Nowcasting tail risk to economic activity at a weekly frequency

Published Version:
Carriero, A., Clark, T.E., Marcellino, M. (2022). Nowcasting tail risk to economic activity at a weekly frequency. JOURNAL OF APPLIED ECONOMETRICS, 37(5), 843-866 [10.1002/jae.2903].

Availability:
This version is available at: https://hdl.handle.net/11585/900991 since: 2022-11-09

Published:
DOI: http://doi.org/10.1002/jae.2903

Nowcasting Tail Risk to Economic Activity at a Weekly Frequency*

Andrea Carriero
Queen Mary, University of London
a.carriero@qmul.ac.uk

Todd E. Clark
Federal Reserve Bank of Cleveland
todd.clark@clev.frb.org

Massimiliano Marcellino
Bocconi University, IGIER and CEPR
massimiliano.marcellino@unibocconi.it

This draft: January 16, 2022

Abstract

This paper focuses on nowcasts of tail risk to GDP growth, with a potentially wide array of monthly and weekly information used to produce nowcasts on a weekly basis. We consider Bayesian mixed frequency regressions with stochastic volatility and Bayesian quantile regressions. Our results show that, within some limits, more information helps the accuracy of nowcasts of tail risk to GDP growth. Accuracy typically improves as time moves forward within a quarter, making additional data available, with monthly data more important to accuracy than weekly data. Accuracy also typically improves with the use of financial indicators in addition to a base set of macroeconomic indicators.

Keywords: forecasting, pandemics, mixed frequency, quantile regression
JEL classification codes: C53, E17, E37, F47

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System. We gratefully acknowledge expert help from Jared Steinberg and Stephanie Tulley with some of the data used and helpful comments from the Editor Barbara Rossi, three anonymous referees, Knut Are Aastveit, Laurent Ferrara, Daniel Lewis, James Mitchell, Andrew Patton, and Saeed Zaman. Marcellino thanks for financial support MIUR – PRIN Bando 2017 – prot. 2017TA7TYC.

1 Introduction

Nowcasting is commonly viewed as an important and unique forecasting problem; see, e.g., Banbura, Giannone, and Reichlin (2011), Banbura, et al. (2013), and Giannone, Reichlin, and Small (2008). It is important because current-quarter forecasts of GDP growth and inflation provide useful summaries of recent news on the economy and because these forecasts are commonly used as inputs to forecasting models, such as some of the DSGE models in use at central banks, that are effective in medium-term forecasting but not necessarily short-term forecasting. As studies such as Faust and Wright (2009, 2013) have emphasized, initial-quarter forecasts often play a key role in the accuracy of forecasts at subsequent horizons. A key challenge is dealing with the differences in data release dates that cause the available information set to differ over points in time within the quarter — what Wallis (1986) refers to as the "ragged edge" of data.

Much (although not all) of the nowcasting literature has focused on data available at a monthly and quarterly frequency. In part, this may reflect data availability: the histories of weekly indicators of economic activity are in many cases not all that long, constraining formal evaluation of forecasts obtained from estimated models. The literature's limited treatment of weekly data may also in part reflect the finding by Banbura, et al. (2013) that higher frequency information does not seem to be especially useful for nowcasting US GDP growth (except perhaps in a continuous monitoring context). That said, higher frequencies have not been entirely ignored; for example, since 2008, the Federal Reserve Bank of Philadelphia has published a weekly index of economic activity that makes use of weekly data on initial claims for unemployment insurance, as developed by Aruoba, Diebold, and Scotti (2009). Other examples using weekly or daily data include Aastveit, Foroni, and Ravazzolo (2017), Andreou, Ghysels, and Kourtellos (2013), and Ferrara and Simoni (2019).[1]

Since early 2020, the shutdown of significant portions of the economy to restrain the outbreak of the COVID-19 pandemic has raised practical interest in high-frequency indicators of economic activity in the US and other economies. For example, it was clear by mid-March 2020 that much of consumer spending would be shutting down and would lead to large drops in employment and GDP in at least the first and second quarters of the year. Yet, in the second half of March, the usual monthly indicators of economic activity were only available for the month of February. Weekly indicators, such as initial

[1] Aastveit, Foroni, and Ravazzolo (2017) use weekly indexes of economic activity and financial conditions published by the Federal Reserve Bank of Chicago to nowcast and forecast GDP growth. Andreou, Ghysels, and Kourtellos (2013) apply MIDAS methods to daily financial data to nowcast GDP growth. Ferrara and Simoni (2019) use search data from Google to produce and examine nowcasts (point nowcasts for the euro area) on a weekly basis.

claims for unemployment insurance, weekly retail sales from Redbook, raw steel production, and output of electric utilities, began to draw attention for the light they could more quickly shed on the emerging downturn. Lewis, Mertens, Stock, and Trivedi (2021) develop a weekly economic index formed as a principal component of 10 underlying series, now published by the New York Fed.

Apart from nowcasting considerations, a rapidly growing body of research has examined tail risks in macroeconomic outcomes, typically at a horizon of one quarter or one year ahead. Most of this work has focused on the risks of significant declines in GDP, and has relied on quantile regression methods to estimate tail risks, as developed in Adrian, Boyarchenko, and Giannone (2019), Adrian, et al. (2020), De Nicolo and Lucchetta (2017), and Giglio, Kelly, and Pruitt (2016) and extended to vector autoregressive models in Chavleishvili and Manganelli (2019). Reichlin, Ricco, and Hasenzagl (2020) propose using leverage indicators to obtain earlier signals of economic vulnerabilities.

In this context, this paper assesses the ability of models to produce accurate nowcasts of tail risk to GDP, with information sets that can include a number of indicators. Particularly in light of the potential value of high-frequency indicators for nowcasting in the face of rapid changes like those that occurred following the outbreak of the COVID-19 pandemic, our assessment includes as a case study the evolution of nowcasts over 2020. Our modeling starting point is the mixed frequency regression setup of Carriero, Clark, and Marcellino (2015) (henceforth, CCM). In this CCM setup, for nowcasting GDP growth within a quarter, each time series of monthly indicators is transformed into three quarterly time series, each containing observations for, respectively, the first, second, or third month of the quarter. At the moment in time that the forecast is formed, the model includes only the quarterly series without missing observations, which addresses the ragged edge of the data. In this paper, we extend the setup to accommodate weekly indicators of economic and financial conditions. With the number of model predictors sometimes large under our mixed frequency treatment, we consider Bayesian shrinkage and the reduction of variable sets to a common factor. Bayesian methods are used to estimate our models, which facilitates providing shrinkage on estimates of a model that can be quite large, conveniently generates predictive densities, and readily allows for stochastic volatility in a linear regression.

Our paper makes three primary contributions. The first consists of extending the CCM forecast calendar setup: to use 15 different weeks as forecast origins for a quarter's nowcast rather than four months. This setup permits an assessment of the evolution of forecasts with the week by week flow of information in the quarter. A second contribution is to consider higher frequency data — in this paper a number of indicators at a weekly frequency and not just monthly indicators as in CCM. We extend

CCM by including in the quarterly regression weekly indicators available at the time of the forecast origin. Our third and key contribution is that we examine nowcasts of tail risk to economic activity, including over the course of 2020, the year of the COVID-19 pandemic. Following precedents such as Adrian, Boyarchenko, and Giannone (2019) and Adrian, et al. (2020), we use a 10 percent quantile forecast as the measure of tail risk, which we evaluate with the quantile score (tick loss function). We also consider expected shortfall and jointly evaluate the quantile and shortfall forecasts. We compare tail risk nowcasts from Bayesian regressions with stochastic volatility and Bayesian quantile regression, as well as a simple average of forecasts.

Our results show that, within some limits, more information helps the accuracy of tail risk forecasts. Over samples ending before the COVID-19 outbreak induced unprecedented volatility in economic activity, real-time forecast accuracy typically improves as time moves forward from week to week within a quarter, making additional data available, with monthly data more important to accuracy than weekly data. In a given week, models with a wider array of indicators often forecast as well as or better than small models, but again within some limits. In our results, there is a benefit to adding a set of financial indicators (consisting of stock returns, a term spread, a credit spread, and the Chicago Fed's index of financial conditions) to the base set of macro indicators. Adding other weekly indicators of economic activity does not have much effect on forecast accuracy, either to help or harm. Our two model formulations — Bayesian regression with stochastic volatility and Bayesian quantile regression — perform comparably; Bayesian shrinkage is effective with large sets of predictors. Notably, this pattern also means that, for nowcasting tail risk, our regression implementation with stochastic volatility performs as well as quantile regression. An equally-weighted average of the forecasts from our model formulations estimated with a base set of macro and financial indicators performs just as well and can be seen as a robust approach to improving tail nowcast accuracy.

In light of the aforementioned volatility in 2020 and the challenges of nowcasting in the aftermath of the pandemic, we separately examine nowcasts in 2020. In this year, as in our earlier samples for formal evaluation, nowcasts obtained with a base set of macro and financial indicators are hard to improve upon, with our three model specifications (and a simple average forecast) achieving similar accuracy for the year. The timeliness of information played an important role in a number of respects. For example, given the sudden shift in activity that occurred late in 2020:Q1 following the pandemic's outbreak in the US, it was only late in the quarter that the models had enough information showing a downturn to be able to project a significant fall in GDP.

Studies closely related to our approach include Schorfheide and Song (2015) and McCracken, Owyang, and Sekhposyan (2021), which develop mixed frequency Bayesian vector autoregressions (BVARs). Schorfheide and Song (2020) examine forecasting during the pandemic with a mixed frequency BVAR and the monthly and quarterly indicators of Schorfheide and Song (2015). Ferrara, Mogliani, and Sahuc (2021) use the quantile regression setup of Adrian, Boyarchenko, and Giannone (2019) and a Bayesian analogue with MIDAS to nowcast euro area GDP growth with an indicator of financial conditions updated on a daily basis. By comparison, our paper deploys a richer set of predictors and model specifications, along with an alternative approach to accommodating mixed frequency data. Mitchell, Poon, and Mazzi (2020) use Bayesian quantile regression methods to form density nowcasts of euro area GDP growth. We build on their work by focusing on tail risk forecasts, a wider information set, and methods other than quantile regression and quantile regression with Lasso (both estimated by Bayesian methods in their analysis). Plagborg-Moller, et al. (2020) re-examine the ability to forecast and nowcast tail risk to GDP growth and conclude that the evidence of such predictability is weak. The modestly more favorable results we report are likely due to our use of some different indicators, a different nowcast calendar with more of a weekly breakdown, and different measures of tail risk accuracy. Finally, while this paper focuses on nowcasting with an array of economic and financial indicators on the quarter, Carriero, Clark, and Marcellino (2020b) examine tail risk forecasts 1- and 4-quarters-ahead obtained using quarterly data to estimate small BVARs with stochastic volatility and simple quantile regressions.

The paper proceeds as follows. Sections 2 through 4 detail the models, data (including the release calendar setup), and forecast metrics, respectively. Section 5 provides our empirical results. Section 6 concludes. A supplemental appendix provides some additional estimation details and empirical results.

2 Models

Our model exposition relies on the following notation. The vector $X_{w,t}$ contains the available predictors at the time the forecast is formed, t is measured in quarters, and w indicates a week within or shortly beyond the quarter. As detailed below, given a set of indicators to be used, there is a different regressor set $X_{w,t}$ for each week $w = 1, \ldots, 15$ a nowcast is formed, reflecting data availability. The regressors of $X_{w,t}$ consist of selected monthly and weekly variables available in week w, along with a constant and GDP growth in period $t - p$. As detailed below, depending on timing and data availability, the GDP lag p is either 1 or 2.

