Particle Filtering, Learning, and Smoothing for Mixed-Frequency State-Space Models

Leippold, Markus ; Yang, Hanlin

DOI: https://doi.org/10.1016/j.ecosta.2019.07.001

Particle Filtering, Learning, and Smoothing for Mixed-Frequency State-Space Models[1]

Markus Leippold[2] and Hanlin Yang[3]

Abstract

A particle filter approach for general mixed-frequency state-space models is considered. It employs a backward smoother to filter high-frequency state variables from low-frequency observations. Moreover, it preserves the sequential nature of particle filters, allows for non-Gaussian shocks and nonlinear state-measurement relation, and alleviates the concern over sample degeneracy. Simulation studies show that it outperforms the commonly used state-augmented approach for mixed-frequency data for filtering and smoothing. In an empirical exercise, predictive mixed-frequency regressions are employed for Treasury bond and US dollar index returns with quarterly predictors and monthly stochastic volatility. Stochastic volatility improves model inference and forecasting power in a mixed-frequency setup but not for quarterly aggregate models.

Keywords: Mixed-Frequency, State-Space Models, Particle Filtering, Backward Smoothing, Stochastic Volatility, Return Predictability

JEL Classification: C13, C32, C53.

[1]We thank seminar participants at the 2016 Swiss Finance Institute Workshop in Finance (Zurich) and 2017 Econometric Society European Meeting (Lisbon).

[2]Department of Banking and Finance, University of Zurich, Switzerland, email: markus.leippold@bf.uzh.ch.

[3]Department of Banking and Finance, University of Zurich, Switzerland, email: hanlin.yang@bf.uzh.ch.

1. Introduction

In economics and finance, state-space models have become a popular modeling device over the last two decades. When estimating these models, we face the tightly coupled problems of state and parameter inference. Often we are interested in sequential real-time estimation, where estimates of state variables and forecasts are formed as new information arrives. However, at least two dilemmas arise in state-space modeling. First, economic data are often sampled at different frequencies. For instance, GDP growth rates and surveys of macroeconomic variables are available only at a quarterly frequency, whereas other variables, including asset returns and interest rates, are available at much higher frequencies. Second, even for the same variable, the sampling frequency may differ across periods. For instance, the real consumption expenditure per capita is collected annually before 1947, quarterly until 1959, and monthly afterward. Simple treatments of such unbalanced datasets either adopt a time aggregation to match a common low sampling frequency, or they exclusively focus on a uniform subset of the entire dataset, or even use a rather crude data augmentation procedure. However, such simplifications may destroy useful information that we could potentially gain from high-frequency observations. Well-known high-frequency empirical features include, for instance, time-varying volatility. Consequently, ignoring this information may give rise to inconsistency and biased model inference, and incur economic losses for investors and policymakers who make decisions based on these data.

In light of these drawbacks, there is a rising interest in econometric methods designed to handle mixed-frequency data. We contribute to this growing body of literature by proposing a simple yet general particle filtering framework for mixed-frequency state-space models (MFSSMs). The central issue that lies in the estimation of MFSSMs is the filtering of state variables. Among all filtering techniques, the Kalman filter, due to its analytical tractability, is the most commonly used approach; see, for instance, Harvey (1989), Mariano and Murasawa (2003), Giannone et al. (2008), Aruoba et al. (2009), and Durbin and Koopman (2012).

However, the Kalman filter builds on the assumptions of normality and the linearity of the state-measurement relation. These assumptions are too restrictive given the empirically observed skewness and fat tails of economic data, as well as the time variation in their volatility. Efficient alternatives include the Markov-Chain Monte Carlo (MCMC) method, which has been used by Schorfheide and Song (2015) and Marcellino et al. (2016), but are not tailor-made for sequential real-time inference. In contrast, particle filters alleviate all these concerns. When the filtering density is not available in closed form, the sequential inference is achieved by using Monte-Carlo samples to approximate the filtering distributions. Without loss of generality and given its well-known advantage of improving the filtering efficiency, we build the mixed-frequency particle filter (MFPF) on the resample-propagate scheme of Carvalho et al. (2010).

The problem of mixed sampling frequencies poses additional difficulties in sequential filtering. Simple implementations employ a state space augmentation step to convert MFSSMs to synchronously-evolving models; see, for instance, Schorfheide et al. (2018) and Leippold and Yang (2019). However, they are potentially hampered by sample degeneracy, which refers to the notorious fact that state variables are improperly represented by only a small effective number of Monte-Carlo samples; see Doucet et al. (2000) and Godsill et al. (2004) for a discussion.

We take a novel view toward particle filtering for MFSSMs, which mitigates sample degeneracy. We realize that, when we incorporate low-frequency observations into the state-space models, they tend to be jointly driven by lagged state variables. When low-frequency observations arrive, forward filtering requires the joint smoothing distribution of lagged state variables. We thus employ a smoother when low-frequency observations become available. In our implementation, we apply the backward smoother, which has been used by Carter and Kohn (1994), Frühwirth-Schnatter (1994), and Godsill et al. (2004) and preserves the sequential nature of particle filters. As the notion of smoothing-based filtering is not restricted to a

specific filter or smoother, it can be immediately extended to more general setups of MFPFs. We also discuss extensions that allow for sequential parameter learning.

We then conduct a simulation study to illustrate the advantage of our approach. We find that when we equip the MFPF with the backward smoother, it delivers smaller smoothing and filtering errors than the commonly used state-augmented MFPF. This outperformance becomes significant when the low-frequency observations are only sparsely observed. The advantage of the backward smoothing-based MFPF comes from the fact that the backward smoother mitigates the concern of sample degeneracy when dealing with mixed-frequency data.

We further conduct an empirical study to motivate the suitability of MFSSMs and MFPFs. We use the Survey of Professional Forecasters (SPF) to predict US Treasury bond returns and trade-weighted dollar index returns. The SPF is provided by the Philadelphia Fed and aggregates professional forecasts of key economic variables. Recent literature, including Chun (2010), Chernov and Mueller (2012), Eriksen (2017), and Foroni et al. (2018), shows that survey data predict key financial asset returns and yield changes. These survey variables are only updated at a quarterly frequency. For prediction, the usual practice is to temporally aggregate monthly returns to a quarterly frequency. However, asset returns exhibit time-varying volatility. This time variation becomes weaker at lower frequencies such as quarterly. However, ignoring time-varying volatility might give biased model inference and incur economic losses for investors who actively manage portfolios and volatility at a relatively high frequency. Thus, there is a practical demand for incorporating quarterly survey variables in monthly-evolving predictive regressions that possibly preserve the high-frequency nature of time-varying volatility.

We use the quarterly survey of the growth rates of industrial production, real consumption expenditure per capita, and CPI inflation rates as the predictors, with the model specification following the mixed-frequency predictive regressions of Leippold and Yang (2019). Incorpo-

rating time-varying volatility results in state-space nonlinearity and prohibits the use of the Kalman filter. For model estimation, we embed the MFPF into a random-walk Metropolis-Hasting algorithm, where we use the MFPF for filtering and likelihood computation and an MCMC iterator to generate parameter posteriors. Empirically, we find that, at a monthly frequency, incorporating stochastic volatility improves return forecasts by the quarterly survey variables, with more favorable density forecasts, Akaike information criteria (AIC), and prediction R^2-values. This finding justifies the usefulness of employing mixed-frequency models to forecast monthly returns using quarterly predictors.

Further, we also examine quarterly aggregate models, in which we use quarterly predictors to forecasting quarterly returns. We find that, at a quarterly frequency, stochastic volatility does not improve model inference or return forecasts. Economically, this implies that investors cannot take advantage of volatility timing at a quarterly frequency, which motivates the use of MFPFs for MFSSMs rather than temporal aggregation.

This paper contributes to the literature of Bayesian mixed-frequency approaches along various dimensions. Most of the literature considers modeling mixed-frequency data by linear-Gaussian models. Applications include, e.g., Mariano and Murasawa (2003) who construct a monthly GDP series from quarterly GDP and business cycle variables, and Aruoba et al. (2009) who develop an economic activity index in real time from various mixed-frequency series. Giannone et al. (2008) evaluate the marginal impact of monthly data releases on nowcasts of quarterly real GDP growth rates. Estimation of these models enjoys closed-form solutions known as the Kalman filter; see Harvey (1989) and Durbin and Koopman (2012) for a textbook treatment. Marcellino et al. (2016) develop a mixed-frequency GDP forecasting model with stochastic volatility and employ an MCMC approach for estimation. Schorfheide and Song (2015) develop a Bayesian mixed-frequency VAR, coupled with the Minnesota prior and MCMC estimators, to forecast quarterly macroeconomic variables. This paper proposes a flexible sequential particle filtering framework for MFSSMs that nests all

these model specifications and further allows for non-Gaussian and nonlinear dynamics.

To implement MFPFs, we take a novel approach that mitigates sample degeneracy due to mixed-frequency data. Among the Bayesian literature, Mariano and Murasawa (2003) modify the measurement equation by setting the loadings on state variables to zero. Giannone et al. (2008) set the measurement error of low-frequency observations to infinity. Aruoba et al. (2009) and Schorfheide and Song (2015) vary the dimension of the observations. In identifying long-run risks from mixed-frequency consumption data and in predicting stock returns, Schorfheide et al. (2018) and Leippold and Yang (2019) employ a state space augmentation procedure to convert their models to synchronously-evolving models. We employ the backward smoother to deal with mixed-frequency data, given that the lagged state variables jointly drive low-frequency observations.

Another branch of literature dealing with mixed-frequency data takes an observation-driven approach; see the seminal work of Ghysels et al. (2004) and Ghysels et al. (2007) for MIDAS. Bai et al. (2013) examine the relation between MIDAS and state-space models. Ghysels (2016) generalizes MIDAS to a VAR setup. Further, it allows the prediction of both high- and low-frequency observations from mixed-frequency data. Pettenuzzo et al. (2016) generalize MIDAS to account for mixed frequencies in both mean forecasts and volatility and find that MIDAS effects in the volatility dynamics improve density forecasts. Valle e Azevedo et al. (2006) propose a factor model that allows for trend-cycle decomposition and mixed-frequency macroeconomic data. Creal et al. (2014) develop an observation-driven mixed-frequency factor model.

Lastly, our paper also contributes to the return predictability literature. In particular, we find that monthly time-varying volatility is a vital model feature for predicting bond returns and foreign exchange rates, which is consistent with extensive empirical evidence. For instance, Johannes et al. (2016), Johnson (2019), and Leippold and Yang (2019) find that incorporating stochastic volatility results in stronger predictability of stock returns. Gargano

et al. (2017) argue that a time-varying volatility specification reconciles the disparity between the statistical and economic performances of bond return predictability. Among many others, Moreira and Muir (2017) empirically find that volatility timing produces significant economic gains.

