations placed every 3rd period
syms lambda rho sigmaKappa sigmaXi
Z = [1, 0, 1, 0];
H = 0;
T = blkdiag([1 1; 0 1], rho .* [cos(lambda), sin(lambda); -sin(lambda) cos(lambda)]);
R = [zeros(1, 3); eye(3)];
Q = diag([sigmaXi; sigmaKappa; sigmaKappa]);
ssE = StateSpaceEstimation(Z, H, T, Q, 'R', R);

accum = Accumulator.GenerateRegular(y, {'avg'}, 1);
ssEA = accum.augmentStateSpaceEstimation(ssE);

% Estimate model at a monthly frequency with starting values from Harvey & Jaeger,
% adjusting for the difference in frequency.
ssEA.ThetaMapping = ssEA.ThetaMapping.addStructuralRestriction(rho, 0, 1);
ssEA.ThetaMapping = ssEA.ThetaMapping.addStructuralRestriction(lambda, pi/72, pi/9);
ssML = ssEA.estimate(y, [0.0943; 0.9610; log(0.00003379); log(0.0000003789)]);
alpha = ssML.smooth(y);
```

Code Example 2: Estimating a Mixed Frequency Stochastic Trend-Cycle Decomposition

the system into state space form, we have

$$y_t = \begin{bmatrix}1 & 0 & 1 & 0\end{bmatrix} \tag{30}$$

$$\alpha_t = \begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \rho\cos\lambda & \rho\sin\lambda\\ 0 & 0 & -\rho\sin\lambda & \rho\cos\lambda\end{bmatrix}\alpha_{t-1} + \begin{bmatrix}0\\ \xi_t\\ \kappa_t\\ \kappa_t^*\end{bmatrix}. \tag{31}$$

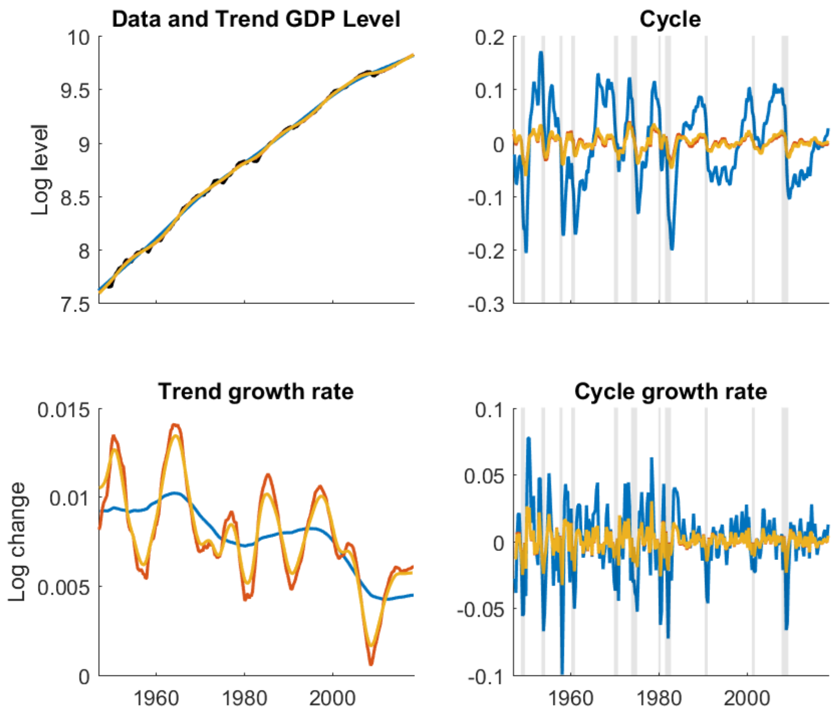
The 4 structural parameters in this model are stacked to form $\theta = [\rho, \lambda, \log\sigma_\xi^2, \log\sigma_\kappa^2]'$ and transformed to their reduced form $\psi = [\rho\cos(\lambda), \rho\sin(\lambda), -\rho\sin(\lambda), \log\sigma_\xi^2, \log\sigma_\kappa^2]^\top$. The user needs only to write out the model in terms of the underlying structural parameters to be estimated, as Code Example 2 demonstrates.

Figure 3 displays the mixed frequency version of the Harvey and Jaeger (1993) trend and cycle decomposition of GDP (top panels) as well as their growth rates (bottom panels), using GDP data from 1947q1-2018q2. For comparison purposes, we also report the same time series components for the Harvey and Jaeger (1993) trend-cycle decomposition estimated at the quarterly frequency, as well as the Hodrick-Precott (HP) filtered estimate of the decomposition (Hodrick and Prescott (1997)).[11] An appealing fact involving these comparisons is the strong similarity across each of the different models estimates. Their similarities suggest that the mixed frequency approach's ability to provide an estimate of the underlying monthly decomposition (something the other two models cannot provide) of the trend and cycle might lead practitioner's to favor the mixed frequency approach.

Though the three models do largely interpret the trend and cycle similarly, one interesting departure of the mixed frequency version is the smaller variation in the trend for GDP (see bottom left panel) relative to both of the quarterly frequency models. This departure while influencing the relative amplitude of the cycle component of GDP (see right panels), largely

[11]For the Hodrick-Prescott (HP) filter we use a smoothing parameter of 1600 which is standard for quarterly frequency series.

Figure 3: Trend-Cycle Decomposition



This figure plots the components of the three trend cycle decompositions estimated for GDP. Across all of the panels, the mixed-frequency Harvey-Jaeger decomposition is reported in blue, while the quarterly (single-frequency) Harvey-Jaeger decomposition is reported in orange. The Hodrick-Precott filter trend-cycle decomposition is reported in yellow. Recessions as defined by the NBER are shaded gray for the panels involving the cycle component (right panels).

does not change the strong coincidence of the cycle with NBER defined business cycles (shaded in gray). Additionally, the mixed frequency model's parameter estimates (e.g. the variance of the shock to trend (σ_ξ^2)) are more natural candidates for semi-calibrated models of the economy modeled at the monthly frequency.

4.4. Estimating the Natural Rate of Interest

Our final application showcases the **MFSS** toolbox's ability to calculate the full set of Kalman filter and smoother weights to create useful measures of how each series contributes to the movements in the underlying latent state. To showcase this functionality, we estimate the natural rate of interest (e.g. Laubach and Williams (2003))–an inherently latent object of substantial economic importance for policymakers. What makes this effort particularly suitable for a mixed frequency setup is that many of the variables incorporated in the estimation of the natural rate of interest (e.g. GDP, inflation, and the federal funds rate) are observed at very different frequencies, ranging from the daily frequency (e.g. the federal funds rate) to the quarterly frequency (e.g. GDP). To date, what this has meant is that estimates of the natural rate of interest available to policymakers have been restricted to a quarterly frequency. By fully leveraging the capabilities of the **MFSS** toolbox, we provide the first estimate, to our knowledge, of the natural rate of interest at the *monthly* frequency.

In order to estimate this model at a monthly frequency, the state space specification of Laubach and Williams (2003) must be adapted so that monthly GDP is a latent variable.[12] The estimation code is given in Code Example 3. We initialize the structural parameters at the estimated values found by Laubach and Williams (2003) and attempt to mimic their handling of the initial state. Further details can be found in the accompanying User Guide.

Figure 4 reports state decompositions using the filter/smoother weights for each of the time series involved in the estimation of the natural rate of interest over the full period of the estimation sample (1959q1-2018q1).[13] For comparison purposes, we report along with the mixed frequency (monthly) estimate of the natural rate of interest (top panel), an estimate from the traditional quarterly frequency model (bottom panel). In each panel, the contributions from each series involved in the estimation of the natural rate of interest (in black) is reported (e.g. GDP, inflation, real rate of interest, oil import prices, and import prices).

Across both panels, it is clear that the primary drivers of the fluctuations in the natural rate of interest have been the real rate of interest and GDP. While the real rate of interest has supported higher levels of the natural rate of interest in some periods, 1980s–2000s, GDP has largely been bringing the natural rate of interest lower since the 1960s. While both versions of the model are in agreement with the countervailing roles of GDP's and the real rate of interest's influence on the natural rate of interest, the mixed frequency version interprets a stronger divergence between the two series in the 1980-2000 period and less divergence in the period since 2010 compared to the quarterly version of the model.

The capabilities of the **MFSS** package to incorporate the filter/smoother weight calculations for many types of variables within a mixed frequency state space formulation are underscored

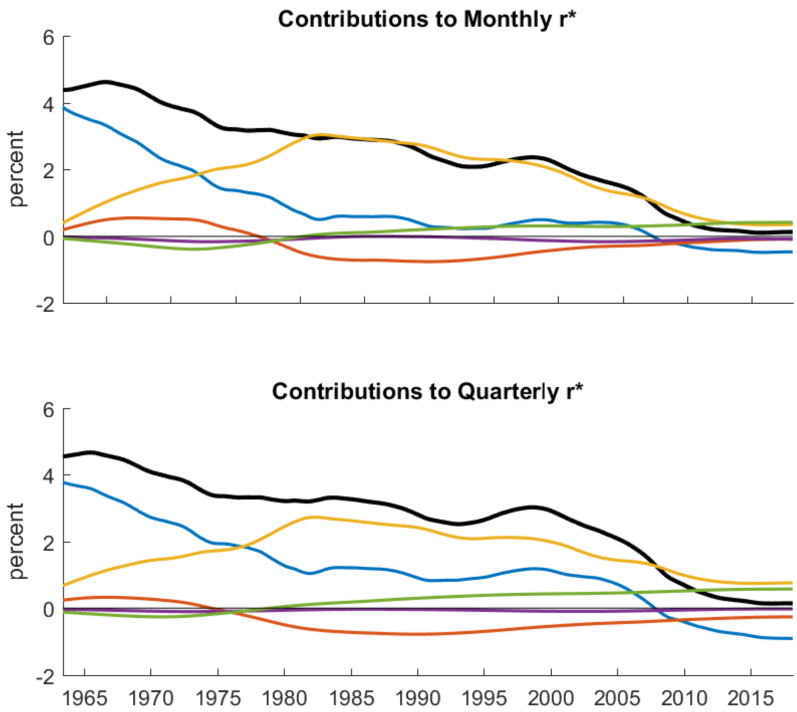
[12] The monthly data for estimating the natural rate of interest follows the data of Laubach and Williams (2003) using quarterly GDP, monthly changes in the PCE price index, the monthly real interest rate constructed using inflation expectations from an AR model of observed inflation, and interpolated measures of import prices and import oil prices.