2.1 Regressions with stochastic volatility

Drawing on our prior work on nowcasting (CCM), one of our models is Bayesian mixed frequency regression with stochastic volatility (BMF-SV). We consider nowcasting the quarterly growth rate of GDP in week w of the current quarter based on the following regression with stochastic volatility, in which the log of the conditional variance of the error term follows a random walk process:

$$\begin{aligned} y_t &= X'_{w,t}\beta_w + v_{w,t} \\ v_{w,t} &= \lambda_{w,t}^{0.5}\epsilon_{w,t},\ \epsilon_{w,t} \sim i.i.d.\ N(0,1) \\ \log(\lambda_{w,t}) &= \log(\lambda_{w,t-1}) + \nu_{w,t},\ \nu_{w,t} \sim\ i.i.d.\ N(0,\phi_w). \end{aligned} \tag{1}$$

In a vector autoregressive context, studies such as Clark (2011), D'Agostino, Gambetti, and Giannone (2013), and Clark and Ravazzolo (2015) have found that this type of stochastic volatility formulation improves the accuracy of both point and density forecasts. In light of the extreme volatility of GDP growth in 2020, for forecasting in that year we extend the BMF-SV specification to add an additional latent state for volatility (independently distributed over time) that can capture short-lived, sharp spikes in the standard deviations of innovations to the regression (in contrast, the stochastic volatility state $\lambda_{w,t}$ has high persistence). This outlier-augmented stochastic volatility specification — detailed in the supplemental appendix — follows the formulations of Stock and Watson (2016) for an unobserved components model and Carriero, et al. (2021) for a BVAR.

The specification of the regressor vector $X_{w,t}$ is partly a function of the way we sample the monthly and weekly variables. For each monthly variable, we first transform it at a monthly frequency as necessary to achieve stationarity. At a quarterly frequency, we then define three different variables, by sampling the monthly series separately for each month of the quarter. The availability of these variables for forecasting GDP in quarter t as of week w drives whether they appear in the forecasting model for that forecast origin. Exactly which variables are included in $X_{w,t}$ depends on when in the quarter the forecast is formed, in line with a direct approach to forecasting.

Regarding the prior and estimation method, we use independent priors for the coefficients and volatility components. The normal priors on the coefficient vector β_w have mean 0 (for all coefficients) and variance that takes a diagonal, Minnesota-style form. The prior variance is Minnesota style in the sense that we take account of the relative scales of variables and it is governed by a small number of hyperparameters. The supplemental appendix provides details. The model is estimated with a Gibbs sampler, using the approach of Kim, Shephard, and Chib (1998) to draw the volatility states. Forecast results from BMF-SV specifications are based on samples of 5000 retained draws, obtained by sampling a total of 30,000 draws, discarding the first 5000, and retaining every 5th draw of the post-burn sample.

2.2 Bayesian quantile regression

In light of our interest in nowcasting tail risk and the prevalence of quantile regression in the tail risk literature, we also consider Bayesian quantile regressions (BQR). In our setting with a regressor set that can be large, Bayesian shrinkage is helpful to mitigate imprecision in coefficient estimates and associated noise in forecasts. Yu and Moyeed (2001) established that quantile regression has a convenient mixture representation that enables Bayesian estimation. With GDP growth in quarter t to be forecast as of week k for quantile τ, our BQR formulation takes the form

$$y_t = X'_{w,t}\beta_{\tau,w} + \sigma_{\tau,w}\epsilon_{\tau,w,t}, \tag{2}$$

where $\epsilon_{\tau,w,t}$ has a mixture representation. For each model at quantile τ and week w, the representation includes $z_{\tau,w,t}$, which is exponentially distributed with scale parameter $\sigma_{\tau,w}$. The mixture representation of the quantile regression model can be written as

$$y_t = X'_{w,t}\beta_{\tau,w} + \theta z_{\tau,w,t} + \kappa\sqrt{\sigma_{\tau,w} z_{\tau,w,t}}\, u_{\tau,w,t}, \tag{3}$$

where θ and κ are fixed parameters as functions of the quantile τ and $u_{\tau,w,t}$ is i.i.d. standard normal.

For the BQR specifications, we use an independent Normal-Gamma prior, with a normal distribution for the regression coefficients and a Gamma distribution for the scale parameter. The prior variance for the coefficients takes the same Minnesota-type form used with the BMF-SV model. We estimate the Bayesian quantile regression with the three-step Gibbs sampler of Khare and Hobert (2012). The first step samples the mixture state time series z from an inverse Gaussian distribution. The second draws the scale parameter $\sigma_{\tau,w}$ from its inverse Gamma conditional posterior. In the third step, the regression parameter vector $\beta_{\tau,w}$ is drawn from its Normal conditional posterior. For each quantile, forecast results from BQR specifications are based on samples of 1000 retained draws, obtained by sampling a total of 6000 draws, discarding the first 1000, and retaining every 5th draw of the post-burn sample. For each retained coefficient draw $\hat{\beta}^{(m)}_{\tau,w}$, $m = 1, \ldots, 1000$, the quantile forecast draw is formed as $X'_{w,t}\hat{\beta}^{(m)}_{\tau,w}$.

Some of the forecast metrics detailed below require an entire predictive distribution and not just a quantile forecast. So following studies including Korobilis (2017) and Mitchell, Poon, and Mazzi (2020), all of our reported BQR results are based on combined empirical densities.[2] In particular, for each of 19 quantiles $\tau = i \cdot 0.05$, $i = 1, \ldots, 19$, we first estimate BQR models. For the resulting set of 19,000 retained quantile forecast draws, we then use kernel smoothing to fit an empirical density (the appendix provides details). Finally, from that empirical forecast density, we compute quantiles and the other forecast objects of interest. We have verified that quantile results are qualitatively the same (e.g., the time profiles of quantiles reported below are essentially the same to the eye) if the combination and smoothing step are skipped in favor of using the posterior mean BQR quantiles directly.[3] Similarly, other results are unchanged with different choices for the estimation of the empirical density.

2.3 Partial quantile regression

As another approach to nowcasting tail risk with large sets of regressors, in some results we consider dimension reduction via the use of a common factor in monthly and weekly indicators. Specifically, we consider the partial quantile regression (PQR) method of Giglio, Kelly, and Pruitt (2016, GKP). GKP characterize partial quantile regression as an adaptation of partial least squares to a quantile regression framework. PQR is targeted to quantile regression in that it uses quantile regression in the factor estimation. In our implementation, we follow GKP in using a single factor specification, and we depart from GKP in estimating the quantile regression used in the forecasting step with Bayesian methods. While detailed results for this model are presented in the supplemental appendix (its nowcast accuracy is comparable to but not better than that of the BMF-SV and BQR models), we include its forecasts in the model average results that are presented in the paper.

At each forecast origin, consider the vector of variables $X_{w,t}$ included in the specifications above, associated with week w. For each quantile τ, we follow the quantile regression-based approach of GKP to obtain a time series of a scalar factor $f_{\tau,w,t}$ from the monthly and weekly indicators of $X_{w,t}$.[4] We

[2]Mitchell, Poon, and Mazzi (2020) find that the combination approach yields density nowcasts more accurate than those obtained with the Adrian, Boyarchenko, and Giannone (2019) approach of fitting a skewed-t density to several quantile estimates.

[3]In fact, the quantile and quantile score results in Carriero, Clark, and Marcellino (2020b) use the direct quantile forecasts rather than empirical densities.

[4]In the first stage quantile regression used to obtain the factor, we include a constant, lagged GDP growth, and one of the components of the monthly and weekly indicators of $X_{w,t}$, on a one-at-a-time basis.

then estimate the mixed frequency quantile regression

$$y_t = Z'_{w,t}\beta_{\tau,w,Z} + f_{\tau,w,t}\beta_{\tau,w,f} + \sigma_{\tau,w}\epsilon_{\tau,w,t} \tag{4}$$

and form draws of the PQR nowcast for quantile τ with the resulting coefficient estimates.[5]

We produce the PQR forecast results with the same approach described above for the BQR specification. For each of 19 quantiles $\tau = i \cdot 0.05$, $i = 1, \ldots, 19$, we estimate PQR models. We then use kernel smoothing to fit an empirical density to the set of 19,000 retained quantile forecast draws, from which we in turn compute quantiles and the other forecast objects of interest.

2.4 Other forecasts

In our forecast evaluation, as a baseline we use a simple AR(2) model with stochastic volatility. This model takes the same basic form given in (1), with $X_{w,t}$ defined to include just a constant and two lags of GDP growth. We generate AR-based forecasts of GDP growth at each of the 15 forecast origin weeks of the quarter, based on the data available in real time as of week w of the quarter.[6] We estimate these AR-SV models with very large prior variances, so as to impose little shrinkage on the AR coefficients.

With one of our better-performing variable sets detailed in the next section, we also consider an equally-weighted average of forecasts from the BMF-SV, BQR, and PQR specifications. We use equal weights partly for their simplicity and partly out of consideration of the tail risk results in Taylor (2020), in which a simple average sometimes performs as well as a more sophisticated combination. For those forecast metrics detailed below that require an entire predictive distribution and not just a quantile forecast, we obtain the predictive density for the average forecast as an equally-weighted linear combination of the predictive densities from the BMF-SV, BQR, and PQR specifications (the appendix details the approach).

An earlier version of this paper (Carriero, Clark and Marcellino (2020b)) considers nowcasts obtained from a number of other approaches, including: frequentist quantile regression (QR); QR with Lasso penalty; BQR with Lasso penalty; alternative factor-based specifications, with a few principal components of the variables used in BMF-SV and BQR specifications; and a variety of forecast averages.

[5]Because PQR features factor reduction, we use a modestly looser prior for the regression coefficients of the PQR specification than for the coefficients of the BQR specification, as detailed in the supplemental appendix.

[6]The models in weeks 4-15 of the quarter are all conventional AR(2) specifications, and for a given quarter, these model estimates and forecasts differ across weeks as the GDP data available are updated. However, in weeks 1-3 of the quarter, the model takes a direct multi-step form, relating GDP in quarter t to GDP in quarters $t-2$ and $t-3$, and the forecast horizon is in effect 2 quarters, not 1 quarter.

As the results were not better than those for BMF-SV and BQR, for brevity this paper focuses on the latter two models and one simple average forecast.

3 Data

As noted above, we focus on current-quarter forecasting of real GDP growth in real time. This section first explains the general design of our forecast calendar and then details the data used.

3.1 General design of the forecast calendar and data set

Whereas most of the nowcasting literature (including CCM) focuses on a monthly calendar of data releases and forecast origins for nowcasting, we consider a weekly calendar of data releases and forecast origins.[7] With this weekly calendar, we consider monthly data as well as weekly data. Our forecast calendar includes 15 weeks for each quarter, beginning with the first full week of the quarter and ending with the third week of the following quarter (the last week before the initial estimate of GDP for the prior quarter is published). For each indicator we consider, we assign it a typical release or availability week based on its usual publication schedule. As examples, at the end of week 1 of a quarter, a forecaster has available data on employment, initial claims for unemployment insurance, interest rates, and stock prices for the prior month, as well as interest rates and stock prices for the first week of the quarter. At the end of week 2, a forecaster also has available (in addition to the data of week 1) retail sales for the prior month, claims for week 1 of the quarter, and interest rates and stock prices for week 2 of the quarter.

We report results for a total of 6 different combinations of macroeconomic and financial indicators. The weekly activity indicators used correspond to most of those in the weekly economic activity index of Lewis, Mertens, Stock, and Trivedi (2021). Table 1 lists the variables and our calendar assumptions.

Drawing on the small-model specification of CCM, our starting point is the *base M* ("M" for macro) set of 6 macroeconomic activity indicators. This set includes 5 monthly indicators broadly informative about economic conditions, selected with some eye to timeliness: payroll employment, industrial production, real retail sales (nominal deflated by the CPI), housing starts, and the manufacturing index from purchasing managers published by the Institute for Supply Management (ISM). It also includes

[7]Some studies consider a higher frequency calendar of nowcast updates. For example, Aastveit, et al. (2014) consider 15 dates for data releases — most monthly or quarterly — in the three months of the quarter and the first month of the following quarter.

initial claims for unemployment insurance, using both monthly and weekly observations as available. Initial claims are commonly considered to be a leading indicator of the business cycle and have the advantage of being available weekly with a fairly short lag (one week).