2. Mixed-Frequency State-Space Models

In what follows, we introduce the state-space model framework that allows for observations sampled at different frequencies. We start with the synchronously evolving state-space models and discuss how we incorporate low-frequency data.

2.1. Model Setup

The standard state-space model takes the form

$$\begin{aligned} Y_{t+1}^H &\sim p\left(Y_{t+1}^H \middle| X_{t+1}, \Theta\right), \\ X_{t+1} &\sim p\left(X_{t+1} \middle| X_t, \Theta\right), \end{aligned} \tag{1}$$

where X is the latent state variable driving the stochastic dynamics, Y^H is the stream of observations available per unit of time, and Θ is the set of parameters and fixed as constant. In particular, H is used as the superscript to emphasize that Y^H is observed at the basis frequency, namely, the frequency at which the state variable evolves. Literally, H denotes a relatively high frequency, which we will distinguish from a lower frequency denoted by L.

Researchers often want to model the joint dynamics of both low- and high-frequency variables. The combination of observations available at different frequencies is motivated by the fact that low (high)-frequency variables can often improve the forecast of high (low)-frequency variables. We denote the low-frequency observation by Y^L, which is assumed to be available per M units of time only. Bayesian approaches involve incorporating Y^L in a way that preserves the high-frequency dynamics in equation (1). Often, these approaches imply that Y^L

is jointly driven by the entire historical path of the state variable over each non-overlapping M-period interval $[t+1, t+M]$

$$Y_{t+M}^{L} \sim p\left(Y_{t+M}^{L} \middle| X_{t+1:t+M}, \Theta\right), \tag{2}$$

with $X_{t+1:t+M} = (X_{t+1}, ..., X_{t+M})$. We call the joint equations (1) and (2) mixed-frequency state-space models (MFSSMs). We use Y_t to denote the set of all observations at time t, and Y^t the set of all observations available up to time t.

It is noteworthy that equations (1) and (2) place minimum model restrictions and nest various state-space models examined by the literature. See, e.g., Mariano and Murasawa (2003), Aruoba et al. (2009), Schorfheide and Song (2015), Marcellino et al. (2016), Schorfheide et al. (2018), and Leippold and Yang (2019). In presenting our approaches, we do not impose any parametric structure. With slight modifications, the MFSSM framework can also account for more general situations such as time-varying sampling frequencies or temporally missing data, but this paper focuses on the general form in equations (1) and (2) to illustrate the notion of smoothing-based particle filtering.

There exists well-established literature that addresses the issue of mixed sampling frequencies in the state-space framework. Among them, the Kalman filter is the most commonly used approach. However, the model specifications for which the Kalman filter is applicable are too restrictive. First, it assumes the random shocks in the state-space models to be Gaussian. However, this assumption is most often violated in economics, as skewed distributions and heavy tails regularly characterize the data. Second, it imposes a linear state-measurement dependence structure that prohibits one from incorporating more flexible yet realistic model features such as time-varying volatility. In contrast, particle filters allow all these features by using Monte-Carlo samples as proxies for the filtering distributions that are not available in closed form. Efficient alternatives include the Markov Chain Monte Carlo (MCMC) method (see, for instance, Marcellino et al. (2016)), but are often not tailor-made for sequential real-

time inference. In what follows, we present a way to handle mixed-frequency data in a particle filter framework.

3. Particle Filtering

Estimating MFSSMs hinges on filtering of state variables when observations are available at different frequencies. In this regard, the literature takes diverging views. For example, Mariano and Murasawa (2003) modify the measurement equations by setting the loadings on the state variables to zero. Giannone et al. (2008) set the measurement error of low-frequency observations to infinity. Aruoba et al. (2009), Durbin and Koopman (2012), and Schorfheide and Song (2015) vary the vector of observations in different periods. Schorfheide et al. (2018) and Leippold and Yang (2019) augment the state space to convert MFSSMs to standard SSMs. For efficient implementation of particle filters, we adopt a novel view. Because the distribution of the low-frequency variables depends on a series of lagged state variables, as indicated by equation (2), we can use efficient smoothing techniques to facilitate forward filtering when mixed-frequency observations arrive simultaneously. The advantage of the smoothing-based approach, as will be shown, is that it greatly mitigates sample degeneracy, a severe problem that may negatively affect the performance of particle filters. Moreover, it preserves the sequential nature of particle filters and is convenient for sequential inference.

3.1. Resample-Propagate Approach

To fix the idea, we restrict our analysis to the resample-propagate framework of Carvalho et al. (2010). As smoothing is largely independent of the particular choice of particle filters, our approach can be extended immediately to other particle filters, including those proposed by Gordon et al. (1993), Liu and West (2001), and Storvik (2002), to name a few.

Particle filtering of MFSSMs is achieved by modifying the standard resample-propagate filter when mixed-frequency observations arrive simultaneously. We start with particle filtering

of the synchronously-evolving model in equation (1). The resample-propagate procedure gives the time-$(t+1)$ filtering distribution based on Bayes' rule

$$\begin{aligned} p\left(X_{t+1}\middle|Y^{t+1},\Theta\right) &\propto \int p\left(Y_{t+1}\middle|X_t,\Theta\right) p\left(X_{t+1}\middle|X_t,Y_{t+1},\Theta\right) \\ &\times \; dp\left(X_t\middle|Y^t,\Theta\right), \end{aligned} \tag{3}$$

where $p\left(Y_{t+1}\middle|X_t,\Theta\right)$ is the predictive likelihood and $p(X_{t+1}|X_t,Y_{t+1},\Theta)$ is the density of X_{t+1} conditional on X_t and Y_{t+1}. Equation (3) is fully adapted as the filtering distribution at time $t+1$ can be obtained directly from that in the last period. Monte-Carlo samples are used to approximate the state densities used in equation (3) if they are not available in closed form. Often, this is true in the presence of non-Gaussian shocks or nonlinearity of the state-space models. To implement the resample-propagate approach, we resample random draws from $p\left(X_t\middle|Y^t,\Theta\right)$ with weights proportional to the predictive likelihood. We obtain $p\left(X_{t+1}\middle|Y^{t+1},\Theta\right)$ by further sampling from the conditional state density associated with each resampled draw. To filter the entire path of state variables, we repeat this procedure until the end of the sample period.

The resample-propagate procedure is a slight modification of the auxiliary particle filter (APF) of Pitt and Shephard (1999). The APF uses an importance function $p\left(Y_{t+1}|\alpha_{t+1}=g(x_t)\right)$ in the resampling procedure based on a best guess of X_{t+1} defined by $\alpha_{t+1}=g(x_t)$, whereas the resample-propagate procedure relies on the predictive likelihood and the conditional state density. If they are available for evaluation and sampling, the efficiency of particle filtering can be greatly improved. The reason for this is that we can use the new observation Y_{t+1} in both resampling and propagation, thereby directing the random draws of the state variables towards the observations as the particles filter proceeds.

3.2. Mixed-Frequency Particle Filtering

To illustrate the smoothing-based approach to dealing with mixed sampling frequencies, also called the mixed-frequency particle filter (MFPF), we focus on an M-period interval $[t+1, t+M]$, where the low-frequency variable Y^L is only observed at time $t+M$. Prior to time $t+M$, the situation is exactly the same as the synchronously-evolving state-space models, as only Y^H is observed. Therefore, for $l = 0, ..., M-2$ and starting from the time-t filtering distribution, we can immediately obtain, by adapting equation (3) to MFSSMs, the filtering distribution in the next period

$$\begin{aligned} p\left(X_{t+l+1}\middle|Y^{t+l+1},\Theta\right) &\propto \int p\left(Y_{t+l+1}\middle|X_{t+l},\Theta\right) p\left(X_{t+l+1}\middle|X_{t+l},Y_{t+l+1},\Theta\right) \\ &\times\ dp\left(X_{t+l}\middle|Y^{t+l},\Theta\right). \end{aligned} \tag{4}$$

However, when proceeding from time $t+M-1$ to $t+M$, equation (4) often cannot be implemented as the conditional likelihood of Y_{t+M} and the conditional density of X_{t+M} do not only depend on the previous state X_{t+M-1}, but also on its entire path within the entire interval, i.e., on $X_{t+1:t+M-1}$. Fortunately, in most of the applications examined by the literature, the filtering algorithm can still proceed if the state space is augmented by lagged state variables. Specifically, if we view $X_{t+1:t+M-1}$ as the set of all state variables at time $t+M-1$, the MFSSM can be written as a synchronously-evolving model, and the filtering equation transforms to

$$\begin{aligned} p\left(X_{t+M}\middle|Y^{t+M},\Theta\right) &\propto \int p\left(Y_{t+M}\middle|X_{t+1:t+M-1},\Theta\right) p\left(X_{t+M}\middle|X_{t+1:t+M-1},Y_{t+M},\Theta\right) \\ &\times\ dp\left(X_{t+1:t+M-1}\middle|Y^{t+M-1},\Theta\right). \end{aligned} \tag{5}$$

Casting the state-augmented models into particle filters is practically equivalent to the fixed-lag smoother of Kitagawa (1996), employed by Schorfheide et al. (2018) to estimate long-run risks models and by Leippold and Yang (2019) to estimate predictive regressions for

stock returns. To obtain the joint distribution $p\left(X_{t+1:t+M-1}\middle|Y^{t+M-1}\right)$, draws of lagged state variables are resampled over time together with the forward filter until time $t+M-1$. However, this simplified algorithm may suffer severely from the problem of sample degeneracy, particularly when M is large and observations are featured with rare values. The reason is that, as the filter proceeds, draws of lagged state variables are resampled sequentially in time without rejuvenation, thereby making the smoothing state distribution improperly represented by only a small number of heterogeneous draws. In light of this drawback, we are motivated to employ more efficient smoothers to obtain the smoothing distribution $p\left(X_{t+1:t+M-1}\middle|Y^{t+M-1},\Theta\right)$.

