Federal Reserve Bank of Cleveland Working Paper Series

Addressing COVID-19 Outliers in BVARs with Stochastic Volatility

Andrea Carriero, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens

Working Paper No. 21-02R

August 2021

Suggested citation: Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens. 2021. "Addressing COVID-19 Outliers in BVARs with Stochastic Volatility." Federal Reserve Bank of Cleveland, Working Paper No. 21-02R. https://doi.org/10.26509/frbc-wp-202102r.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Addressing COVID-19 Outliers in BVARs with Stochastic Volatility*

Andrea Carriero,[1] Todd E. Clark,[2]
Massimiliano Marcellino,[3] and Elmar Mertens[4]

[1]*Queen Mary University of London,* [2]*Federal Reserve Bank of Cleveland,*
[3]*Bocconi University, IGIER, and CEPR,* [4]*Deutsche Bundesbank*

August 9, 2021

Abstract

The COVID-19 pandemic has led to enormous movements in economic data that strongly affect parameters and forecasts obtained from standard VARs. One way to address these issues is to model extreme observations as random shifts in the stochastic volatility (SV) of VAR residuals. Specifically, we propose VAR models with outlier-augmented SV that combine transitory and persistent changes in volatility. The resulting density forecasts for the COVID-19 period are much less sensitive to outliers in the data than standard VARs. Evaluating forecast performance over the last few decades, we find that outlier-augmented SV schemes do at least as well as a conventional SV model. Predictive Bayes factors indicate that our outlier-augmented SV model provides the best data fit for the period since the pandemic's outbreak, as well as for earlier subsamples of relatively high volatility.

Keywords: Bayesian VARs, stochastic volatility, outliers, pandemics, forecasts

JEL classification codes: C53, E17, E37, F47

*We gratefully acknowledge helpful suggestions and comments received from Michele Lenza, Frank Schorfheide, and participants at the NBER Summer Institute, the 11th ECB Conference on Forecasting Techniques, the 2021 IAAE conference, as well as seminars at the Federal Reserve Bank of New York, University of Pennsylvania, ECB, Bocconi University, Bank of England, and the Deutsche Bundesbank. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland, the Federal Reserve System, the Eurosystem, or the Deutsche Bundesbank. Additional results are provided in a supplementary online appendix, and replication codes are available at `https://github.com/elmarmertens/CCMMoutlierVAR-code`. Corresponding author: Todd E. Clark, todd.clark@researchfed.org.

1 Introduction

Bayesian VARs have a successful track record in point and density forecasting, the measurement of tail risks, and structural analysis. However, incoming data in 2020 posed some basic challenges for estimation and inference with VARs. The economic turbulence created by the ongoing COVID-19 pandemic is reflected in extreme realizations for a number of macroeconomic and financial series for the US, as shown in Figure 1. For example, payroll employment plummeted by about 15 percent from March to April 2020, a decline nearly 16 times as large as the previous largest monthly decline, and real income rose by about 12 percent in the month, an increase 3 times larger than the previous record growth rate.[1] Since then, real income has continued to fluctuate strongly, recording further record rates of increase and decline in early 2021. Measured by the business conditions index of Aruoba, Diebold, and Scotti (2009), the drop in real activity recorded in 2020 is more than 5 times as deep as in any other recession since 1960, so that the previous Great Recession of 2007-09 "appears minor by comparison" as noted by Diebold (2020). These extreme realizations can have strong effects on parameter estimates and forecasts generated by conventional constant-parameter VARs. In response, Schorfheide and Song (2020) suggest ignoring the recent data in estimating VAR parameters, whereas Lenza and Primiceri (2020) propose a specific form of heteroskedasticity, tuned to the COVID-19 data, to down-weight observations since March 2020 in the estimation.

Prior to the COVID-19 era, heteroskedastic VAR models, in particular models with stochastic volatility (SV), have been shown to provide more accurate point and density forecasts than constant-parameter models (see, e.g., Clark (2011), Clark and Ravazzolo (2015), and D'Agostino, Gambetti, and Giannone (2013)). SV models generate time variation in predictive densities through changes in the variance-covariance matrix of the VAR's forecast errors over time, with potential

[1] These calculations use log growth rates and data from the April 2021 vintage of FRED-MD. The rise in measured income from March to April also reflects payouts of government stimulus in that month. In contrast, over the following month, real income fell by about 4.5 percent, the then second-highest drop in our data (the largest drop in real income, by about 5 percent, that occurred in January 2013).

benefits for the accuracy of density forecasts (Clark, McCracken, and Mertens (2020)). In addition, heteroskedasticity affects the estimation of slope coefficients in each VAR equation (at least in finite samples). As an application of generalized least squares, when extreme realizations are modeled as sudden increases in volatility, heteroskedastic VARs will down-weight the associated observations when estimating parameters; in the limit, outliers associated with infinite volatility would be discarded.[2]

A typical SV model assumes changes in volatility to be highly persistent. However, by definition, extreme observations are more reflective of short-lived spikes, not permanent increases, in volatility. Like Schorfheide and Song (2020) and Lenza and Primiceri (2020), we view the extreme observations of the COVID-19 period as possible outliers that are characterized by transient and infrequent increases in volatility, in which case it may be desirable to reduce their influence on model estimates and forecast distributions. In earlier work, the conventional SV model has already been extended to feature fat-tailed, instead of normal, errors, as in Jacquier, Polson, and Rossi (2004), henceforth denoted "SV-t." More recently, Stock and Watson (2016) designed a discrete mixture representation of rare but large outliers in volatility (henceforth "SVO"), as observed, for example, during the global financial crisis.

In this paper, we introduce a novel combination of (1) an SV model with large but infrequent volatility outliers with (2) an SV-t model that generates fat-tailed errors to an otherwise standard Bayesian VAR (BVAR). The resulting SVO-t model nests the Stock-Watson SVO approach and the fat-tail SV-t model of Jacquier, Polson, and Rossi (2004). We show that a BVAR with SVO(-t) is highly effective in filtering outliers from data associated with the COVID-19 pandemic, to produce better-behaved forecast densities.

[2]For example, when applied to data samples starting in the 1960s or 1970s, VARs with SV tend to discount data points prior to the onset of the low-volatility period known as the Great Moderation that started in the mid-1980s (Perez-Quiros and McConnell (2000)). Of course, the distinction between generalized and ordinary least squares matters only in finite samples, as both converge to the same asymptotic limit (to which a Bayesian estimate would also converge). But as demonstrated by the COVID-19 episode, common samples of macroeconomic data are still sufficiently finite for (huge) outliers to matter.

The Stock-Watson volatility model with outliers, SVO, augments the standard SV specification of a highly persistent volatility state with an outlier volatility state that infrequently and temporarily jumps to values above 1. In its original form, first considered by Stock and Watson (2016) in unobserved component models of inflation, the model has Gaussian errors. We combine the outlier-augmented SV process with VARs that have Gaussian (SVO) or t-distributed (SVO-t) errors, and also consider the case of t-distributed errors without volatility outliers (SV-t). In light of the heterogeneous occurrence of outliers that is visible, for example, in Figure 1, outlier states are variable-specific in our baseline model.

We demonstrate that SVO, SVO-t, and SV-t share the same latent state representation where residuals are written as the product of a normally distributed shock and a set of outlier states, but differ in the assumed densities for the outlier states. In particular, SVO puts more mass on outliers being large events that increase volatility by more than twofold, whereas SV-t sees outliers as more moderately sized, and SVO-t is a combination of the two. Conventional procedures for estimation of BVAR-SV models can easily be extended to handle SVO(-t). Specifically, we show that the standard MCMC algorithm used for estimation of BVAR-SV models can still be used, but with the addition of two extra steps. First, realized outlier states need to be drawn from their posteriors, conditional on draws for each variable's outlier probability. Second, the outlier probability for each variable is drawn from a (conditional posterior) distribution conditional on the draws of the time series of outlier states.

Throughout our empirical analysis, we use a medium-sized data set of 16 monthly variables, motivated by research that has found that larger BVARs tend to forecast more accurately than smaller BVARs, while going beyond medium-sized models adds little (e.g., Bańbura, Giannone, and Reichlin (2010), Carriero, Clark, and Marcellino (2019), and Koop (2013)). Although at this point we are comfortable viewing the extreme realizations of the COVID-19 period as outliers, we should emphasize that our approach is data-based: Our model provides a probabilistic assessment of timing and scale of realized outliers in the data; we are not simply deeming (i.e., restricting) recent observations to be outliers.

As a starting point for our empirical work, we confirm the findings of Lenza and Primiceri (2020) and Schorfheide and Song (2020) that forecasts generated since March 2020 from homoskedastic BVARs are often distorted. For example, suppose one uses monthly data through April 2020 to estimate a medium-sized BVAR and forecast payroll employment growth starting in May 2020. In light of the suggestion of Schorfheide and Song (2020), we also consider forecasts for the same period but using parameter estimates based on data ending in February.[3] The forecasts turn out to be strikingly different. In general, the recent outliers cause the forecast paths of some variables to become extreme by historical standards. Instead, we find that BVARs with SV or SVO specifications generated better-behaved forecasts than in the constant-variance case. Both SV and SVO estimates register increases in forecast uncertainty. But, while the SV specification sees all shocks to forecast uncertainty as permanent, the SVO(-t) model explicitly allows for one-off spikes in volatility, resulting in estimates of forecast uncertainty that are still elevated but, in our subjective assessment, appear less extreme and more reasonable. So, in our assessment, the SVO and SVO-t specifications offer an effective approach for managing infrequent outliers with BVARs used for forecasting.

As an alternative, we also consider relying on a standard BVAR-SV but treating as missing data those observations identified ex-ante as extreme. The methods discussed so far adjust parameters (including the volatility states) but not the data vector used at the forecast origin in forming a prediction; treating observations as missing data also alters the jumping-off point of the forecasts. To identify extreme observations as outliers, we use an ex-ante criterion known from the literature on dynamic factor models that is based on the distance of a given data point from the time-series median.[4] This approach differs from the SVO approach, which estimates the occurrence of outliers jointly with the VAR, by treating the dates of outliers as known ex-ante. In addition, echoing

[3]Forecasts for April 2020, obtained with parameter estimates based on pre-COVID-19 data, are documented in the supplementary online appendix.

[4]Following Stock and Watson (2002), applications of dynamic factor models have considered observations to be outliers when they are some multiple of the inter-quartile range away from the series median; among others, see Artis, Banerjee, and Marcellino (2005) and McCracken and Ng (2016).

arguments developed in Mitchell and Weale (2021), the missing-data treatment remains agnostic about the specific stochastic properties of those observations that are pre-selected as outliers. In the COVID-19 period, this approach also produces much better-behaved forecasts than a constant-variance BVAR. In forecasting, the biggest difference with the outlier-augmented SV procedures is that conditioning on the incidence of outliers, while otherwise ignoring any signal from their specific realization, leads to predictive densities that can be considerably tighter than those from SVO(-t) (or SV-t).[5]

Although to this point we have focused on the efficacy of methods for reducing distortions to forecast distributions in the presence of outliers, to be broadly effective, it is important that a given method not only helps reduce such distortions but also performs effectively in forecasting over long periods of time less affected by outliers. Accordingly, we conduct a quasi-real-time evaluation of forecast performance using monthly data with an evaluation window starting in 1975 and ending in 2017, comparing the accuracy of point and density forecasts from our proposed SVO and SVO-t specifications and the alternatives discussed above. It turns out that pre-COVID data include outliers; indeed, SVO(-t) detects pre-COVID-19 outliers in macroeconomic and financial time series, whose existence had been noted before by, among others, Stock and Watson (2002). In historical forecast accuracy for the 1975-2017 sample, the SVO approach marginally outperforms SV. The SVO-t model, which features stochastic volatility, fat tails, and the outlier state treatment, yields even better forecasts. However, the alternative approach of treating outliers as missing data in an otherwise conventional VAR with SV performs about as well.

All told, the use of VARs with time-varying volatility, like SV and SVO, broadly mitigates the drastic effects that outliers can have on forecasts. But only an outlier-augmented SV specification, like SVO or SVO-t, or the alternative of treating extreme observations as missing, prevents the width of predictive densities from blowing up as they would in the SV case. Importantly, the

[5] Even tighter densities (around similar point forecasts) are obtained when using deterministic dummies for each month of the COVID-19 period. The use of dummies essentially removes pandemic observations from the estimation sample, which reduces uncertainty for exogenous reasons.

added value of SVO and SVO-t also holds up over a longer sample outside the recent COVID-19 episode.

To evaluate which model best characterizes the data in the COVID-19 period, we employ predictive Bayes factors (which are based on sums of predictive likelihoods). By this measure, our SVO specification fits the COVID-19 sample the best, with SVO-t next. In earlier samples, the SVO and SVO-t specifications also fare well in model fit. The advantages of these models are driven by the subsamples of relatively high volatility; the baseline SV model fits best in the Great Moderation years of 1985 through 2007.

The remainder of this paper proceeds as follows. Section 2 briefly reviews the related literature not covered above. Section 3 introduces the SVO and SVO-t models and alternative specifications to handle outliers, and describes their estimation. Section 4 describes the data used. Section 5 provides our results, including a forecast comparison between the various models over a long pre-COVID-19 sample, details about estimated outlier states before and during the COVID-19 episode, the evolution of forecasts made over the course of 2020 and early 2021, and model fit. Section 6 summarizes robustness checks provided in our supplementary online appendix. Section 7 concludes.

2 Related literature

In addition to Lenza and Primiceri (2020), Schorfheide and Song (2020), and our paper, the arrival of COVID-19 has prompted a number of studies to consider treating the extreme observations of the COVID-19 period as outliers. A particular contribution of our paper is the comprehensive analysis of model fit and forecast performance over a wide set of macroeconomic and financial variables of BVARs with and without outlier-augmented SV. By studying model performance over a relatively long sample of post-war US data, we can also document the recurring benefits of outlier treatments at times of crisis or other economic upheavals.