Our second variable set — *base M-F* — adds to the base M variable set a set of monthly and weekly financial indicators ("F" for financial), selected with an eye toward those that have been found in the literature to have some predictive content for output: the Chicago Fed's national financial conditions index (NFCI), stock returns as measured with the S&P 500 index, the term spread between the 10-year and 1-year Treasury yields (constant maturity), and the credit spread between Moody's Baa corporate yield and the 10-year Treasury yield.

The variable set *base M + small weekly* puts together the base M variables and a small set of weekly indicators of economic activity with data available back to the mid-1980s: continuing claims for unemployment insurance, consumer comfort, steel production, and electric utility output. The variable set *base M + large weekly* puts together the base M variables and a large set of weekly indicators of economic activity consisting of (i) those in the small set plus (ii) loadings of railroad cars, total fuel sales, and Redbook same-store retail sales. For these additional indicators, data are available back to the mid-1990s. Similarly, the variable sets *base M-F + small weekly* and *base M-F + large weekly* put the base M-F variables together with the small and large sets of weekly indicators, respectively.

Our model specifications reflect additional choices regarding transformations and treatment of data frequency. As Table 1 indicates, with variables subject to trends, such as GDP, employment, or stock prices, we use growth rates. For variables available at a daily frequency (interest rates and stock prices), we use monthly averages and weekly averages as our monthly and weekly observations. At a monthly frequency, the growth rate of the S&P 500 is the percent change in the month-average index values. At a weekly frequency, to smooth out some of the higher frequency noise in stock prices, we use a 4-week percent change in the weekly index. We smooth the consumer comfort measure by using a 4-week average of the weekly data. In light of noisiness and strong seasonality in weekly indicators of steel production, utility output, car loadings, fuel sales, and Redbook retail sales, we follow Ferrara and Simoni (2019) and Lewis, Mertens, Stock, and Trivedi (2021) and rely on 52-week growth rates.

3.2 Details of data used

Quarterly real-time data on GDP are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM).

For the predictors subject to significant revisions — payroll employment, industrial production, retail sales, and housing starts — we use real-time data, obtained from the RTDSM (employment, industrial production, and housing starts) or the Federal Reserve Bank of St. Louis' ALFRED database (retail sales). For the CPI used to deflate retail sales, we use the 1967-base-year CPI available from the BLS rather than a real-time series; Kozicki and Hoffman (2004) show that the 1967-base-year series is very similar to real-time CPI inflation. For the other variables, subject to either small revisions or no revision, we simply use the currently available time series, obtained from either the Federal Reserve Board's FAME database or Haver Analytics.[8]

The full forecast evaluation period runs from 1985:Q1 through 2020:Q4, which involves real-time data vintages from January 1985 through March 2021. For each forecast origin t starting in the first week of 1985:Q1, we use the real-time data vintage t to estimate the forecast models (recursively, allowing the sample to expand as forecasting moves forward in time) and construct forecasts of GDP growth in the quarter. In forming the data set used to estimate the forecasting models at each point in time, we use the monthly vintages of (quarterly) GDP available from the RTDSM, taking care to make sure the GDP time series used in the regression is the one available at the time the forecast is being formed. The starting point of the model estimation sample varies across some of our specifications due to differences in data availability. With the base M and base M-F variable sets, we use a common start date of 1971:Q2, the soonest possible given the availability all indicators (the NFCI series begins in early 1971) and lags allowed in models. Adding some more weekly indicators of economic activity pulls the estimation sample start up to 1987:Q1, and adding the full set moves the estimation start date to 1996:Q3. In these cases, we shorten the evaluation samples to start in 2000:Q1 and 2007:Q1, respectively.

To evaluate the accuracy of the real-time forecasts, we follow studies such as Clark (2011) and Faust and Wright (2009) and use the second available estimates of GDP in the quarterly vintages of the RTDSM as actuals in evaluating accuracy.

3.3 Indicators used

In the results reported in this paper, for the most part we only include in the model values of these variables for the current quarter t, the quarter for which GDP growth is being forecast. Our rationale is primarily that, in the simpler monthly setup of CCM, we did not find any payoff to longer lags.

[8] Supplemental Appendix Table A1 gives the source from which we obtained each series.

All model specifications include in the regressor set $X_{w,t}$ a constant and one lag of GDP growth. In most cases, this means the models include GDP growth in period $t-1$. However, in the case of models used to forecast in the first few weeks of the quarter, because the value of GDP growth in period $t-1$ is not actually available in real time, we include in the models GDP growth in period $t-2$. (This is consistent with our general direct multi-step treatment of the forecasting models.)

As noted above, depending on the week of the quarter that the forecast is being formed, exactly which variables are in the forecasting models (that is, in $X_{w,t}$) varies, reflecting real-time data availability and the usual publication schedules of the indicators. Table 2 details the model specifications (and variable timing) we use, covering, for simplicity, just a few of our variable sets. For each week indicated in the first column, the table has three rows of entries, with the first listing the relevant base M indicators, the second row covering the finance indicators, and the third listing the small weekly indicators included in the given week's models. The variable sets *base M*, *base M-F*, and *base M-F + small weekly* combine these predictors as indicated. Models for the *base M-F + large weekly* variable set include the first, second, and third row indicators plus three additional weekly indicators with the same specification shown for the variables in the third row.

Regarding model details, the dependent variable of the model is GDP growth in quarter t. Subscripts of t, $t-1$, and $t-2$ refer to the current, once lagged, and twice lagged quarters, respectively. Months and weeks within the quarter are indicated by superscripts containing $m1$, $m2$, $m3$ for the first to third months of the quarter and containing $w1$, $w2$, ..., $w12$ for weeks 1 through 12 of the quarter and $w13$ through $w15$ for the first few weeks of the next quarter. For a given variable in a given week, the table shows in the superscripted notation which months or weeks of the variable in question are available and included in the model. For example, in week 9 of the quarter, we have available and include in the model employment data for the first two months of the quarter and retail sales for just the first month of the quarter. The table indicates this aspect of the specification with the week 9, row 1 entries of $\text{emp}_t^{(m1,m2)}$ and $\text{retail}_t^{(m1)}$.

With the weekly indicators of unemployment claims and financial conditions, in light of their overlap with monthly data, our models reflect some specific choices on the timing or selection of which readings are included. For these variables, once a full month is available, we include the full month average in the model and not weekly observations from that month. With weekly observations that are included, we take an average across the weeks available for the month and put that average in the model and not each week's reading. For example, with stock prices and spreads, at the end of the third week of a month, we have available readings for weeks 1 through 3, and we enter the three-week average in the model. In the table, this is denoted with a superscript showing $w1 + w2 + w3$. For instance, in the specification for week 10 using the base M-F variable set, the BMF-SV and BQR specifications include variables for each of the month 1, month 2, and week 9 readings of the NFCI (indicated by the row 2 entry $\text{NFCI}_t^{(m1,m2,w9)}$) and variables for each of the month 1, month 2, and weeks 9-10 average reading of the term spread (indicated by $\text{SP}_t^{(m1,m2,w9+w10)}$).

With other weekly indicators of economic activity, in light of the underlying transformations used to reduce their noise (52-week percent changes in most cases), we only include in the model a single weekly reading that is the most recent available. For example, in the specification for week 10 including the small set of weekly economic activity indicators, the predictors include the week 9 readings on consumer sentiment (the 4-week average), steel production (52-week percent change), and electric utility output (52-week percent change), indicated by the table entries $\text{sment}_t^{(w9)}$, $\text{steel}_t^{(w9)}$, and $\text{util}_t^{(w9)}$.

Across variable sets and forecast origins, our forecasting models vary widely in size. In some cases (base M, week 5), the model is relatively small, with six predictors. In many other cases, the models have a healthy number of regressors without necessarily being large (e.g., the base M-F model in week 7 has 17 predictors). In some settings, the model becomes quite large: the base M-F + large weekly model peaks at 47 regressors (in week 15). With models of these medium to large sizes, our Bayesian approaches to estimation incorporate shrinkage help to limit the effects of parameter estimation error on forecast accuracy.

4 Forecast Metrics

In assessing the efficacy of the models described in the previous section, we consider a range of forecast metrics. As a broad starting point, we evaluate overall density accuracy with the continuous ranked probability score (CRPS) developed in Gneiting and Raftery (2007):

$$\text{CRPS}_t = \int_{-\infty}^{\infty} \left(F(z) - \mathbf{1}_{\{y_t \leq z\}}\right)^2 dz, \tag{5}$$

where F denotes the cumulative distribution function associated with the predictive density f and $\mathbf{1}_{\{y_t \leq z\}}$ denotes an indicator function taking value 1 if $y_t \leq z$ and 0 otherwise. Computing the CRPS requires a complete predictive density. The estimation of the BMF-SV models yields such a density.

As noted in section 2.2, we obtain complete predictive densities for BQR specifications by combining posterior draws across a range of quantiles. For all models, with samples from the predictive densities, we compute the CRPS using an approach (their equation (3)) recommended by Jordan, Krueger, and Lerch (2019).

Our evaluation of tail risk forecasts includes a few different forecast metrics.[9] We focus on the 10 percent left tail. We have verified (see the online appendix for detailed results) that our results on lower tail forecasts are robust to instead using the 5 percent quantile.

The first tail metric is the quantile score, commonly associated with the tick loss function (see, e.g., Giacomini and Komunjer (2005)). The quantile score (QS) is computed as

$$\mathrm{QS}_{\tau,t} = (y_t - Q_{\tau,t})(\tau - \mathbf{1}_{\{y_t \leq Q_{\tau,t}\}}), \tag{6}$$

where y_t is the actual outcome for GDP growth, $Q_{\tau,t}$ is the forecast of quantile τ, and the indicator function $\mathbf{1}_{\{y_t \leq Q_{\tau,t}\}}$ has a value of 1 if the outcome is at or below the forecast quantile and 0 otherwise. For the BMF-SV models, the quantile is simply estimated as the associated percentile of the simulated predictive distribution. For the BQR models, the quantile is computed using the empirical distribution of the combined set of BQR forecast draws.

Of course, the tail quantile forecast corresponds to the value at risk (VaR) forecast.[10] Conceptually, VaR has a number of disadvantages, leading some (e.g., Basel standards on financial risk management), to prefer expected shortfall (ES). In our context, at the 10 percent level, ES is a measure of the average growth rate that would be observed if growth were in that tail of the distribution. However, as explained in Fissler and Ziegel (2016), expected shortfall by itself is not an elicitable risk measure (i.e., the correct forecast need not be the unique minimizer of the loss function). Instead, VaR and ES can be jointly elicited, and Fissler and Ziegel (2016) derive a general class of such scoring functions. Studies such as Patton, Ziegel, and Chen (2019) and Taylor (2019) develop specific functions within this general class. However, some of these functions are designed with asset returns in mind and embed a restriction that ES is strictly negative.[11] Tail forecasts of GDP growth often (in periods of economic expansion) violate such a restriction, with a positive ES.

[9] We have also considered the left-tail weighted quantile score-based version of the CRPS developed in Gneiting and Ranjan (2011). The supplemental appendix provides these results, which are very similar to those for the standard CRPS.

[10] Adrian, et al. (2020) coined the term "growth at risk" for GDP growth forecasts, and De Nicolo and Lucchetta (2017) coined similar terms for industrial production and employment.

[11] For example, while the FZ0 function of Patton, Ziegel, and Chen (2019) has the advantage of homogeneity, it requires that ES be negative.