[13] For an explicit formulation of the model equations involved with estimating the natural rate of interest, see the accompanying code in addition to Laubach and Williams (2003).

```
% Estimated parameters from Laubach-Williams code
thetaR = [1.564 -0.609 -0.055 0.574 0.369 0.041 0.002 0.035 1.414 ...
  log(0.346^2) log(0.760^2) log(0.597^2)]';

% Define model
syms a1 a2 a3 b1 b2 b3 b4 b5 c sigma2Ystar sigma2IS sigma2PC
Z = [0 0 1 0 zeros(1,5);
    -b3 0 b3 0 zeros(1,5)];
beta = [zeros(1,5); b1 b2 (1-b1-b2) b4 b5];
H = diag([0 sigma2PC]);
T = [1 0 0 0 1 0 0 0 0; 1 zeros(1,8);
  1-a1 -a2 a1 a2 1-(12*c*a3/2) -12*c*a3/2 -a3/2 -a3/2 0; 0 0 1 0 zeros(1,5);
  zeros(1,4) 1 zeros(1,4); zeros(1,4) 1 zeros(1,4);
  zeros(1,6) 1 zeros(1,2); zeros(1,6) 1 zeros(1,2);
  0 0 0 0 12*c 0 1 0 0];
gamma = [zeros(2); a3/2 a3/2; zeros(6,2)];
R = [1 0 0 0; zeros(1,4);
     1 0 1 0 ; zeros(1,4);
     0 lambda_g/3 0 0; zeros(1,4);
     0 0 0 lambda_z/(3*a3); zeros(1,4);
     zeros(1,4)];
Q = diag([sigma2Ystar sigma2Ystar sigma2IS sigma2IS]);

ssE = StateSpaceEstimation(Z, H, T, Q, 'beta', beta, 'gamma', gamma, 'R', R);

% Bounds on the slopes of the IS and Phillips curves
ssE.ThetaMapping = ssE.ThetaMapping.addStructuralRestriction(a3, [], -0.0025);
ssE.ThetaMapping = ssE.ThetaMapping.addStructuralRestriction(b3, 0.025, []);

% Augment to match mixed frequency
accum = Accumulator.GenerateRegular(YM, {'avg', ''}, [1 0]);
ssEA = accum.augmentStateSpaceEstimation(ssE);

% Initial state
interpGDP = interp1(3:3:size(gdpM,1), gdpM(3:3:end), 1:size(gdpM))';
lagGDP = 100*interpGDP(25:26)';
gInit = 0.9394933/3;
ystarInit = [809.7823659 809.7823659-gInit];
rstarInit = 12*gInit;
ssE.a0 = [ystarInit, lagGDP, repmat(gInit, [1 2]), zeros(1, 2), rstarInit]';
ssE.P0 = 0.2*eye(9);
ssE.P0(3,3) = 0; ssE.P0(4,4) = 0;
ssEA.a0 = [ssE.a0; 100*mean(lagGDP)];
ssEA.P0 = blkdiag(ssE.P0, 0.2);

% Estimate
[ssMOpt, ~, thetaMOpt] = ssEA.estimate(Y, thetaR, X, W);

alphaM = ssMOpt.smooth(Y, X, W);
[yContribM, paramContribM, inflContribM, ratesContribM] = ...
  ssMOpt.decompose_smoothed(Y, X, W);
```

Code Example 3: Estimating the Natural Rate of Interest

Figure 4: Natural Rate of Interest Model



This figure reports contributions of each time series to an estimate of the natural rate of interest ($r^\star$). The smoothed estimate of the natural rate of interest is reported in black, while contributions to it driven by (i) GDP are reported in blue, (ii) inflation are reported in orange, (iii) the real rate of interest are reported in yellow, (iv) oil import prices are reported in purple, and (v) import prices are reported in green.

by this application. As the development of higher frequency identification strategies become more prevalent in economic and financial applications, these capabilities of the **MFSS** package will only further elucidate identification challenges and implications of modeling assumptions.

5. Conclusion

Mixed frequency modeling through the use of state space methods is a valuable tool for a wide array of economic and financial applications. Previously, using these methods required substantial effort by practitioners to implement even in the simplest of models. It is our hope that this article together with the availability of our **MFSS** toolbox will increase the accessibility of mixed frequency state space modeling and facilitate its adoption in a wider array of empirical applications. By outlining its use across four very different types of applications, we were able to highlight the unique set of capabilities of the **MFSS** toolbox. In each case, adopting a mixed frequency framework was shown to facilitate better variable selection, more parsimonious modeling choices, as well as a wider variety of inference possibilities.

The User's Guide accompanying this article contains many more examples that may be of interest to readers interested in using these methods. In the future, it will be important to extend the capabilities of this package to incorporate the frontier in shrinkage estimation strategies given the growth of densely parameterized models. Furthermore, as micro-economic applications gain more access to higher frequency panel datasets, it will also be likely that the relevant applications for this suite of methods will extend to these additional fields within economics and other social sciences.

6. Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The opinions expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System. We would like to thank Bill Kluender for excellent research assistance.

References

Antolin-Diaz J, Drechsel T, Petrella I (2017). "Tracking the Slowdown in Long-Run GDP Growth." *The Review of Economics and Statistics*, **99**(2), 343–356. doi:10.1162/REST_a_00646. https://doi.org/10.1162/REST_a_00646, URL https://doi.org/10.1162/REST_a_00646.

Aruoba SB, Diebold FX, Scotti C (2009). "Real-Time Measurement of Business Conditions." *Journal of Business & Economic Statistics*, **27**(4), 417–427.

Bańbura M, Giannone D, Reichlin L (2010). "Large Bayesian Vector Auto Regressions." *Journal of Applied Econometrics*, **25**(1), 71–92.

Brave S, Butters A (2012). "Diagnosing the Financial System: Financial Conditions and Financial Stress." *International Journal of Central Banking*, **8**(2), 191–293. ISSN 1815-4654.

Brave S, Butters RA, Justiniano A (2016). "Forecasting Economic Activity with Mixed Frequency Bayesian VARs." *Working Paper Series WP-2016-5*, Federal Reserve Bank of Chicago.

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software, Articles*, **76**(1), 1–32. ISSN 1548-7660.

Commandeur J, Koopman S, Ooms M (2011). "Statistical Software for State Space Methods." *Journal of Statistical Software, Articles*, **41**(1), 1–18. ISSN 1548-7660. doi:10.18637/jss.v041.i01. URL https://www.jstatsoft.org/v041/i01.

Doornik JA (2010). *OxMetrics*. URL http://www.doornik.com/products.html.

Durbin J, Koopman SJ (2012). *Time Series Analysis by State Space Methods: Second Edition (Oxford Statistical Science Series)*. Oxford University Press. ISBN 019964117X.

Foroni C, Marcellino M (2014). "Mixed-Frequency Structural Models: Identification, Estimation, and Policy Analysis." *Journal of Applied Econometrics*, **29**(7), 1118–1144.

Ghysels E, Santa-Clara P, Valkanov R (2004). "The MIDAS Touch: Mixed Data Sampling Regression Models." *CIRANO Working Papers 2004s-20*, CIRANO.

Harvey A, Chung CH (2000). "Estimating the Underlying Change in Unemployment in the UK." *Journal of the Royal Statistical Society Series A*, **163**(3), 303–309.

Harvey A, Jaeger A (1993). "Detrending, Stylized Facts and the Business Cycle." *Journal of Applied Econometrics*, **8**(3), 231–47.

Harvey AC (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey AC, Pierse RG (1984). "Estimating Missing Observations in Economic Time Series." *Journal of the American Statistical Association*, **79**(385), 125–131.

Helske J (2018). ***KFAS**: Kalman Filter and Smoother for Exponential Family State Space Models*. URL https://cran.r-project.org/package=KFAS.

Hodrick RJ, Prescott EC (1997). "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking*, **29**(1), 1–16. URL https://ideas.repec.org/a/mcb/jmoncb/v29y1997i1p1-16.html.

Kim TB (2010). "Bayesian Mixed Frequency Estimation of DSGE models." *Technical report*, Duke University.

Koop GM (2013). "Forecasting with Medium and Large Bayesian VARs." *Journal of Applied Econometrics*, **28**(2), 177–203.

Koopman SJ, Doornik JA, Shephard N (2010a). *SsfPack*. URL http://www.ssfpack.com/.

Koopman SJ, Durbin J (2000). "Fast Filtering and Smoothing for Multivariate State Space Models." *Journal of Time Series Analysis*, **21**(3), 281–296. ISSN 1467-9892.

Koopman SJ, Durbin J (2003). "Filtering and Smoothing of State Vector for Diffuse State-Space Models." *Journal of Time Series Analysis*, **24**(1), 85–98.

Koopman SJ, Harvey A (2003). "Computing Observation Weights for Signal Extraction and Filtering." *Journal of Economic Dynamics and Control*, **27**(7), 1317 – 1333. ISSN 0165-1889.

Koopman SJ, Harvey A, Doornik JA, Shephard N (2010b). *STAMP*. URL http://www.stamp-software.com/.