Antolín-Díaz, Drechsel, and Petrella (2021) develop a dynamic factor model for nowcasting in the US, evaluated in real time from 2000 onward, where outliers are modeled as additive measurement errors that have t-distributions. Focusing on euro area inflation, Bobeica and Hartwig (2021) document that pandemic observations can shift parameter estimates and find some benefits to allowing fat tails in a VAR for the euro area.[6] In another application to euro area data, Alvarez and Odendahl (2021) find that the pandemic's outliers distort VAR estimates and consider alternative approaches to modeling volatility outliers.

Prior to the arrival of COVID-19, some studies had already considered VAR specifications with fat-tailed error distributions. For example, t-distributed shocks were used in BVAR-SV models by Chiu, Mumtaz, and Pintér (2017) and Clark and Ravazzolo (2015) and estimated DSGE models, with and without SV, by Cúrdia, Del Negro, and Greenwald (2014) and, more recently, by Chib, Shin, and Tan (2021). Karlsson and Mazur (2020) and Chan (2020) provide general treatments of heteroskedasticity in BVAR models with and without SV and fat-tailed error distributions.

Other more recent analyses have proposed approaches more geared to specific circumstances of the pandemic. In part, these approaches go beyond our forecasting application and consider the estimation of causal (or structural) dependencies. For example, Primiceri and Tambalotti (2020) and Ng (2021) argue for seeing the COVID-19 period as adding a new type of shock to the dynamic system of the economy. Based on the assumption that the new COVID-19 shock has been the dominant source of variation since early 2020, Primiceri and Tambalotti (2020) derive a set of conditional forecasts for different scenarios of future developments. Instead, Ng (2021) uses pandemic indicators to "de-covid" data prior to estimation of time series models. Specifically, in application to a structural VAR, Ng (2021) shows that after accounting for exogenous COVID-19-related indicators, dynamic responses to other shocks appear similar pre- and post-COVID-19.[7]

[6] In a related study, Hartwig (2021) considers US data.

[7] In another application, Ng (2021) augments a dynamic factor model with COVID-19-related indicators to update uncertainty measures from Jurado, Ludvigson, and Ng (2015).

3 BVAR models

We study VAR models of the following form:

$$y_t = \Pi_0 + \Pi(L)y_{t-1} + v_t\,, \qquad\qquad v_t \sim N(0, \Sigma_t) \tag{1}$$

where y_t is a vector of N observables, $\Pi(L) = \sum_{i=1}^{p} \Pi_i \; L^{i-1}$ is a pth order lag polynomial of VAR coefficients, and v_t denotes the VAR's residuals. We denote the vector of stacked coefficients contained in $\{\Pi_i\}_{i=0}^{p}$ as Π. Throughout, we maintain the assumption of time-invariant transition coefficients Π, which is commonly (and so far successfully) used in forecasting.[8] All models are specified with non-conjugate priors for Π and Σ_t.

The models differ mainly in whether the residuals are homoskedastic, or in the form of their heteroskedasticity. Heteroskedasticity in VAR residuals has important effects on the estimation of Π, in particular when there are outliers with large residual volatility. Intuitively, observations with higher residual volatility receive less weight in the estimation of VAR coefficients. For the sake of illustration, consider an AR(1) model without intercept: $y_t = \pi y_{t-1} + v_t$, $v_t \sim N(0, \sigma_t^2)$ with σ_t^2 known, and a prior conditional on past data $\pi | y_{t-1} \sim N(\underline{\pi}, \underline{\omega}^2)$. This is a signal extraction problem where y_t serves as a noisy signal about the unknown π, with a signal-to-noise ratio that is decreasing in σ_t^2. Accordingly, the posterior mean for π is a weighted average of the prior mean, $\underline{\pi}$, and the data-driven OLS estimate, π^{OLS}, with the weight decreasing in σ_t^2. In the case of observing

[8] These linear models remain the workhorse of applied forecasting in policy analysis and a benchmark for use in research. Beyond linear VARs, Guerrón-Quintana and Zhong (2017) and Huber, et al. (forthcoming) employ semi- and non-parametric methods to better allow forecasting relationships to adapt to changing conditions, in particular at times of crisis. Having said that, our proposed approach to outliers could also be incorporated into VARs that feature time-varying regression parameters in the smaller specification and estimation approach of D'Agostino, Gambetti, and Giannone (2013) and the larger specification and estimation approach of Chan (2019). Results reported in the supplementary online appendix indicate little to no variation in the parameters of the conditional mean in heteroskedastic VAR models since early 2020, but sizable variation in a homoskedastic VAR.

a single observation y_t, these are:

$$E(\pi | y_t, y_{t-1}) = (1-\kappa) \cdot \underline{\pi} + \kappa \cdot \pi^{OLS}, \quad \text{with} \quad \pi^{OLS} = \frac{y_t \, y_{t-1}}{y_{t-1}^2}, \quad \text{and} \quad \kappa = \frac{\underline{\omega}^2}{\underline{\omega}^2 + \frac{\sigma_t^2}{y_{t-1}^2}}.$$

Recursive application of the above extends the example to multiple periods. In addition, the logic of down-weighting observations subject to high residual variance carries over to the multivariate case, as described, for example, in Koop (2003, Chapter 6).

As argued above, time-varying volatility in the VAR residuals, v_t, can help to insulate estimation of the transition coefficients Π from the effects of extreme outliers. However, density forecasts will crucially depend on the assumed dynamics of the variances in Σ_t, and we further consider different forms of persistence in variance changes below.

Down-weighting extreme observations in the estimation of Π will not completely insulate the resulting forecasts from outliers. Consider again the case of the AR(1) without intercept, where the h-step-ahead forecast is given by $y_{t+h|t} = \pi^h \, y_t$ and y_t was an outlier. Even if the outlier were excluded from estimation of π, it would still have a direct effect on the forecast $y_{t+h|t}$.[9] To address these concerns, we consider a variant of the SV model that treats pre-specified outliers as missing values. To identify extreme observations as outliers, we use an ex-ante criterion taken from the literature on dynamic factor models that is based on the distance between a given data point and the time-series median. Treating pre-identified outlier observations as missing data also avoids specification of their exact stochastic distribution; see, for example, Mitchell and Weale (2021) for a related discussion.

[9]In VAR (or AR) models with higher lag orders, the forecast would not singularly depend on the outlier y_t but also preceding values that are not necessarily outliers. Nevertheless, outliers in the "jump-off" data point, y_t, may unduly influence the forecast.

3.1 Model specification

We consider the following six variants of the VAR model (1). The first five differ in the specified process for the residuals v_t, whereas the last variant treats pre-specified outliers as missing data. Section 6 includes a summary of robustness checks conducted with some other specifications, described there.

1) CONST: A homoskedastic VAR with $v_t \sim N(0, \Sigma)$.

2) SV: In this baseline SV model, the VAR residuals can be written as

$$v_t = A^{-1} \Lambda_t^{0.5} \varepsilon_t \,, \qquad \text{with} \quad \varepsilon_t \sim N(0, I) \,, \tag{2}$$

where A^{-1} is a unit-lower-triangular matrix, $\Lambda_t^{0.5}$ is a diagonal matrix of stochastic volatilities, and the reduced-form variance-covariance matrix of innovations is $\Sigma_t = A^{-1} \Lambda_t (A^{-1})'$. The vector of logs of the diagonal elements of Λ_t, denoted $\log \lambda_t$, evolves as a random walk with correlated errors:

$$\log \lambda_t = \log \lambda_{t-1} + e_t \,, \qquad \text{with} \quad e_t \sim N(0, \Phi). \tag{3}$$

3) SVO-t: The SVO-t model is intended to capture two kinds of outliers that are both modeled as transitory changes in volatility: The first kind captures rare jumps in volatility. The second kind occurs more often, but is less extreme in impact (consistent with draws from the tails of a fat-tailed distribution). Each kind of outlier enters the model in the form of a diagonal matrix of scale factors, denoted O_t and Q_t, with diagonal elements $o_{j,t}$ and $q_{j,t}$, respectively, that are mutually $i.i.d.$ over all j and t.

The first kind of outlier, $o_{j,t}$, has a two-part distribution that distinguishes between regular observations and outliers. When variable j has a regular observation in period t we have $o_{j,t} =$

1, while for an outlier it is $o_{j,t} \geq 2$.[10] Outliers in variable j occur with probability p_j and the distribution for $o_{j,t}$ is:

$$o_{j,t} = \begin{cases} 1 & \text{with probability} \quad 1 - p_j \\ U(2, 20) & \text{with probability} \quad p_j \end{cases}$$

for $j = 1, \ldots, N$ and where $U(2, 20)$ denotes a uniform distribution with support between 2 and 20.

The second, less extreme, type of outlier in the SVO-t model is equivalent to having t-distributed VAR residuals (conditional on Λ_t and O_t). Following Jacquier, Polson, and Rossi (2004), we let the squares of the diagonal elements of Q_t, $q_{j,t}$, have inverse-gamma distributions:

$$q_{j,t}^2 \sim IG\left(\frac{d_j}{2}, \frac{d_j}{2}\right).$$

The vector of VAR residuals in the SVO-t model is written as

$$v_t = A^{-1}\, \Lambda_t^{0.5}\, O_t\, Q_t\, \varepsilon_t,$$

with A^{-1} and $\Lambda_t^{0.5}$ specified as before. The j^{th} residual $q_{j,t} \cdot \varepsilon_{j,t}$ (adjusted for the rotation by A^{-1} and scaling by $\Lambda_t^{0.5}\, O_t$), has a student-t distribution with d_j degrees of freedom, since $\varepsilon_{j,t} \sim N(0, 1)$ and $d_j / q_{j,t} \sim \chi^2_{d_j}$. Since O_t, Q_t, and Λ_t are diagonal, they commute, and the time-varying variance-covariance matrix of the VAR residuals can conveniently be expressed as $\Sigma_t = A^{-1}\, O_t\, Q_t\, \Lambda_t\, Q_t'\, O_t'$ $(A^{-1})'$.

We place a beta prior on the outlier probability p_j that corresponds to 10 years' worth of prior data, centered at a mean consistent with one outlier per decade. For the t-component of the SVO-t model, we follow Jacquier, Polson, and Rossi (2004) and estimate the degrees of freedom d_j for

[10] The lower bound of 2 on the scale shift in outliers is motivated by seeing outliers as events firmly outside the typical mass of their otherwise Gaussian distribution (conditional on $o_{j,t}$).

each variable using a uniform discrete prior with a range of 3 to 40.

4) SVO: We also consider a simplified version of the SVO-t model, denoted SVO, where $Q_t = I$ so that the VAR residuals are Gaussian (conditional on O_t and Λ_t). In this case, the time-varying variance-covariance matrix of the VAR residuals is given by $\Sigma_t = A^{-1} O_t \Lambda_t O_t' (A^{-1})'$. The SVO model is similar to the treatment of volatility outliers by Stock and Watson (2016) in an unobserved component model of inflation.[11] As in Stock and Watson (2016), we place a beta prior on the outlier probability p_j so that the prior mean implies an outlier frequency of once every 4 years in monthly data (and precision consistent with 10 years' worth of prior observations). As illustrated in the supplementary online appendix, the prior mean of $p_j = 1/(4 \cdot 12)$ implies about the same variance of $o_{j,t}$ in the SVO model as do our prior means of p_j and d_j in the SVO-t model for the combined outlier states $o_{j,t} \cdot q_{j,t}$.

5) SV-t: The SV model with t-distributed errors, SV-t, is a simplified version of the SVO-t model where $O_t = I$, so that the time-varying variance-covariance matrix of the VAR residuals is given by $\Sigma_t = A^{-1} Q_t \Lambda_t Q_t' (A^{-1})'$. The SV-t specification corresponds to the fat-tailed SV model of Jacquier, Polson, and Rossi (2004), where the standard-normal shocks ε_t driving the VAR residuals in (2) are replaced by t-distributed shocks. For our estimation, the degrees of freedom of the t distribution are estimated as in Jacquier, Polson, and Rossi (2004), using a uniform discrete prior with a range of 3 to 40. Overall, estimates not reported in the interest of brevity confirm that SVO is more geared than SV-t toward generating sizable outliers at a variable-specific rate of occurrence p_j that is directly governed by an explicit prior, and SVO-t adds to that the flexibility of a fat-tailed error distribution.

6) SV-OutMiss: This model applies the standard SV specification for Σ_t, but ignores a given set of outlier observations in the VAR estimation altogether by treating them as missing data. The approach builds on a practice known from the literature on dynamic factor models (DFM), in which

[11] In an application to inflation data, Stock and Watson (2016) use a $U(2, 10)$ distribution for $o_{j,t} > 1$.

input data are pruned of extreme observations that are multiples times the inter-quartile range away from the series median. Typical values for the multiple used in the literature vary from 5 to 10, and we adopt a threshold factor of 5 as a baseline (we obtain very similar results using a factor of 10). Our supplementary online appendix provides an overview of which observations in our data qualify as outliers according to this criterion. Apart from readings for employment, consumption, income, and stock returns in 2020-21, and the fairly frequent occurrence of outliers in income throughout the sample seen also in Panel (a) of Figure 1, further outliers are recorded in industrial production, inflation, and stock returns during the recession of 2007-09. as well as exchange rates during the 1970s.[12]

The DFM literature replaces extreme observations by the time-series median or a similar moment of central tendency. We adopt the same ex-ante criterion for the identification of outliers, but we instead treat these as missing data in estimation and forecasting. In the limit, the missing data approach corresponds to a version of attaching additive measurement error to specific observations, but with infinite variance, whereas the remaining observations are observed without error. For each missing value, our Bayesian methods generate a posterior distribution that also informs the resulting forecasts. Formally, denote the history of y_t after pruning outliers as z^t, and continue the AR(1) example introduced above: Forecasts are then generated by $y_{t+1|t} = \pi^h \, E(y_t|z^t)$ where $E(y_t|z^t)$ is identical to y_t in the no-outlier case. Similarly, forecast uncertainty is generated based on estimates of SV that condition only on z^t, not potential outliers in the history of y_t.