3.3. Backward Smoothing

We show that in attempting to obtain $p\left(X_{t+1:t+M-1}\middle|Y^{t+M-1},\Theta\right)$, the backward smoother of Carter and Kohn (1994), Frühwirth-Schnatter (1994), and Godsill et al. (2004) serves as a convenient alternative to the fixed-lag forward smoother and preserves the sequential nature of particle filters. The backward smoother relies on Bayes' rule and the Markovian structure of the state process to obtain the following backward recursive representation

$$\begin{aligned} p\left(X_{t+1:t+M-1}\middle|Y^{t+M-1},\Theta\right) &= p\left(X_{t+M-1}\middle|Y^{t+M-1},\Theta\right) \\ &\times \prod_{l=1}^{M-2} p\left(X_{t+l}\middle|X_{t+l+1},Y^{t+l},\Theta\right). \end{aligned} \tag{6}$$

To convert equation (6) into a numerically implementable form, we exploit Bayes' rule once more and express the conditional state density on the right-hand side as

$$p\left(X_{t+l}\middle|X_{t+l+1},Y^{t+l},\Theta\right) \propto p\left(X_{t+l+1}\middle|X_{t+l},\Theta\right) p\left(X_{t+l}\middle|Y^{t+l},\Theta\right). \tag{7}$$

Thus, conditional on X_{t+l+1} and Y^{t+l}, the state density in the previous period is proportional to the state transition density times the filtering density, which is directly available from the MFPF. In the context of particle filtering, given each draw of $X_{t+(l+1)}$, the back-

ward smoother can be performed by recursively resampling from the filtering distribution $p\left(X_{t+l}\middle|Y^{t+l},\Theta\right)$, with sampling weights proportional to $p\left(X_{t+l+1}\middle|X_{t+l},\Theta\right)$. Repeating this exercise backward until time $t+1$ and piecing all resampled draws together give the smoothing density $p(X_{t+1:t+M-1}|Y^{t+M-1},\Theta)$ from $p(X_{t+M-1}|Y^{t+M-1},\Theta)$. The advantage is that random draws of each lagged state variable are resampled only once from their respective filtering distributions, in contrast to the fixed-lag smoother that does consecutive resampling. Hence, the heterogeneity of Monte-Carlo samples approximating the smoothing distribution is greatly improved as is widely recognized by the Bayesian literature; see Doucet et al. (2000) and Godsill et al. (2004) for a discussion. The backward smoother imposes minimum model assumptions and can handle all applications for which the state transition density is available. Moreover, the backward smoother does not build on a specific filter and can thus be fitted to other filters.

3.4. Implementation

We denote the set of N random draws from the filtering distribution $p\left(X_t\middle|Y^t,\Theta\right)$ by $\left\{X_t^{(i)}\right\}_{i\in\mathcal{N}}$, where $\mathcal{N}=\{1,2,...,N\}$ is the index set. Further, for notational convenience, we use $p^N\left(X_t\middle|Y^t,\Theta\right)$ to denote the empirical distribution formed by these draws.

Step 1 (Filter X_{t+l+1} sequentially for $l=0,...,M-2$). This step executes equation (4) numerically.

> **Resample.** Draw a size-N set of index $\left(k^{(i)}\right)_{i\in\mathcal{N}}$ from $\mathcal{N}$, with weights proportional to the predictive likelihood
>
> $$w_{t+l+1|t+l}^{(i)} \propto p\left(Y_{t+l+1}\middle|X_{t+l}^{(i)},\Theta\right), \tag{8}$$
>
> where $Y_{t+l+1}=Y_{t+l+1}^H$ and $k^{(i)}$ is the index from the i-th draw.

Propagate. For each draw $k^{(i)}$, draw $X_{t+l+1}^{(i)}$ from the conditional density

$$X_{t+l+1}^{(i)} \sim p\left(X_{t+l+1}\middle|X_{t+l}^{\left(k^{(i)}\right)}, Y_{t+l+1}, \Theta\right). \tag{9}$$

The set $\left\{X_{t+l+1}^{(i)}\right\}_{i\in\mathcal{N}}$ represents N Monte-Carlo draws from $p\left(X_{t+l+1}\middle|Y_{t+l+1}, \Theta\right)$ and is denoted by $p^N\left(X_{t+l+1}\middle|Y_{t+l+1}, \Theta\right)$.

Step 2 (Smooth $X_{t+1:t+M-1}$). This step executes the backward smoother in equations (6)-(7). Set $\widetilde{X}_{t+M-1}^{(i)} = X_{t+M-1}^{(i)}$ to initialize. Smoothing is achieved by the following resampling scheme backward in time for $l = M-2, M-3, ..., 1$.

Resample. For each $i \in \mathcal{N}$, draw a sample, denoted by $\widetilde{X}_{t+l}^{(i)}$, from the empirical time-$(t+l)$ filtering distribution $\left\{X_{t+l}^{(j)}\right\}_{j\in\mathcal{N}}$ with weights defined as follows

$$w_{t+l|t+l+1}^{(j)} \propto p\left(\widetilde{X}_{t+l+1}^{(i)}\middle|X_{t+l}^{(j)}, \Theta\right). \tag{10}$$

The resampling procedure gives $\left\{\widetilde{X}_{t+l}^{(i)}\right\}_{i\in\mathcal{N}}$. After reaching time $t+1$ through backward smoothing, we piece these samples together to obtain

$$\left\{\widetilde{X}_{t+1:t+M-1}^{(i)}\right\}_{i\in\mathcal{N}} = \left\{\left(\widetilde{X}_{t+1}^{(i)}, ..., \widetilde{X}_{t+M-1}^{(i)}\right)\right\}_{i\in\mathcal{N}},$$

which is also the empirical smoothing distribution $p^N\left(X_{t+1:t+M-1}\middle|Y^{t+M-1}, \Theta\right)$.

Step 3 (Filter X_{t+M}). Implementing once more the standard resample-propagate procedure gives $p^N\left(X_{t+M}\middle|Y^{t+M}, \Theta\right)$ and completes the loop.

Resample. Draw a size-N set of index $\left(k^{(i)}\right)_{i\in\mathcal{N}}$ from $\mathcal{N}$, where $k^{(i)}$ is the index from the i-th draw, with weights proportional to the joint likelihood of $Y_{t+M} = \left(Y_{t+M}^H, Y_{t+M}^L\right)$:

$$w_{t+M|t+M-1}^{(i)} \propto p\left(Y_{t+M}\middle|\widetilde{X}_{t+1:t+M-1}^{(i)}, \Theta\right). \tag{11}$$

Propagate. For each draw $k^{(i)}$, draw $X_{t+M}^{(i)}$ from the conditional distribution

$$X_{t+M}^{(i)} \sim p\left(X_{t+M}\middle|\widetilde{X}_{t+1:t+M-1}^{\left(k^{(i)}\right)}, Y_{t+M}, \Theta\right). \tag{12}$$

The set $\left\{X_{t+M}^{(i)}\right\}_{i\in\mathcal{N}}$ represents N samples from $p\left(X_{t+M}\middle|Y_{t+M},\Theta\right)$ and is denoted by $p^N\left(X_{t+M}\middle|Y_{t+M},\Theta\right)$.

3.5. Extentions: Parameter Learning

Literature contains various particle filters developed to achieve more general tasks such as sequential parameter learning. To illustrate the flexibility of the smoothing-based approach to filtering, we examine the particle learning method of Carvalho et al. (2010) for MFSSMs. Similar extensions can be made to the sequential parameter learning algorithms of Liu and West (2001), Storvik (2002), and many others.

In particle learning, the parameter prior is usually assumed to be conjugate. Under this assumption, the prior and posterior are distributions of the same type, and can be fully determined by their sufficient statistics. Sequential parameter estimation is achieved by augmenting the state space by sufficient statistics. To illustrate, we denote the set of sufficient statistics at time t by s_t. The state-space model is fully characterized by the augmented state variable (X, s). Prior to time $t+M$, particle learning takes the standard resample-propagate procedure

$$\begin{aligned} p\left(X_{t+l+1}, s_{t+l+1}, \Theta\middle|Y^{t+l+1}\right) \;\propto\; & \int p\left(Y_{t+l+1}\middle|X_{t+l},\Theta\right) p\left(X_{t+l+1}\middle|X_{t+l},\Theta,Y_{t+l+1}\right) \\ & \times\; p\left(s_{t+l+1}\middle|X_{t+l},X_{t+l+1},s_{t+l},Y_{t+l+1}\right) \\ & \times\; p\left(\Theta\middle|s_{t+l+1}\right) dp\left(X_{t+l},s_{t+l},\Theta\middle|Y^{t+l}\right), \end{aligned} \tag{13}$$

where $l = 0, ..., M-2$. $p\left(Y_{t+l+1}\middle|X_{t+l},\Theta\right)$ is the predictive likelihood, $p\left(X_{t+l+1}\middle|X_{t+l},\Theta,Y_{t+l+1}\right)$ is the conditional state density, and $p\left(s_{t+l+1}\middle|X_{t+l},X_{t+l+1},s_{t+l},Y_{t+l+1}\right)$ describes the sufficient statistics updating equation. The dynamic updating of parameter posteriors, $p\left(\Theta\middle|s_{t+l+1}\right)$, is

tracked by sufficient statistics. Similar to particle filters, the smoothed state path $X_{t+1:t+M-1}$ is needed to evaluate the predictive likelihood and the conditional state density. We follow Carvalho et al. (2010) and employ the backward smoother for particle learning

$$\begin{aligned} p\left(X_{t+1:t+M-1}, s_{t+M-1}, \Theta \middle| Y^{t+M-1}\right) &= p\left(X_{t+M-1}, s_{t+M-1}, \Theta \middle| Y^{t+M-1}\right) \\ &\times \prod_{l=1}^{M-2} p\left(X_{t+l} \middle| X_{t+l+1}, Y^{t+l}, \Theta\right), \end{aligned} \tag{14}$$

with

$$p\left(X_{t+l} \middle| X_{t+l+1}, Y^{t+l}, \Theta\right) \propto p\left(X_{t+l+1} \middle| X_{t+l}, \Theta\right) p\left(X_{t+l} \middle| Y^{t+l}, \Theta\right), \quad l = M-2, ..., 1. \tag{15}$$

We now assume that the empirical filtering distribution $p^N\left(X_{t+l}, s_{t+l}, \Theta \middle| Y^{t+l}\right)$, for $l = 1, ..., M-1$, has been obtained from the forward filter. By definition, each of them is characterized by a sample set $(X_{t+l}, s_{t+l}, \Theta)^{(i)}_{i \in \mathcal{N}}$. Carvalho et al. (2010) mistakenly argue that the backward smoother works trivially in the same way as for particle filters. Specifically, they argue that the i-th sample of the smoothing distribution of X_{t+l} shall be drawn from the marginal distribution $p^N\left(X_{t+l} \middle| Y^{t+l}\right)$. However, according to equation (15), smoothing shall be based on $p\left(X_{t+l} \middle| Y^{t+l}, \Theta^{(i)}\right)$, which can be obtained only through refiltering for each fixed parameter draw, $\Theta^{(i)}$. To address the issue of smoothing, Yang et al. (2018) link the unknown but necessary distribution, $p\left(X_{t+l} \middle| Y^{t+l}, \Theta\right)$, to $p\left(X_{t+l}, \Theta \middle| Y^{t+l}\right)$, which is available from particle learning, through a multivariate normal approximation. Our view is that by directly employing the MFPF for $p\left(X_{t+l} \middle| Y^{t+l}, \Theta^{(i)}\right)$, we avoid the error stemming from multivariate normal approximations.