Kuzin V, Marcellino M, Schumacher C (2011). "MIDAS vs. Mixed-frequency VAR: Nowcasting GDP in the Euro Area." *International Journal of Forecasting*, **27**(2), 529–542.

Laubach T, Williams JC (2003). "Measuring the Natural Rate of Interest." *The Review of Economics and Statistics*, **85**(4), 1063–1070. doi:10.1162/003465303772815934. https://doi.org/10.1162/003465303772815934, URL https://doi.org/10.1162/003465303772815934.

Marcellino M (1999). "Some Consequences of Temporal Aggregation in Empirical Analysis." *Journal of Business & Economic Statistics*, **17**(1), 129–136.

Mariano RS, Murasawa Y (2003). "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series." *Journal of Applied Econometrics*, **18**(4), 427–443.

Mariano RS, Murasawa Y (2010). "A Coincident Index, Common Factors, and Monthly Real GDP." *Oxford Bulletin of Economics and Statistics*, **72**(1), 27–46.

Mitchell J, Smith RJ, Weale MR, Wright S, Salazar EL (2005). "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth." *The Economic Journal*, **115**(501), F108–F129. ISSN 1468-0297.

Perktold J, Seabold S, Taylor J (2017). *StatsModels*. URL https://www.statsmodels.org/stable/index.html.

Proietti T (2006). "Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited." *The Econometrics Journal*, **9**(3), 357–372.

Sargent TJ, Stachurski J (2018). "A First Look at the Kalman Filter." https://lectures.quantecon.org/py/kalman.html.

Schorfheide F, Song D (2015). "Real-Time Forecasting With a Mixed-Frequency VAR." *Journal of Business & Economic Statistics*, **33**(3), 366–380.

Shumway RH, Stoffer DS (1982). "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm." *Journal of Time Series Analysis*, **3**(4), 253–264. doi:10.1111/j.1467-9892.1982.tb00349.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1982.tb00349.x.

A. Estimating the Latent State

We are interested in the estimate of the latent state from a state space model with observed data $\{y_t, x_t, w_t\}$ and time-varying parameters of the form

$$y_t = Z_t\alpha_t + d_t + \beta_t x_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, H_t) \tag{32}$$

$$\alpha_t = T_t\alpha_{t-1} + c_t + \gamma_t w_t + R_t\eta_t \qquad \eta_t \sim \mathcal{N}(0, Q_t) \tag{33}$$

$$\alpha_1 \sim \mathcal{N}(a_1, P_1) \tag{34}$$

A.1. Multivariate Filter

As shown in Durbin and Koopman (2012) (§4.3), the filtered estimates of the state, $a_t = \mathbb{E}(\alpha_t|Y_t)$ and $P_t = \mathrm{Var}(\alpha_t|Y_t)$, are given by

$$a_{t+1} = T_{t+1}a_t + c_{t+1} + \gamma_{t+1}w_{t+1} + K_t v_t \tag{35}$$

$$F_t = Z_t P_t Z_t^\top + H_t \tag{36}$$

$$P_{t+1} = T_{t+1}P_t L_t^\top + R_{t+1}Q_{t+1}R_{t+1}^\top \tag{37}$$

$$K_t = T_{t+1}P_t Z_t^\top F_t^{-1} \tag{38}$$

$$v_t = y_t - Z_t a_t - d_t - \beta_t x_t \tag{39}$$

$$L_t = T_{t+1} - K_t Z_t \tag{40}$$

and the smoothed estimates, $\hat{\alpha}_t = \mathbb{E}(\alpha_t|Y_n)$ and $V_t = \mathrm{Var}(\alpha_t|Y_n)$ are given by

$$\hat{\alpha}_t = a_t + P_t r_t \tag{41}$$

$$V_t = P_t - P_t N_t P_t \tag{42}$$

$$r_t = Z_t^\top F_t^{-1} v_t + L_t^\top r_{t+1} \tag{43}$$

$$N_t = Z_t^\top F_t^{-1} Z_t + L_t^\top N_{t+1} L_t \tag{44}$$

where $r_{n+1} = 0$ and $N_{n+1} = 0$.

A.2. Univariate Filter

Substantial computational gains are available using the univariate Kalman filter by avoiding the inversion of the F_t matrix above (as well as enabling the use of the exact initial filter, see below). For more details, see Durbin and Koopman (2012) (§6.4).

When any H_t is non-diagonal, the observation equation is transformed by taking the LDL factorization of the H_t matrices. The transformed parameters are marked with ^{u} to denote that they have been transformed to be suitable for the univariate filter.

$$y_t^u = Z_t^u\alpha_t + d_t^u + \beta_t^u x_t + \varepsilon_t^u \qquad \varepsilon_t^u \sim N(0, H_t^u) \tag{45}$$

$$y_t^u = C_t^{-1}y_t \qquad Z_t^u = C_t^{-1}Z_t \qquad d_t^u = C_t^{-1}d_t \qquad \beta_t^u = C_t^{-1}\beta_t \qquad \varepsilon_t^u = C_t^{-1}\varepsilon_t \tag{46}$$

$$H_t = C_t H_t^u C_t^\top \tag{47}$$

The multivariate filter and smoother is then modified by computing several quantities via the univariate filter on the transformed system:

$$a_{t,i+1} = a_{t,i} + K_{t,i}v_{t,i} \tag{48}$$
$$P_{t,i+1} = P_{t,i} - K_{t,i}F_{t,i}K_{t,i}^\top \tag{49}$$
$$v_{t,i} = y_{t,i} - Z_{t,i}a_{t,i} - d_{t,i} - \beta_{t,i}x_t \tag{50}$$
$$F_{t,i} = Z_{t,i}P_{t,i}Z_{t,i}^\top + H_{t,i} \tag{51}$$
$$K_{t,i} = P_{t,i}Z_{t,i}^\top F_{t,i}^{-1} \tag{52}$$
$$a_{t+1,1} = T_{t+1}a_{t,p+1} + c_{t+1} + \gamma_{t+1}w_{t+1} \tag{53}$$
$$P_{t+1,1} = T_{t+1}P_{t,p+1}T_{t+1}^\top + R_{t+1}Q_{t+1}R_{t+1}^\top \tag{54}$$

where $y_{i,t}$ is the ith element of y_t^u, $Z_{t,i}$ is the ith row of Z_t^u, $d_{t,i}$ is the ith element of d_t^u, $\beta_{t,i}$ is the ith row of β_t^u, and $H_{t,i}$ is the ith diagonal element of H_t^u. The filtered estimates of the state $a_t = a_{t,1}$ and $P_t = P_{t,1}$ are equivalent to those computed in the multivariate filter above. Note that $F_{t,i}$ is a scalar, so that $F_{t,i}^{-1}$ is simply scalar division instead of more computationally expensive matrix inversion. However, this comes at the cost of there being no convenient transformation of the quantities $v_{t,i}$, $F_{t,i}$ or $K_{t,i}$ that recovers the multivariate versions computed above.

Similarly, the univariate smoother provides $r_t = r_{t,0}$ and $N_t = N_{t,0}$:

$$r_{t,i-1} = Z_{t,i}^\top F_{t,i}^{-1}v_{t,i} + L_{t,i}^\top r_{t,i} \tag{55}$$
$$N_{t,i-1} = Z_{t,i}^\top F_{t,i}^{-1}Z_{t,i} + L_{t,i}^\top N_{t,i}L_{t,i} \tag{56}$$
$$L_{t,i} = I_m - K_{t,i}Z_{t,i} \tag{57}$$
$$r_{t-1,p} = T_t^\top r_{t,0} \tag{58}$$
$$N_{t-1,p} = T_{t-1}^\top N_{t,0}T_{t-1} \tag{59}$$

where $r_{t+1,p} = 0$ and $N_{t+1,p} = 0$.

A.3. Exact Initial Kalman Filter

When the state α_t is stationary, the initial values a_1 and P_1 can be computed as the unconditional mean and variance of the state given the system parameters by inverting the transition equation.

To handle cases where some states are non-stationary, the state is separated into states with known variance (those that are stationary) and those that are initialized as diffuse (the non-stationary states). Let $\tilde{R}$ be a selection matrix with columns from the identity such that the initial shock η_1 is applied to the states with known variance. The selection matrix A is composed of the columns of the identity matrix associated with the diffuse states such that taking the limit as $\kappa \to \infty$ allows the diffuse states to have infinite initial variance where the initial values are given by

$$a_1 = a + \tilde{R}\eta_1 + A\delta \qquad \eta_1 \sim N(0, \tilde{Q}) \qquad \delta \sim N(0, \kappa I) \tag{60}$$
$$P_1 = P_{*,1} + \kappa P_{\infty,1} \qquad P_{*,1} = \tilde{R}\tilde{Q}\tilde{R}^\top \qquad P_{\infty,1} = AA^\top \tag{61}$$

The unconditional mean, a, and variance, $P_{*,1}$, of the state are computed by inverting the stationary portion of the system. Any elements of a associated with non-stationary or diffuse states are set to 0. Letting $\tilde{T} = \tilde{R}^\top T_1 \tilde{R}$ and $\tilde{c} = \tilde{R}^\top c_1$, this is accomplished by

$$\tilde{R}^\top a = (I_m - \tilde{T})^{-1}\tilde{c} \tag{62}$$

$$\text{vec}(\tilde{R}\tilde{Q}\tilde{R}^\top) = (I_{m^2} - \tilde{T}\otimes\tilde{T})^{-1}\text{vec}(\tilde{R}^\top R_1 Q_1 R_1^\top \tilde{R}) \tag{63}$$

When all states are stationary, this initialization collapses down to the simple case where $\alpha_1 \sim N(a_1, P_1)$, where a_1 and P_1 are the unconditional mean and variance of the state, determined by inverting the full system.