3.2 Model estimation

Each of our models is estimated with an MCMC sampler, based on the methods of Carriero, Clark, and Marcellino (2019) (henceforth "CCM") for estimating large BVARs, but as corrected in Carriero, et al. (2021a). As in CCM, we use a Minnesota prior for the VAR coefficients Π and follow their other choices for priors as far as applicable, too. Throughout, we use $p = 12$ lags in a monthly

[12] In recognition of the end of the Bretton-Woods system, outliers in exchange rates are ignored for estimation windows ending prior to 1985.

data set, which is described in further detail in Section 4.

Here we briefly explain the algorithm adjustments needed for the version of the model with constant variance and the alternative with outlier volatility states. The algorithm includes all of the same steps given in CCM (as corrected in Carriero, et al. (2021a)), except for necessary adjustments to account for the two alternative cases. For the constant-volatility model, an inverse-Wishart prior for Σ, with a (conditionally) conjugate inverse-Wishart updating step for the MCMC sampler, replaces the SV block of the model.[13]

For the SVO-t variant, the following extra steps are added to the original BVAR-SV setup: Realized outlier states $o_{j,t}$ and $q_{j,t}$ need to be drawn from their posteriors. The step for $o_{j,t}$ conditions on draws for the outlier probability p_j and proceeds analogously to the sampling of the mixture states needed with the Kim, Shephard, and Chib (1998) approach to the stochastic volatility states $\log \lambda_t$. The step for $q_{j,t}$ takes a draw from an inverse Gamma distribution. A further additional step draws the outlier probability p_j for each variable from a (conditional posterior) beta distribution conditional on the draws of the time series of outlier states. The algorithms for SVO and SV-t are simplified versions of that for SVO-t.[14]

For the SV-OutMiss model, which treats pre-specified outliers as missing values, the MCMC sampler for the standard SV model is augmented by an additional step that draws the missing values from a state-space representation of the VAR system using the disturbance smoothing algorithm of Durbin and Koopman (2002). Computational cost increases substantially with the SV-OutMiss model, as it requires an additional sequence of Kalman filtering and smoothing steps. In contrast, the added cost of computing SVO-t (or SVO or SV-t) over standard SV is small, since this model

[13]The prior for Σ in the constant-variance model is uninformative; that is, we use an improper Wishart with zero degrees of freedom and scale matrix equal to zero.

[14]The ordering of steps in our MCMC sampler reflects the recommendations of Del Negro and Primiceri (2015) as implemented also by Cúrdia, Del Negro, and Greenwald (2014) (for SV-t) and Stock and Watson (2016) (for SVO). Specifically, the t-error states, $q_{j,t}$, are sampled before the SV mixture states of Kim, Shephard, and Chib (1998), while draws from $o_{j,t}$ condition on those mixture states so that $o_{j,t}$ and p_j are sampled after the SV steps known from Kim, Shephard, and Chib.

adds only steps for sampling the $i.i.d.$ outlier states.

All results in the paper are based on 1,000 retained draws, obtained by sampling a total of 1,200 draws with 200 burn-in draws. Unreported comparisons of posteriors obtained under different starting values indicate satisfactory convergence of the MCMC algorithms.

4 Data

Our data set consists of monthly observations for 16 macroeconomic and financial variables for the sample 1959:M3 to 2021:M3, taken from the April 2021 vintage of the FRED-MD database maintained by the Federal Reserve Bank of St. Louis. The variables and their transformation to logs or log-differences are listed in Table 1. To avoid issues related to the effective lower bound (ELB) on nominal interest rates, the data set includes only longer-term interest rates and omits a policy rate measure, like the federal funds rate, which was constrained by the ELB from late 2008 to 2016, and then again starting in March 2020.[15]

A few selected series are shown in Figure 1. Observations are marked as outliers in red if their distance from the series median exceeds 5 times the inter-quartile range (IQR), where median and IQR are computed from the pre-COVID-19 sample. As discussed in the introduction, similar definitions of outliers have been used in the literature on factor models. Real personal income, shown in Panel (a) of the figure, has regularly displayed outliers over the post-war sample. Many other series, like payroll growth shown in Panel (b), exhibit such outliers only over the recent COVID-19 period, whereas a few others, like returns on the S&P500, in Panel (c), inflation, or the exchange rate between the US dollar and pound sterling, displayed large outliers only on earlier

[15]The related paper by Lenza and Primiceri (2020) does not include any interest rates in its VAR setup. When simulating forecasts for our longer-rate measures, the 5- and 10-year Treasury yields, individual draws have fallen below the ELB as well, and the predictive densities were truncated at the ELB in these cases. Due to the dynamic nature of the forecast simulation, this truncation also has indirect effects on the predictive densities of other variables. In companion work (Carriero, et al. (2021b)), we focus on the estimation of VARs that model nominal interest rates as censored variables.

occasions. Some variables, like the unemployment rate in Panel (d), have registered outstanding changes since the pandemic's outbreak, but without registering explicit outliers by this metric. In some cases, outliers may be attributed to unusual events. For example, in results not shown, industrial production registers a positive outlier in December 1959, when production bounced back following a strike in the steel industry from mid-July through early November. More recently, income transfers from the CARES Act caused growth in personal income to surge in April 2020.

5 Results

This section presents results on forecast performance pre-COVID-19, outlier estimates, forecasts made in 2020-21, and model fit.

5.1 Forecast performance pre-COVID-19

Applicability of the outlier-augmented BVAR-SVO and BVAR-SVO-t models is not necessarily specific to data resulting from the current COVID-19 pandemic. As noted above, individual data series have exhibited occasional outliers before, leading to some earlier studies of the potential benefits of modeling fat-tailed error distributions and other forms of outliers.[16] On the other hand, to the extent our models are motivated by a goal to accommodate the extremes of the COVID-19 period, one might be concerned that a model successful in this period could in some sense overfit earlier data. Motivated by these considerations, in this section, we evaluate the forecast performance of the alternative BVAR specifications described in Section 3 when applied to data prior to the onset of COVID-19.

We conduct an out-of-sample forecast evaluation in quasi-real time, where we simulate forecasts made from 1975:M1 through 2017:M12.[17] For every forecast origin, each model is re-

[16]See, for example, Chiu, Mumtaz, and Pintér (2017), Clark and Ravazzolo (2015), and Cúrdia, Del Negro, and Greenwald (2014) for the use of SV-t specifications in VARs or DSGE models and Stock and Watson (2016) for the use of SVO in unobserved component models.

[17]The end of our evaluation window has been chosen to avoid overlap with COVID-19-related realizations; how-

estimated based on growing samples of data that start in 1959:M3. All data are taken from the April 2021 vintage of FRED-MD; we abstract from issues related to real-time data collection. The forecast horizons considered extend from 1 to 24 months. We evaluate point and density forecasts based on root-mean-squared errors (RMSE) and continuous ranked probability scores (CRPS), respectively, as described in, among others, Clark and Ravazzolo (2015) and Krüger, et al. (forthcoming). Statistical significance of differences in loss functions is evaluated using the Diebold and Mariano (1995) and West (1996) test.

Tables 2 and 3 compare point and density forecasts generated by a homoskedastic BVAR and BVARs with SV and SVO-t specifications, taking the SV model as the benchmark. In Table 2's point forecasts, across variables and horizons SV and CONST are broadly comparable in accuracy. For some variable-horizon combinations, one may be a little better or worse than the other (with RMSE ratios ranging between roughly 0.97 and 1.03), but the differences are generally immaterial. Density results as gauged by the CRPS are broadly comparable, with SV sometimes modestly better than CONST and sometimes modestly worse. However, at the 24-months-ahead horizon, the performance of the SV model deteriorates, with the CONST specification offering gains often on the order of 20 to 25 percent.

The SVO-t specification could be expected to capture better the occasional outliers in pre-COVID-19 data, but possibly also at the expense of overfitting elsewhere. However, such concerns are not borne out by our forecast evaluation. In terms of both point and density forecasts, SVO-t typically performs as well as, and at times even better than, SV, as well as the CONST model. Point forecasts generated by the SVO-t model over the post-war period (and pre-COVID) are generally on par with those from the SV model, with RMSE ratios in some cases a little below or above 1 but often very close to 1. With density forecast accuracy as gauged by the CRPS, at shorter horizons the SVO-t specification performs very similarly to the SV baseline, with CRPS ratios very close to 1, occasionally a bit lower. At the 12 months horizon, SVO-t yields larger gains

ever, we obtain very similar results when the evaluation window is extended through the end of our data sample in 2020.

over SV, ranging from 2 to 6 percent. Bigger gains in accuracy occur at the horizon of 24 months, with improvements as large as 15 percent. At this horizon, SVO-t improves forecast accuracy for variables including consumption, industrial production, employment, hours, and stock returns. The SVO-t gains are largest for real income, the variable most prone to outliers. Overall, the evidence suggests that consistent use of SVO-t over the post-war sample improves on the commonly used SV specification, in particular in terms of density forecasts and for those variables more subject to frequent outliers, such as personal income.

Tables 4 and 5 compare SVO-t against versions of the model that strip out the t-distributed component (SVO) or the Stock-Watson outlier state (SV-t) as well as the SV-OutMiss approach, which treats pre-specified outliers as missing data as described in Section 3. Note that these comparisons take SVO-t as the baseline, so that a ratio less (more) than 1 means the alternative model is more (less) accurate than the baseline. Point forecasts from the SVO and SV-t alternatives are quite similar in accuracy to those from the SVO-t specification. Differences in relative RMSE are never more than 4 percent and typically just 1 percent or less. In density accuracy, the SV-t model is similar to our preferred SVO-t specification with CRPS ratios between 0.98 and 1.0 in nearly all cases. At horizons of 12 months or less, the SV-OutMiss specification yields density accuracy very similar to that of the SVO-t baseline. But at 24 months, SV-OutMiss forecasts can be modestly less accurate (e.g., industrial production and hours). The more noticeable differences in CRPS accuracy occur with the SVO model. Although at shorter horizons SVO accuracy is quite similar to SVO-t accuracy, at longer horizons SVO-t provides the more accurate forecasts, often by a statistically significant margin, reaching 10 to 12 percent for consumption, industrial production, employment, and hourly earnings.

5.2 Outlier estimates in 2020-21 and before

As described in Section 3, the SVO-t approach extends the baseline SV model by adding latent outlier states $o_{j,t}$ and $q_{j,t}$ for each variable $j = 1, \ldots, N$, with the former uniformly distributed and squares of the latter having an inverse Gamma distribution. The outlier states enrich the dynamics of the time-varying variance-covariance matrix, Σ_t, so that volatility can change due to transitory changes in $o_{j,t}$ and $q_{j,t}$, as well as the persistent variations induced through the log-SV terms $\log\lambda_t$. The SVO model adds just the state $o_{j,t}$ to an SV model, whereas the SV-t specification adds just the state $q_{j,t}$. In each case, the additional latent states serve to pick up on temporary increases in volatility that would be ill-represented by the more persistent variations modeled via the conventional SV processes for $\log\lambda_t$.

To show how our models view the evidence on outlier probabilities over the full historical sample, Table 6 reports posterior means and 68 percent credible sets for the probabilities of large outliers in the SVO and SVO-t models and for the degrees of freedom for the fat-tail components of the SV-t and SVO-t specifications, along with the corresponding priors. In the SVO model, the posterior mean probability of a large outlier is greatest for real income, at 3.19 percent, and ranges from about 0.3 percent (housing starts) to 1.1 percent (nonfarm payrolls and hours) for other variables. In the SV-t specification, the posterior mean estimate of the degrees of freedom is 3 for one-half of the model's variables — implying frequent small outliers — but above 20 (near-Gaussian) for six other variables. In the SVO-t model that allows for both small and large outliers, the estimated degrees of freedom are quite similar to those of the restricted SV-t specification, whereas the estimated probabilities of large outliers are sharply lower than in the SVO model.

We can also provide a closer comparison of the volatility and outlier estimates obtained from SVO-t, SVO, and SV-t. Focusing on just real income and S&P500 returns in the interest of chart readability, Figure 2 displays posterior medians of the SV component (i.e., $\lambda_{j,t}^{0.5}$) and outlier estimates ($o_{j,t}$ and $q_{j,t}$) obtained over the full sample, with the SV component captured in solid black lines and the outlier components in dotted colored lines. Echoing our discussion of each model's properties in Section 3, these results show that, when SV-t (first column) and SV (second column) are compared, SV-t tends to see outliers as being more moderately sized but occurring also more regularly than SVO. For example, in the real income estimates, SV-t shows a relatively large number of outliers in the 1970s and 1980s, whereas SVO shows fewer outliers that are larger in size. With S&P500 returns, SVO shows few outliers before the COVID-19 period, whereas the SV-t estimates yield relatively regular, small outliers, with more variability in the SV estimate ($\lambda_{j,t}^{0.5}$) in the SVO case than the SV-t case. Our preferred SVO-t specification captures aspects of both SV-t-type outliers and SVO-type outliers. With SVO-t, the probability of a small outlier is a little lower than in the SV-t case, with some probability mass shifted to a large outlier. Similarly, with SVO-t, the probability of a large outlier is lower than in the SVO results, with some probability mass shifted to a small outlier.