4. Simulation Studies

To demonstrate the empirical advantage of the backward smoothing-based particle filter over the commonly used state-augmented filter for MFSSMs, we conduct an intensive simula-

tion study.

4.1. Model Specification

For our simulation study, we use a linear-Gaussian model that takes the following form for each non-overlapping M-period interval $[t+1, t+M]$

$$\begin{aligned} Y^H_{t+l+1} &= K_{Y,0} + K_{Y,1} X_{t+l+1} + \sigma_Y \epsilon^Y_{t+l+1}, \\ Y^L_{t+M} &= X_{t+1} + \cdots + X_{t+M}, \\ X_{t+l+1} &= K_{X,0} + K_{X,1} X_{t+l} + \sigma_X \epsilon^X_{t+l+1}, \quad l = 0, ..., M-1, \end{aligned} \tag{16}$$

where ϵ_X and ϵ_Y are identically and normally distributed with zero mean and unit variance. Y^H is observed in each period. X is the latent state variable, whose temporal sum Y^L is observed per M units of time. The approach of linear sum aggregation in equation (16) is often used when researchers want to incorporate the low-frequency variable Y^L in a relatively high-frequency model. Schorfheide and Song (2015) use a monthly VAR to forecast quarterly macroeconomic variables. Schorfheide et al. (2018) estimate a monthly long-run risk model based on annually aggregated consumption data. Leippold and Yang (2019) incorporate the quarterly trend deviation in the consumption-wealth ratio in a monthly predictive regressions for stock returns. Other applications include, for instance those by Mariano and Murasawa (2003), Aruoba et al. (2009), and Marcellino et al. (2016). In principle, using equation (16) we can account for more sophisticated features such as time-varying volatility, however for the simulation study, we focus on the linear-Gaussian form as it enjoys a closed form of the filtering and smoothing distributions, which allows for convenient evaluation of the performance.

4.2. The Filtering and Smoothing Performance

We draw 1,000 simulated sample paths using equation (16) with $K_{Y,0} = 0$, $K_{Y,1} = 1$, $\sigma_Y = 0.1$, $K_{X,0} = 0.01$, $K_{X,1} = 0.8$, $\sigma_X = 0.1$, and $X_0 \sim \mathcal{N}(5 \times 10^{-2}, 25 \times 10^{-6})$, where

each simulated path consists of 300 observations of Y^H. We assume that, for each simulated path, the temporal sum of X, denoted by Y^L, is observed per M units of time only, with $M = 3, 4, 5$, and 6, respectively. Thus, the simulation study covers examples which include both sparsely and frequently observed low-frequency observations. For each simulated path, we implement the MFPFs with the fixed-lag forward smoother and backward smoother, as discussed in Section 3. For each simulation, we further compute the absolute error of the estimated mean relative to its standard deviation (AEM)

$$\mathrm{AEM}_{t|s} = \frac{\left|\widehat{\mathbb{E}}\left[X_t \middle| Y^s\right] - \mathbb{E}\left[X_t \middle| Y^s\right]\right|}{\sqrt{\mathrm{Var}\left(X_t \middle| Y^s\right)}}. \tag{17}$$

The AEM measures the filtering error for $t = s$ and smoothing error for $t < s$. Additionally, we also consider the absolute error of the estimated standard deviation (AESD)

$$\mathrm{AESD}_{t|s} = \left| \frac{\sqrt{\widehat{\mathrm{Var}}\left[X_t \middle| Y^s\right]}}{\sqrt{\mathrm{Var}\left(X_t \middle| Y^s\right)}} - 1 \right|, \tag{18}$$

and the effective number of samples for $p^N\left(X_t \middle| Y^s\right)$ in percentage (EN)

$$\mathrm{EN}_{t|s} = \frac{1}{N} \cdot \left(\sum_{i=1}^{\widehat{N}} \left(\widehat{w}_{t|s}^{(i)} \right)^2 \right)^{-1}, \tag{19}$$

where $\widehat{N}$ is the total number of heterogeneous samples and $\widehat{w}_{t|s}^{(i)}$ is the frequency of all particles taking exactly the same value as the i-th heterogeneous draw. If each particle occurs only once, the effective number of particles equals one; otherwise, it will be strictly smaller. Because the advantage of our backward smoothing-based MFPF is the alleviation of sample degeneracy, it works well compared to other approaches when the number of samples N is not large. Thus, we set $N = 1,000$.

For each M and each time in history, the above measures of the filtering and smoothing

errors in equations (17) and (18) are averaged over all simulations. The backward smoothing-based MFPF, labeled BS, is compared with the forward smoothing-based MFPF, labeled FS, which equivalently implements the state space augmentation procedure. We would like to explore whether the backward smoother makes a difference in smoothing and further filtering when mixed-frequency data arrive. Figure 1 displays the time series of the average error for the smoothing distribution in each cycle, namely, $p\left(X_{t+1:t+M-2}\middle|Y^{t+M-1}\right)$ for $t = 0, M, 2M, ..., 300 - M$. The smoothing distribution is of particular interest because it is a necessary ingredient for the MFPF to proceed, as indicated by equation (5). For $M = 3$, only the smoothing distribution of X_{t+1} is needed, whereas for $M = 6$, the joint smoothing distribution of $X_{t+1:t+4}$ is needed. There are two striking patterns. First of all, for any frequency considered in the simulation study, the backward smoother consistently delivers a significantly smaller error than the commonly used forward smoother. Second, as M increases from 3 to 6, the outperformance by the backward smoother becomes increasingly striking. For $M = 4, 5$, and 6, the error given by the forward smoother is serrated: The smoothing error for $p\left(X_{t+s}\middle|Y^{t+M-1}\right)$ increases as s goes back from time $t+M-2$ to $t+1$, because state variables in earlier dates are resampled more often and thus subject to sample degeneracy. This pattern is also reflected by the results in Table 1, which displays related summary statistics. In contrast, the error of the backward smoother remains much more stable over time because each lagged state variable is only resampled once from the respective filtering distribution.

Figure 2 displays the AEM for the filtering distribution in the end of each observation cycle, namely, $p\left(X_{t+M}\middle|Y^{t+M}\right)$ for $t = 0, M, 2M, ..., 300 - M$. Table 2 displays related summary statistics. The filtering distribution is directly obtained from the smoothing distribution $p\left(X_{t+1:t+M-1}\middle|Y^{t+M-1}\right)$. Not surprisingly, the outperformance by the backward smoother further translates into a smaller filtering error when the low-frequency observation arrives. Again, as M becomes large, the filtering error associated with the forward smoother becomes larger, whereas that associated with the backward smoother stays at a level of 5%. The filtering

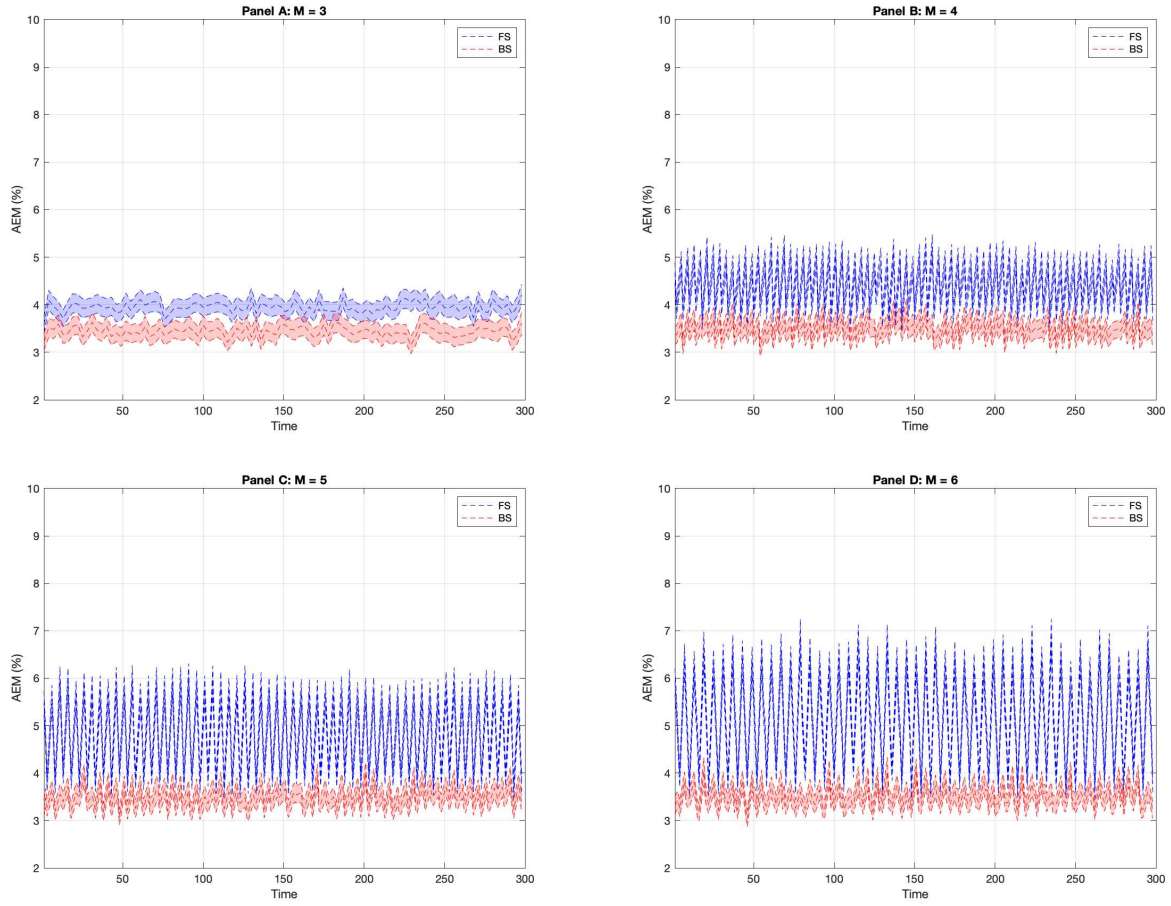

Figure 1: This figure displays the absolute error of the smoothed mean relative to the standard deviation ($\text{AEM}_{t+1:t+M-2|t+M-1}$) for $t = 0, M, 2M, ..., 300 - M$ and $M = 3, 4, 5, 6$ respectively. The blue (red) interval is the pointwise 95%-confidence interval of the sample mean of $\text{AEM}_{t+1:t+M-2|t+M-1}$ for the forward (FS) and backward (BS) smoothers, respectively. The sample mean is obtained from 1,000 simulated paths.

distribution for other periods is less affected by the arrival of mixed-frequency observations and hence has a relatively smaller and stable error, as illustrated by Table 2.