Given this initialization, the univariate filter recursions must be altered to separate the states with finite and infinite variances (see Durbin and Koopman (2012) (§5.2) and Koopman and Durbin (2000)):

$$F_{*,t,i} = Z_{t,i}P_{*,t,i}Z_{t,i}^\top + H_{t,i} \qquad F_{\infty,t,i} = Z_{t,i}P_{\infty,t,i}Z_{t,i}^\top \tag{64}$$

$$K_{*,t,i} = P_{*,t,i}Z_{t,i}^\top F_{*,t,i}^{-1} \qquad K_{\infty,t,i} = P_{\infty,t,i}Z_{t,i}^\top F_{\infty,t,i}^{-1} \tag{65}$$

$$a_{t,i+1} = \begin{cases} a_{t,i} + K_{*,t,i}v_{t,i} & F_{\infty,t,i} = 0 \\ a_{t,i} + K_{\infty,t,i}v_{t,i} & F_{\infty,t,i} \neq 0 \end{cases} \tag{66}$$

$$P_{*,t,i+1} = \begin{cases} P_{*,t,i} - K_{*,t,i}K_{*,t,i}^\top F_{*,t,i} & F_{\infty,t,i} = 0 \\ P_{*,t,i} - (K_{*,t,i}K_{\infty,t,i}^\top + K_{\infty,t,i}K_{*,t,i}^\top - K_{\infty,t,i}K_{\infty,t,i}^\top)F_{*,t,i} & F_{\infty,t,i} \neq 0 \end{cases} \tag{67}$$

$$P_{\infty,t,i+1} = \begin{cases} P_{\infty,t,i} & F_{\infty,t,i} = 0 \\ P_{\infty,t,i} - K_{\infty,t,i}K_{\infty,t,i}^\top F_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases} \tag{68}$$

$$a_{t+1,1} = T_{t+1}a_{t,p+1} + c_{t+1} + \gamma_{t+1}w_{t+1} \tag{69}$$

$$P_{\infty,t+1,1} = T_{t+1}P_{\infty,t,p+1}T_{t+1}^\top \qquad P_{*,t+1,1} = T_{t+1}P_{*,t,p+1}T_{t+1}^\top + R_{t+1}Q_{t+1}R_{t+1}^\top \tag{70}$$

For any set of system parameters where the state can be identified, there exists some time d such that $F_{\infty,d,i} = 0$ for all i. For time $t \geq d$, the simpler Kalman filter recursion above can be employed with $P_{t,i} = P_{*,t,i}$.

The smoother must be similarly altered so that beginning at $t = d$, the computation of $r_{t,i}$ is expanded to account for the initialization:

$$L_{\infty,t,i} = I_m - K_{\infty,t,i}Z_{t,i} \qquad L_{*,t,i} = I_m - K_{*,t,i}Z_{t,i} \tag{71}$$

$$L_{t,i}^{(0)} = (K_{\infty,t,i} - K_{*,t,i})Z_{t,i}F_{*,t,i}F_{\infty,t,i}^{-1} \tag{72}$$

$$r_{t,i-1}^{(0)} = \begin{cases} Z_{t,i}^\top F_{*,t,i}^{-1}v_{t,i} + L_{*,t,i}^\top r_{t,i}^{(0)} & F_{\infty,t,i} = 0 \\ L_{\infty,t,i}^\top r_{t,i}^{(0)} & F_{\infty,t,i} \neq 0 \end{cases} \tag{73}$$

$$r_{t,i-1}^{(1)} = \begin{cases} r_{t,i}^{(1)} & F_{\infty,t,i} = 0 \\ Z_{t,i}^\top F_{\infty,t,i}^{-1}v_{t,i} + L_{t,i}^{(0)\top}r_{t,i}^{(0)} + L_{\infty,t,i}^\top r_{t,i}^{(1)} & F_{\infty,t,i} \neq 0 \end{cases} \tag{74}$$

$$r^{(0)}_{t-1,p} = T_t^\top r^{(0)}_{t,0} \qquad r^{(1)}_{t-1,p} = T_t^\top r^{(1)}_{t,0} \tag{75}$$

$$\hat{\alpha}_t = a_t + P_{*,t,1} r^{(0)}_{t,0} + P_{\infty,t,1} r^{(1)}_{t,0} \tag{76}$$

where $r^{(0)}_{d,p} = r_{d,p}$ and $r^{(1)}_{d,p} = 0$. Note that $L_{\infty,t,i}$ and $L^{(0)}_{t,i}$ only need to be computed when $F_{\infty,t,i} \neq 0$ and $L_{*,t,i}$ only needs to be computed when $F_{\infty,t,i} = 0$.

For the smoothed variance of the state,

$$V_t = P_{*,t,1} - P_{*,t,1} N^{(0)}_{t,0} P_{*,t,1} - \left(P_{\infty,t,1} N^{(1)}_{t,0} P_{*,t,1}\right)^\top - P_{\infty,t,1} N^{(1)}_{t,0} P_{*,t,1} - P_{\infty,t,1} N^{(2)}_{t,0} P_{\infty,t,1} \tag{77}$$

$$N^{(0)}_{t-1,p} = T_t^\top N^{(0)}_{t,0} T_t \qquad N^{(1)}_{t-1,p} = T_t^\top N^{(1)}_{t,0} T_t \qquad N^{(2)}_{t-1,p} = T_t^\top N^{(1)}_{t,0} T_t \tag{78}$$

where when $F_{\infty,t,i} = 0$,

$$N^{(0)}_{t,i-1} = Z_{t,i}^\top F^{-1}_{*,t,i} Z_{t,i} + L^\top_{*,t,i} N^{(0)}_{t,i} L_{*,t,i} \tag{79}$$

$$N^{(1)}_{t,i-1} = N^{(1)}_{t,i} \tag{80}$$

$$N^{(2)}_{t,i-1} = N^{(2)}_{t,i} \tag{81}$$

and when $F_{\infty,t,i} \neq 0$,

$$N^{(0)}_{t,i-1} = L^\top_{\infty,t,i} N^{(0)}_{t,i} L_{\infty,t,i} \tag{82}$$

$$N^{(1)}_{t,i-1} = Z^{*\top}_{t,i} F^{-1}_{\infty,t,i} Z_{t,i} + L^\top_{\infty,t,i} N^{(0)}_{t,i} L^{(0)}_{t,i} + L^\top_{\infty,t,i} N^{(1)}_{t,i} L_{\infty,t,i} \tag{83}$$

$$\begin{aligned} N^{(2)}_{t,i-1} = {} & Z^\top_{t,i} F^{-2}_{\infty,t,i} Z_{t,i} F_{*,t,i} + L^{(0)\top}_{t,i} N^{(1)}_{t,i} L^{(0)}_{t,i} \\ & + L^\top_{\infty,t,i} N^{(1)}_{t,i} L^{(0)}_{t,i} + L^{(0)\top}_{t,i} N^{(1)}_{t,i} L_{\infty,t,i} + L^\top_{\infty,t,i} N^{(2)}_{t,i} L_{\infty,t,i} \end{aligned} \tag{84}$$

where $N^{(0)}_{d,p} = N_{d,p}$ and $N^{(1)}_{d,p} = N^{(2)}_{d,p} = 0$.

B. Parameter Estimation

B.1. General Maximum Likelihood Estimation

The likelihood of of data $y_1, \ldots, y_n$ in the standard filter as shown in Durbin and Koopman (2012) (§7.2) is given by the prediction error decomposition:

$$\log L(Y_n) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \left(\log |F_t| + v_t^\top F^{-1}_{t,i} v_t \right) \tag{85}$$

For the univariate filter, the same decomposition works in the univariate context:

$$\log L(Y_n) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \sum_{i=1}^{p} \log F_{t,i} + v^2_{t,i} / F_{t,i} \tag{86}$$

The likelihood for the exact initial filter allows for some simplifications in $F_{\infty,t,i}$:

$$\log L_d(Y_n) = -\frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{p}\iota_{t,i}\log 2\pi - \frac{1}{2}\sum_{t=1}^{d}\sum_{i=1}^{p} w_{t,i} - \frac{1}{2}\sum_{t=d}^{n}\sum_{i=1}^{p}\iota_{t,i}\left(\log F_{t,i} + v_{t,i}^2/F_{t,i}\right) \quad (87)$$

where $\iota_{t,i} = 1$ if $F_{*,t,i} \neq 0$ or $t > d$, and

$$w_{t,i} = \begin{cases} \iota_{t,i}(\log(F_{*,t,i}) + v_{t,i}^{(0)2}/F_{*,t,i}) & F_{\infty,t,i} = 0 \\ \log F_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases} \quad (88)$$

Since these quantities are naturally produced by the Kalman filter, this is the preferred method to calculate the likelihood.

For parameter estimation, the elements of the parameter matrices Z_t, d_t, β_t, H_t, T_t, c_t, γ_t, R_t, and Q_t are divided into those that are known and those that are to be estimated as a function of the parameter vector θ. We are interested in the maximum likelihood estimate of those parameters given data $Y_n = \{y_1, \ldots, y_n\}$.

We are interested in the maximum likelihood estimate of a set of *structural parameters* (θ) given data $y_1, \ldots, y_n$ where elements of the *state space parameters* (Z_t, d_t, β_t, H_t, T_t, c_t, γ_t, R_t, and Q_t) depend on θ. For simplicity, the state space parameters will be restricted such that each of their scalar elements must be a function of a single element of a vector of the *reduced form parameters* (ψ) which depend on θ. The reduced form parameters are unrestricted in how they may depend on θ, allowing for a rich specification of parameter restrictions. In many estimations, there will be the same number of structural parameters and reduced form parameters.