Time variation in Σ_t affects our forecasts through two channels: first, the estimation of VAR coefficients Π as discussed in Section 3; and second, the projection of uncertainty about future shocks v_t that arises when simulating forward the dynamics of $\log(\lambda_t)$, as given in (2), to construct predictive densities. The forecast results we have seen so far, for 1975 to 2017, seem to suggest that the latter channel is more relevant than the former, as the RMSE differences between SV and SVO-t are very small, while those in CRPS are sometimes larger. The outlier states in SVO-t (as well as SVO and SV-t) allow for spikes in volatility to occur without having to project a persistent increase in uncertainty into the future as SV would be required to do. To illustrate the effects of this feature, we compare trajectories of time-varying volatility as estimated in quasi-real time over the course of 2020 and early 2021.[18]

Focusing on the example of payroll growth to limit charts, Figure 3 reports estimates of time variation in the volatility of forecast errors generated by SV and SVO-t, as well as the persistent components of Σ_t imputed from SVO-t when the effects of the outlier states $o_{j,t}$ and $q_{j,t}$ are ignored. (The online appendix provides results for other variables.) For this counterfactual, we compute $\tilde{\Sigma}_t = A^{-1}\,\Lambda_t\,(A^{-1})'$ based on the SVO-t estimates for Λ_t and A^{-1}.[19] In addition, we consider the corresponding measures of residual volatility obtained from the SV-OutMiss model, described in Section 3, that treats pre-specified outliers as missing data. These estimates show that, over

[18]The reported trajectories of volatilities in the VAR residuals, v_t, reflect smoothed estimates of the square roots of the diagonal elements of Σ_t computed from MCMC estimates for different end-points of the data (that correspond to different forecast origins in our out-of-sample forecast evaluation).

[19]In contrast, $\Sigma_t = A^{-1}\,O_t\,Q_t\,\Lambda_t\,Q_t'\,O_t'\,(A^{-1})'$ is the actual variance-covariance matrix of forecast errors in the SVO-t model that accounts also for the effects from realized outlier states O_t and Q_t.

the COVID-19 period, the SVO-t model clearly differentiates between increases in uncertainty that are short- and longer-lived, which the SV model cannot do. Volatility estimates from the SV model, shown in Panel (a) of the figure, reflect the impact of COVID-19 in the spring with a strong increase, which leveled off somewhat over the summer, but remained substantially elevated in the fall.

In contrast, SVO-t proves more nimble in accounting for the extreme data seen in the spring with a big jump in overall volatility in April as shown in Panel (b) of the figure. However, as revealed by comparison with Panel (d), this jump is largely seen as a transitory result of an outlier (both as it occurred in the spring and with the hindsight of estimates constructed based on data for the fall). In contrast, the persistent component of volatility in the case of SVO-t is seen to have risen no more than 8-fold over the course of the year. That is, the SVO-t estimates yield a much smaller rise in the persistent component of volatility than do the estimates from the SV model. The SV-OutMiss model yields an even smaller increase in the persistent component of volatility (the only component of volatility in that model); the estimates from SV-OutMiss shown in Panel (c) have risen by less than 5 times their level at the beginning of the year.

The more moderate rise in estimates of the persistent volatility component obtained with the SVO-t specification yields noticeably narrower (and arguably less extreme) uncertainty bands around forecasts compared to the SV model. In contrast, forecasts that condition on knowledge of when outliers occurred, but otherwise ignore any further information from their realization (as in the SV-OutMiss case), lead to particularly narrow uncertainty bands. As discussed next, the aforementioned pattern in volatility estimates shown in Figure 3 is mirrored in out-of-sample forecast densities generated over the course of 2020 and early 2021.

5.3 Forecasts made in 2020-21

In the months immediately preceding the COVID-19 outbreak, such as January 2020, predictive densities generated from the CONST and SV models differ a little, but not markedly so for most variables.[20] As we now detail, the picture changed significantly in subsequent months.

Over the course of March and April, the COVID-19 pandemic sharply affected the US economy, most visibly with the introduction of lockdown measures in the second half of March 2020, resulting in strong swings, particularly among measures of real activity, in subsequent months. Figure 4 displays the evolution of forecasts for real income and payroll growth over the months of March, April, and June generated from our alternative BVAR models.[21] As noted by Lenza and Primiceri (2020) and Schorfheide and Song (2020), forecasts generated by homoskedastic BVARs, like our CONST specification, can display extreme behavior since the spring of 2020.[22] For example, Panel (h) shows that, following the drop in payroll growth in March and April, the CONST model's posterior median forecast for May is about -136 percent (at an annualized rate) and between -64 and -124 percent for the next few months. The model's estimated forecast uncertainty is immense, with a 68 percent uncertainty band that widens to 100 percentage points or more by the 12-months-ahead horizon.

In contrast, the reaction of point and density forecasts generated by the SV and SVO-t specifications to the incoming data in spring 2020 is better behaved, particularly with SVO-t. Considering again the payroll growth forecasts shown in Figure 4, the SV model yields very negative point forecasts for May and the next few months, but not nearly as negative as those from the CONST model (e.g., the posterior median forecast for May is -17.8 percent and -20.1 percent for the SV and SVO-t models, respectively). The SVO-t model yields point forecasts fairly similar to those of the SV model, for most variables and forecast origins — consistent with our comparison of fore-

[20] Forecasts from the other alternatives, notably SVO and SV-t, are similar to those generated by the SV model in January 2020.

[21] For brevity, our discussion will abstract from nuances of the real-time data flow, and simply refer to forecasts being "made" at (or even "in" the month of) a particular forecast origin, even though the underlying data would have been available in FRED-MD only in a subsequent month.

[22] Lenza and Primiceri (2020) consider a slightly smaller VAR system (with six variables covering mostly employment and price data and observations starting only in 1988) where problems related to COVID-19 already become apparent with data for March 2020; in our 16-variable system case estimated from data starting in 1959, the effects of outliers become most apparent by April.

cast performance pre-COVID-19 in Table 2. That said, the SV model is prone to some distortion of its estimated forecast uncertainty, particularly early in the COVID-19 period. In March, April, and June of 2020, the uncertainty bands of the predictive densities obtained with SV are typically wider than those of not only the SVO-t but also the CONST specifications. In keeping with the volatility comparisons provided above, while the observations of 2020 widen the predictive densities of both SV and SVO-t forecasts, their impact is much greater for the former than for the latter. As indicated in rows 1 and 3 of the figure, SVO-t generates much narrower bands than SV. Moreover, the SVO-t bands also remain narrower for forecasts made in subsequent months, such as June 2020.

Rows 2 and 4 of Figure 4 compare our preferred SVO-t results to those for the more restrictive SVO and SV-t specifications. As expected, while the point forecasts of these specifications are difficult to distinguish, bigger differences are evident in the predictive densities. The predictive densities are generally the narrowest with the SVO-t forecasts. The SVO model generally yields wider densities, although in most cases the differences are less stark in June compared to March and April. Estimates are more varied with the SV-t model. In some cases (e.g., for payroll growth at the March 2020 forecast origin), the SV-t forecast intervals are very similar to the SVO-t estimates. But, in other cases, the SV-t intervals are wider than the SVO-t estimates; one example occurs with real income in the April 2020 forecasts, when the incoming data displayed a particularly large jump.

In additional forecast results for the pandemic period, we compare results from the SVO-t specification (which treats outliers as unknown and estimates them) to results from the SV-OutMiss approach that conditions on knowledge of when and which outliers occurred in the data. As described above, outliers are observations that are more than 5 times the inter-quartile range away from their sample median.[23] We then treat these observations as missing data in an otherwise standard VAR-SV model. In addition to omitting outlier data from the estimation of parameters and volatility states, SV-OutMiss replaces the outliers in the data vectors used to simulate predictive

[23] We obtain similar results with a threshold of 10 times the inter-quartile range.

densities at every forecast origin.

For these specifications, Figure 5 provides predictive densities for more recent forecast origins, ranging from September 2020 to March 2021, for not only growth in real income and payrolls but also S&P500 returns and the unemployment rate. Even almost a year after the onset of the COVID-19 pandemic impacted economic data, uncertainty bands from SVO-t remain noticeably wider than before the pandemic (results omitted in the interest of brevity). In most cases, forecast densities obtained from SV-OutMiss, which treats the timing of outliers as known, are relatively tight. However, exceptions are evident in the unemployment rate forecasts provided in the bottom row, with the SV-OutMiss bands wider than those of SVO-t for forecasts made with data in September and December 2020. Although harder to discern in the wide scales of the charts necessitated by the extreme realizations of actual data, the point forecasts produced by the alternative methods tend to be broadly similar at longer forecast horizons, although more sizable differences can occur at shorter horizons.

5.4 Model fit

So which model best characterizes the data in the COVID-19 period? The COVID-19 sample is too short to permit meaningful inference on the average accuracy of out-of-sample forecasts. Drawing on precedents such as Geweke and Amisano (2010), we instead consider the basic metric of predictive Bayes factors: the sums of 1-step-ahead predictive likelihoods. In these comparisons, we take the SV specification as the baseline and report sums of differences in predictive likelihoods, such that the more positive (negative) the number, the better (worse) the fit of a given specification compared to SV. Particularly with unusual observations, some care in computing predictive scores is warranted. We follow the recommendations of Krüger, et al. (forthcoming) and use what they characterize as a mixture-of-parameters approach. As an instance of Rao-Blackwellization, the approach relies as far as possible on the availability of analytical expressions for predictive likelihoods conditional on parameter values and latent SV states at each MCMC draw. In computational accuracy, we find it to be particularly important to integrate out future values of the transitory outlier states, instead of characterizing their arrival via Monte Carlo simulation. The supplementary online appendix provides further details on the calculations for each model.

A first issue is how the models compare by this measure of model fit over the COVID-19 sample of March 2020 through February 2021. These estimates are provided in the last row of Table 7. Over this sample, the best fitting model is SVO, followed by the SVO-t specification. In an overall fit sense, the data seem to favor a specification allowing infrequent, large outliers, and the data imply that the fit gain over the SV baseline is large. The SV-t model comes in third, with a relative score at least 30 log points lower. The SV-OutMiss approach that rests on identifying outliers ex-ante fits the data of the COVID-19 period much worse, with a score difference on the order of -950 log points. Perhaps not surprising, the CONST specification fares the worst over this volatile period.

The consideration of model fit over the COVID-19 period of course raises the question of how, earlier in time, the specifications compare by the same metric. For the sample running from 1975 (when our out-of-sample forecast evaluation of Section 5.1 began) into 2021, the patterns in model fit line up with those for the COVID-19 period, but with a bigger advantage of the SVO model. The SVO model also fares best in two other periods known for relatively high economic volatility: the 1975-1984 period coinciding with what some have referred to as the Great Inflation and the 2008-2014 sample of the Great Recession and ensuing slow recovery. The SVO-t model again has the second best score in the 2008-2014 period, but slips to fourth best in the 1975-1984 sample. In contrast, over the relatively tranquil period of 1985-2007, key years of the Great Moderation, the benchmark SV specification fits the data best, beating all of the other specifications. SV-OutMiss fits the data next-best, because there are few outliers, so that this approach is a small departure from SV. Among the models featuring some form of SV, allowing frequent, small outliers in the SV-t and SVO-t models fits the data worst, with SVO and its large, infrequent outliers not as far off the SV benchmark. Overall, our approach of extending an SV model to allow infrequent outliers works well by the metric underlying predictive Bayes factors, achieving its gains in the several historical subsamples that have featured high volatility.

Although model fit as assessed through predictive likelihoods is elemental to Bayesian evaluation of models, results on fit can differ from some results on out-of-sample forecast accuracy. In our case, admittedly, while our 1975-2017 forecast results are favorable to our proposed specifications, they are not necessarily as sharp as this section's results on model fit. One explanation is that the predictive likelihoods (log-scores) are more responsive to outcomes in the tails; forecast metrics including RMSE and CRPS are relatively insensitive to outcomes in the tails. Our SVO estimates appear to assign a little more predictive mass in the tails compared to other models. Given the strong curvature of the log score loss function (predictive likelihood), this difference in tail mass makes much more of a difference for log scores than CRPS.[24]

6 Robustness checks

This section provides an overview of several model robustness checks we have conducted. We also note that adjusting the forecast and model evaluation samples to start in 1985 instead of 1975 yields the same qualitative results; the supplementary online appendix provides tables with these results.

Common outlier: With the COVID-19 pandemic inducing extreme volatility in a number of variables, some may view it as plausible that the outlier is common to all variables, rather than independent across variables as in the SVO specification. Some other work, such as Lenza and Primiceri (2020), has developed models in which the pandemic induces a common shift in volatility in an otherwise homoskedastic VAR. Accordingly, we have also considered a specification in which the outlier state is common to all variables, in which case the time-varying variance-covariance matrix of the VAR residuals is given by $\Sigma_t = \bar{o}_t^2 A^{-1}\,\Lambda_t\,(A^{-1})'$, where $\bar{o}_t$ denotes a scalar outlier state.

In results detailed in the supplementary online appendix, making the outlier common seems to have no advantages. In historical estimates, the common-outlier specification registers virtu-

[24]Note that, due to the $i.i.d.$ nature of the outlier states, the SVO log scores depend only on the vector of outlier probabilities, not fitted values of past outliers.

ally no outliers prior to the COVID-19 pandemic. Instead, the common-outlier specification sees outliers only in the early stages of the pandemic period, from March through June 2020, when a good number of variables experienced enormous realizations at the same time, but none in late 2020 or early 2021. Imposing the same outlier on all variables during COVID-19 leads to some marked differences in the width of predictive densities compared to the SVO(-t) models that feature variable-specific outliers, but fairly identical performance in point forecasts over the pandemic period. In general, making outliers common in some cases makes forecasts slightly less accurate compared to the SVO specification that models outliers as independent across variables.