Apart from the AEM, Tables 1 and 2 also report the AESD and EN for the smoothing and filtering distributions. In Table 1, the AESD accumulates backward in time for the forward smoother but does not do so for the backward smoother, which is consistent with the AEM. Table 2 further shows that the filtering variance for $p\left(X_{t+M}\middle|Y^{t+M}\right)$ is more accurate for the backward smoother. Lastly, Table 1 also confirms that the backward smoother mitigates sample degeneracy: The EN stays around 46% for the backward smoother. However, the EN

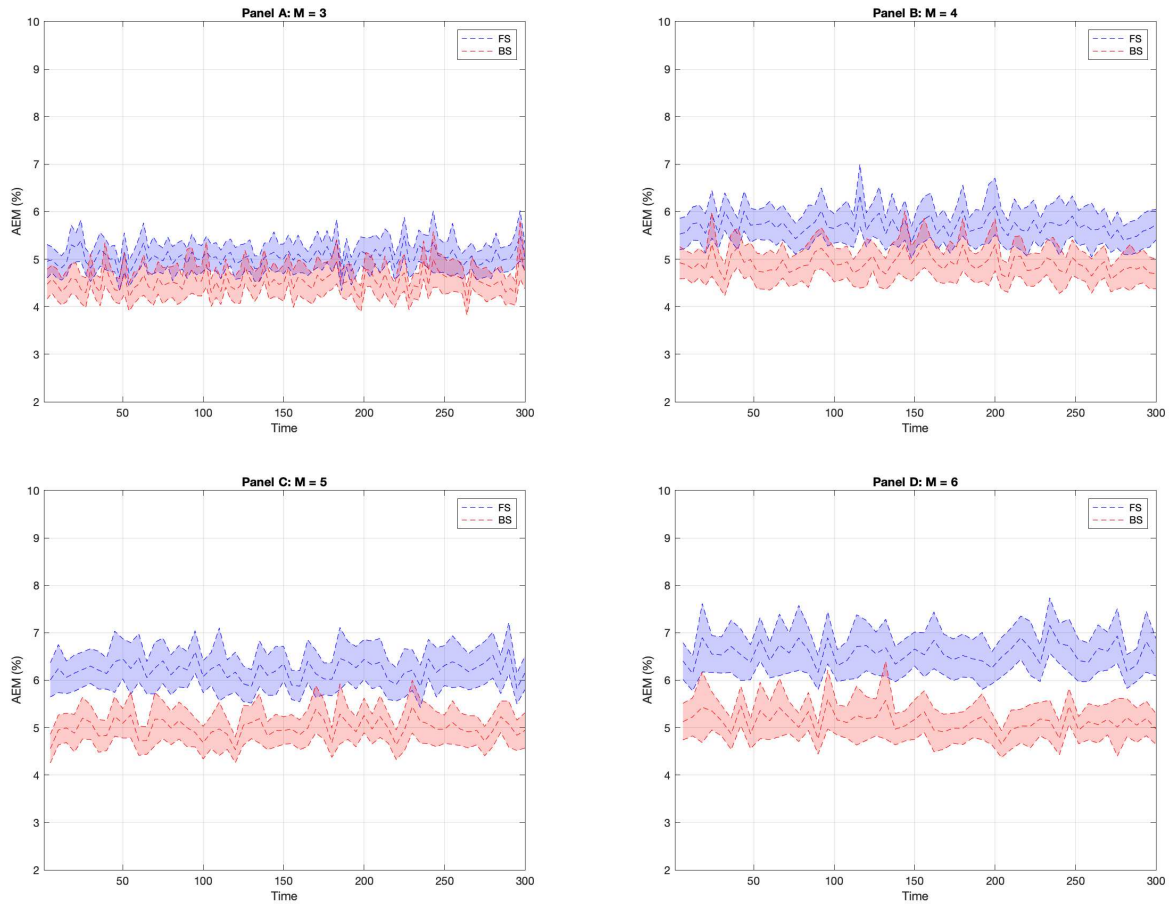

Figure 2: This figure displays the absolute error of the filtered mean relative to the standard deviation ($\text{AEM}_{t+M|t+M}$) for $t = 0, M, 2M, ..., 300 - M$ and $M = 3, 4, 5, 6$ respectively. The blue (red) interval is the pointwise 95%-confidence interval of the sample mean of $\text{AEM}_{t+M|t+M}$ for the forward (FS) and backward (BS) smoothers, respectively. The sample mean is obtained from 1,000 simulated paths.

is much smaller for the forward smoother and decays further for lagged state variables at the very beginning of each cycle, which is consistent with the forward nature of the algorithm.

5. Empirical Studies

As documented in previous literature, macroeconomic survey data predict key financial variables such as bond and foreign currency returns and yield changes. See, e.g., Chun (2010), Chernov and Mueller (2012), Eriksen (2017), and Foroni et al. (2018). The survey variables are forward-looking and have been shown to have additional explanatory power that is empirically

Table 1: Simulation Results for Smoothing Distributions

AEM for Smoothing Distributions (%)								
	FS				BS			
$M =$	3	4	5	6	3	4	5	6
$X_{t+M-2}\|Y^{t+M-1}$	3.95	3.87	3.84	3.81	3.44	3.34	3.30	3.30
$X_{t+M-3}\|Y^{t+M-1}$		4.96	4.79	4.75		3.64	3.46	3.43
$X_{t+M-4}\|Y^{t+M-1}$			5.77	5.51			3.69	3.48
$X_{t+M-5}\|Y^{t+M-1}$				6.46				3.78
AESD for Smoothing Distributions (%)								
	FS				BS			
$M =$	3	4	5	6	3	4	5	6
$X_{t+M-2}\|Y^{t+M-1}$	2.60	2.60	2.60	2.61	2.07	2.07	2.09	2.08
$X_{t+M-3}\|Y^{t+M-1}$		3.27	3.23	3.24		2.08	2.08	2.10
$X_{t+M-4}\|Y^{t+M-1}$			3.80	3.79			2.09	2.10
$X_{t+M-5}\|Y^{t+M-1}$				4.27				2.09
EN for Smoothing Distributions (%)								
	FS				BS			
$M =$	3	4	5	6	3	4	5	6
$X_{t+M-2}\|Y^{t+M-1}$	40	39	39	39	47	46	46	46
$X_{t+M-3}\|Y^{t+M-1}$		30	30	30		46	46	46
$X_{t+M-4}\|Y^{t+M-1}$			22	22			46	46
$X_{t+M-5}\|Y^{t+M-1}$				18				46

Simulation results for smoothing distributions of MFPFs with the forward (FS) and backward (BS) smoothers. The estimates are based on 1000 simulations, each of which has 300 observations. For each simulated path, the M-period sum of X is observed per M units of time. The average absolute error of the filtered mean and standard deviation, and the effective number of samples are reported. These variables are averaged over $t = 0, M, 2M, ..., 300 - M$ and over all simulated paths. 1000 particles are used for MFPFs.

not reflected in traditional predictors. In this section, we apply the MFPF to predictive regressions in which macroeconomic survey data are used to forecast the returns on Treasury bonds and US trade-weighted dollar index. The survey data are available only at quarterly intervals, and the usual practice employed in the literature is to temporally aggregate all variables to a quarterly frequency.

However, high-frequency empirical features such as time-varying volatility tend to be

Table 2: Simulation Results for Filtering Distributions

AEM for Filtering Distributions (%)								
	FS				BS			
$M =$	3	4	5	6	3	4	5	6
$X_{t+M}\|Y^{t+M}$	5.05	5.70	6.22	6.57	4.54	4.90	4.99	5.13
$X_{t+M-1}\|Y^{t+M-1}$	2.75	2.76	2.75	2.74	2.23	2.22	2.21	2.20
$X_{t+M-2}\|Y^{t+M-2}$	2.92	2.80	2.73	2.73	2.40	2.26	2.22	2.20
$X_{t+M-3}\|Y^{t+M-3}$		3.13	2.83	2.75		2.54	2.26	2.23
$X_{t+M-4}\|Y^{t+M-4}$			3.28	2.82			2.61	2.27
$X_{t+M-5}\|Y^{t+M-5}$				3.38				2.68
AESD for Filtering Distributions (%)								
	FS				BS			
$M =$	3	4	5	6	3	4	5	6
$X_{t+M}\|Y^{t+M}$	3.06	3.34	3.56	3.73	2.54	2.65	2.76	2.86
$X_{t+M-1}\|Y^{t+M-1}$	1.80	1.80	1.80	1.81	1.27	1.27	1.27	1.27
$X_{t+M-2}\|Y^{t+M-2}$	1.80	1.80	1.80	1.80	1.27	1.27	1.27	1.28
$X_{t+M-3}\|Y^{t+M-3}$		1.82	1.80	1.79		1.28	1.28	1.28
$X_{t+M-4}\|Y^{t+M-4}$			1.82	1.80			1.30	1.27
$X_{t+M-5}\|Y^{t+M-5}$				1.84				1.30
EN for Filtering Distributions (%)								
	FS				BS			
$M =$	3	4	5	6	3	4	5	6
$X_{t+M}\|Y^{t+M}$	28	25	22	21	28	25	22	21
$X_{t+M-1}\|Y^{t+M-1}$	100	100	100	100	100	100	100	100
$X_{t+M-2}\|Y^{t+M-2}$	100	100	100	100	100	100	100	100
$X_{t+M-3}\|Y^{t+M-3}$		100	100	100		100	100	100
$X_{t+M-4}\|Y^{t+M-4}$			100	100			100	100
$X_{t+M-5}\|Y^{t+M-5}$				100				100

Simulation results for filtering distributions of MFPFs with the forward (FS) and backward (BS) smoothers. The estimates are based on 1000 simulations, each of which has 300 observations. For each simulated path, the M-period sum of X is observed per M units of time. The average absolute error of the smoothed mean and standard deviation, and the effective number of samples are reported. These variables are averaged over $t = 0, M, 2M, ..., 300 - M$ and over all simulated paths. 1000 particles are used for MFPFs.

smoothed out by the above temporal aggregation. In particular, a time-varying volatility setup has been shown to be crucial for predicting returns. For instance, Johannes et al.