More formally, let $\theta^U \in \mathbb{R}^{n_\theta}$ and $\psi \in \mathbb{R}^{n_\psi}$ where $n_\theta \leq n_\psi$. Define a set of bounds $\underline{\theta}$ and $\bar{\theta}$ such that each element θ_i of $\theta \in \mathbb{R}^{n_\theta}$, $\underline{\theta}_i \leq \theta_i \leq \bar{\theta}_i$. Define a set of functions $\Psi_i : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ such that each element of ψ is a function of the structural parameters, $\psi_i = \Psi_i(\theta)$. Except in cases of cross-parameter restrictions, $n_\theta = n_\psi$ and $\Psi_i(\theta) = \theta_i$ so that $\psi = \theta$.

Additionally define a set of functions $\tau_{X_{i,j}} : \mathbb{R} \to \mathbb{R}$ for $X \in \{Z_t, d_t, \beta_t, H_t, T_t, c_t, \gamma_t, R_t, Q_t\}$ such that each element of the parameter matrices to be estimated is a transformation of an element of ψ. Common $\tau_{X_{i,j}}$ transformations include the identity, exponential, negative exponential, and logistic transformations to allow for an unbounded $\psi \in \mathbb{R}^{n_\psi}$ while maintaining bounds on the parameter matrices. When estimating models with mixed-frequency data, all state space parameters shared between the high-frequency and low-frequency states depend on the same reduced form parameters simply by using different $\tau_{X_{i,j}}$ functions. In almost all cases, the specification of the $\tau_{X_{i,j}}$ functions will be done automatically to account for state space parameter bounds and accumulator definitions. User definitions of $\tau_{X_{i,j}}$ should be rare.

With the likelihood of an unconstrained parameter vector θ^U defined, the optimization can be performed using gradient ascent. In practice, it has been observed that further improvements in the likelihood are possible using simplex based optimization methods once the gradient ascent method has converged. Due to this, the two methods are repeated until neither is able to improve the likelihood using standard convergence criteria.

B.2. VAR Estimation

For VAR models, estimation via the EM algorithm of Shumway and Stoffer (1982) is computationally more convenient. To do so, we need to compute $J_t = \text{Cov}(\alpha_t, \alpha_{t+1}|Y_n)$ as shown in

Durbin and Koopman (2012) (§4.7),

$$J_t = P_t L_t^{\dagger\top}(I_m - N_{t+1}P_{t+1}) \tag{89}$$

where $L_t^\dagger$ is defined in C.1. Using this and quantities from A, the algorithm requires iterating between the two steps:

1. Calculate $\hat{\alpha}_t^{i+1}$, V_t^{i+1}, J_t^{i+1} using parameters T^i, c^i, Q^i and a_0^i.
2. Calculate $a_0^{i+1} = \hat{\alpha}_0^i$ and estimate state space parameters T^{i+1}, c^{i+1}, Q^{i+1} given the estimates of the state $\hat{\alpha}_t^{i+1}$, V_t^{i+1}, J_t^{i+1} according to a modified OLS estimate.

This method avoids the computation of gradients of the likelihood at the expense of computing the variance and covariance of the state. As a result, it scales much better as the size of the state increases. While it does require a fixed value for the variance of the initial state, if P_0 needs to be estimated the optimal values from the EM algorithm can be used as a starting point for general maximum likelihood estimation.

C. State Decompositions

Since all of the Kalman filter and smoother calculations are linear, we can decompose the estimated states by the effects of the data and parameters on the state in each period. In general, we will be creating a four-dimensional object since each state $k \in \{1, \ldots, m\}$ at time t is affected by data series $i \in \{1, \ldots, p\}$ from time j. Depending on the use of the decompositions, this 4D object will be collapsed in different ways. For example, if we were suspicious of the validity of an observation, we could inspect how much it affects a state variable of interest, in which case we could select a particular j, k, and i and plot across t. More commonly, we will be interested in plots of how a given state variable is determined by the data, in which case we sum across the origin dates of the data j and plot the effect of each data series on a given state i across time t.

C.1. Filtered State Decompositions

For the filter, we are interested in

$$a_t = \mathbb{1}\omega_t^{a_0} + \mathbb{1}\omega_{tt}^c + \sum_{j=1}^{t-1}\left(\mathbb{1}\omega_{tj}^c + \mathbb{1}\omega_{tj}^d + \mathbb{1}\omega_{tj}^x + \mathbb{1}\omega_{tj}^w + \mathbb{1}\omega_{tj}\right) \tag{90}$$

where

- $\omega_t^{a_0}$ is an $(m \times m)$ matrix of the effect of the initial conditions on a_t,
- ω_{tj}^c is an $(m \times m)$ matrix of the effect of c_j on a_t,
- ω_{tj}^d is an $(m \times p)$ matrix of the effect of d_j on a_t, and
- ω_{tj}^x is an $(m \times k)$ matrix of the effect of x_j on a_t, and
- ω_{tj}^w is an $(m \times l)$ matrix of the effect of w_j on a_t, and

- ω_{tj} is an $(m \times p)$ matrix of the effect of y_j on a_t.

In each case, the period of the state being affected is denoted by t and the period of the observation of parameter causing the effect is denoted by j.

We begin by incorporating the observed data from a single period. Using the filtering equations from above, we examine the integration of information contained in y_t on $a_{t+1} = a_{t+1,1}$ given $a_{t,1}$:

$$\begin{aligned}
a_{t,2} &= a_{t,1} + K_{t,1}v_{t,1} \\
&= a_{t,1} + K_{t,1}(y_{t,1} - Z_{t,1}a_{t,1} - d_{t,1} - \beta_{t,1}x_t) \\
&= L_{t,1}a_{t,1} + K_{t,1}(y_{t,1} - d_{t,1} - \beta_{t,1}x_t) \\
a_{t,3} &= a_{t,2} + K_{t,2}v_{t,2} \\
&= a_{t,2} + K_{t,2}(y_{t,2} - Z_{t,2}a_{t,2} - d_{t,2} - \beta_{t,2}x_t) \\
&= L_{t,2}(L_{t,1}a_{t,1} + K_{t,1}(y_{t,1} - d_{t,1} - \beta_{t,1}x_t)) + K_{t,2}(y_{t,2} - d_{t,2} - \beta_{t,2}x_t) \\
&= L_{t,2}L_{t,1}a_{t,1} + K_{t,2}(y_{t,2} - d_{t,2} - \beta_{t,2}x_t) + L_{t,2}K_{t,1}(y_{t,1} - d_{t,1} - \beta_{t,1}x_t) \\
&\vdots
\end{aligned}$$

$$a_{t,p+1} = \left(\prod_{k=p}^{1} L_{t,k}\right) a_{t,1} + \sum_{i=1}^{p}\left(\prod_{k=p}^{i+1} L_{t,k}\right) K_{t,i}(y_{t,i} - d_{t,i} - \beta_{t,i}x_t) \tag{91}$$

$$\begin{aligned}
a_{t+1,1} &= T_{t+1}a_{t,p+1} + c_{t+1} + \gamma_{t+1}w_{t+1} \\
&= T_{t+1}\left(\prod_{k=p}^{1} L_{t,k}\right) a_{t,1} + c_{t+1} + \gamma_{t+1}w_{t+1} + T_{t+1}\sum_{i=1}^{p}\left(\prod_{k=p}^{i+1} L_{t,k}\right) K_{t,i}y_{t,i} \\
&\quad - T_{t+1}\sum_{i=1}^{p}\left(\prod_{k=p}^{i+1} L_{t,k}\right) K_{t,i}d_{t,i} - T_{t+1}\sum_{i=1}^{p}\left(\prod_{k=p}^{i+1} L_{t,k}\right) K_{t,i}\beta_{t,i}x_t \\
&= L_t^{\dagger}a_{t,1} + c_{t+1} + \gamma_{t+1}w_{t+1} + K_t^{\dagger}y_t - K_t^{\dagger}d_t - K_t^{\dagger}\beta_t x_t
\end{aligned} \tag{92}$$

where the products over $L_{t,k}$ proceed from higher to lower values of k and

$$L_t^{\dagger} = T_{t+1}\prod_{i=p}^{1} L_{t,i} \tag{93}$$

$$K_t^{\dagger} = T_{t+1}\left[\left(\prod_{i=p}^{2} L_{t,i}\right) K_{t,1} \quad \dots \quad L_{t,p}K_{t,p-1} \quad K_{t,p}\right] \tag{94}$$

To accommodate the exact initialization, when $t \leq d$ we adjust these expressions for the condition on $F_{\infty,t,i}$ by creating selection objects

$$L_t^{\dagger} = T_{t+1}\left(\prod_{i=p}^{1} S_{t,i}^{*}\right) \tag{95}$$

$$K_t^{\dagger} = T_{t+1}\left[\left(\prod_{i=p}^{2} L_{t,i}\right) \tilde{K}_{t,1} \quad \dots \quad L_{t,p}\tilde{K}_{t,p-1} \quad \tilde{K}_{t,p}\right] \tag{96}$$

$$S^*_{t,i} = \begin{cases} L_{*,t,i} & F_{\infty,t,i} = 0 \\ L_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases} \qquad \tilde{K}_{t,i} = \begin{cases} K_{*,t,i} & F_{\infty,t,i} = 0 \\ K_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases} \tag{97}$$

which allows the earlier recursion on $a_{t,1}$ to hold.