Accordingly, rather than rely on an existing scoring function designed with asset returns in mind, we consider a different implementation — new to this paper — of a VaR-ES scoring function that allows ES to be positive or negative.[12] In particular, letting $\text{ES}_{\tau,t}$ denote the expected shortfall estimate at quantile τ (for BQR, we estimate the ES using complete predictive densities obtained by combining posterior draws across a range of quantiles), our VaR-ES scoring function takes the form

$$\text{S}_{\tau,t} = b + (Q_{\tau,t} - y_t)\left(\mathbf{1}_{\{y_t \leq Q_{\tau,t}\}} - \tau\right) + \frac{1}{\tau} e^{a^{-1}\text{ES}_{\tau,t}} \left(\mathbf{1}_{\{y_t \leq Q_{\tau,t}\}} (Q_{\tau,t} - y_t) + \tau \left(\text{ES}_{\tau,t} - Q_{\tau,t} - a\right)\right). \quad (7)$$

In the general notation of Fissler and Ziegel (2016), the scoring function in (7) uses $G_1(x) = x$ and $G_2(x) = e^{a^{-1}x}$. Because the $G_2(x)$ function is not homogenous, the scoring function values are specific to the units of the variable of interest, which in our case is annualized (quarter-on-quarter) GDP growth. In implementation, the scoring function coefficient a is set to 4, and the constant b is set to 6 to ensure a positive score (for simplicity in reporting some results; this setting is irrelevant for the score differences across models of interest).

To gauge statistical significance of differences in CRPS, QS, and the VaR-ES score, we estimate Diebold and Mariano (1995)–West (1996) t-tests for equality of the average loss. We conduct these tests on a one-sided basis, such that the alternative hypothesis is that the indicated forecast is more accurate than the AR-SV benchmark.[13]

As a final check on tail risk forecasts, we apply Christoffersen's (1998) conditional coverage test to the 10 percent tail forecast. The conditional coverage test combines likelihood ratio tests of (1) independence and (2) unconditional coverage to provide a general conditional efficiency criterion for interval forecasts. The conditional test has a $\chi^2(2)$ distribution.

5 Empirical Results

This section presents our results on the accuracy of out-of-sample nowcasts of GDP growth. We first examine the accuracy of forecasts in data through 2019:Q4 and then consider as a case study the evolution of forecasts over the course of 2020. We treat 2020 separately because of the extreme volatility induced by the COVID-19 outbreak and subsequent recovery.

[12] We thank Andrew Patton for tremendously helpful advice that resulted in our function choice.

[13] We compute the variance underlying the test statistics using the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992).

5.1 Nowcast accuracy through 2019:Q4

As a starting point, we assess how increasing the basic information set over the weeks of the quarter affects forecast accuracy. Figure 1 reports CRPSs, 10 percent quantile scores, and 10 percent VaR-ES scores from the BMF-SV model estimated with the base M variable set (results are similar with the base M-F variable set), for evaluation samples of 1985:Q1-2019:Q4 (upper panel) and 2000:Q1-2019:Q4 (lower panel). To facilitate comparisons across weeks, for each measure we normalize the accuracy in a given week by the accuracy in week 1, so that in week 1 the observation is equal to 1.0. For both samples, the CRPS, 10 percent QS, and 10 percent VaR-ES score fairly steadily improve by the week; additional information on the quarter materially improves the accuracy of both density and tail risk forecasts. From week 1 to week 15, the QS and VaR-ES score improve by 25 to 35 percent, and the CRPS improves by about 20 percent.

To compare real-time accuracy across models and variable sets, Tables 3 through 6 provide results for CRPS, QS, VaR-ES score, and conditional coverage for evaluation samples ending with 2019:Q4. To facilitate comparisons across models, results for CRPS and QS are reported as ratios relative to the AR-SV benchmark (ratios less than 1 indicate a given model is more accurate than the baseline), and results for the VaR-ES score are reported as differences relative to the AR-SV model (positive entries mean a given model is more accurate).[14]

Starting with Table 3's CRPS results on overall density accuracy, the AR-SV benchmark is easily beaten by most of the specifications making use of monthly or weekly indicators, in keeping with the results in CCM using just monthly data and the BMF-SV specification. There are two respects in which the results indicate that having more data is helpful to nowcast accuracy. First, having more recent data on the quarter helps. Some improvement in forecast accuracy is achieved with weekly indicators in the first few weeks of the month, before monthly indicators for the quarter are yet available. For example, in the results for 1985:Q1-2019:Q4 using the base M-F variable set, as of week 3 of the quarter the BQR and BMF-SV specifications yield CRPS improvements of a little more than 10 percent. The relative accuracy of specifications using monthly and weekly indicators also improves as additional data on the quarter becomes available across weeks. Continuing with the same example, the BQR and BMF-SV gains increase to nearly 20 percent by week 11. The second respect in which more data is helpful to nowcast accuracy is that models including financial indicators commonly improve (albeit sometimes

[14]For all measures, the accuracy of the AR-SV benchmark improves early in the quarter when the estimate of GDP for the previous quarter is first released but changes little from week 5 to week 15.

just slightly) on models with just macro indicators. In particular, comparing the base M and base M-F results, for a given model, forecasts obtained with the latter variable set are more accurate than those obtained with the former. That being said, the payoff to additional indicators has some limits. Notably, from a baseline of the base M set of indicators, adding small or large sets of weekly activity indicators fails to yield improvements in density accuracy. On the other hand, from a baseline of the base M-F set of indicators, adding small or large sets of weekly activity indicators often does not harm density accuracy much, either. That is, given a model and evaluation sample, CRPS ratios for the base M-F+small weekly and base M-F+large weekly variable sets are often similar to those for the base M-F variable set.

Finally, regarding model performance for a given variable set, the BMF-SV and BQR specifications yield quite similar accuracy. The gains achieved by the BMF-SV and BQR specifications relative to the AR-SV baseline are often statistically significant, somewhat more so in the longer evaluation samples of 1985:Q1-2019:Q4 and 2000:Q1-2019:Q4 than the shorter sample of 2007:Q1-2019:Q4. A simple average of the BMF-SV, BQR, and PQR forecasts also performs relatively well. Given the base M-F variable set, while the relative accuracy of the BMF-SV and BQR forecasts changes across forecast origin weeks and evaluation samples, the average is always very close to the accuracy of whichever forecast is most accurate. In this sense, the simple average may be seen as a robust forecast.

The results on accuracy of tail risk forecasts display similar patterns. Table 4 provides results for the 10 percent QS. By this tail risk metric, the AR-SV benchmark is easily beaten by most of the specifications making use of monthly or weekly indicators. Again, in two respects having more data is helpful to nowcast accuracy. First, the QS ratios typically decline across weeks of the quarter. For example, in the results for 1985:Q1-2019:Q4 using the base M-F variable set, as of week 3 of the quarter the BQR and BMF-SV specifications yield QS improvements of 12 to 13 percent, and by week 13, the corresponding gains are 23 to 27 percent. Second, the gains are typically larger with the base M-F variable set than the base M variable set. As of week 3 (week 13) of the quarter, the QS ratios for the BQR and BMF-SV specifications in the 1985:Q1-2019:Q4 sample are 1.01 and 0.85 (0.92 and 0.77), respectively, with the base M variable set. The finding that financial indicators are helpful to tail risk nowcast accuracy is consistent with the emphasis of much of the recent literature on tail risks (e.g., Adrian, Boyarchenko, and Giannone (2019)). The results for QS are also similar to those for CRPS in that including weekly indicators of economic activity (in the small and large sets of indicators) does not improve accuracy in the 2000:Q1-2019:Q4 and 2007:Q1-2019:Q4 samples. On the other hand, adding these indicators does not harm the accuracy of tail risk nowcasts as measured by QS.

Focusing on a comparison of models, with most variable set choices, the BMF-SV and BQR specifications achieve similar accuracy (with the exception of the base M variable set in the 1985:Q1-2019:Q4 sample). Our finding that a linear regression specification with stochastic volatility nowcasts tail risk as well as quantile regression-based methods echoes the finding of Carriero, Clark, and Marcellino (2020a) for 1- and 4-quarters-ahead forecasts from BVAR specifications as compared to simple quantile regression. Once again, with the base M-F variable set, the simple average forecast is consistently close to the most accurate forecast and offers some robustness in that sense. All this being said, comparing the QS results in Table 4 to the CRPS results in Table 3 indicates that statistical significance of gains is harder to achieve for tail risk accuracy than density accuracy. For example, by the QS measure, the gains in accuracy achieved with the base M-F variable set in weeks 7 through 15 are consistently statistically significant in the 1985:Q1-2019:Q4 sample but not the shorter samples.

Table 5's results for the VaR-ES score are broadly similar to those for QS. Again, accuracy relative to the AR-SV baseline generally improves as more information on monthly and weekly indicators becomes available across weeks of the quarter, and including financial indicators in the base M-F variable set commonly improves on the accuracy of forecasts obtained with just the base M variable set. As an example of the latter regularity, with the BQR model and the 1985:Q1-2019:Q4 sample, the score differential with the base M-F variable set averages 0.63 across the weeks shown, compared to an average of 0.11 with the base M variable set. In addition, the BMF-SV and BQR specifications achieve comparable accuracy gains, with statistical significance a little more prevalent in results with the VaR-ES score than the QS.

Finally, to provide a general conditional efficiency evaluation of the tail quantile nowcasts, Table 6 reports p-values for Christoffersen's (1998) conditional coverage test applied to the 10 percent interval forecast.[15] In the case of the AR-SV benchmark, the null of correct conditional coverage is never rejected, implying that, conditional on just past GDP, the nowcasts from this model yield a tail interval forecast that has correct unconditional coverage and passes the independence test component. Making use of monthly and weekly indicators to improve the tail forecast accuracy as measured by the metrics discussed above can yield some rejections of the null of correct conditional coverage. The rejections

[15]In some cases, we made some simple adjustments to the test statistic calculation to avoid numerical problems. In particular, for some forecasts, there are no back-to-back observations below the tail quantile. To avoid the independence piece of the test statistic including the log of 0, we replaced these instances of zeros with a very small positive number in order to compute the test statistic. In these cases, the overall test statistic is driven by the unconditional coverage piece, with very little influence from the independence piece of the test statistic.

occur primarily in the 1985:Q1-2019:Q4 sample with the base M variable set. These rejections are driven by the unconditional coverage component of the test, with the models yielding a tail estimate that is too low for actual outcomes to fall in the interval with an average frequency of 10 percent (the supplemental appendix provides the unconditional coverage rates). The specifications that include financial indicators yield fewer rejections of correct conditional coverage. Over the evaluation samples that start in 2000:Q1 or 2007:Q1, rejections are rare, reflecting unconditional coverage rates that are closer to the nominal 10 percent than is sometimes the case in the sample starting in 1985:Q1.

To help shed some light on the patterns documented above, we conclude our evidence with a few examples of the historical time series of forecasts for samples ending in 2019:Q4. Figure 2 provides forecasts from the BMF-SV and BQR specifications estimated with the base M-F variable set. These charts include the point, 10 percent quantile, and 90 percent quantile forecasts produced at a set of selected forecast origins within the quarter, along with the actual GDP growth outcome.[16] As indicated in this figure, with the BMF-SV model and the base M-F variable set, the 10 percent and 90 percent forecast quantiles tend to move together. We do not seem to obtain the asymmetric moves in quantiles (with the downside moving down more than the upside does, around the times of recessions) evident in the 1-quarter-ahead and 4-quarter-ahead results of Carriero, Clark, and Marcellino (2020a). With the BQR specification, the estimates display some asymmetry in the sense that the lower tail forecast is modestly more volatile than the upper tail forecast. However, the asymmetry is less striking than in the results of Adrian, Boyarchenko, and Giannone (2019) for a sample (of in-sample forecasts) going back to the 1970s.

There appear to be a few key drivers of the finding that asymmetries are reduced in our nowcasting setting than in the forecast settings of Adrian, Boyarchenko, and Giannone (2019) and Carriero, Clark, and Marcellino (2020a). One is the sample: Asymmetries are more striking around the recessions of the 1970s (included in the in-sample results of Adrian, Boyarchenko, and Giannone (2019)) than in samples starting in the mid-1980s, as in this paper. The other key drivers are the conditioning information and forecast or nowcast horizon. If we reduce the set of predictors to the NFCI and examine in-sample nowcasts, asymmetries are sharp in the first few weeks of the quarter but less sharp (although still present) in subsequent weeks. Adding our other base macro and financial indicators to the model further reduces (but does not eliminate in the BQR case) the asymmetry in fluctuations of the lower

[16]The point forecast is the posterior mean for the BMF-SV model and the posterior mean of the combined empirical distributions for the BQR and PQR specifications.

quantile compared to the upper quantile. In particular, with more information used to form the nowcast, the upper quantile becomes more variable, showing sharper reductions around the time of recessions.