(2014), Johnson (2019), and Leippold and Yang (2019) find that time-varying volatility models have improved predictive power for stock returns. Gargano et al. (2017) find that it is also critical to reconciling the disparity between the statistical and economic measures of bond return predictability. Economically, a monthly-evolving model can preserve the high-frequency nature of the volatility dynamics and allows for volatility timing, which is economically significant as empirically shown by Moreira and Muir (2017). In practice, forecasting asset returns at a monthly frequency is common among practitioners who manage portfolios and volatility at a relatively high frequency such as monthly.

Thus, we are aiming at incorporating quarterly survey variables into monthly-evolving predictive regressions. We will conduct a comprehensive specification analysis and evaluate the performance of various model specifications. In particular, we highlight the empirical advantage of our backward smoothing-based MFPF and explore the predictive power of quarterly predictors and the role of stochastic volatility.

5.1. Model Specification

The growth rates of industrial production (IP), real personal consumption expenditures (PCE), and CPI inflation rates (CPI) for the next quarter, which we use as the predictors are drawn from the SPF. The complete dataset is provided by the Philadelphia Fed and dates back to 1968:Q4. However, it has been handled by the Philadelphia Fed and released in real time only since 1990:Q2. Thus, we only consider the sample period of 1990:Q2 to 2018:Q3. The goal is to predict monthly returns. The quarterly survey data are reported in annualized terms. For our studies, we convert them to quarterly terms to match their sampling frequency. Figure 3 displays the time series of IP, PCE, and CPI in percentage. All variables appear to be mean-reverting and persistent and drop during all periods of market turmoils including the burst of the dotcom bubble in 2001 and the 2008 Great Recession.

We consider a variety of model specifications following Leippold and Yang (2019). They incorporate the quarterly trend deviation in the consumption-wealth ratio in both mixed-

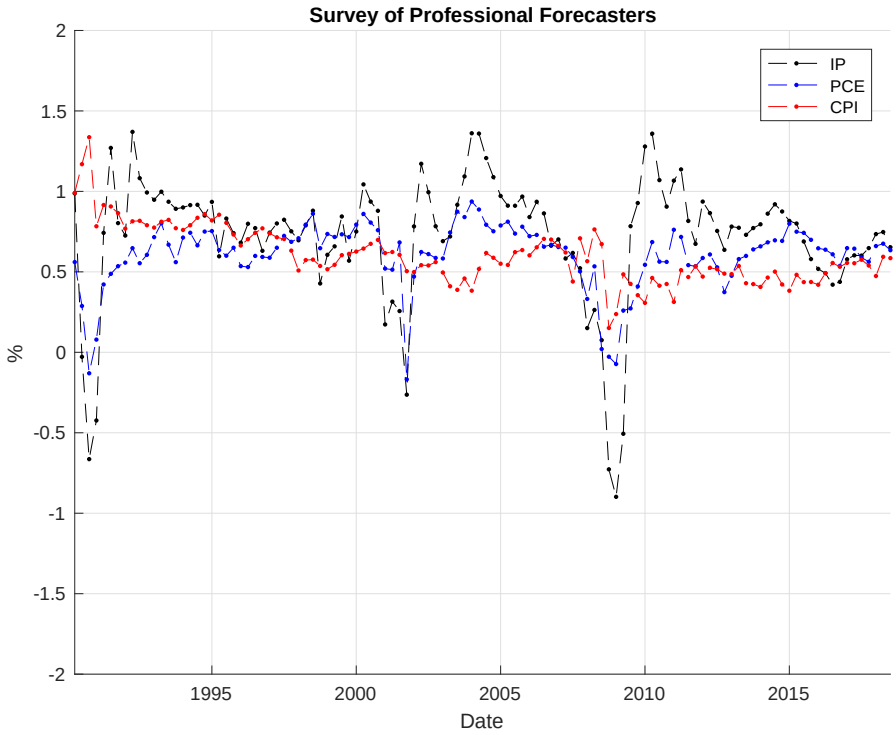

Figure 3: This figure displays the time series of IP, PCE, and CPI in the Survey of Professional Forecasters. All variables are quarterly sampled. The sample period extends from 1990:Q3 to 2018:Q3.

frequency and quarterly aggregate predictive regressions for stock returns. Specifically, to incorporate these quarterly predictors, each of them, denoted by Z, is assumed to be the quarterly sum of a monthly-evolving linear-Gaussian process X

$$\begin{aligned} Z_{t+3} &= X_{t+1} + X_{t+2} + X_{t+3}, \\ X_{t+1} &= K_{X,0} + K_{X,1} X_t + \sigma_X \epsilon^X_{t+1}, \end{aligned} \tag{20}$$

where ϵ^X_{t+1} is i.i.d. standard normal. X is not observed, whereas Z is observed only in the end of each quarter. Economically, X can be understood as the monthly forecast implied by the quarterly survey. Using a linear-Gaussian model is consistent with the stylized empirical fact that the survey variables are mean-reverting and persistent, as illustrated by Figure 3. To forecast asset returns in the next month, we use X instead of the quarterly observations

directly

$$r_{t+1} \quad = \quad K_{r,0} + K_{r,1} X_{t+1} + \sigma_r \epsilon^r_{t+1}, \tag{21}$$

where r_{t+1} is the monthly return and $K_r = (K_{r,0}, K_{r,1})$ is the prediction coefficient. The forecasting error is assumed to be i.i.d. normal, with a standard deviation of σ_r. For Treasury bonds, r_{t+1} is the average logarithmic return of 5-, 10-, 15-, 20-, 25-, and 30-year bonds in excess of the 1-month bond return. The bond returns are obtained from the yield curve data from Gürkaynak et al. (2007). For the dollar index, r_{t+1} is the monthly logarithmic return and obtained from Federal Reserve Economic Data. The specification exactly follows the simulation study in Section 4. Indeed, casting the predictive regressions in state-space form, Z and r play the roles of low- and high-frequency observations, respectively. Notable, the time-$(t+1)$ return is assumed to be driven by X_{t+1}. In actual prediction, because the filtering distribution of X_{t+1} is not available up to time t, we use

$$K_{Y,0} + K_{Y,1} \left(K_{X,0} + K_{X,1} \mathbb{E} \left[X_t \middle| Y^t \right] \right) \tag{22}$$

to predict the time-$(t+1)$ return. The motivation for using X_{t+1} instead of X_t to formulate the predictive regression is that r_{t+1} enters into both the likelihood and filtering distribution, which makes particle filtering practically more efficient. We assume all parameters in equations (20) and (21) to be constant. We call this model the mixed-frequency constant volatility (MF-CV) model.

There is rich evidence of time variation in return volatility at monthly or higher frequencies. Thus, we follow Johannes et al. (2014) and Leippold and Yang (2019) and also consider mixed-

frequency predictive regressions with stochastic volatility (MF-SV)

$$\begin{aligned} r_{t+1} &= K_{r,0} + K_{r,1} X_{t+1} + \sqrt{V_t}\epsilon^r_{t+1}, \\ \ln V_{t+1} &= K_{V,0} + K_{V,1} \ln V_t + \sigma_V \epsilon^V_{t+1}. \end{aligned} \tag{23}$$

The instantaneous variance is assumed to be a log-linear Gaussian process, with the volatility leverage effect captured by a correlation between shocks to returns and variance $\rho = \text{corr}(\epsilon^r_{t+1}, \epsilon^V_{t+1})$. The log-linear specification has several advantages compared to other volatility models considered in the literature. First of all, it is the most parsimonious specification for modeling mean reversion and persistence of the volatility dynamics. Second, it is able to generate sufficient skewness and kurtosis as empirically observed in the data when compared to other volatility models such as the model of Heston (1993). Third, it guarantees that the instantaneous variance remains strictly positive, a necessity which is not satisfied by many other models in a discrete-time setup.

5.2. Model Estimation

MF-SV models are nonlinear in state variables and thus cannot be estimated by the Kalman filter. We take a particle Markov-Chain Monte Carlo (MCMC) approach to making inference about the set of all parameters Θ for each model. Specifically, the inference procedure builds on Bayes' rule

$$p\left(\Theta \middle| Y^T\right) \propto p\left(Y^T \middle| \Theta\right) p(\Theta), \tag{24}$$

where $p(\Theta)$ is the parameter prior. Equation (24) states that the parameter posterior is available if one can evaluate the joint likelihood of all observations for each choice of Θ. To this end, we use the backward smoothing-based MFPF to compute the joint likelihood, which

can be further factorized into the conditional likelihood of observations in each period

$$p\left(Y^T|\Theta\right) = \prod_{t=1}^{T} p\left(Y_t|Y_{t-1},\Theta\right). \tag{25}$$

The conditional likelihood can be evaluated sequentially, given the filtering distribution obtained from the MFPF. For example, for each $M(=3)$ units of time, we have

$$\begin{aligned} p\left(Y_{t+l+1}|Y_{t+l},\Theta\right) &\approx \int p\left(Y_{t+l+1}|X_{t+l},\Theta\right) dp^N\left(X_{t+l}|Y^{t+l},\Theta\right), \quad l=0,...,M-2, \\ p\left(Y_{t+M}|Y_{t+M-1},\Theta\right) &\approx \int p\left(Y_{t+M}|X_{t+1:t+M-1},\Theta\right) dp^N\left(X_{t+1:t+M-1}|Y^{t+M-1},\Theta\right). \end{aligned} \tag{26}$$

Given the approximate likelihood, we then employ a random-walk Metropolis-Hasting algorithm to draw sample parameters from the posterior. Andrieu et al. (2010) show that using the approximate likelihood instead of the exact likelihood still delivers draws from the actual posterior as the Markov chain evolves. Further, the prior is assumed to be flat, namely $p(\Theta) \propto 1$ on its support. Both assumptions simplify the sampling procedure to the greatest extent. Lastly, each model is estimated by 10,000 sequential MCMC draws of Θ with a sufficiently long burn-in process and a sample number of 1,000 for particle filtering and smoothing.