Capturing the pandemic period with dummy variables: As another simple approach to conditioning on knowledge of when and which outliers occurred in the data, particularly the timing of the COVID-19 pandemic, we consider an otherwise standard BVAR-SV model with separate dummy variables added to represent each month of the sample since COVID's outbreak in March 2020.[25] Predictive densities for selected forecast origins in 2020-21 are provided in the supplementary online appendix. By soaking up all information contained in data since the onset of the pandemic, the dummy approach generates point forecasts comparable to our outlier-augmented SV models. But because the dummy approach is conditioned on ex-ante knowledge that all COVID-19-related data points are highly unusual, its forecast densities are much tighter than those derived from our more agnostic outlier-augmented SV models or the SV-OutMiss specification.

AR(1) processes for SV: Our SV specifications treat log volatility as a random walk, following studies such as Cogley and Sargent (2005), Stock and Watson (2007), Justiniano and Primiceri (2008), and Clark (2011). We have also considered results from models in which SV is a persistent

[25]Wide priors are assigned to each dummy coefficient. Denote the dummy coefficient for each month $t \geq$ 2020:03 by δ_t. The prior for each δ_t is a mean-zero normal distribution, with a large variance set equal to $1/\varepsilon$, where ε is a small number chosen as a function of machine precision (identical to the output of the `eps` function in MATLAB). For $t \geq$ 2020:03, only the sum of δ_t and the residual v_t are identified.

AR(1) process rather than a random walk.[26] These results are qualitatively similar to the random walk-based results presented above. For example, the SVO-t specification forecasts at least as well as SV in historical data and in the COVID-19 period, and the predictive Bayes factors show overall similar patterns as reported for our random walk-based specifications.

VARs in levels: Our model specifications use log growth rates of trending variables (income, consumption, employment, etc.). Some work in the forecasting literature instead uses log levels in VARs, including examples such as Bańbura, Giannone, and Reichlin (2010) and Lenza and Primiceri (2020). In a factor model setting, Antolín-Díaz, Drechsel, and Petrella (2021) model outliers as occurring in levels and not growth rates. In the supplementary online appendix results, we have verified that VAR models with log levels rather than growth rates yield results showing overall the same patterns as detailed above.

Variable ordering: In VARs with stochastic volatility specified as in equations (1) through (3), variable ordering affects estimates. In practice, some work, such as Cogley and Sargent (2005), has found that results do not depend much on ordering. But recent work by Arias, Rubio-Ramirez, and Shin (2021) and Hartwig (2020) has shown that ordering choices in VARs with time-varying parameters and SV can affect out-of-sample forecasts. In particular, in their results, ordering has little effect on point forecasts but measurable effects on density-related measures, including the standard deviation of the predictive density and the length of prediction intervals.

The relatively large number of variables in our model means a very large number of possible orderings. Accordingly, we have investigated sensitivity to variable ordering with an approach meant to be broad but streamlined to be computationally tractable (if still demanding). Our basic metric for sensitivity is the distance between predictive densities obtained in one ordering versus another. We assess the distance and its significance with the potential scale reduction factor (PSRF) of Gelman and Rubin (1992). In particular, we compare predictive densities generated from the

[26]Clark and Ravazzolo (2015) find that random walk and AR(1) specifications yield relatively similar forecast performance in post-war US data.

VAR-SVO model at different forecast origins around and during the onset of the COVID-19 pandemic, December 2019, March and April 2020, and March 2021. For each of these origins, we randomly draw 640 different orderings of the model's 16 variables, estimate each model, and form forecast densities. We then compute a Gelman-Rubin scale reduction test for each variable at each horizon (1 to 24 months ahead).

Overall, these results, detailed in the supplementary online appendix, suggest small ordering effects in our forecasts. The vast majority of Gelman-Rubin statistics are under 1.2. Only in April 2021 does a handful get as high as about 1.3, indicating some small to modest differences in densities, typically for economic activity variables at medium forecast horizons. Of course, we chose these forecast origins to reflect different conditions in the COVID-19 period, with the economy near its depths in April 2020 and almost a year into recovery as of March 2021. For forecast origins in late 2019 or early 2021, the Gelman-Rubin statistics show no significant differences in densities across variable orderings. The differences that are detected occur with the April 2020 forecast origin.

Model stability: The unusual developments of the pandemic inevitably raise a question as to whether it represents a break in conventional business cycle dynamics and time series models. Our results treat the VARs as stable, taking various steps to limit the influence that extreme observations can have on model estimates. Of course, although it would be ideal to be able to formally test model stability, the sample since March 2020 is too short to permit formal inference with conventional tests or metrics.

As a simple and feasible alternative, we examine the stability of recursive estimates of the VAR from January 2020 through the end of our sample in 2021. To assess the significance of a change in each coefficient, we take the January 2020 posterior for each coefficient as a reference point. For the sake of comparability, we standardize the change in the posterior means obtained at subsequent forecast origins, by dividing the changes by January's posterior standard deviation.[27]

[27] The resulting statistics are similar in spirit to a frequentist t-statistic, though without necessarily identical interpretation in the context of our Bayesian application.

The supplementary online appendix provides charts of these statistics over time, for each variable and month, for the intercepts and lag coefficients. We consider stability for all of the paper's specifications: CONST, SV, SV-OutMiss, SVO-t, SVO, and SV-t specifications. Broadly, these results indicate that — except for the CONST case — there are at most only fairly limited changes in some coefficients, while the vast majority of coefficients show little change. By our simple measures of significance, the CONST specification is quite prone to some coefficient change, most sizably for some economic activity indicators. In the SV specifications, coefficient change appears much less significant. The SVO and SVO-t models show changes in intercepts for some variables, but otherwise, estimates look to be broadly stable over the period.

7 Conclusion

We study the use of an outlier-augmented stochastic volatility specification for Bayesian VARs. Our work is prompted by the enormous realizations of many macroeconomic time series witnessed over the course of 2020-21 as COVID-19 impacted many economies across the world. As recognized by other recent studies such as Lenza and Primiceri (2020) and Schorfheide and Song (2020), these outliers have strong, and sometimes outsized, effects on forecasts made with standard constant-variance VARs. Our SVO(-t) approach extends to BVARs the earlier work of Stock and Watson (2016) in the context of unobserved component models of inflation, and it is related to SV models with t-distributed errors developed by Jacquier, Polson, and Rossi (2004). The SVO model features stochastic volatility, and an outlier-state treatment, and the SVO-t specification augments SVO with fat-tailed shocks.

Although VARs with time-varying volatility tend to down-weight high-volatility observations in the construction of parameter estimates, the resulting forecasts can be better insulated from outliers. As shown in Section 5.3, BVARs with time-varying volatility generate point forecasts that are less distorted than in the constant-variance case. But a conventional SV model expects all changes in volatility to be persistent, so that it extrapolates huge forecast uncertainty from the initial COVID-19 shocks. In contrast, SVO(-t) allows the model to fit sharp spikes in current volatility while adapting its uncertainty forecasts more moderately.

An alternative approach could be to pre-screen the data to identify outliers in individual variables based on a simple measure of historical norms, and then treat these variable-specific outliers as missing observations in an otherwise conventional VAR with SV. Forecasts from this alternative missing-data approach (SV-OutMiss) neglect the possible arrival of future outliers. In contrast, our outlier-augmented SV models provide a coherent treatment of extremes in the data by modeling the occurrence of outliers as stochastic events, with unknown timing. Accordingly, the resulting forecast densities fully reflect the uncertainty emanating from the presence of outliers in the data. As a result, the outlier-augmented SV models are particularly attractive for continued use over the yet-unknown course of economic developments related to the COVID-19 pandemic as well as future events.

In light of the potential for our approaches to overfit data predating the pandemic period, we conduct a quasi-real-time evaluation of forecast performance using monthly data with an evaluation window starting in 1975 and ending in 2017. We compare the accuracy of point and density forecasts, as measured by RMSE and CRPS, from standard VARs against our proposed SVO(-t) specifications. Even in the pre-COVID-19 period, our outlier-augmented SV models forecast the data, on balance, a little better than a conventional VAR-SV model. The SV-OutMiss specification delivers a performance competitively similar to that of SVO(-t).

To evaluate which model best characterizes the data in the COVID-19 period, forecast accuracy could, of course, be a natural metric. However, the sample is too short to support formal inference on the basis of average forecast accuracy. Instead, we employ predictive Bayes factors. By this measure, our SVO specification fits the COVID-19 sample the best, with SVO-t next. The neglected arrival of future outliers in the SV-OutMiss model incurs a sizable penalty in the predictive Bayes factors. Over the entire evaluation sample since 1975, the SVO specification again fares best. The gains of the outlier-augmented SV model are driven by various episodes of relatively high volatility in the data; in contrast, the baseline SV model fits well only in the Great

Moderation years of 1985 through 2007.

.

References

[1] Alvarez, Luis J., and Florens Odendahl (2021), "COVID-19 and Bayesian VARs in the Euro Area," *mimeo*, July.

[2] Antolín-Díaz, Juan, Thomas Drechsel, and Ivan Petrella (2021), "Advances in nowcasting economic activity: Secular trends, large shocks and new data," Discussion Paper 15926, CEPR.

[3] Arias, Jonas E., Juan F. Rubio-Ramirez, and Minchul Shin (2021), "Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models," Working Paper 21-21, Federal Reserve Bank of Philadelphia, `https://doi.org/10.21799/frbp.wp.2021.21`.

[4] Artis, Michael J., Anindya Banerjee, and Massimiliano Marcellino (2005), "Factor forecasts for the UK," *Journal of Forecasting*, 24, 279–298, `https://doi.org/10.1002/for.957`.

[5] Aruoba, S. Boragan, Francis X. Diebold, and Chiara Scotti (2009), "Real-time measurement of business conditions," *Journal of Business & Economic Statistics*, 27, 417–427, `https://doi.org/10.1198/jbes.2009.07205`.

[6] Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010), "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92, `https://doi.org/10.1002/jae.1137`.

[7] Bobeica, Elena, and Benny Hartwig (2021), "The COVID-19 shock and challenges for time series models," Working Paper 2021/2558, ECB.

[8] Carriero, Andrea, Joshua Chan, Todd E. Clark, and Massimiliano Marcellino (2021a), "Corrigendum to: Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors," *manuscript*.

[9] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019), "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154, https://doi.org/10.1016/j.jeconom.2019.04.024.

[10] Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2021b), "Forecasting with Shadow-Rate VARs," Working Paper 2021-09, Federal Reserve Bank of Cleveland, https://doi.org/10.26509/frbc-wp-202109.

[11] Chan, Joshua C.C. (2019), "Large hybrid time-varying parameter VARs," CAMA Working Paper 2019-77, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

[12] ——— (2020), "Large Bayesian VARs: A flexible Kronecker error covariance structure," *Journal of Business & Economic Statistics*, 38, 68–79, https://doi.org/10.1080/07350015.2018.1451336.

[13] Chib, Siddhartha, Minchul Shin, and Fei Tan (2021), "DSGE-SVt: An econometric toolkit for high-dimensional DSGE models with SV and t errors," Working Paper 21-02, Federal Reserve Bank of Philadelphia, https://doi.org/10.21799/frbp.wp.2021.02.

[14] Chiu, Ching-Wai (Jeremy), Haroon Mumtaz, and Gábor Pintér (2017), "Forecasting with VAR models: Fat tails and stochastic volatility," *International Journal of Forecasting*, 33, 1124–1143, https://doi.org/10.1016/j.ijforecast.2017.03.001.

[15] Clark, Todd E. (2011), "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business and Economic Statistics*, 29, 327–341, https://doi.org/10.1198/jbes.2010.09248.

[16] Clark, Todd E., Michael W. McCracken, and Elmar Mertens (2020), "Modeling time-varying uncertainty of multiple-horizon forecast errors," *The Review of Economics and Statistics*, 102, 17–33, https://doi.org/10.1162/rest_a_00809.

[17] Clark, Todd E., and Francesco Ravazzolo (2015), "Macroeconomic forecasting performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, 30, 551–575, https://doi.org/10.1002/jae.2379.

[18] Cogley, Timothy, and Thomas J. Sargent (2005), "Drifts and volatilities: Monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302, https://doi.org/10.1016/j.red.2004.10.009.

[19] Cúrdia, Vasco, Marco Del Negro, and Daniel L. Greenwald (2014), "Rare shocks, great recessions," *Journal of Applied Econometrics*, 29, 1031–1052, https://doi.org/10.1002/jae.2395.

[20] D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101, https://doi.org/10.1002/jae.1257.

[21] Del Negro, Marco, and Giorgio E. Primiceri (2015), "Time varying structural vector autoregressions and monetary policy: A corrigendum," *Review of Economic Studies*, 82, 1342–1345, https://doi.org/10.1093/restud/rdv024.

[22] Diebold, Francis X. (2020), "Real-time real economic activity entering the Pandemic Recession," *Covid Economics*, 62, 1–19, CEPR.

[23] Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–63, https://doi.org/10.2307/1392185.

[24] Durbin, J., and S.J. Koopman (2002), "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89, 603–615, https://doi.org/10.1093/biomet/89.3.603.

[25] Gelman, Andrew, and Donald B. Rubin (1992), "Inference from iterative simulation using multiple sequences," *Statistical Science*, 7, 457–472, https://doi.org/10.1214/ss/1177011136.

[26] Geweke, John, and Gianni Amisano (2010), "Comparing and evaluating Bayesian predictive distributions of asset returns," *International Journal of Forecasting*, 26, 216 – 230, https://doi.org/10.1016/j.ijforecast.2009.10.007, Special Issue: Bayesian Forecasting in Economics.