5.2 Nowcasts in 2020

We conclude our analysis with an examination of tail risk nowcasts in the pandemic-distorted year of 2020. We treat 2020 separately out of two considerations. First, as a case study, it is useful to examine the evolution of nowcasts with additional data on each quarter in the context of the swift changes that occurred in growth over the course of the year. Second, GDP growth's unprecedented lows and highs during the year would cause forecasts and outcomes in 2020 to dominate and possibly distort a formal historical evaluation of a longer period that included the year.

Table 7 reports CRPS, 10 percent QS, and VaR-ES score results for 2020:Q1-2020:Q4. In light of the very small sample and extreme volatility, we just report the statistics without assessing statistical significance. As noted in section 2.1, the AR-SV and BMF-SV forecasts for 2020 are based on models augmented to include an outlier state for volatility that is independently distributed over time.[17] As expected, the AR-SV scores indicate that accuracy was orders of magnitude worse in 2020 than in the samples ending with 2019. For example, the QS of the AR-SV benchmark was more than 11 in 2020, compared to roughly 0.4 or 0.5 in the samples ending with 2019. Using monthly and weekly indicators to nowcast in 2020 yielded material improvements in accuracy, to a greater extent than observed in the earlier samples. For instance, in 2020, the BMF-SV and BQR specifications improved the CRPS by roughly 50 percent and the QS by 70 to 80 percent. Once again, a number of choices of model-variable combinations perform relatively well, although it is hard to beat using the base M-F variable set with the BMF-SV and BQR models, without any one model clearly better than another. For example, as of week 11, with the base M-F variable set, the CRPS ratios are 0.48 and 0.49 for the BMF-SV and BQR models, respectively, and the corresponding QS ratios are 0.22 and 0.20. The average forecast achieves CRPS and QS ratios of 0.48 and 0.20, respectively.

Figure 3 reports the evolution of nowcasts for 2020, covering the base M-F variable set and the BMF-SV, BQR, and average forecasts. The charts include not only the 10 percent quantile forecasts (orange and pink lines) but also point forecasts (blue and gray lines) and the actual growth outcomes (black dots). As indicated in the nowcasts in the left-most portion of Figure 3, early in 2020:Q1, before the pandemic spread and shutdowns began in mid- to late March, not only the point nowcasts but also

[17]Forecast results for these augmented models over the samples ending in 2019 are indistinguishable from those reported.

the nowcasts of the 10 percent quantile were above 0. It was around week 8 that stock prices started to register a falloff in response to global news on the pandemic's outbreak. It took a little more time for indicators of economic activity to reflect the shutdown. In most cases, the nowcasts turned significantly negative in week 12, and all ended up being below the actual growth outcome.

For 2020:Q2, the 10 percent quantile nowcasts were sharply negative from the beginning, with the week 1 projections ranging from -6.6 percent to -15.3 percent. The nowcasts turned sharply more negative in week 5, with the availability of the first monthly economic indicators on the second quarter; over the ensuing several weeks, the 10 percent quantile nowcasts showed some improvement, but remained extremely low. The actual outcome for GDP growth in the second quarter (-37.7 percent) turned out to be comparable to the tail risk nowcasts from the last few forecast origins for the quarter.

As the economy began to recover in the second half of 2020, the nowcasts showed variability across weeks and models before stabilizing. In 2020:Q3, the nowcasts correctly detected a better outlook for GDP growth compared to the prior quarter, but generally failed to detect — in both point forecasts and the tail risk nowcast — the big positive swing in growth in 2020:Q3 (actual growth of 28.8 percent). That said, without any precedent for the turn of events with the pandemic and recovery, one might subjectively argue that it was extremely difficult to assess downside risks to the economy in the third quarter. That may be particularly true given that most high frequency data primarily concern goods and not services, and the pandemic was highly unusual by historical standards in that it most affected services rather than goods. In 2020:Q4, with growth of economic activity slowing to a more conventional rate (4.2 percent), the nowcasts were more stable across weeks of the quarter and less extreme than often the case in the previous two quarters.[18]

6 Conclusions

This paper examines nowcasts of tail risk to GDP growth, with a potentially wide array of monthly and weekly information used to produce nowcasts on a weekly basis. We consider different models, consisting of Bayesian mixed frequency regressions with stochastic volatility, Bayesian quantile regressions, and Bayesian partial quantile regression, the last of which incorporates data reduction through a common

[18]Without the outlier state included in the model, the tail quantile nowcasts from the SV specification were considerably more negative, reflecting a substantial widening of the predictive distribution stemming from a large increase in the estimated level of volatility that occurred in the prior two quarters, following the pandemic's outbreak. Compared to its level two years earlier, the volatility $\lambda_t^{0.5}$ rose by a multiple of 4 to 5 in the end-of-sample estimates for 2020:Q3 and 2020:Q4 nowcasts. Having the outlier state in the model mitigates this rise in the persistent volatility state.

factor. Our analysis includes a case study of nowcasting in the pandemic year of 2020.

Our results show that, within some limits, more information helps the accuracy of tail risk forecasts. Tail forecast accuracy generally improves as additional observations on the quarter become available across weeks. In addition, extending the base macro variable set to include financial indicators improves accuracy. Adding just the small or large sets of weekly activity indicators to the base macro variable set does not help accuracy, but as long as financial indicators are in the model, adding weekly activity indicators does not harm accuracy. As to model or approach choice, the three specifications we consider perform comparably, offering solid gains in forecast accuracy (relative to a baseline of an AR model with stochastic volatility), with benefits more or less maximized when financial indicators are included in the model. An equally-weighted average of the forecasts from our three model formulations estimated with a base set of macro and financial variables can be seen as a robust approach to improving tail nowcast accuracy. The role of timely information was also evident in nowcasts made 2020: For example, given the sudden shift in activity that occurred late in 2020:Q1 following the pandemic's outbreak, it was only late in the quarter that the models had enough information showing a downturn to be able to project a significant fall in GDP.

References

[1] Aastveit, K. A., Foroni, C., & Ravazzolo, F. (2017). Density forecasts with Midas models. *Journal of Applied Econometrics* 32, 783-801. `https://doi.org/10.1002/jae.2545`

[2] Aastveit, K. A., Gerdrup, K. R., Jore A. S., & Thorsrud, L. A. (2014). Nowcasting GDP in real time: a density combination approach. *Journal of Business and Economic Statistics* 32, 48-68. `https://doi.org/10.1080/07350015.2013.844155`

[3] Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review* 109, 1263-1289. `https://doi.org/10.1257/aer.20161923`

[4] Adrian, T., Grinberg, F., Liang, N., Malik, S., & Yu, J. (2020). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics*, forthcoming. `https://doi.org/`

[5] Andrews, D. W. K., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953-966. `https://doi.org/10.2307/2951574`

[6] Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business and Economic Statistics* 31, 240-251. `https://doi.org/10.1080/07350015.2013.767199`

[7] Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business and Economic Statistics* 27, 417-427. https://doi.org/10.1198/jbes.2009.07205

[8] Banbura, M., Giannone, D., & Reichlin, L. (2011). Nowcasting. In M. P. Clements & D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting* (pp. 63-90). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0008

[9] Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Nowcasting and the real time data flow. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2* (pp. 195-237). Amsterdam: North Holland. https://doi.org/10.1016/B978-0-444-53683-9.00004-9

[10] Carriero, A., Clark, T. E., & Marcellino, M. (2015). Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A* 178, 837-862. https://doi.org/10.1111/rssa.12092

[11] Carriero, A., Clark, T. E., & Marcellino, M. (2020a). Capturing macroeconomic tail risks with Bayesian vector autoregressions. Federal Reserve Bank of Cleveland, Working Paper No. 20-02. https://doi.org/10.26509/frbc-wp-202002

[12] Carriero, A., Clark, T. E., & Marcellino, M. (2020b). Nowcasting tail risks to economic activity with many indicators. Federal Reserve Bank of Cleveland, Working Paper No. 20-13R2. https://doi.org/10.26509/frbc-wp-202013r2

[13] Carriero, A., Clark, T. E., Marcellino, M., & Mertens, E. (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

[14] Chavleishvili, S., & Manganelli, S. (2019). Forecasting and stress testing with quantile vector autoregression. ECB Working Paper No. 2330.

[15] Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* 39, 841-862. https://doi.org/10.2307/2527341

[16] Clark, T. E. (2011). Real-time density forecasts from BVARs with stochastic volatility. *Journal of Business and Economic Statistics* 29, 327-341. https://doi.org/10.1198/jbes.2010.09248

[17] Clark, T. E., & Ravazzolo, F. (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] Cook, T., & Doh, T. (2019). Assessing macroeconomic tail risks in a data-rich environment. Federal Reserve Bank of Kansas City, Research Working Paper no. 19-12. https://doi.org/10.18651/RWP2019-12

[19] D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28, 82-101. https://doi.org/10.1002/jae.1257

[20] De Nicolo, G., & Lucchetta, M. (2017). Forecasting tail risks. *Journal of Applied Econometrics* 32, 159-170. https://doi.org/10.1002/jae.2509

[21] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253-263. https://doi.org/10.2307/1392185

[22] Faust, J., & Wright, J. H. (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics* 27, 468-479. `https://doi.org/10.1198/jbes.2009.07214`

[23] Faust, J., & Wright, J. H. (2013). Forecasting inflation. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2* (pp. 2-56). Amsterdam: North Holland. `https://doi.org/10.1016/B978-0-444-53683-9.00001-3`

[24] Ferrara, L., Mogliani, M., & Sahuc, J-G. (2021). High-frequency monitoring of growth-at-risk. *International Journal of Forecasting*, forthcoming. `https://doi.org/10.1016/j.ijforecast.2021.06.010`

[25] Ferrara, L., & Simoni, A. (2019). When are Google data useful to nowcast GDP? An approach via pre-selection and shrinkage. Banque de France, Working Paper No. 717.

[26] Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics* 44, 1680-1707. `https://doi.org/10.1214/16-AOS1439`

[27] Foroni, C., Ghysels, E., & Marcellino, M. (2013). Mixed-frequency vector autoregressive models. In T. Fomby, L. Kilian, & A. Murphy (Eds.), *Advances in Econometrics: VAR Models in Macroeconomics, Financial Econometrics, and Forecasting — New Developments and Applications* (pp. 247-272). `https://doi.org/10.1108/S0731-9053(2013)0000031007`

[28] Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics* 23, 416-431. `https://doi.org/10.1198/073500105000000018`

[29] Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting GDP and inflation: The real-time informational content of macroeconomic data releases. *Journal of Monetary Economics* 55, 665-676. `https://doi.org/10.1016/j.jmoneco.2008.05.010`

[30] Giglio, S., Kelly, B., & Pruitt, S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics* 119, 457-471. `https://doi.org/10.1016/j.jfineco.2016.01.010`

[31] Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102, 359-378. `https://doi.org/10.1198/016214506000001437`

[32] Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business and Economic Statistics* 29, 411-422. `https://doi.org/10.1198/jbes.2010.08110`

[33] Jordan, A., Krueger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software* 90, 1-37. `https://doi.org/10.18637/jss.v090.i12`

[34] Khare, K., & Hobert, J. P. (2012). Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression. *Journal of Multivariate Analysis* 112, 108-116. `https://doi.org/10.1016/j.jmva.2012.05.004`

[35] Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361-393. `https://doi.org/10.1111/1467-937X.00050`

[36] Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting* 33, 11-20. https://doi.org/10.1016/j.ijforecast.2016.07.005

[37] Kozicki, S., & Hoffman, B. (2004). Rounding error: A distorting influence on index data. *Journal of Money, Credit and Banking* 36, 319-338. https://www.jstor.org/stable/383897

[38] Lewis, D., Mertens, K., Stock, J., & Trivedi, M. (2021). Measuring real activity using a weekly economic index. *Journal of Applied Econometrics*, forthcoming.