To find how information propagates across time, we use this new formulation of the filter, starting with the first period:

$$\begin{aligned}
a_1 =& L_0^\dagger a_0 + c_1 + \gamma_1 w_1 \\
a_2 =& L_1^\dagger a_1 + c_2 + \gamma_2 w_2 + K_1^\dagger y_1 - K_1^\dagger d_1 - K_1^\dagger \beta_1 x_1 \\
=& L_1^\dagger (L_0^\dagger a_0 + c_1 + \gamma_1 w_1) + c_2 + \gamma_2 w_2 + K_1^\dagger y_1 - K_1^\dagger d_1 - K_1^\dagger \beta_1 x_1 \\
=& L_1^\dagger L_0^\dagger a_0 + L_1^\dagger c_1 + L_1^\dagger \gamma_1 w_1 + c_2 + \gamma_2 w_2 + K_1^\dagger y_1 - K_1^\dagger d_1 - K_1^\dagger \beta_1 x_1 \\
a_3 =& L_2^\dagger a_2 + c_3 + \gamma_3 w_3 + K_2^\dagger y_2 - K_2^\dagger d_2 - K_2^\dagger \beta_2 x_2 \\
=& L_2^\dagger L_1^\dagger L_0^\dagger a_0 + L_2^\dagger L_1^\dagger c_1 + L_2^\dagger c_2 + c_3 + L_2^\dagger L_1^\dagger \gamma_1 w_1 + L_2^\dagger \gamma_2 w_2 + \gamma_3 w_3 + \\
& L_2^\dagger K_1^\dagger y_1 + K_2^\dagger y_2 - L_2^\dagger K_1^\dagger d_1 - K_2^\dagger d_2 - L_2^\dagger K_1^\dagger \beta_1 x_1 - K_2^\dagger \beta_2 x_2
\end{aligned}$$

From which the recursion becomes clear and we can see that information propagates with the product of the $L_t^\dagger$ matrix.

When using the univariate treatment of multivariate series, each series element of y_t can potentially affect other elements of y_t^u. To account for this, we could repeatedly consider the effects of each observation where each element of y_t is transformed and considered in turn. To do so in a single operation, we can simply consider the product $C_t^{-1}\text{diag}(y_t)$, which gives us the full contribution from each element of y_t across all transformed series y_t^u.

From the above expressions, we can infer that the weights are

$$\begin{aligned}
\omega_{tj} &= \left(\prod_{k=t-1}^{j+1} L_k^*\right) K_j^* C_j^{-1} \text{diag}(y_j) & \omega_{tj}^x &= -\left(\prod_{k=t-1}^{j+1} L_k^*\right) K_j^* C_j^{-1} \beta_j \text{diag}(x_j) \\
\omega_{tj}^d &= -\left(\prod_{k=t-1}^{j+1} L_k^*\right) K_j^* C_j^{-1} \text{diag}(d_j) & \omega_{tj}^w &= \left(\prod_{k=t-1}^{j} L_k^*\right) \gamma_j \text{diag}(w_j) \\
\omega_{tj}^c &= \left(\prod_{k=t-1}^{j} L_k^*\right) \text{diag}(c_j) & \omega_t^{a_0} &= \left(\prod_{k=t-1}^{0} L_k^*\right) \text{diag}(a_0)
\end{aligned} \tag{98}$$

completing the earlier desired decomposition of the filtered state.

C.2. Smoothed State Decompositions

For the smoother, we similarly want to compute quantities $\{\hat{\omega}_{tj}, \hat{\omega}_t^c, \hat{\omega}_t^d, \hat{\omega}_t^x, \hat{\omega}_t^w, \hat{\omega}_t^{a_0}\}$ such that

$$\hat{\alpha}_t = \mathbb{1}\hat{\omega}_t^{a_0} + \sum_{j=1}^{T} \left(\mathbb{1}\hat{\omega}_{tj} + \mathbb{1}\hat{\omega}_{tj}^c + \mathbb{1}\hat{\omega}_{tj}^d + \mathbb{1}\hat{\omega}_t^x + \mathbb{1}\hat{\omega}_t^w \right) \tag{99}$$

We first begin by breaking this expression into contributions from a_t and r_t,

$$
\begin{aligned}
r_t &= \mathbb{1}\,\overset{r}{\omega}{}_t^{a_0} + \sum_{j=t}^{T}\left(\mathbb{1}\,\overset{r}{\omega}{}_{tj}^{c} + \mathbb{1}\,\overset{r}{\omega}{}_{tj}^{d} + \mathbb{1}\,\overset{r}{\omega}{}_{tj}^{x} + \mathbb{1}\,\overset{r}{\omega}{}_{tj}^{w} + \mathbb{1}\,\overset{r}{\omega}_{tj}\right) && (100)\\
\hat{\alpha}_t &= a_t + P_t r_t \\
&= \mathbb{1}\omega_t^{a_0} + \mathbb{1}P_t\,\overset{r}{\omega}{}_t^{a_0} + \sum_{j=1}^{T}\Big(\mathbb{1}\omega_{tj}^{c} + \mathbb{1}P_t\,\overset{r}{\omega}{}_{tj}^{c} + \mathbb{1}\omega_{tj}^{w} + \mathbb{1}P_t\,\overset{r}{\omega}{}_{tj}^{w} \\
&\qquad + \mathbb{1}\omega_{tj}^{d} + \mathbb{1}P_t\,\overset{r}{\omega}{}_{tj}^{d} + \mathbb{1}\omega_{tj}^{x} + \mathbb{1}P_t\,\overset{r}{\omega}{}_{tj}^{x} + \mathbb{1}\omega_{tj} + \mathbb{1}P_t\,\overset{r}{\omega}_{tj}\Big) \\
&= \mathbb{1}\hat{\omega}_t^{a_0} + \sum_{j=1}^{T}\left(\mathbb{1}\hat{\omega}_{tj}^{c} + \mathbb{1}\hat{\omega}_{tj}^{w} + \mathbb{1}\hat{\omega}_{tj}^{d} + \mathbb{1}\hat{\omega}_{tj}^{x} + \mathbb{1}\hat{\omega}_{tj}\right) && (101)
\end{aligned}
$$

where

$$
\begin{aligned}
&\hat{\omega}_{tj} = \omega_{tj} + P_t\,\overset{r}{\omega}_{tj} &\quad& \hat{\omega}_{tj}^{d} = \omega_{tj}^{d} + P_t\,\overset{r}{\omega}{}_{tj}^{d} &\quad& \hat{\omega}_{tj}^{c} = \omega_{tj}^{c} + P_t\,\overset{r}{\omega}{}_{tj}^{c} \qquad (102)\\
&\hat{\omega}_{tj}^{x} = \omega_{tj}^{x} + P_t\,\overset{r}{\omega}{}_{tj}^{x} && \hat{\omega}_{tj}^{w} = \omega_{tj}^{w} + P_t\,\overset{r}{\omega}{}_{tj}^{w} && \hat{\omega}_{t}^{a_0} = \omega_{t}^{a_0} + P_t\,\overset{r}{\omega}{}_{t}^{a_0}
\end{aligned}
$$

Before examining the smoother recursion, we first rewrite the observation errors in matrix form,

$$
\begin{aligned}
v_t^u &= \begin{bmatrix} y_{t,1} - Z_{t,1}a_{t,1} - d_{t,1} - \beta_{t,1}x_t \\ y_{t,2} - Z_{t,2}a_{t,2} - d_{t,2} - \beta_{t,2}x_t \\ y_{t,3} - Z_{t,3}a_{t,3} - d_{t,3} - \beta_{t,3}x_t \\ \vdots \\ y_{t,p} - Z_{t,p}a_{t,p} - d_{t,p} - \beta_{t,p}x_t \end{bmatrix} \\
&= y_t^u - d_t^u - \begin{bmatrix} Z_{t,1}a_{t,1} \\ Z_{t,2}[L_{t,1}a_{t,1} + K_{t,1}(y_{t,1} - d_{t,1} - \beta_{t,1}x_t)] \\ Z_{t,3}[L_{t,2}L_{t,1}a_{t,1} + L_{t,2}K_{t,1}(y_{t,1} - d_{t,1} - \beta_{t,1}x_t) + K_{t,2}(y_{t,2} - d_{t,2} - \beta_{t,2}x_t)] \\ \vdots \\ Z_{t,p}\left[\left(\prod_{k=p-1}^{1} L_{t,k}\right)a_{t,1} + \sum_{i=1}^{p-1}\left(\prod_{k=p-1}^{i+1} L_{t,k}\right)K_{t,i}(y_{t,i} - d_{t,i} - \beta_{t,i}x_t)\right] \end{bmatrix} \\
&= y_t^u - d_t^u - \beta_t^u x_t - \tilde{A}_t^y(y_t^u - d_t^u - \beta_t^u x_t) - A_t^a a_t \\
&= A_t^y C_t^{-1}(y_t - d_t - \beta_t x_t) - A_t^a a_t && (103)
\end{aligned}
$$

where $A_t^y = I_p - \tilde{A}_t^y$ and

$$\tilde{A}_t^y = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ Z_{t,2}K_{t,1} & 0 & \dots & 0 & 0 & 0 \\ Z_{t,3}L_{t,2}K_{t,1} & Z_{t,3}K_{t,2} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ Z_{t,p-1}(\prod_{k=p-2}^{2} L_{t,k})K_{t,1} & Z_{t,p-1}(\prod_{k=p-2}^{3} L_{t,k})K_{t,2} & \dots & Z_{t,p-1}K_{t,p-2} & 0 & 0 \\ Z_{t,p}(\prod_{k=p-1}^{2} L_{t,k})K_{t,1} & Z_{t,p}(\prod_{k=p-1}^{3} L_{t,k})K_{t,2} & \dots & Z_{t,p}L_{t,p-1}K_{p-2} & Z_{t,p}K_{t,p-1} & 0 \end{bmatrix} \tag{104}$$

$$A_t^a = \begin{bmatrix} Z_{t,1} \\ Z_{t,2}L_{t,1} \\ \vdots \\ Z_{t,p}\left(\prod_{k=p-1}^{1} L_{t,k}\right) \end{bmatrix} \tag{105}$$