5.3. The Filtering and Smoothing Performance

Tables 3 and 4 display the parameter estimates for MF-CV and MF-SV models when each survey variable is used to predict bond and dollar index returns. Before analyzing the economic implications, we explore the filtering and smoothing performance of the MFPF. We use parameter estimates from the MCMC exercises and apply the state augmentation approach (FS) and the backward smoother (BS) to MFPFs. To reduce sampling biases, we use 10,000 samples for filtering. For MF-CV models, the filtering and smoothing distributions have a closed form, against which the MFPF is benchmarked. Because MF-SV models do not have a closed form for filtering and smoothing, we run the MFPF with BS and 100,000

Table 3: Mixed-Frequency Predictive Regressions for Bond Returns

	IP		PCE		CPI	
	CV	SV	CV	SV	CV	SV
$K_{r,0}$	0.35	−0.96	−0.58	−0.48	−0.62	−0.64
	(0.40)	(0.40)	(0.10)	(0.19)	(0.74)	(0.48)
$K_{r,1}$	0.79	6.35	6.64	4.80	4.94	5.85
	(1.42)	(1.36)	(0.22)	(0.19)	(3.38)	(2.14)
σ_r	3.94		3.82		3.90	
	(0.15)		(0.15)		(0.18)	
$K_{X,0}$	0.04	0.04	0.03	0.03	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$K_{X,1}$	0.84	0.84	0.83	0.82	0.94	0.91
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
σ_X	0.07	0.07	0.04	0.05	0.03	0.03
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
$K_{V,0}$		−4.07		−3.73		−3.21
		(1.05)		(1.23)		(1.43)
$K_{V,1}$		0.41		0.45		0.52
		(0.16)		(0.13)		(0.14)
σ_V		0.55		0.51		0.48
		(0.11)		(0.13)		(0.10)
ρ		0.49		0.54		0.52
		(0.18)		(0.14)		(0.19)
Log likelihood	3.33	3.36	3.55	3.55	3.66	3.68
AIC$\times 10^{-3}$	−2.27	−2.28	−2.41	−2.41	−2.49	−2.50
$R^2(\%)$	0.51	2.67	0.00	0.55	0.86	1.12

Estimation results for predicting monthly bond returns using quarterly SPF. Bond returns are the monthly average returns of 5-, 10-, 15-, 20-, 25-, and 30-year bonds in excess of 1-month bond yields. Log likelihood is the time-series average of log likelihood. AIC is the Akaike information criteria. The sample period extends from 1990:Q3 to 2018:Q3.

particles as the benchmark. For each cycle ($M = 3$), we examine the AEM for the smoothing distribution $p\left(X_{t+1:t+M-1}\middle|Y^{t+M-1}\right)$ and the filtering distribution $p\left(X_{t+M}\middle|Y^{t+M}\right)$ as they are necessary components for MFPFs to proceed.

To provide some indication about the size of the filtering and smoothing errors relative to their standard deviations, we display in Figure 4 the average errors for the bond and dollar index return predictions for the MF-SV models. Clearly, the error increases when the movements of the survey variables become extreme, mostly during market turmoils. Such a

Table 4: Mixed-Frequency Predictive Regressions for Dollar Index Returns

	IP		PCE		CPI	
	CV	SV	CV	SV	CV	SV
$K_{r,0}$	0.35	0.55	0.62	0.20	0.06	−0.22
	(0.12)	(0.19)	(0.08)	(0.19)	(0.15)	(0.16)
$K_{r,1}$	−0.84	−1.67	−2.50	−0.07	0.70	1.96
	(0.42)	(0.68)	(0.32)	(0.90)	(0.68)	(0.74)
σ_r	1.23		1.24		1.29	
	(0.05)		(0.05)		(0.06)	
$K_{X,0}$	0.03	0.03	0.04	0.04	0.02	0.02
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
$K_{X,1}$	0.87	0.84	0.82	0.81	0.88	0.89
	(0.03)	(0.03)	(0.01)	(0.02)	(0.03)	(0.03)
σ_X	0.07	0.09	0.05	0.05	0.03	0.03
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
$K_{V,0}$		−1.96		−1.65		−2.01
		(2.18)		(1.78)		(1.26)
$K_{V,1}$		0.78		0.75		0.68
		(0.24)		(0.20)		(0.14)
σ_V		0.35		0.40		0.62
		(0.12)		(0.14)		(0.10)
ρ		0.20		0.25		0.18
		(0.21)		(0.24)		(0.21)
Log likelihood	4.48	4.50	4.70	4.72	4.81	4.83
AIC$\times 10^{-3}$	−3.05	−3.06	−3.20	−3.21	−3.28	−3.29
$R^2(\%)$	0.62	1.30	1.19	0.01	0.36	1.10

Estimation results for predicting dollar index returns using quarterly SPF. Dollar index returns are the monthly trade-weighted US dollar index returns. Log likelihood is the time-serie average of log likelihood. The sample period extends from 1990:Q3 to 2018:Q3.

pattern is a common phenomenon when Gaussian particle filters are applied to time series with non-Gaussian shocks as they may lead to sample degeneracy. Nevertheless, the backward smoother tends to give smaller errors with a few exceptions including MF-CV-IP model for bond returns and MF-SV-PCE model for dollar index returns. For the MF-SV models, the improvements are from 3.24% to 2.23% and from 5.86% to 4.29% for bond returns and dollar index returns, respectively. Similarly, the improvements for MF-CV models are from 5.90% to 5.17% and from 4.86% to 4.77% for bond and index returns, respectively. These improvements come from the better performance of the backward smoother as demonstrated in Section 5.

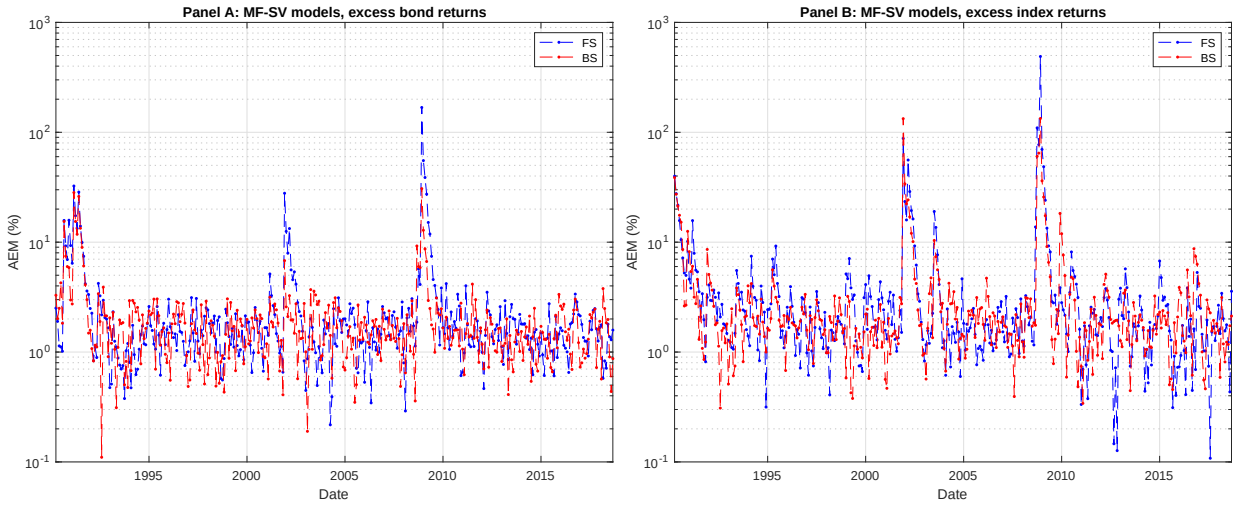

Figure 4: This figure displays the absolute error of the filtered and smoothed mean relative to the standard deviation ($\text{AEM}_{t+1:t+M-1|t+M-1}$ and $\text{AEM}_{t+M|t+M}$) with $M = 3$. The blue (red) curve corresponds to the forward (FS) and backward (BS) smoothers, respectively. Panel A is based on monthly bond excess returns and Panel B on monthly dollar index returns. The sample period extends from 1990:Q3 to 2018:Q3.

As a result, the backward smoother further delivers better estimates of expected returns.

5.4. Empirical Performance of Return Predictability

This section further compares the predictive performance of MF-CV and MF-SV models documented in Tables 3 and 4. For both specifications, the estimates for the predictor dynamics are similar. However, incorporating time-varying volatility significantly affects the estimates of the prediction coefficients, $K_{r,0}$ and $K_{r,1}$, and hence the return forecasts. Moreover, the time variation and predictability in the volatility dynamics are statistically significant, as indicated by the estimate of $K_{V,1}$. We use the average monthly logarithmic likelihood, AIC, and the R^2-value to measure the forecasting performance. The model likelihood can also be viewed as the density forecast of observations.

All evidence points to the fact that MF-SV models empirically outperform MF-CV models in predicting monthly returns, and thus highlights the suitability of MF-SV models and MFPFs for prediction. First of all, for almost all pairs of models, incorporating time-varying volatility leads to more significant estimates of the prediction coefficient, $K_{r,1}$, except when PCE is

used to predict dollar index returns. Second, MF-SV models deliver a larger likelihood, which is further translated, by adjusting the number of parameters, into a smaller value of AIC. Lastly, the prediction R^2 becomes larger when time-varying volatility is incorporated. For each predictor, the R^2 increases from 0.51% to 2.67%, from 0.86% to 0.55%, and from 0.00% and 1.12%, respectively. In particular, there is no return predictability by PCE at a monthly frequency, but incorporating stochastic volatility raises the forecasting power slightly.

We can draw similar conclusions for dollar index returns with PCE being the only exception. The R^2 for IP and CPI increases from 0.62% to 1.30% and from 0.36% to 1.10%, whereas the R^2 for PCE decreases from 1.19% to 0.01%. Summarizing all results, we can draw two conclusions. First, quarterly survey variables predict monthly bond and dollar index returns, which justifies the use of mixed-frequency models when forecasting monthly returns is the objective. Second, mixed-frequency models preserve the high-frequency evolution of volatility, which in return improves return predictability. The results are consistent with evidence supporting time-varying volatility models in stock and bond literature.

5.5. Quarterly Aggregate Models

If time-varying volatility is a critical model feature for predicting stock returns at a monthly frequency, it is also interesting to explore whether it remains as important for quarterly aggregate models. This is also practically relevant, because if quarterly aggregate models work very well, then there is less reason for developing econometric approaches for MFSSMs. The literature already shows that temporal aggregation may cause a loss of information from high-frequency return movements. For instance, Leippold and Yang (2019) find that quarterly SV models are misspecified in the evolution frequency of volatility and underperform in comparison with quarterly CV models in predicting stock returns. Thus we draw a comparison between the quarterly aggregate CV and SV models. Specifically, at the end of each quarter,

CV models use the survey data to forecast the next quarter's return

$$\begin{aligned} r_{t+1,t+3} &= r_{t+1} + r_{t+2} + r_{t+3}, \\ r_{t+1,t+3} &= K_{r,0} + K_{r,1} Z_t + \sigma_r \epsilon^r_{t+1,t+3}. \end{aligned} \tag{27}$$

For SV models, the instantaneous volatility is assumed to take the same form as equation (23) but evolves at a quarterly frequency. We call these models Q-CV and Q-SV models, respectively.