[39] McCracken, M. W., Owyang, M. T., & Sekhposyan, T. (2021). Real-time forecasting and scenario analysis using a large mixed frequency Bayesian VAR. *International Journal of Central Banking* 18, 327-367.

[40] Mitchell, J., Poon, A., & Mazzi , G. L. (2020). Nowcasting euro area GDP growth using Bayesian quantile regression. Forthcoming, *Advances in Econometrics*

[41] Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics* 211, 388-413. https://doi.org/10.1016/j.jeconom.2018.10.008

[42] Plagborg-Moller, M., Reichlin, L., Ricco, G., & Hasenzagl, T. (2020). When is growth at risk? *Brookings Papers on Economic Activity*, Spring, 167-220.

[43] Reichlin, L., Ricco, G., & Hasenzagl, T. (2020). Financial variables as predictors of real growth vulnerability. Deutsche Bundesbank, Discussion Paper No. 05/2020.

[44] Schorfheide, F., & Song, D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business and Economic Statistics* 33, 366-380. https://doi.org/10.1080/07350015.2014.954707

[45] Schorfheide, F., & Song, D. (2020). Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic. Manuscript, University of Pennsylvania.

[46] Stock, J. H., & Watson, M. W. (2016). Core inflation and trend inflation. *Review of Economics and Statistics* 98, 770-784. https://doi.org/10.1162/REST_a_00608

[47] Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business and Economic Statistics* 37, 121-133. https://doi.org/10.1080/07350015.2017.1281815

[48] Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting* 36, 428-441. https://doi.org/10.1016/j.ijforecast.2019.05.014

[49] Wallis, K. F. (1986). Forecasting with an econometric model: The 'ragged edge' problem. *Journal of Forecasting* 5, 1-13. https://doi.org/10.1002/for.3980050102

[50] West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067-1084. https://doi.org/10.2307/2171956

[51] Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* 54, 437-447. https://doi.org/10.1016/S0167-7152(01)00124-9

Table 1: **Variables used**

indicator	*mnemonic (transformation)*	*frequency (release timing/delay)*
real GDP	GDP (400Δ ln)	quarterly (4)
payroll employment	emp (Δ ln)	monthly (1)
ISM purchasing managers index, manufacturing	ISM	monthly (1)
retail sales (nominal/CPI)	retail (Δ ln)	monthly (2)
industrial production	IP (Δ ln)	monthly (3)
housing starts	starts (ln)	monthly (3)
initial claims for unemployment insurance	claims	monthly (1), weekly (1)
continuing claims for unemployment insurance	cclaims	monthly (1), weekly (2)
Chicago Fed index of financial conditions	NFCI	monthly (1), weekly (1)
S&P index of stock prices	SP (Δ ln)	monthly (1), weekly (0)
term spread: 10-year less 1-year Treasury rates	TS	monthly (1), weekly (0)
credit spread: Moody's Baa yield less 10-year Treasury	CS	monthly (1), weekly (0)
Bloomberg index of consumer comfort	sment	weekly (1)
raw steel production	steel (Δ ln, 52 week)	weekly (1)
electric utility output	util (Δ ln, 52 week)	weekly (1)
loadings of railroad cars	loads (Δ ln, 52 week)	weekly (1)
fuel sales	fuel (Δ ln, 52 week)	weekly (1)
Redbook same-store retail sales	rbook (%Δ, 52 week)	weekly (1)

Notes: The first column lists the variables included in our models. The second column gives the indicator names used, along with any transformations made of the data. Note that because Redbook sales are reported as a 52-week percent change, for this indicator we used the simple percent change rather than the log growth rate applied to other trending variables. The third column indicates the frequency of the underlying data used and information on release timing assumptions (our dating is based on typical publication schedules and end-of-week availability), which determine which variables enter our models at each forecast origin. For quarterly and monthly indicators, the entries in parentheses give the week in a given month in which an estimate for the prior quarter or month is initially released. As examples, GDP for quarter $t-1$ is typically reported in week 4 of month 1 of quarter t, employment for month $t-1$ is normally published in week 1 of month t, and the average credit spread for the month of $t-1$ is available by the end of week 1 of month t. For indicators available at the weekly frequency, the entries in parentheses give the delay (in weeks) of publication or availability of the release for week t of the month. For example, Treasury yields and stock prices for week t are available at the end of week t, so the delay is 0. As indicated, for most of the other weekly indicators, the publication lag is 1 week.

Table 2: **Specifications of BMF models of GDP growth**

week of qrtr. (forecast origin)	*variables (in addition to constant)*
1 (qrtr. t)	GDP_{t-2}, $\text{emp}_{t-1}^{(m1,m2,m3)}$, $\text{ISM}_{t-1}^{(m1,m2,m3)}$, $\text{retail}_{t-1}^{(m1,m2)}$, $\text{IP}_{t-1}^{(m1,m2)}$, $\text{starts}_{t-1}^{(m1,m2)}$, $\text{claims}_{t-1}^{(m1,m2,m3)}$ $\text{NFCI}_{t-1}^{(m1,m2,m3)}$, $\text{SP}_{t-1}^{(m1,m2,m3)}$, $\text{TS}_{t-1}^{(m1,m2,m3)}$, $\text{CS}_{t-1}^{(m1,m2,m3)}$, $\text{SP}_{t}^{(w1)}$, $\text{TS}_{t}^{(w1)}$, $\text{CS}_{t}^{(w1)}$ $\text{cclaims}_{t-1}^{(m1,m2,m3)}$
2 (qrtr. t)	GDP_{t-2}, $\text{emp}_{t-1}^{(m1,m2,m3)}$, $\text{ISM}_{t-1}^{(m1,m2,m3)}$, $\text{retail}_{t-1}^{(m1,m2,m3)}$, $\text{IP}_{t-1}^{(m1,m2)}$, $\text{starts}_{t-1}^{(m1,m2)}$, $\text{claims}_{t-1}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(w1)}$ $\text{NFCI}_{t-1}^{(m1,m2,m3)}$, $\text{SP}_{t-1}^{(m1,m2,m3)}$, $\text{TS}_{t-1}^{(m1,m2,m3)}$, $\text{CS}_{t-1}^{(m1,m2,m3)}$, $\text{NFCI}_{t}^{(w1)}$, $\text{SP}_{t}^{(w1+w2)}$, $\text{TS}_{t}^{(w1+w2)}$, $\text{CS}_{t}^{(w1+w2)}$ $\text{cclaims}_{t-1}^{(m1,m2,m3)}$, $\text{sment}_{t}^{(w1)}$, $\text{steel}_{t}^{(w1)}$, $\text{util}_{t}^{(w1)}$
3 (qrtr. t)	GDP_{t-2}, $\text{emp}_{t-1}^{(m1,m2,m3)}$, $\text{ISM}_{t-1}^{(m1,m2,m3)}$, $\text{retail}_{t-1}^{(m1,m2,m3)}$, $\text{IP}_{t-1}^{(m1,m2,m3)}$, $\text{starts}_{t-1}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(w1+w2)}$ $\text{NFCI}_{t}^{(w1+w2)}$, $\text{SP}_{t}^{(w1+w2+w3)}$, $\text{TS}_{t}^{(w1+w2+w3)}$, $\text{CS}_{t}^{(w1+w2+w3)}$ $\text{cclaims}_{t}^{(w1)}$, $\text{sment}_{t}^{(w2)}$, $\text{steel}_{t}^{(w2)}$, $\text{util}_{t}^{(w2)}$
4 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t-1}^{(m1,m2,m3)}$, $\text{ISM}_{t-1}^{(m1,m2,m3)}$, $\text{retail}_{t-1}^{(m1,m2,m3)}$, $\text{IP}_{t-1}^{(m1,m2,m3)}$, $\text{starts}_{t-1}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(w1+w2+w3)}$ $\text{NFCI}_{t}^{(w1+w2+w3)}$, $\text{SP}_{t}^{(w1+w2+w3+w4)}$, $\text{TS}_{t}^{(w1+w2+w3+w4)}$, $\text{CS}_{t}^{(w1+w2+w3+w4)}$ $\text{cclaims}_{t}^{(w1+w2)}$, $\text{sment}_{t}^{(w3)}$, $\text{steel}_{t}^{(w3)}$, $\text{util}_{t}^{(w3)}$
5 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1)}$, $\text{ISM}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1)}$ $\text{NFCI}_{t}^{(m1)}$, $\text{SP}_{t}^{(m1,w5)}$, $\text{TS}_{t}^{(m1,w5)}$, $\text{CS}_{t}^{(m1,w5)}$ $\text{cclaims}_{t}^{(m1)}$, $\text{sment}_{t}^{(w4)}$, $\text{steel}_{t}^{(w4)}$, $\text{util}_{t}^{(w4)}$
6 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1)}$, $\text{ISM}_{t}^{(m1)}$, $\text{retail}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,w5)}$ $\text{NFCI}_{t}^{(m1,w5)}$, $\text{SP}_{t}^{(m1,w5+w6)}$, $\text{TS}_{t}^{(m1,w5+w6)}$, $\text{CS}_{t}^{(m1,w5+w6)}$ $\text{cclaims}_{t}^{(m1)}$, $\text{sment}_{t}^{(w5)}$, $\text{steel}_{t}^{(w5)}$, $\text{util}_{t}^{(w5)}$
7 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1)}$, $\text{ISM}_{t}^{(m1)}$, $\text{retail}_{t}^{(m1)}$, $\text{IP}_{t}^{(m1)}$, $\text{starts}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,w5+w6)}$ $\text{NFCI}_{t}^{(m1,w5+w6)}$, $\text{SP}_{t}^{(m1,w5+w6+w7)}$, $\text{TS}_{t}^{(m1,w5+w6+w7)}$, $\text{CS}_{t}^{(m1,w5+w6+w7)}$ $\text{cclaims}_{t}^{(m1,w5)}$, $\text{sment}_{t}^{(w6)}$, $\text{steel}_{t}^{(w6)}$, $\text{util}_{t}^{(w6)}$
8 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1)}$, $\text{ISM}_{t}^{(m1)}$, $\text{retail}_{t}^{(m1)}$, $\text{IP}_{t}^{(m1)}$, $\text{starts}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,w5+w6+w7)}$ $\text{NFCI}_{t}^{(m1,w5+w6+w7)}$, $\text{SP}_{t}^{(m1,w5+w6+w7+w8)}$, $\text{TS}_{t}^{(m1,w5+w6+w7+w8)}$, $\text{CS}_{t}^{(m1,w5+w6+w7+w8)}$ $\text{cclaims}_{t}^{(m1,w5+w6)}$, $\text{sment}_{t}^{(w7)}$, $\text{steel}_{t}^{(w7)}$, $\text{util}_{t}^{(w7)}$
9 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2)}$, $\text{ISM}_{t}^{(m1,m2)}$, $\text{retail}_{t}^{(m1)}$, $\text{IP}_{t}^{(m1)}$, $\text{starts}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,m2)}$ $\text{NFCI}_{t}^{(m1,m2)}$, $\text{SP}_{t}^{(m1,m2,w9)}$, $\text{TS}_{t}^{(m1,m2,w9)}$, $\text{CS}_{t}^{(m1,m2,w9)}$ $\text{cclaims}_{t}^{(m1,m2)}$, $\text{sment}_{t}^{(w8)}$, $\text{steel}_{t}^{(w8)}$, $\text{util}_{t}^{(w8)}$
10 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2)}$, $\text{ISM}_{t}^{(m1,m2)}$, $\text{retail}_{t}^{(m1,m2)}$, $\text{IP}_{t}^{(m1)}$, $\text{starts}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,m2,w9)}$ $\text{NFCI}_{t}^{(m1,m2,w9)}$, $\text{SP}_{t}^{(m1,m2,w9+w10)}$, $\text{TS}_{t}^{(m1,m2,w9+w10)}$, $\text{CS}_{t}^{(m1,m2,w9+w10)}$ $\text{cclaims}_{t}^{(m1,m2)}$, $\text{sment}_{t}^{(w9)}$, $\text{steel}_{t}^{(w9)}$, $\text{util}_{t}^{(w9)}$
11 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2)}$, $\text{ISM}_{t}^{(m1,m2)}$, $\text{retail}_{t}^{(m1,m2)}$, $\text{IP}_{t}^{(m1,m2)}$, $\text{starts}_{t}^{(m1,m2)}$, $\text{claims}_{t}^{(m1,m2,w9+w10)}$ $\text{NFCI}_{t}^{(m1,m2,w9+w10)}$, $\text{SP}_{t}^{(m1,m2,w9+w10+w11)}$, $\text{TS}_{t}^{(m1,m2,w9+w10+w11)}$, $\text{CS}_{t}^{(m1,m2,w9+w10+w11)}$ $\text{cclaims}_{t}^{(m1,m2,w9)}$, $\text{sment}_{t}^{(w10)}$, $\text{steel}_{t}^{(w10)}$, $\text{util}_{t}^{(w10)}$
12 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2)}$, $\text{ISM}_{t}^{(m1,m2)}$, $\text{retail}_{t}^{(m1,m2)}$, $\text{IP}_{t}^{(m1,m2)}$, $\text{starts}_{t}^{(m1,m2)}$, $\text{claims}_{t}^{(m1,m2,w9+w10+w11)}$ $\text{NFCI}_{t}^{(m1,m2,w9+w10+w11)}$, $\text{SP}_{t}^{(m1,m2,w9+w10+w11+w12)}$, $\text{TS}_{t}^{(m1,m2,w9+w10+w11+w12)}$, $\text{CS}_{t}^{(m1,m2,w9+w10+w11+w12)}$ $\text{cclaims}_{t}^{(m1,m2,w9+w10)}$, $\text{sment}_{t}^{(w11)}$, $\text{steel}_{t}^{(w11)}$, $\text{util}_{t}^{(w11)}$
13 (qrtr. $t+1$)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2,m3)}$, $\text{ISM}_{t}^{(m1,m2,m3)}$, $\text{retail}_{t}^{(m1,m2)}$, $\text{IP}_{t}^{(m1,m2)}$, $\text{starts}_{t}^{(m1,m2)}$, $\text{claims}_{t}^{(m1,m2,m3)}$ $\text{NFCI}_{t}^{(m1,m2,m3)}$, $\text{SP}_{t}^{(m1,m2,m3,w13)}$, $\text{TS}_{t}^{(m1,m2,m3,w13)}$, $\text{CS}_{t}^{(m1,m2,m3,w13)}$ $\text{cclaims}_{t}^{(m1,m2,m3)}$, $\text{sment}_{t}^{(w12)}$, $\text{steel}_{t}^{(w12)}$, $\text{util}_{t}^{(w12)}$
14 (qrtr. $t+1$)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2,m3)}$, $\text{ISM}_{t}^{(m1,m2,m3)}$, $\text{retail}_{t}^{(m1,m2,m3)}$, $\text{IP}_{t}^{(m1,m2)}$, $\text{starts}_{t}^{(m1,m2)}$, $\text{claims}_{t}^{(m1,m2,m3,w13)}$ $\text{NFCI}_{t}^{(m1,m2,m3,w13)}$, $\text{SP}_{t}^{(m1,m2,m3,w13+w14)}$, $\text{TS}_{t}^{(m1,m2,m3,w13+w14)}$, $\text{CS}_{t}^{(m1,m2,m3,w13+w14)}$ $\text{cclaims}_{t}^{(m1,m2,m3)}$, $\text{sment}_{t}^{(w13)}$, $\text{steel}_{t}^{(w13)}$, $\text{util}_{t}^{(w13)}$
15 (qrtr. $t+1$)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2,m3)}$, $\text{ISM}_{t}^{(m1,m2,m3)}$, $\text{retail}_{t}^{(m1,m2,m3)}$, $\text{IP}_{t}^{(m1,m2,m3)}$, $\text{starts}_{t}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(m1,m2,m3,w13+w14)}$ $\text{NFCI}_{t}^{(m1,m2,m3,w13+w14)}$, $\text{SP}_{t}^{(m1,m2,m3,w13+w14+w15)}$, $\text{TS}_{t}^{(m1,m2,m3,w13+w14+w15)}$, $\text{CS}_{t}^{(m1,m2,m3,w13+w14+w15)}$ $\text{cclaims}_{t}^{(m1,m2,m3,w13)}$, $\text{sment}_{t}^{(w14)}$, $\text{steel}_{t}^{(w14)}$, $\text{util}_{t}^{(w14)}$