To adjust for the diffuse filter, this must be altered so that the lower-diagonal elements of $\tilde{A}_t^y$ are adjusted depending on the condition on $F_{\infty,t,i}$ and the propagation of $a_{t,0}$ respects the selection between $L_{*,t,i}$ and $L_{\infty,t,i}$:

$$\tilde{A}^y_{t\,(i,j)} = Z_{t,i}\left(\prod_{k=i-1}^{j} S^*_{t,k}\right)\tilde{K}_{t,j} \tag{106}$$

$$A_t^{a*} = \begin{bmatrix} Z_{t,1} \\ Z_{t,2}S^*_{t,2} \\ \vdots \\ Z_{t,p}\left(\prod_{k=p-1}^{1} S^*_{t,k}\right) \end{bmatrix} \tag{107}$$

Similarly to how a_t was handled for the filter, the incorporation of the information at time t to form $r_{t,0}$ given $r_{t+1,0}$ can be written as a single operation in matrix for as

$$\begin{aligned} r_{t,p} &= T_{t+1}^\top r_{t+1,0} \\ r_{t,p-1} &= Z_{t,p}^\top F_{t,p}^{-1} v_{t,p} + L_{t,p}^\top T_{t+1}^\top r_{t+1,0} \\ r_{t,p-2} &= Z_{t,p-1}^\top F_{t,p-1}^{-1} v_{t,p-1} + L_{t,p-1}^\top r_{t,p-1} \\ &= Z_{t,p-1}^\top F_{t,p-1}^{-1} v_{t,p-1} + L_{t,p-1}^\top Z_{t,p}^\top F_{t,p}^{-1} v_{t,p} + L_{t,p-1}^\top L_{t,p}^\top T_{t+1}^\top r_{t+1,0} \\ &\vdots \\ r_{t,0} &= M_t^\dagger v_t^u + L_t^{\dagger\top} r_{t+1,0} \\ &= M_t^\dagger A_t^y C_t^{-1} y_t - M_t^\dagger A_t^y C_t^{-1} d_t - M_t^\dagger A_t^y C_t^{-1}\beta_t x_t - M_t^\dagger A_t^a a_t + L_t^{\dagger\top} r_{t+1,0} \end{aligned} \tag{108}$$

where

$$\begin{aligned} M_t^\dagger &= \begin{bmatrix} Z_{t,1}^\top F_{t,1}^{-1} & L_{t,1}^\top Z_{t,2}^\top F_{t,2}^{-1} & \dots & \left(\prod_{i=1}^{p-1} L_{t,i}^\top\right) Z_{t,p}^\top F_{t,p}^{-1} \end{bmatrix} \\ &= A_t^{a\top} \operatorname{diag}\left(\begin{bmatrix} F_{t,1}^{-1} & F_{t,2}^{-1} & \dots & F_{t,p}^{-1} \end{bmatrix}\right) \end{aligned} \tag{109}$$

Noting the similarity between propagation through time like a_t, we have that

$$
\begin{aligned}
\overset{r}{\omega}_{tj} &= L_t^{\dagger\top}\, \overset{r}{\omega}_{t+1,j} - \left\{ M_t^\dagger A_t^a \omega_{tj} \right\}_{t>j} + \left\{ M_t^\dagger A_t^y C_t^{-1} \text{ diag}(y_t) \right\}_{t=j} \\
\overset{r\,d}{\omega}_{tj} &= L_t^{\dagger\top}\, \overset{r\,d}{\omega}_{t+1,j} - \left\{ M_t^\dagger A_t^a \omega_{tj}^d \right\}_{t>j} - \left\{ M_t^\dagger A_t^y C_t^{-1} \text{ diag}(d_t) \right\}_{t=j} \\
\overset{r\,x}{\omega}_{tj} &= L_t^{\dagger\top}\, \overset{r\,x}{\omega}_{t+1,j} - \left\{ M_t^\dagger A_t^a \omega_{tj}^x \right\}_{t>j} - \left\{ M_t^\dagger A_t^y C_t^{-1} \beta_t \text{ diag}(x_t) \right\}_{t=j} \\
\overset{r\,c}{\omega}_{tj} &= L_t^{\dagger\top}\, \overset{r\,c}{\omega}_{t+1,j} - \left\{ M_t^\dagger A_t^a \omega_{tj}^c \right\}_{t\geq j} \\
\overset{r\,w}{\omega}_{tj} &= L_t^{\dagger\top}\, \overset{r\,w}{\omega}_{t+1,j} - \left\{ M_t^\dagger A_t^a \omega_{tj}^w \right\}_{t\geq j} \\
\overset{r\,a_0}{\omega}_{t} &= L_t^{\dagger\top}\, \overset{r\,a_0}{\omega}_{t+1} - M_t^\dagger A_t^a \omega_t^{a_0}
\end{aligned}
\tag{110}
$$

To accommodate the exact initialization, we expand r_t as as done in the smoother:

$$
\hat{\alpha}_t = a_t + P_{*,t,1} r_{t,0}^{(0)} + P_{\infty,t,1} r_{t,0}^{(1)} \tag{111}
$$

where

$$
r_{t,i-1}^{(0)} = \begin{cases} Z_{t,i}^\top F_{*,t,i}^{-1} v_{t,i} + L_{*,t,i}^\top r_{t,i}^{(0)} & F_{\infty,t,i} = 0 \\ L_{\infty,t,i}^\top r_{t,i}^{(0)} & F_{\infty,t,i} \neq 0 \end{cases} \tag{112}
$$

$$
r_{t,i-1}^{(1)} = \begin{cases} r_{t,i}^{(1)} & F_{\infty,t,i} = 0 \\ Z_{t,i}^\top F_{\infty,t,i}^{-1} v_{t,i} + L_{\infty,t,i}^\top r_{t,i}^{(1)} + L_{t,i}^{(0)\top} r_{t,i}^{(0)} & F_{\infty,t,i} \neq 0 \end{cases} \tag{113}
$$

To handle the selection in these equations based on the condition $F_{\infty,t,i} = 0$, define a set of selection quantities:

$$
S_{t,i}^* = \begin{cases} L_{*,t,i} & F_{\infty,t,i} = 0 \\ L_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases} \qquad S_{t,i}^\infty = \begin{cases} I_m & F_{\infty,t,i} = 0 \\ L_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases} \qquad S_{t,i}^{(0)} = \begin{cases} 0_{m\times m} & F_{\infty,t,i} = 0 \\ L_{t,i}^{(0)} & F_{\infty,t,i} \neq 0 \end{cases} \tag{114}
$$

$$
M_{t,i}^* = \begin{cases} Z_{t,1}^\top F_{*,t,1}^{-1} & F_{\infty,t,i} = 0 \\ 0_{m\times 1} & F_{\infty,t,i} \neq 0 \end{cases} \qquad M_{t,i}^\infty = \begin{cases} 0_{m\times 1} & F_{\infty,t,i} = 0 \\ Z_{t,1}^\top F_{\infty,t,1}^{-1} & F_{\infty,t,i} \neq 0 \end{cases} \tag{115}
$$

This allows the recursion of the diffuse smoother to be written as

$$
r_{t,i-1}^{(0)} = M_{t,i}^* v_{t,i} + S_{t,i}^{*\top} r_{t,i}^{(0)} \tag{116}
$$

$$
r_{t,i-1}^{(1)} = M_{t,i}^\infty v_{t,i} + S_{t,i}^{\infty\top} r_{t,i}^{(1)} + S_{t,i}^{(0)\top} r_{t,i}^{(0)} \tag{117}
$$

Using these quantities, we can rewrite the recursion for $r_{t,0}^{(0)}$ as we did $r_{t,0}$. To arrive at the similar recursion for $r_{t,0}^{(1)}$, first make the simplifications that were done for $r_{t,0}$:

$$
r_{t,0}^{(1)} = \tilde{M}_t^\infty v_t^u + L_t^{\infty\top} r_{t+1,0}^{(1)} + \sum_{i=1}^{p} \left(\prod_{k=1}^{i-1} S_{t,k}^{\infty\top} \right) S_{t,i}^{(0)\top} r_{t,i}^{(0)} \tag{118}
$$

Also note that we can infer from the manipulation of $r_{t,0}$ earlier that

$$r_{t,i} = \left(\prod_{k=i+1}^{p} L_{t,k}^{\top}\right) T_{t+1}^{\top} r_{t+1,0} + \sum_{j=i+1}^{p} \left(\prod_{k=i+1}^{j-1} L_{t,k}^{\top}\right) Z_{t,j}^{\top} F_{t,j}^{-1} v_{t,j} \tag{119}$$

Making the required notation adjustments, these two equations give the recursion

$$\begin{aligned} r_{t,0}^{(1)} &= \tilde{M}_t^{\infty} v_t^u + L_t^{\infty\top} r_{t+1,0}^{(1)} \\ &\quad + \sum_{i=1}^{p} \left(\prod_{k=1}^{i-1} S_{t,k}^{\infty\top}\right) S_{t,i}^{(0)\top} \left(\prod_{k=i+1}^{p} S_{t,k}^{*\top}\right) T_{t+1}^{\top} r_{t+1,0}^{(0)} \\ &\quad + \sum_{i=1}^{p} \left(\prod_{k=1}^{i-1} S_{t,k}^{\infty\top}\right) S_{t,i}^{(0)\top} \sum_{j=i+1}^{p} \left(\prod_{k=i+1}^{j-1} S_{t,k}^{*\top}\right) M_{t,j}^{*} v_{t,j} \\ &= (\tilde{M}_t^{\infty} + M_t^{(0)}) v_t^u + L_t^{(0)\top} r_{t+1,0}^{(0)} + L_t^{\infty\top} r_{t+1,0}^{(1)} \end{aligned} \tag{120}$$

This allows us to write the recursions required for the diffuse smoother as

$$r_{t,0}^{(0)} = M_t^{*} v_t^u + L_t^{*\top} r_{t+1,0}^{(0)} \tag{121}$$

$$r_{t,0}^{(1)} = M_t^{\infty} v_t^u + L_t^{(0)\top} r_{t+1,0}^{(0)} + L_t^{\infty\top} r_{t+1,0}^{(1)} \tag{122}$$

$$L_t^{*} = T_{t+1} \left(\prod_{i=p}^{1} S_{t,i}^{*}\right) \tag{123}$$