[27] Guerrón-Quintana, Pablo, and Molin Zhong (2017), "Macroeconomic forecasting in times of crises," Finance and Economics Discussion Series 2017-018, Board of Governors of the Federal Reserve System (U.S.), https://doi.org/10.17016/FEDS.2017.018.

[28] Hartwig, Benny (2020), "Robust inference in time-varying structural VAR models: The DC-Cholesky multivariate stochastic volatility model," Discussion Paper 34/2020, Deutsche Bundesbank.

[29] ——— (2021), "Bayesian VARs and prior calibration in times of COVID-19," https://doi.org/10.2139/ssrn.3792070, *mimeo*.

[30] Huber, Florian, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner (forthcoming), "Nowcasting in a pandemic using non-parametric mixed frequency VARs," *Journal of Econometrics*, https://doi.org/10.1016/j.jeconom.2020.11.006.

[31] Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (2004), "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122, 185 – 212, https://doi.org/10.1016/j.jeconom.2003.09.001.

[32] Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng (2015), "Measuring uncertainty," *American Economic Review*, 105, 1177–1216, https://doi.org/10.1257/aer.20131193.

[33] Justiniano, Alejandro, and Giorgio E. Primiceri (2008), "The time-varying volatility of macroeconomic fluctuations," *The American Economic Review*, 98, 604–641, https://doi.org/10.1257/aer.98.3.604.

[34] Karlsson, Sune, and Stepan Mazur (2020), "Flexible fat-tailed vector autoregression," Working Paper 2020:5, Örebro University, School of Business.

[35] Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *The Review of Economic Studies*, 65, 361–393, https://doi.org/10.1111/1467-937X.00050.

[36] Koop, Gary (2003), *Bayesian Econometrics*: Wiley-Interscience.

[37] Koop, Gary M. (2013), "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28, 177–203, https://doi.org/10.1002/jae.1270.

[38] Krüger, Fabian, Sebastian Lerch, Thordis L. Thorarinsdottir, and Tilmann Gneiting (forthcoming), "Predictive inference based on Markov Chain Monte Carlo output," *International Statistical Review*, https://doi.org/10.1111/insr.12405.

[39] Lenza, Michele, and Giorgio E. Primiceri (2020), "How to estimate a VAR after March 2020," NBER Working Paper 27771, National Bureau of Economic Research, Inc, https://doi.org/10.3386/w27771.

[40] McCracken, Michael W., and Serena Ng (2016), "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34, 574–589, https://doi.org/10.1080/07350015.2015.1086655.

[41] Mitchell, James, and Martin Weale (2021), "Censored density forecasts: Production and evaluation," Working Paper 2021-12, Federal Reserve Bank of Cleveland, `https://doi.org/10.26509/frbc-wp-202112`.

[42] Ng, Serena (2021), "Modeling macroeconomic variations after COVID-19," Working Paper 29060, National Bureau of Economic Research, `https://doi.org/10.3386/w29060`.

[43] Perez-Quiros, Gabriel, and Margaret M. McConnell (2000), "Output fluctuations in the United States: What has changed since the early 1980's?" *American Economic Review*, 90, 1464–1476, `https://doi.org/10.1257/aer.90.5.1464`.

[44] Primiceri, Giorgio E., and Andrea Tambalotti (2020), "Macroeconomic forecasting in the time of covid-19," June, *mimeo*.

[45] Schorfheide, Frank, and Dongho Song (2020), "Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic," Working Paper 20-26, Federal Reserve Bank of Philadelphia, `https://doi.org/10.21799/frbp.wp.2020.26`.

[46] Stock, James H., and Mark W. Watson (2002), "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, 20, 147–162, `https://doi.org/10.1198/073500102317351921`.

[47] ——— (2007), "Why has U.S. inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39, 3–33, `https://doi.org/10.1111/j.1538-4616.2007.00014.x`.

[48] ——— (2016), "Core inflation and trend inflation," *The Review of Economics and Statistics*, 98, 770–784, `https://doi.org/10.1162/REST_a_00608`.

[49] West, Kenneth D. (1996), "Asymptotic inference about predictive ability," *Econometrica*, 64, 1067–1084, `https://doi.org/10.2307/2171956`.

Table 1: List of variables

Variable	FRED-MD code	transformation	Minnesota prior
Real Income	RPI	$\Delta \log(x_t) \cdot 1200$	0
Real Consumption	DPCERA3M086SBEA	$\Delta \log(x_t) \cdot 1200$	0
IP	INDPRO	$\Delta \log(x_t) \cdot 1200$	0
Capacity Utilization	CUMFNS		1
Unemployment Rate	UNRATE		1
Nonfarm Payrolls	PAYEMS	$\Delta \log(x_t) \cdot 1200$	0
Hours	CES0600000007		0
Hourly Earnings	CES0600000008	$\Delta \log(x_t) \cdot 1200$	0
PPI (Fin. Goods)	WPSFD49207	$\Delta \log(x_t) \cdot 1200$	1
PCE Prices	PCEPI	$\Delta \log(x_t) \cdot 1200$	1
Housing Starts	HOUST	$\log(x_t)$	1
S&P 500	SP500	$\Delta \log(x_t) \cdot 1200$	0
USD / GBP FX Rate	EXUSUKx	$\Delta \log(x_t) \cdot 1200$	0
5-Year Yield	GS5		1
10-Year Yield	GS10		1
Baa Spread	BAAFFM		1

Note: Data obtained from the 2021-04 vintage of FRED-MD. Monthly observations from 1959:M03 to 2021:M03. Entries in the column "Minnesota prior" report the prior mean on the first own-lag coefficient of the corresponding variable in each BVAR. Prior means on all other VAR coefficients are set to zero.

Table 2: RMSE (baseline comparisons)

					Relative to SV ...							
	SV				CONST				SVO-t			
Variable / Horizons	1	3	12	24	1	3	12	24	1	3	12	24
Real Income	7.11	7.11	7.10	7.78	1.00	1.01	1.01	0.93*	1.01	1.00	1.01**	0.94
Real Consumption	5.79	5.89	5.82	5.93	0.99	1.00	1.00	1.00	1.00	1.01	1.00	1.00
IP	7.26	7.95	8.21	8.99	1.03*	1.01	1.01	0.99	1.00	0.99	0.99	0.97**
Capacity Utilization	0.51	1.06	2.93	4.41	1.03*	1.02	1.03	0.97	1.01	1.00	0.97	0.96
Unemployment Rate	0.16	0.27	0.82	1.26	1.00	1.00	1.03	1.07*	1.00	0.99	0.98	0.98
Nonfarm Payrolls	1.66	1.87	2.13	2.51	1.01	0.99	1.03	0.98	0.99	0.99	1.00	0.99
Hours	0.24	0.28	0.45	0.48	1.04**	1.01	1.02	1.04	1.00	0.99	0.96*	0.98
Hourly Earnings	3.01	3.02	3.25	3.78	1.02	1.00	1.00	1.04	1.00	1.00	1.01*	1.02**
PPI (Fin. Goods)	6.63	6.79	7.16	7.64	1.00	1.01	1.05*	1.07**	1.00	1.00	1.00	1.01
PCE Prices	2.11	2.49	2.81	3.41	1.00	1.02	1.10*	1.16***	1.01**	1.01	1.01*	1.03**
Housing Starts	0.08	0.11	0.24	0.36	1.00	0.99	1.04	1.04	1.00	1.00	1.00	1.02***
S&P 500	43.48	44.61	43.65	43.04	1.00	1.02	1.00	1.00	1.01	1.00	1.01	1.01***
USD / GBP FX Rate	27.86	29.82	29.34	33.16	1.03**	1.03**	1.01	0.89	1.00	1.00	1.00	0.88
5-Year yield	0.33	0.71	1.45	1.98	1.01	1.06**	1.07	1.09	1.01**	1.00	1.00	0.98*
10-Year yield	0.29	0.63	1.34	1.78	1.00	1.05*	1.09	1.14	1.00	1.00	1.00	0.99
Baa Spread	0.52	1.11	1.78	1.88	1.02	1.06*	1.08	1.12*	0.99**	1.00	1.00	0.99

Note: Comparison of “SV” (baseline, in denominator of relative comparisons) against “CONST” and “SVO-t.” Values below 1 indicate improvement over baseline. Evaluation window from 1975:M01 through 2017:M12. Significance assessed by Diebold-Mariano-West test using Newey-West standard errors with $h+1$ lags.

Table 3: Avg CRPS (baseline comparisons)

					Relative to SV ...							
	SV				CONST				SVO-t			
Variable / Horizons	1	3	12	24	1	3	12	24	1	3	12	24
Real Income	3.08	3.18	3.55	4.41	1.07***	1.00	0.92***	0.75***	0.99	0.96***	0.94***	0.85***
Real Consumption	3.03	3.12	3.46	4.40	1.02*	1.03**	0.94***	0.75***	1.00	1.00	0.97***	0.90***
IP	3.95	4.27	4.91	6.33	1.04***	1.02	0.93***	0.77***	1.00	0.99*	0.96***	0.88***
Capacity Utilization	0.28	0.56	1.63	2.85	1.03**	1.04**	1.00	0.85***	1.00	1.00	0.99	0.95**
Unemployment Rate	0.09	0.15	0.42	0.75	0.99	1.01	1.04	0.98	1.00	1.00	1.00	0.98
Nonfarm Payrolls	0.85	0.96	1.31	1.88	1.06***	1.03**	0.94**	0.73***	0.99	0.99	0.97***	0.91***
Hours	0.12	0.15	0.25	0.34	1.06***	1.01	0.99	0.83***	0.99**	0.98**	0.96***	0.90***
Hourly Earnings	1.60	1.65	1.92	2.65	1.04***	1.01	0.93***	0.80***	1.00	0.99**	0.98***	0.92***
PPI (Fin. Goods)	3.45	3.60	4.05	4.97	1.02	1.01	1.00	0.89***	1.00	0.99	0.98***	0.94***
PCE Prices	1.13	1.32	1.58	2.14	1.02	1.02	1.05	0.97	1.01**	1.01	1.00	0.97***
Housing Starts	0.04	0.06	0.13	0.21	0.99	0.99	1.03	0.95	1.00	1.01	1.01**	1.00
S&P 500	23.19	24.03	26.09	31.68	1.00	1.01	0.91***	0.74***	1.00	0.99**	0.97***	0.90***
USD / GBP FX Rate	15.46	16.46	17.45	20.06	1.04***	1.03*	0.94***	0.81***	0.99*	0.99***	0.96***	0.89***
5-Year yield	0.17	0.37	0.76	1.04	1.04**	1.09***	1.10**	1.08	1.01**	1.01	1.00	1.00
10-Year yield	0.15	0.33	0.71	0.97	1.02	1.07***	1.11**	1.09	1.01**	1.01	1.00	1.01
Baa Spread	0.21	0.46	0.95	1.28	1.18***	1.20***	1.06	0.93	0.99**	1.00	1.00	0.97***

Note: Comparison of “SV” (baseline, in denominator of relative comparisons) against “CONST” and “SVO-t.” Values below 1 indicate improvement over baseline. Evaluation window from 1975:M01 through 2017:M12. Significance assessed by Diebold-Mariano-West test using Newey-West standard errors with $h + 1$ lags.

Table 4: Relative RMSE (outlier-augmented models)

	SVO				SV-t				SV-OutMiss			
Variable / Horizon	1	3	12	24	1	3	12	24	1	3	12	24
Real Income	0.99**	0.99**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01
Real Consumption	0.99*	0.99*	1.00	0.99**	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00
IP	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.01
Capacity Utilization	1.00	1.01	1.01	1.02	1.00	1.00	1.00	0.99	0.99	1.00	0.99	1.01
Unemployment Rate	1.00	1.00	1.01	1.02	1.00	1.00	1.00	1.00	0.99	1.01	1.01	1.01
Nonfarm Payrolls	1.01	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	0.99
Hours	1.00	1.01	1.02	1.02	1.00	1.00	1.01	1.00	1.00	1.00	1.02	1.00
Hourly Earnings	1.00	1.00	0.98***	0.96***	1.00	1.00	1.00	1.00*	1.00	1.00	0.99**	0.97**
PPI (Fin. Goods)	1.00	1.00	1.00	0.99**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
PCE Prices	0.99*	0.99	0.98**	0.97***	1.00	1.00	1.00	1.00	0.99	0.99**	0.99	0.98**
Housing Starts	1.00	1.00	0.99**	0.98*	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98***
S&P 500	1.00	1.00	1.00	0.99***	1.00	1.00*	1.00*	1.00	1.00	1.00	1.00	0.99
USD / GBP FX Rate	1.00*	1.00	1.00	1.00	1.00	1.00	1.00	1.01**	1.01	0.99	1.00	0.99
5-Year Yield	0.99**	0.99***	0.99	1.01	0.99*	0.99	1.00	1.00	0.99**	0.99	0.99	1.00
10-Year Yield	0.99*	0.99**	0.99	1.01	0.99*	0.99	1.00	1.00	0.99*	1.00	1.00	1.00
Baa Spread	1.00	1.00	0.99	1.01	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.01

Note: Comparison of "SVO-t" (baseline, in denominator) against "SVO," "SV-t," and "SV-OutMiss." Values below 1 indicate improvement over baseline. Evaluation window from 1975:M01 through 2017:M12. Significance assessed by Diebold-Mariano-West test using Newey-West standard errors with $h + 1$ lags. Due to the close behavior of some of the models compared, and rounding of the report values, a few comparisons show a significant relative RMSE of 1.00. These cases arise from persistent differences in performance that are, however, too small to be relevant after rounding.