Notes: For each week indicated in the first column, the table has three rows of entries, with the first listing the relevant base M indicators, the second row covering the finance indicators, and the third listing the small weekly indicators included in the given week's models. The variable sets *base M*, *base M-F*, and *base M-F + small weekly* combine these predictors as indicated.

Table 3: **Out-of-sample forecast accuracy: CRPS**

variable set	model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
		1985:Q1-2019:Q4							
	AR-SV	1.12	1.12	1.06	1.06	1.06	1.06	1.06	1.06
base M	BMF-SV	0.91 **	0.89 **	0.92 ***	0.86 ***	0.83 ***	0.81 ***	0.77 ***	0.77 ***
	BQR	0.94	0.92	0.97	0.90 **	0.87 ***	0.84 ***	0.81 ***	0.81 ***
base M-F	BMF-SV	0.91	0.88 **	0.90 **	0.85 **	0.82 **	0.81 ***	0.78 ***	0.78 ***
	BQR	0.93	0.89 *	0.94	0.87 **	0.84 **	0.82 **	0.80 ***	0.79 ***
	avg.	0.91	0.88 *	0.91 *	0.86 **	0.83 **	0.82 **	0.80 ***	0.80 ***
		2000:Q1-2019:Q4							
	AR-SV	1.19	1.19	1.11	1.11	1.09	1.09	1.10	1.10
base M	BMF-SV	0.88 **	0.88 *	0.92 ***	0.87 **	0.84 **	0.84 **	0.79 **	0.78 **
	BQR	0.87 **	0.86 *	0.92 **	0.85 ***	0.83 **	0.82 **	0.78 **	0.77 **
base M-F	BMF-SV	0.85 *	0.85 *	0.86 **	0.83 *	0.80 *	0.81 *	0.77 **	0.77 **
	BQR	0.85 *	0.83 *	0.88 *	0.82 *	0.80 *	0.80 *	0.77 **	0.76 **
	avg.	0.83 **	0.84 *	0.86 *	0.84 *	0.81 *	0.82 *	0.78 **	0.78 **
base M + small weekly	BMF-SV	0.94 *	0.95	0.95 **	0.91 **	0.89 **	0.88 *	0.82 **	0.79 **
	BQR	0.94 *	0.95	0.95 **	0.91 **	0.89 **	0.89 *	0.84 **	0.80 **
base M-F + small weekly	BMF-SV	0.93	0.92	0.90 *	0.88 *	0.85 *	0.84 *	0.79 **	0.78 **
	BQR	0.92	0.91	0.89 *	0.87 *	0.84 *	0.84 *	0.79 **	0.79 **
		2007:Q1-2019:Q4							
	AR-SV	1.26	1.26	1.15	1.15	1.15	1.14	1.16	1.15
base M	BMF-SV	0.88 **	0.85 *	0.91 **	0.83 **	0.81 **	0.81 *	0.76 **	0.74 **
	BQR	0.86 *	0.84	0.91 *	0.81 **	0.80 **	0.80 *	0.75 **	0.74 **
base M-F	BMF-SV	0.83	0.81 *	0.84 *	0.77 *	0.74 *	0.75 *	0.71 **	0.71 *
	BQR	0.82	0.78 *	0.86	0.75 *	0.74 *	0.74 *	0.71 **	0.71 **
	avg.	0.80 *	0.80	0.84	0.78 *	0.75 *	0.77	0.73 *	0.73 *
base M + small weekly	BMF-SV	0.92 *	0.91	0.92 **	0.87 **	0.86 *	0.85	0.78 **	0.73 **
	BQR	0.94	0.93	0.93 **	0.88 *	0.87 *	0.87	0.80 *	0.76 *
base M-F + small weekly	BMF-SV	0.90	0.88	0.86 *	0.83	0.81	0.81	0.75 *	0.74 *
	BQR	0.91	0.88	0.85 *	0.82	0.81	0.82	0.76 *	0.75 *
base M + large weekly	BMF-SV	0.97	0.96	0.91 **	0.92	0.90	0.92	0.88	0.84
	BQR	0.99	0.98	0.92 **	0.92	0.91	0.93	0.88	0.86
base M-F + large weekly	BMF-SV	0.90	0.92	0.85 **	0.86	0.82 *	0.84	0.79 *	0.78 *
	BQR	0.91	0.93	0.85 **	0.85	0.83	0.85	0.81 *	0.80

Notes: The weeks indicated in the columns refer to the weeks of forecast origins for the quarter (omitting even-numbered weeks to reduce the size of the table). In each panel, the top row gives the CRPS from the benchmark model and variable set, and other rows report the ratio of CRPS for the indicated variable set and model to the benchmark (lower is better), for the indicated sample. Statistical significance of differences in quantile scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test, conducted on a one-sided basis, such that the alternative hypothesis is that the indicated forecast is more accurate than the benchmark.

Table 4: **Out-of-sample forecast accuracy: 10% quantile score**

variable set	model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
		1985:Q1-2019:Q4							
	AR-SV	0.41	0.41	0.36	0.36	0.35	0.35	0.35	0.35
base M	BMF-SV	0.91	0.85	0.90 ***	0.81 **	0.81 **	0.80 **	0.77 **	0.74 **
	BQR	1.08	1.01	0.99	0.96	0.96	0.96	0.92	0.90
base M-F	BMF-SV	0.78 *	0.78 *	0.85 *	0.76 **	0.75 **	0.76 **	0.73 **	0.73 **
	BQR	0.82	0.77	0.85	0.74 *	0.77 *	0.77 *	0.77 *	0.77 *
	avg.	0.75 *	0.74	0.76 *	0.72 **	0.74 **	0.75 *	0.74 **	0.73 **
		2000:Q1-2019:Q4							
	AR-SV	0.45	0.45	0.40	0.40	0.38	0.37	0.37	0.38
base M	BMF-SV	0.92	0.83	0.90 **	0.82 *	0.82	0.83	0.80	0.76
	BQR	0.95	0.87	0.83 **	0.81	0.84	0.84	0.81	0.77
base M-F	BMF-SV	0.78	0.75	0.81 *	0.73	0.74	0.75	0.74	0.74
	BQR	0.80	0.71	0.80	0.69	0.72	0.74	0.74	0.73
	avg.	0.73	0.70	0.73	0.68 *	0.71	0.74	0.73	0.72
base M + small weekly	BMF-SV	0.95	0.91	0.95	0.86	0.86	0.84	0.79	0.71 *
	BQR	0.88	0.91	0.93 *	0.83 *	0.83	0.84	0.79	0.73 *
base M-F + small weekly	BMF-SV	0.82	0.81	0.84	0.78	0.80	0.77	0.74	0.73
	BQR	0.77	0.80	0.81	0.75	0.79	0.77	0.75	0.73
		2007:Q1-2019:Q4							
	AR-SV	0.53	0.53	0.46	0.46	0.43	0.43	0.43	0.43
base M	BMF-SV	0.92	0.81	0.86 ***	0.77	0.76	0.76	0.73	0.69
	BQR	0.94	0.84	0.76 **	0.77	0.78	0.78	0.74	0.71
base M-F	BMF-SV	0.75	0.71	0.76	0.67	0.66	0.67	0.66	0.68
	BQR	0.71	0.63	0.75	0.61	0.64	0.64	0.66	0.67
	avg.	0.68	0.65	0.66	0.62	0.62	0.65	0.65	0.66
base M + small weekly	BMF-SV	0.94	0.88	0.90 **	0.80 *	0.78	0.77	0.72	0.66 *
	BQR	0.90	0.90	0.86 ***	0.75 *	0.76 *	0.77	0.72 *	0.68 *
base M-F + small weekly	BMF-SV	0.78	0.75	0.75	0.70	0.73	0.71	0.69	0.70
	BQR	0.74	0.76	0.73 *	0.66	0.72	0.70	0.69	0.71
base M + large weekly	BMF-SV	0.97	0.90	0.92	0.83	0.83	0.83	0.80	0.76
	BQR	0.99	0.99	0.92	0.82	0.81	0.84	0.81	0.76
base M-F + large weekly	BMF-SV	0.79	0.80	0.79	0.72	0.73	0.73	0.73	0.71
	BQR	0.78	0.84	0.77 *	0.70	0.73	0.74	0.75	0.73

Notes: The weeks indicated in the columns refer to the weeks of forecast origins for the quarter (omitting even-numbered weeks to reduce the size of the table). In each panel, the top row gives the 10% quantile scores (QS) from the benchmark model and variable set, and other rows report the ratio of QS for the indicated variable set and model to the benchmark (lower is better), for the indicated sample. Statistical significance of differences in quantile scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test, conducted on a one-sided basis, such that the alternative hypothesis is that the indicated forecast is more accurate than the benchmark.