$$L_t^{\infty} = T_{t+1} \left(\prod_{i=p}^{1} S_{t,i}^{\infty}\right) \tag{124}$$

$$L_t^{(0)} = T_{t+1} \left[\sum_{i=1}^{p} \left(\prod_{k=1}^{i-1} S_{t,k}^{\infty\top}\right) S_{t,i}^{(0)\top} \left(\prod_{k=i+1}^{p} S_{t,k}^{*\top}\right)\right]^{\top} \tag{125}$$

$$M_t^{*} = \begin{bmatrix} M_{t,1}^{*} & S_{t,1}^{*\top} M_{t,2}^{*} & \dots & \left(\prod_{i=1}^{p-1} S_{t,i}^{*\top}\right) M_{t,p}^{*} \end{bmatrix} \tag{126}$$

$$M_t^{\infty} = \tilde{M}_t^{\infty} + M_t^{(0)} \tag{127}$$

$$\tilde{M}_t^{\infty} = \begin{bmatrix} M_{t,1}^{\infty} & S_{t,1}^{\infty\top} M_{t,2}^{\infty} & \dots & \left(\prod_{i=1}^{p-1} S_{t,i}^{\infty\top}\right) M_{t,p}^{\infty} \end{bmatrix} \tag{128}$$

$$M_t^{(0)} = \sum_{i=1}^{p} \left(\prod_{k=1}^{i-1} S_{t,k}^{\infty\top}\right) S_{t,i}^{(0)\top} M_{t,i}^{(0)} \tag{129}$$

$$M_{t,i}^{(0)} = \begin{bmatrix} 0 & \dots & 0 & M_{t,i+1}^{*} & S_{t,i+1}^{*\top} M_{t,i+2}^{*} & \dots & \left(\prod_{k=i+1}^{p-1} S_{t,k}^{*\top}\right) M_{t,p}^{*} \end{bmatrix} \tag{130}$$

Combined with the expressions above gives the recursion

$$r_{t,0}^{(0)} = M_t^{*} A_t^{y} C_t^{-1} y_t - M_t^{*} A_t^{y} C_t^{-1} d_t - M_t^{*} A_t^{a} a_t + L_t^{*\top} r_{t+1,0}^{(0)} \tag{131}$$

$$r_{t,0}^{(1)} = M_t^{\infty} A_t^{y} C_t^{-1} y_t - M_t^{\infty} A_t^{y} C_t^{-1} d_t - M_t^{\infty} A_t^{a} a_t + L_t^{(0)\top} r_{t+1,0}^{(0)} + L_t^{\infty\top} r_{t+1,0}^{(1)} \tag{132}$$

from which the weights for the diffuse smoother are similar to above

$$
\begin{aligned}
\overset{r^{(0)}}{\omega}_{tj} &= L_t^{*\top} \overset{r^{(0)}}{\omega}_{t+1,j} - \{M_t^* A_t^a \omega_{tj}\}_{t>j} + \left\{M_t^* A_t^y C_t^{-1} \operatorname{diag}(y_t)\right\}_{t=j} \\
\overset{r^{(0)^d}}{\omega}_{tj} &= L_t^{*\top} \overset{r^{(0)^d}}{\omega}_{t+1,j} - \left\{M_t^* A_t^a \omega_{tj}^d\right\}_{t>j} - \left\{M_t^* A_t^y C_t^{-1} \operatorname{diag}(d_t)\right\}_{t=j} \\
\overset{r^{(0)^x}}{\omega}_{tj} &= L_t^{*\top} \overset{r^{(0)^x}}{\omega}_{t+1,j} - \left\{M_t^* A_t^a \omega_{tj}^x\right\}_{t>j} - \left\{M_t^* A_t^y C_t^{-1} \beta_t \operatorname{diag}(x_t)\right\}_{t=j} \\
\overset{r^{(0)^c}}{\omega}_{tj} &= L_t^{*\top} \overset{r^{(0)^c}}{\omega}_{t+1,j} - \left\{M_t^* A_t^a \omega_{tj}^c\right\}_{t\geq j} \\
\overset{r^{(0)^w}}{\omega}_{tj} &= L_t^{*\top} \overset{r^{(0)^w}}{\omega}_{t+1,j} - \left\{M_t^* A_t^a \omega_{tj}^w\right\}_{t\geq j} \\
\overset{r^{(0)^{a_0}}}{\omega}_{tj} &= L_t^{*\top} \overset{r^{(0)^{a_0}}}{\omega}_{t+1} - M_t^* A_t^a \omega_t^{a_0} \\
\overset{r^{(1)}}{\omega}_{tj} &= L_t^{\infty\top} \overset{r^{(1)}}{\omega}_{t+1,j} - \{M_t^\infty A_t^a \omega_{tj}\}_{t>j} + \left\{M_t^\infty A_t^y C_t^{-1} \operatorname{diag}(y_t)\right\}_{t=j} + L_t^{(0)} \overset{r^{(0)}}{\omega}_{t+1,j} \\
\overset{r^{(1)^d}}{\omega}_{tj} &= L_t^{\infty\top} \overset{r^{(1)^d}}{\omega}_{t+1,j} - \left\{M_t^\infty A_t^a \omega_{tj}^d\right\}_{t>j} - \left\{M_t^\infty A_t^y C_t^{-1} \operatorname{diag}(d_t)\right\}_{t=j} + L_t^{(0)} \overset{r^{(0)^d}}{\omega}_{t+1,j} \\
\overset{r^{(1)^x}}{\omega}_{tj} &= L_t^{\infty\top} \overset{r^{(1)^x}}{\omega}_{t+1,j} - \left\{M_t^\infty A_t^a \omega_{tj}^x\right\}_{t>j} - \left\{M_t^\infty A_t^y C_t^{-1} \beta_t \operatorname{diag}(x_t)\right\}_{t=j} + L_t^{(0)} \overset{r^{(0)^x}}{\omega}_{t+1,j} \\
\overset{r^{(1)^c}}{\omega}_{tj} &= L_t^{\infty\top} \overset{r^{(1)^c}}{\omega}_{t+1,j} - \left\{M_t^\infty A_t^a \omega_{tj}^c\right\}_{t\geq j} + L_t^{(0)} \overset{r^{(0)^c}}{\omega}_{t+1,j} \\
\overset{r^{(1)^w}}{\omega}_{tj} &= L_t^{\infty\top} \overset{r^{(1)^w}}{\omega}_{t+1,j} - \left\{M_t^\infty A_t^a \omega_{tj}^w\right\}_{t\geq j} + L_t^{(0)} \overset{r^{(0)^w}}{\omega}_{t+1,j} \\
\overset{r^{(1)^{a_0}}}{\omega}_{t} &= L_t^{\infty\top} \overset{r^{(1)^{a_0}}}{\omega}_{t+1} - M_t^\infty A_t^a \omega_t^{a_0} + L_t^{(0)} \overset{r^{(0)^{a_0}}}{\omega}_{t+1}
\end{aligned}
\tag{133}
$$

These can then be used similarly to those above to find weights for α_t for $t \leq d$:

$$
\begin{aligned}
\hat{\omega}_{tj} &= \omega_{tj} + P_t^* \overset{r^{(0)}}{\omega}_{tj} + P_t^\infty \overset{r^{(1)}}{\omega}_{tj} & \hat{\omega}_{tj}^d &= \omega_{tj}^d + P_t^* \overset{r^{(0)^d}}{\omega}_{tj} + P_t^\infty \overset{r^{(1)^d}}{\omega}_{tj} \\
\hat{\omega}_{tj}^x &= \omega_{tj}^x + P_t^* \overset{r^{(0)^x}}{\omega}_{tj} + P_t^\infty \overset{r^{(1)^x}}{\omega}_{tj} & \hat{\omega}_{tj}^c &= \omega_{tj}^c + P_t^* \overset{r^{(0)^c}}{\omega}_{tj} + P_t^\infty \overset{r^{(1)^c}}{\omega}_{tj} \\
\hat{\omega}_{tj}^w &= \omega_{tj}^w + P_t^* \overset{r^{(0)^w}}{\omega}_{tj} + P_t^\infty \overset{r^{(1)^w}}{\omega}_{tj} & \hat{\omega}_{t}^{a_0} &= \omega_{t}^{a_0} + P_t^* \overset{r^{(0)^{a_0}}}{\omega}_{t} + P_t^\infty \overset{r^{(1)^{a_0}}}{\omega}_{t}
\end{aligned}
\tag{134}
$$

C.3. Using the Weights

Once the filter and smoother weights have been computed, we can use them in several ways to summarize the parameters of our model. Most directly comparable to the coefficients in a traditional state space context, we can create a $(p \times m)$ matrix w that summarizes the total absolute contributions for each state from all of the observed data:

$$
w = \sum_{t=1}^{T} \sum_{j=1}^{T} |\hat{\omega}_{tj}| \tag{135}
$$

This matrix is often more informative after it has been row normalized to sum to one so that each element gives the proportion of a given state attributable to a time series.

Additionally, plots of the contributions to the state can help diagnose what data are informative about them. In particular, for dynamic factor models we may be interested in understanding how a given state is affected by different time series so that decomposing the state at time t by all influences from the p time series across all observation periods j is informative. Plots of the cumulative effects on the state are also often informative, especially when the scale of the time series differs substantially.

Affiliation:

Scott Brave, David Kelley
Federal Reserve Bank of Chicago
230 S. LaSalle St., Chicago, IL, 60604, USA

R. Andrew Butters
Business Economics and Public Policy, Kelley School of Business, Indiana University
1309 E. Tenth Street, Bloomington, IN, 47405, USA
E-mail: `rabutter@indiana.edu`