Next, we examine the performance of temporally aggregated models. Table 5 displays the estimation results for Q-CV and Q-SV models. At a quarterly frequency, there is reasonable predictability shown by all three survey variables with R^2 of 4.27%, 0.88%, and 1.75%. However, unlike mixed-frequency models, the improvement from adding stochastic volatility is weaker. First of all, the volatility becomes less persistent and thus less predictable, as indicated by the estimate of $K_{V,1}$. The reason is trivial in that temporal aggregation tends to smooth out the high-frequency variation of volatility. Accordingly, there is no improvement in R^2, though the likelihood becomes slightly larger. AIC does not support Q-SV models, either. To summarize, at a quarterly frequency, the benefit from incorporating stochastic volatility is limited.

Turning to quarterly aggregate models for dollar index returns, we find no evidence of improvement due to stochastic volatility, either. The log likelihood and R^2 are indistinguishable for Q-CV and Q-SV models. The AIC for Q-SV models becomes even worse for IP and PCE, suggesting that stochastic volatility does not improve model inference at quarterly frequency.

6. Conclusion

In this paper, we address the issue of state-space modeling with mixed-frequency data. To this end, we propose a general particle filtering framework. The advantage of the mixed-

Table 5: Quarterly Predictive Regressions for Bond Returns

	IP		PCE		CPI	
	CV	SV	CV	SV	CV	SV
$K_{r,0}$	−1.29	−1.32	−0.65	−2.42	−1.87	−2.90
	(1.40)	(1.48)	(2.21)	(2.13)	(2.44)	(2.26)
$K_{r,1}$	3.79	3.39	3.45	5.92	5.49	6.74
	(1.70)	(1.80)	(3.46)	(3.34)	(3.88)	(3.56)
σ_r	7.23		7.35		7.32	
	(0.49)		(0.49)		(0.49)	
$K_{V,0}$		−8.08		−9.04		−8.91
		(1.87)		(1.92)		(1.91)
$K_{V,1}$		−0.49		−0.67		−0.64
		(0.35)		(0.36)		(0.35)
σ_V		0.46		0.44		0.40
		(0.21)		(0.20)		(0.19)
ρ		0.25		0.45		0.47
		(0.35)		(0.34)		(0.35)
Log likelihood	1.21	1.23	1.20	1.21	1.20	1.21
AIC$\times 10^{-3}$	−0.27	−0.27	−0.26	−0.26	−0.27	−0.26
$R^2(\%)$	4.27	4.27	0.88	0.88	1.75	1.75

Estimation results for predicting monthly bond returns using quarterly SPF. Bond returns are quarterly and aggregated from the monthly returns. Log likelihood is the time-series average of log likelihood. The sample period extends from 1990:Q3 to 2018:Q3.

frequency particle filter (MFPF) is that it allows for non-Gaussian shocks and nonlinear dynamics, for instance, stochastic volatility models. In dealing with mixed-frequency observations, we employ the backward smoother to facilitate forward filtering. The backward smoother mitigates the notorious issue of sample degeneracy that arises from mixed-frequency data. Simulation studies demonstrate that the filtering and smoothing error is lower than that of the commonly used state-augmented approach to MFPFs.

As an empirical application for our MFPF, we examine the forecasting power of quarterly economic surveys for Treasury bond returns and US trade-weighted dollar index. We find that monthly-evolving stochastic volatility models produce better forecasts than constant volatility models, whereas quarterly aggregate models do not. The empirical studies justify the use of mixed-frequency models that preserve the high-frequency nature of volatility dynamics.

Table 6: Quarterly Predictive Regressions for Dollar Index Returns

	IP		PCE		CPI	
	CV	SV	CV	SV	CV	SV
$K_{r,0}$	0.76	0.50	1.11	0.50	−1.11	−1.32
	(0.51)	(0.56)	(0.78)	(0.84)	(0.86)	(0.76)
$K_{r,1}$	−0.39	−0.14	−1.04	−0.33	2.64	2.80
	(0.61)	(0.66)	(1.23)	(1.30)	(1.37)	(1.22)
σ_r	2.62		2.61		2.58	
	(0.18)		(0.18)		(0.17)	
$K_{V,0}$		−14.45		−14.02		−14.66
		(3.24)		(3.19)		(2.63)
$K_{V,1}$		−0.97		−0.92		−0.97
		(0.44)		(0.44)		(0.36)
σ_V		0.19		0.23		0.10
		(0.18)		(0.18)		(0.16)
ρ		0.96		0.98		0.98
		(0.45)		(0.46)		(0.43)
Log likelihood	2.23	2.23	2.23	2.23	2.24	2.24
AIC$\times 10^{-3}$	−0.50	−0.49	−0.50	−0.49	−0.50	−0.50
$R^2(\%)$	0.36	0.36	0.64	0.64	3.22	3.22

Estimation results for predicting quarterly trade-weighted US dollar index returns using quarterly SPF. Dollar index returns are the quarterly trade-weighted US dollar index returns. Log likelihood is the time-series average of log likelihood. The sample period extends from 1990:Q3 to 2018:Q3.

References

Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein, 2010, Particle Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72, 269–342.

Aruoba, S Borağan, Francis X Diebold, and Chiara Scotti, 2009, Real-time measurement of business conditions, *Journal of Business & Economic Statistics* 27, 417–427.

Bai, Jennie, Eric Ghysels, and Jonathan H. Wright, 2013, State space models and MIDAS regressions, *Econometric Reviews* 32, 779–813.

Carter, Chris K., and Robert Kohn, 1994, On gibbs sampling for state space models, *Biometrika* 81, 541–553.

Carvalho, Carlos, Michael S. Johannes, Hedibert F. Lopes, and Nick Polson, 2010, Particle learning and smoothing, *Statistical Science* 25, 88–106.

Chernov, Mikhail, and Philippe Mueller, 2012, The term structure of inflation expectations, *Journal of Financial Economics* 106, 367–394.

Chun, Albert Lee, 2010, Expectations, bond yields, and monetary policy, *The Review of Financial Studies* 24, 208–247.

Creal, Drew, Bernd Schwaab, Siem Jan Koopman, and Andre Lucas, 2014, Observation-driven mixed-measurement dynamic factor models with an application to credit risk, *Review of Economics and Statistics* 96, 898–915.

Doucet, Arnaud, Simon Godsill, and Christophe Andrieu, 2000, On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing* 10, 197–208.

Durbin, James, and Siem Jan Koopman, 2012, *Time series analysis by state space methods*, volume 38 (Oxford University Press).

Eriksen, Jonas N, 2017, Expected business conditions and bond risk premia, *Journal of Financial and Quantitative Analysis* 52, 1667–1703.

Foroni, Claudia, Pierre Guérin, and Massimiliano Marcellino, 2018, Using low frequency information for predicting high frequency variables, *International Journal of Forecasting* 34, 774–787.

Frühwirth-Schnatter, Sylvia, 1994, Data augmentation and dynamic linear models, *Journal of Time Series Analysis* 15, 183–202.

Gargano, Antonio, Davide Pettenuzzo, and Allan Timmermann, 2017, Bond return predictability: Economic value and links to the macroeconomy, *Management Science* 65, 508–540.

Ghysels, Eric, 2016, Macroeconomics and the reality of mixed frequency data, *Journal of Econometrics* 193, 294 – 314.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov, 2004, The MIDAS touch: mixed data sampling regression models, Cirano working papers, CIRANO.

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov, 2007, MIDAS regressions: further results and new directions, *Econometric Reviews* 26, 53–90.

Giannone, Domenico, Lucrezia Reichlin, and David Small, 2008, Nowcasting: the real-time informational content of macroeconomic data, *Journal of Monetary Economics* 55, 665–676.

Godsill, Simon J., Arnaud Doucet, and Mike West, 2004, Monte Carlo smoothing for nonlinear time series, *Journal of the American Statistical Association* 99, 156–168.

Gordon, Neil J., David J. Salmond, and Adrian F.M. Smith, 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, in *IEE Proceedings F-Radar and Signal Processing*, volume 140, 107–113, IET.

Gürkaynak, Refet S, Brian Sack, and Jonathan H Wright, 2007, The US Treasury yield curve: 1961 to the present, *Journal of monetary Economics* 54, 2291–2304.

Harvey, Andrew C., 1989, *Forecasting, structural time series models and the Kalman filter* (Cambridge university press).

Heston, Steven L, 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency, *The Review of Financial Studies* 6, 327–343.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2014, Sequential learning, predictability, and optimal portfolio returns, *The Journal of Finance* 69, 611–644.

Johannes, Michael, Lars A. Lochstoer, and Yiqun Mou, 2016, Learning about consumption dynamics, *The Journal of Finance* 71, 551–600.

Johnson, Travis L, 2019, A fresh look at return predictability using a more efficient estimator, *The Review of Asset Pricing Studies* 9, 1–46.

Kitagawa, Genshiro, 1996, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* 5, 1–25.

Leippold, Markus, and Hanlin Yang, 2019, Mixed-frequency predictive regressions, Working paper, University of Zurich.

Liu, Jane, and Mike West, 2001, Combined parameter and state estimation in simulation-based filtering, in *Sequential Monte Carlo Methods in Practice*, 197–223 (Springer).

Marcellino, Massimiliano, Mario Porqueddu, and Fabrizio Venditti, 2016, Short-term GDP forecasting With a mixed-frequency dynamic factor model With stochastic volatility, *Journal of Business & Economic Statistics* 34, 118–127.

Mariano, Roberto S, and Yasutomo Murasawa, 2003, A new coincident index of business cycles based on monthly and quarterly series, *Journal of Applied Econometrics* 18, 427–443.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *The Journal of Finance* 72, 1611–1644.

Pettenuzzo, Davide, Allan Timmermann, and Rossen Valkanov, 2016, A MIDAS approach to modeling first and second moment dynamics, *Journal of Econometrics* 193, 315–334.

Pitt, Michael K., and Neil Shephard, 1999, Filtering via simulation: auxiliary particle filters, *Journal of the American Statistical Association* 94, 590–599.

Schorfheide, Frank, and Dongho Song, 2015, Real-time forecasting with a mixed-frequency VAR, *Journal of Business & Economic Statistics* 33, 366–380.

Schorfheide, Frank, Dongho Song, and Amir Yaron, 2018, Identifying long-run risks: A Bayesian mixed-frequency approach, *Econometrica* 86, 617–654.

Storvik, Geir, 2002, Particle filters for state-space models with the presence of unknown static parameters, *IEEE Transactions on Signal Processing* 50, 281–289.

Valle e Azevedo, Joao, Siem Jan Koopman, and António Rua, 2006, Tracking the business cycle of the euro area: A multivariate model-based bandpass filter, *Journal of Business & Economic Statistics* 24, 278–290.

Yang, Biao, Jonathan R Stroud, Gabriel Huerta, et al., 2018, Sequential Monte Carlo smoothing with parameter estimation, *Bayesian Analysis* 13, 1137–1161.