Table 5: Relative Avg CRPS (outlier-augmented models)

	SVO				SV-t				SV-OutMiss			
Variable / Horizon	1	3	12	24	1	3	12	24	1	3	12	24
Real Income	1.00	1.01**	1.05***	1.10***	1.00	1.00	0.99***	0.99***	1.00	1.00	1.02*	1.07***
Real Consumption	1.00	1.01**	1.06***	1.10***	1.00	1.00	0.99***	0.98***	0.99	0.99	1.01***	1.08***
IP	1.00	1.02***	1.06***	1.12***	0.99	1.00	0.99***	0.99***	1.00	1.01*	1.03***	1.10***
Capacity Utilization	1.00	1.01	1.03**	1.08***	1.00	1.00	0.99**	0.98***	1.00	1.01	0.99	1.02
Unemployment Rate	1.00	1.00	1.02**	1.06***	1.00	1.00*	0.99**	0.98***	0.99	1.00	0.99	1.01
Nonfarm Payrolls	1.01**	1.01	1.06***	1.11***	1.00	1.00	0.99***	0.98***	1.00	1.00	1.02*	1.07***
Hours	1.01***	1.02***	1.05***	1.10***	1.00*	1.00	0.99**	0.98***	1.01	1.02**	1.03***	1.09***
Hourly Earnings	1.00	1.02***	1.06***	1.10***	1.00	1.00*	0.98***	0.98***	1.00	1.01	1.02**	1.07***
PPI (Fin. Goods)	1.00	1.01***	1.03***	1.07***	1.00	1.00	0.99***	0.99***	0.99	1.00	1.01***	1.04***
PCE Prices	0.99*	1.00	1.02***	1.05***	1.00	1.00	0.99***	0.98***	0.98**	0.98***	0.99	1.01
Housing Starts	1.00	1.01*	1.01**	1.03**	1.00	1.00**	0.99***	0.99**	0.99	0.99	0.99**	0.99
S&P 500	1.00	1.01***	1.05***	1.10***	1.00	1.00	0.99***	0.99***	1.00	1.00	1.02***	1.08***
USD / GBP FX Rate	1.01*	1.00	1.02***	1.04***	1.00	1.00	1.00	1.02**	1.01	1.00	1.02**	1.07***
5-Year Yield	1.00	0.99**	1.00	1.03***	0.99*	0.99**	1.00	0.99***	0.99	0.99	0.99	0.99
10-Year Yield	1.00	0.99	1.00	1.03***	0.99**	0.99**	1.00	0.98***	0.99*	1.00	1.00	0.99**
Baa Spread	1.00	1.00	1.03**	1.08***	1.00	1.00	0.99	0.97***	1.01	0.99	0.99	1.02

Note: Comparison of “SVO-t” (baseline, in denominator) against “SVO,” “SV-t,” and “SV-OutMiss.” Values below 1 indicate improvement over baseline. Evaluation window from 1975:M01 through 2017:M12. Significance assessed by Diebold-Mariano-West test using Newey-West standard errors with $h + 1$ lags. Due to the close behavior of some of the models compared, and rounding of the report values, a few comparisons show a significant relative CRPS of 1.00. These cases arise from persistent differences in performance that are, however, too small to be relevant after rounding.

Table 6: Outlier-Augmented SV Parameters

	SVO	SV-t	SVO-t	
Variable	p_j	ν_j	p_j	ν_j
Prior	1.81	21	0.58	21
	0.85 – 3.32	*9.00 – 34.00*	*0.15 – 1.56*	*9.00 – 34.00*
Real Income	3.19	3	0.35	3
	2.37 – 4.27	*3 – 3*	*0.08 – 0.93*	*3 – 3*
Real Consumption	0.49	3	0.17	3
	0.21 – 0.94	*3 – 3*	*0.04 – 0.46*	*3 – 3*
IP	0.50	3	0.13	3
	0.25 – 0.90	*3 – 3*	*0.04 – 0.32*	*3 – 3*
Capacity Utilization	0.42	26	0.12	25
	0.20 – 0.75	*14 – 36*	*0.03 – 0.34*	*14 – 36*
Unemployment Rate	0.62	21	0.30	31
	0.34 – 1.01	*8 – 34*	*0.13 – 0.55*	*22 – 38*
Nonfarm Payrolls	1.12	3	0.17	3
	0.62 – 1.75	*3 – 3*	*0.04 – 0.52*	*3 – 3*
Hours	1.05	3	0.20	3
	0.56 – 1.72	*3 – 3*	*0.05 – 0.48*	*3 – 3*
Hourly Earnings	0.45	3	0.15	3
	0.21 – 0.85	*3 – 4*	*0.04 – 0.40*	*3 – 3*
PPI (Fin. Goods)	0.50	6	0.12	4
	0.24 – 0.90	*3 – 21*	*0.03 – 0.34*	*3 – 8*
PCE Prices	0.41	14	0.12	9
	0.19 – 0.78	*3 – 32*	*0.03 – 0.34*	*3 – 31*
Housing Starts	0.34	35	0.11	35
	0.16 – 0.63	*28 – 39*	*0.03 – 0.29*	*29 – 39*
S&P 500	0.61	3	0.13	3
	0.27 – 1.11	*3 – 3*	*0.03 – 0.35*	*3 – 3*
USD / GBP FX Rate	0.94	3	0.56	5
	0.56 – 1.40	*3 – 3*	*0.30 – 0.97*	*3 – 17*
5-Year yield	0.36	31	0.12	31
	0.17 – 0.65	*22 – 38*	*0.03 – 0.31*	*20 – 38*
10-Year yield	0.36	35	0.11	35
	0.17 – 0.64	*28 – 39*	*0.03 – 0.30*	*28 – 39*
Baa Spread	0.85	27	0.24	28
	0.49 – 1.29	*15 – 36*	*0.07 – 0.58*	*17 – 36*

Notes: Outlier probability p_j (in percent), and t-distribution's degrees of freedom, ν_j, of orthogonalized residuals of each variable. Median, 15.87% and 84.14% posterior quantiles from full-sample estimation using data from 1959:M03 – 2021:M03. The priors are, $p_j \sim Beta(2.5, 117.5)$ for SVO, and $p_j \sim Beta(1.0, 119.0)$ forSVO-t. and $\nu_j \sim U(3, 40)$ direcretized over an integer-valued grid.

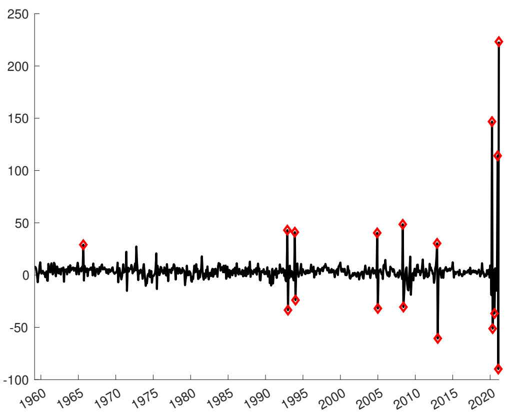
Table 7: Log Bayes Factors Relative to SV

	Models				
Samples	**SVO-t**	**SVO**	**SV-t**	**OutMiss**	**CONST**
Full sample					
1975:01-2021:02	244.11	334.84	195.19	-782.79	-9200.01
G Inflation					
1975-1984	8.25	33.22	10.37	17.38	-250.02
G Moderation					
1985-2007	-41.94	-9.69	-52.00	-6.64	-385.43
GFC					
2008-2014	21.53	29.50	13.69	-56.28	-236.40
COVID-19					
2020:03-2021:02	225.33	232.59	193.28	-739.52	-8167.44

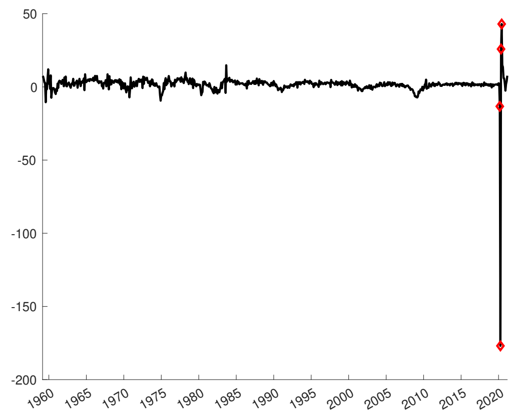
Note: Differences in cumulative log Bayes factors, $\log L(\mathcal{M}_i) - \log L(\mathcal{M}_0)$, where $\log L(\mathcal{M}_i) = \sum_{t=T_0}^{T_1} \log p(y_{t+1}|y^t, \mathcal{M}_i)$ between the different models listed above ($\mathcal{M}_i$) and the SV model ($\mathcal{M}_0$), measured over different subsamples of forecast origins, t. Unless stated otherwise, samples extend from January to December of the years given.

Figure 1: Some selected data series

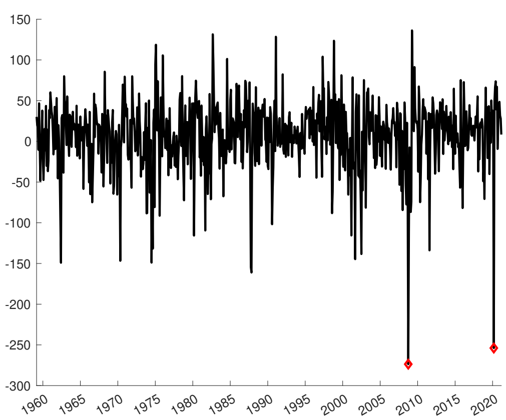
(a) Real personal income



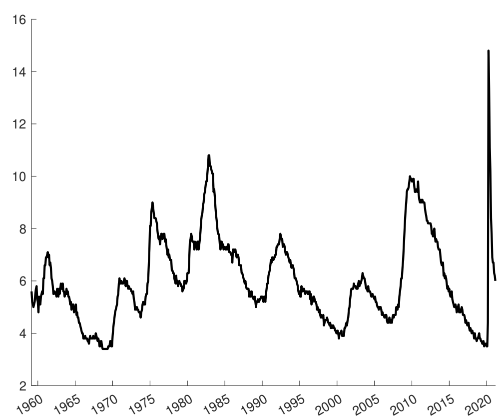
(b) Payroll growth



(c) S&P 500 returns

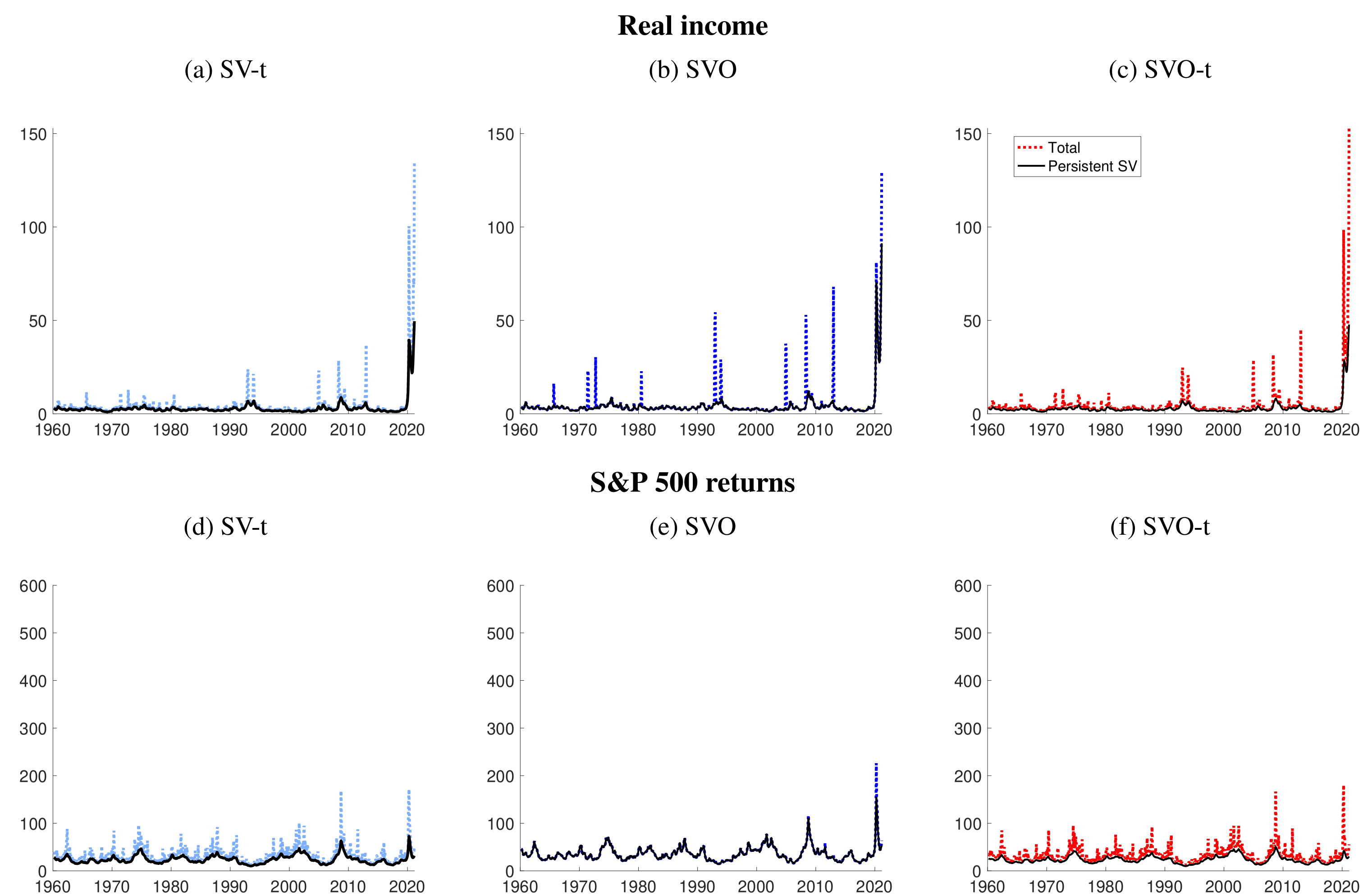


(d) Unemployment rate



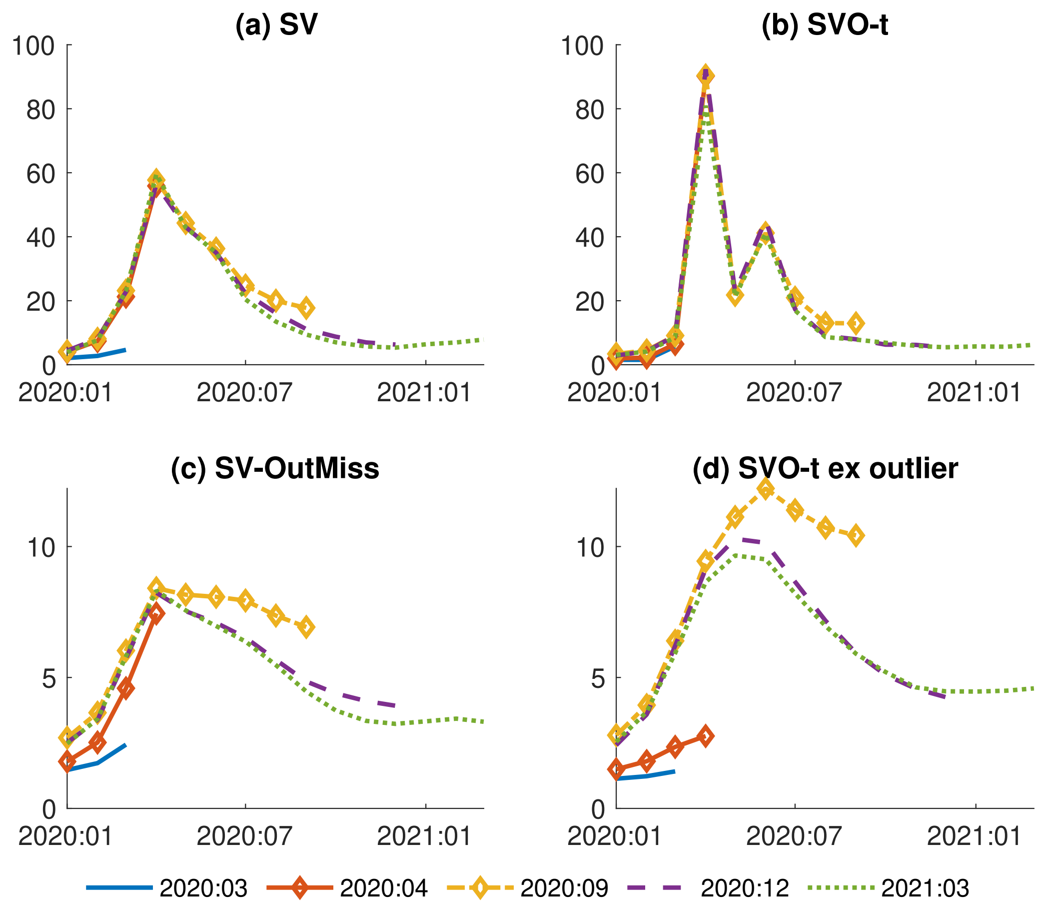
Note: Data for selected time series, with data transformations as listed in Table 1. Red dots denote observations that are more than five times the inter-quartile range away from the series median.

Figure 2: Contributions of outlier adjustments to forecast-error volatilities



Note: Posterior median estimates per March 2021, of time-varying volatilities in forecast errors of the indicated variables in three outlier-augmented versions of the VAR-SV model. Dotted lines depict the actual forecast-error volatility, including the effects of O_t and Q_t as applicable in each model. The solid lines depict the component of each variable's forecast-error volatility due to the persistent SV component. Specifically, for the SVO-t model, the forecast-error volatility is given by the square root of diagonal elements of $\Sigma_t = A^{-1} O_t Q_t \Lambda_t Q_t' O_t A^{-T}$, whereas the contribution from the persistent SV component follows from $\tilde{\Sigma}_t = A^{-1} \Lambda_t A^{-T}$. For SVO and SV-t, corresponding computations are performed using only O_t and Q_t, respectively. These calculations are performed for every MCMC draw, with the resulting medians reported in the figure.

Figure 3: Time-varying volatilities since 2020 of payroll growth

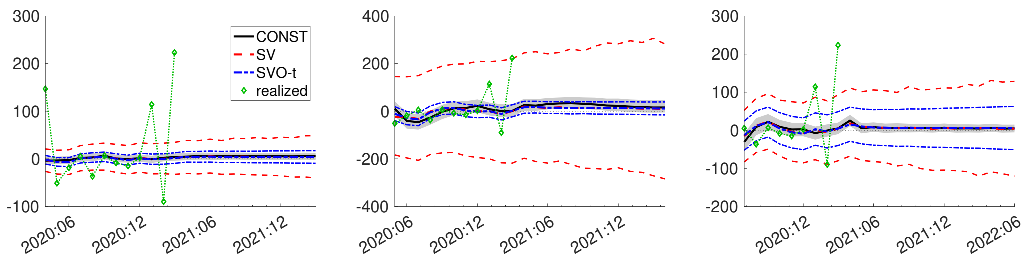


Note: Quasi-real-time trajectories of time-varying volatility in VAR residuals, measured by the diagonal elements of $\text{Var}_t\left(v_t\right) = \Sigma_t$ implied by different models. Medians of (smoothed) posterior obtained from different data samples ending at forecast origins as indicated in the figure legend. Panel (d) displays estimates of stochastic volatility for SVO-t that ignore the contributions from outliers and that are computed from $\tilde{\Sigma}_t = A^{-1}\,\Lambda_t\,A^{-T}$ (i.e., neglecting the O_t and Q_t components in the computation of the uncertainty measures shown here, while including these outliers in estimation of A^{-1}, Λ_t, etc.). Reflecting the sizable differences in the size of estimates resulting with and without outlier treatment, different scales are used in upper- and lower-row panels.

Figure 4: Predictive densities since March 2020

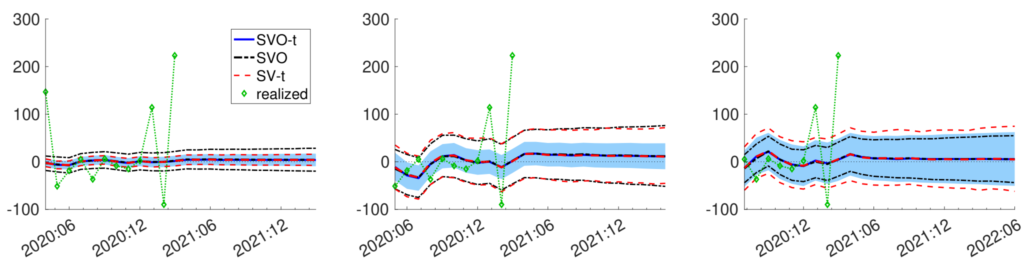
Real income: CONST and SV alternatives

(a) March (b) April (c) June



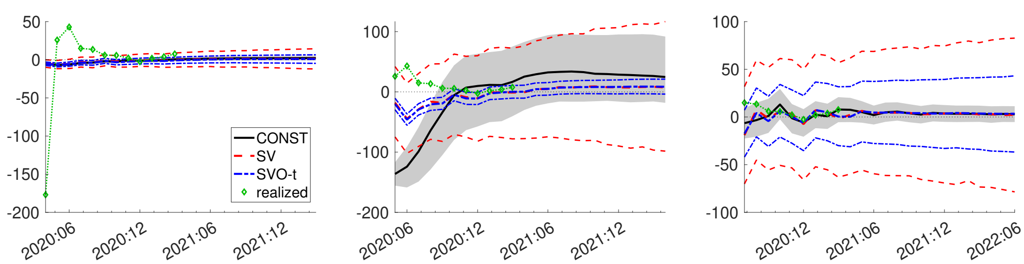
Real income: SV with outlier adjustments

(d) March (e) April (f) June



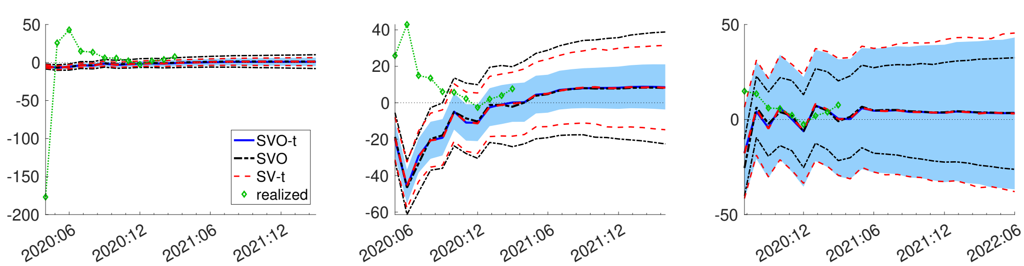
Payroll growth: CONST and SV alternatives

(g) March (h) April (i) June



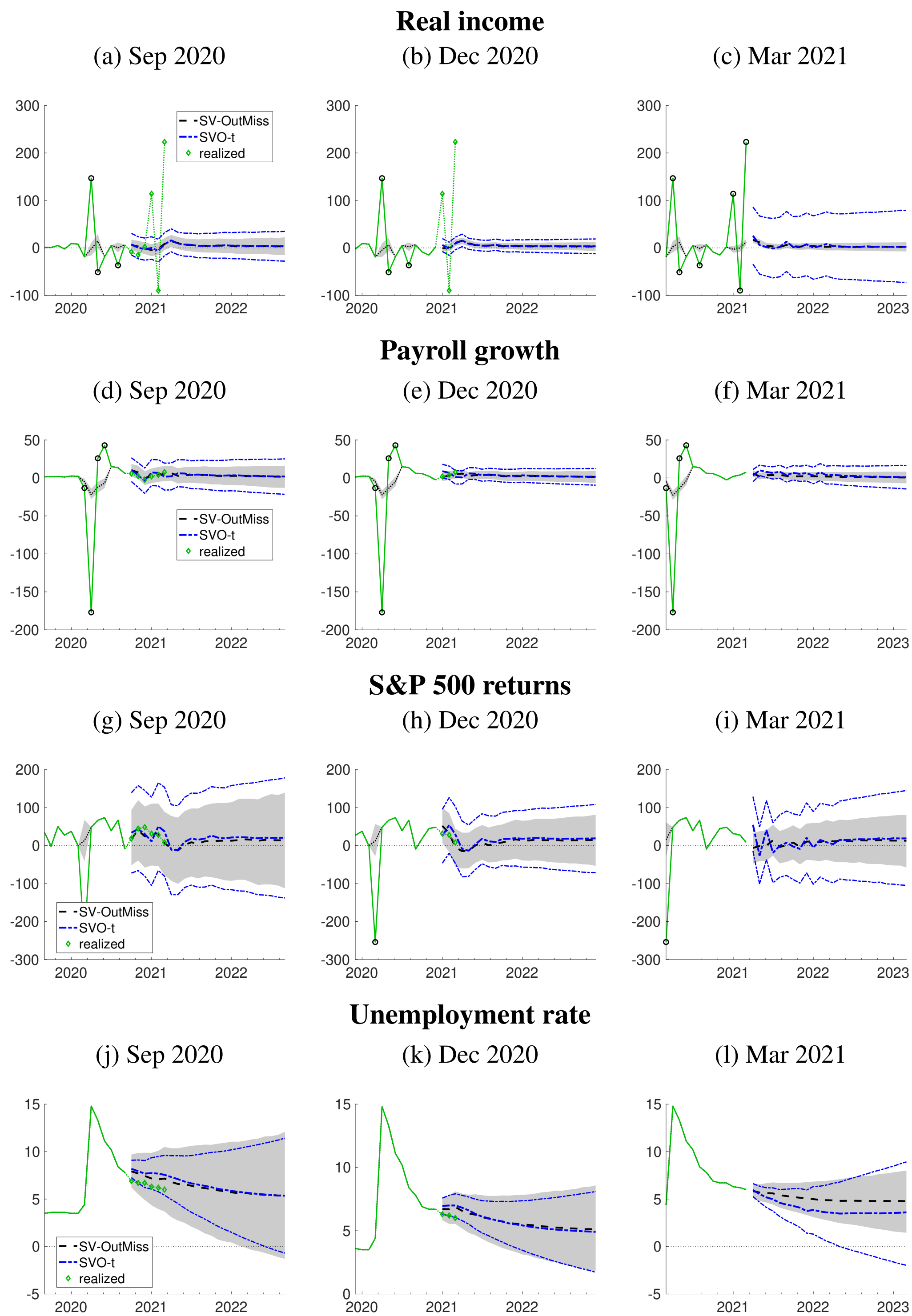
Payroll growth: SV with outlier adjustments

(j) March (k) April (l) June



Note: Medians and 68% uncertainty bands of predictive densities, simulated out-of-sample at various forecast origins as indicated in each panel. The solid green line denotes realized data prior to the forecast origin.

Figure 5: Predictive densities since mid 2020



Note: Medians and 68% uncertainty bands of predictive densities, simulated out-of-sample at various forecast origins as indicated in each panel. Forecasts generated from the SV-OutMiss approach identify observations ex-ante as outliers, based on being more than 5 times the inter-quartile range away from the median; these outliers are indicated with a circle, and the corresponding backcast densities from the SV-OutMiss model are superimposed.