Table 5: **Out-of-sample forecast accuracy: 10% VaR-ES score**

variable set	model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
		1985:Q1-2019:Q4							
	AR-SV	3.61	3.57	3.12	3.11	3.02	2.99	3.04	3.03
base M	BMF-SV	0.41	0.65	0.37 ***	0.64 ***	0.64 **	0.64 **	0.74 ***	0.82 ***
	BQR	-0.15	0.08	-0.02	0.17	0.14	0.12	0.24	0.28
base M-F	BMF-SV	0.89 *	0.86 *	0.51 **	0.82 ***	0.78 **	0.72 **	0.83 ***	0.83 ***
	BQR	0.58	0.72	0.36	0.79 **	0.63 *	0.61 **	0.67 **	0.68 **
	avg.	0.99 *	0.94 *	0.72 **	0.96 ***	0.81 **	0.74 **	0.82 ***	0.85 ***
		2000:Q1-2019:Q4							
	AR-SV	4.03	3.97	3.46	3.43	3.27	3.23	3.27	3.27
base M	BMF-SV	0.52	0.85	0.44 ***	0.68 **	0.62 *	0.53	0.60	0.73 *
	BQR	0.41	0.66	0.62 **	0.76 *	0.61 *	0.52	0.64	0.71 *
base M-F	BMF-SV	0.98	1.05	0.69 *	0.94 *	0.83 *	0.68	0.75 *	0.72 *
	BQR	0.67	0.94	0.50	0.96 *	0.71	0.61	0.72 *	0.72
	avg.	1.17	1.18	0.84 *	1.15 **	0.88 *	0.70 *	0.83 *	0.84 *
base M + small weekly	BMF-SV	0.36	0.56	0.28 *	0.55 *	0.51 *	0.52	0.68 *	0.88 **
	BQR	0.70 *	0.50	0.29	0.60 *	0.62 *	0.52	0.69 *	0.83 *
base M-F + small weekly	BMF-SV	0.83	0.92	0.60	0.78 *	0.62	0.66 *	0.74 *	0.72 *
	BQR	1.08	0.93	0.69 *	0.85 *	0.64	0.62	0.69	0.71
		2007:Q1-2019:Q4							
	AR-SV	5.02	4.93	4.24	4.19	4.01	3.95	4.02	4.02
base M	BMF-SV	0.65	1.14	0.69 ***	0.96 *	0.95 *	0.89 *	0.99 *	1.08 *
	BQR	0.63	1.01	1.14 **	1.01 *	0.91 *	0.84	0.99 *	1.02 *
base M-F	BMF-SV	1.22	1.40	0.97 *	1.24 *	1.11 *	1.00 *	1.06 *	0.92
	BQR	1.22	1.38	0.57	1.38 *	1.09	1.07 *	0.98	0.78
	avg.	1.41	1.49	1.15 *	1.44 *	1.20 *	1.05 *	1.11 *	1.00 *
base M + small weekly	BMF-SV	0.47	0.82	0.56 ***	0.92 **	0.85 *	0.87 *	1.03 *	1.12 *
	BQR	0.62	0.62	0.66 ***	1.08 **	0.98 **	0.84 *	1.04 **	1.06 *
base M-F + small weekly	BMF-SV	1.21	1.38	1.05 *	1.18 *	0.88	0.83	0.91	0.77
	BQR	1.25	1.18	1.14 *	1.26 *	0.81	0.76	0.80	0.63
base M + large weekly	BMF-SV	0.45	0.84	0.42 *	0.77 *	0.68	0.70	0.82	0.89 *
	BQR	0.16	0.28	0.27	0.82 *	0.81 *	0.67	0.80	0.90 *
base M-F + large weekly	BMF-SV	1.20	1.24	0.94 *	1.09 *	0.85	0.82	0.80	0.82
	BQR	1.08	0.95	1.01 *	1.17 *	0.80	0.73	0.61	0.56

Notes: The weeks indicated in the columns refer to the weeks of forecast origins for the quarter (omitting even-numbered weeks to reduce the size of the table). In each panel, the top row gives the 10% VaR-ES score from the benchmark model and variable set, and other rows report differences in scores for the indicated variable set and model relative to the benchmark (higher number is better), for the indicated sample. Statistical significance of differences in scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test, conducted on a one-sided basis, such that the alternative hypothesis is that the indicated forecast is more accurate than the benchmark.

Table 6: **Out-of-sample forecast accuracy: Tests of conditional coverage**

variable set	model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
		1985:Q1-2019:Q4							
	AR-SV	0.25	0.34	0.38	0.38	0.38	0.24	0.52	0.38
base M	BMF-SV	0.08 *	0.24	0.38	0.03 **	0.00 ***	0.01 **	0.01 **	0.04 **
	BQR	0.00 ***	0.00 ***	0.01 ***	0.00 ***	0.00 ***	0.00 ***	0.00 ***	0.00 ***
base M-F	BMF-SV	0.15	0.03 **	0.84	0.29	0.25	0.25	0.25	0.08 *
	BQR	0.85	0.03 **	0.68	0.50	0.25	0.25	0.08 *	0.08 *
	avg.	0.29	0.00 ***	0.66	0.15	0.15	0.15	0.25	0.15
		2000:Q1-2019:Q4							
	AR-SV	0.38	0.38	0.66	0.66	0.66	0.66	0.66	0.66
base M	BMF-SV	0.64	0.73	0.73	0.37	0.24	0.41	0.41	0.58
	BQR	0.16	0.16	0.37	0.24	0.24	0.24	0.24	0.10
base M-F	BMF-SV	0.37	0.37	0.99	0.24	0.58	0.58	0.58	0.41
	BQR	0.99	0.64	0.68	0.52	0.58	0.66	0.41	0.41
	avg.	0.64	0.24	0.68	0.58	0.58	0.58	0.66	0.58
base M + small weekly	BMF-SV	0.38	0.66	0.73	0.99	0.99	0.66	0.58	0.66
	BQR	0.64	0.73	0.30	0.20	0.66	0.66	0.99	0.64
base M-F + small weekly	BMF-SV	0.88	0.88	0.99	0.66	0.52	0.64	0.66	0.58
	BQR	0.64	0.66	0.88	0.93	0.52	0.52	0.66	0.66
		2007:Q1-2019:Q4							
	AR-SV	0.19	0.19	0.44	0.44	0.44	0.44	0.44	0.44
base M	BMF-SV	0.80	0.44	0.47	0.30	0.24	0.52	0.52	0.75
	BQR	0.30	0.30	0.87	0.24	0.24	0.24	0.24	0.24
base M-F	BMF-SV	0.54	0.54	0.72	0.24	0.50	0.50	0.75	0.52
	BQR	0.62	0.24	0.42	0.71	0.50	0.50	0.24	0.24
	avg.	0.54	0.52	0.26	0.50	0.50	0.52	0.75	0.75
base M + small weekly	BMF-SV	0.19	0.44	0.47	0.80	0.80	0.75	0.75	0.80
	BQR	0.62	0.62	0.25	0.53	0.75	0.75	0.75	0.80
base M-F + small weekly	BMF-SV	0.80	0.80	0.87	0.75	0.67	0.67	0.80	0.80
	BQR	0.54	0.54	0.80	0.80	0.67	0.80	0.75	0.80
base M + large weekly	BMF-SV	0.19	0.44	0.47	0.87	0.75	0.75	0.80	0.75
	BQR	0.38	0.38	0.44	0.47	0.80	0.80	0.75	0.75
base M-F + large weekly	BMF-SV	0.80	0.80	0.72	0.54	0.67	0.75	0.80	0.80
	BQR	0.54	0.80	0.44	0.80	0.80	0.80	0.80	0.80

Notes: The weeks indicated in the columns refer to the weeks of forecast origins for the quarter (omitting even-numbered weeks to reduce the size of the table). Each panel reports the p-values of Christoffersen (1998) tests of conditional coverage, for the 10 percent left-tail interval. Statistical significance of rejections is indicated by *** (1%), ** (5%), or * (10%).

Table 7: **Out-of-sample forecast accuracy in 2020**

variable set	model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
		CRPS							
	AR-SV	21.07	21.09	20.69	20.72	20.62	20.59	20.57	20.56
base M	BMF-SV	0.50	0.70	0.55	0.47	0.51	0.51	0.62	0.60
	BQR	0.50	0.54	0.52	0.46	0.49	0.50	0.52	0.52
base M-F	BMF-SV	0.50	0.64	0.52	0.45	0.49	0.48	0.61	0.58
	BQR	0.53	0.53	0.51	0.47	0.50	0.49	0.55	0.54
	avg.	0.56	0.60	0.50	0.46	0.48	0.48	0.52	0.51
base M + small weekly	BMF-SV	0.58	0.69	0.63	0.55	0.60	0.61	0.68	0.73
	BQR	0.58	0.64	0.63	0.57	0.65	0.70	0.77	0.79
base M-F + small weekly	BMF-SV	0.58	0.67	0.61	0.54	0.58	0.60	0.70	0.71
	BQR	0.57	0.65	0.66	0.62	0.69	0.73	0.78	0.80
base M + large weekly	BMF-SV	0.62	0.68	0.67	0.60	0.62	0.64	0.64	0.71
	BQR	0.61	0.66	0.70	0.66	0.69	0.74	0.77	0.79
base M-F + large weekly	BMF-SV	0.60	0.68	0.66	0.62	0.64	0.66	0.74	0.76
	BQR	0.58	0.66	0.74	0.71	0.75	0.79	0.85	0.84
		10% quantile score							
	AR-SV	11.40	11.38	11.06	11.08	11.03	11.02	11.02	11.03
base M	BMF-SV	0.66	0.86	0.23	0.22	0.27	0.27	0.44	0.46
	BQR	0.54	0.36	0.27	0.26	0.22	0.22	0.11	0.14
base M-F	BMF-SV	0.66	0.79	0.23	0.23	0.24	0.22	0.40	0.37
	BQR	0.59	0.45	0.26	0.25	0.21	0.20	0.18	0.18
	avg.	0.70	0.58	0.27	0.26	0.22	0.20	0.23	0.28
base M + small weekly	BMF-SV	0.77	0.83	0.33	0.25	0.41	0.44	0.53	0.65
	BQR	0.69	0.63	0.26	0.26	0.32	0.43	0.52	0.55
base M-F + small weekly	BMF-SV	0.78	0.80	0.35	0.26	0.39	0.41	0.55	0.58
	BQR	0.73	0.67	0.27	0.26	0.42	0.46	0.53	0.52
base M + large weekly	BMF-SV	0.77	0.78	0.40	0.31	0.42	0.46	0.42	0.56
	BQR	0.70	0.65	0.26	0.26	0.33	0.45	0.46	0.49
base M-F + large weekly	BMF-SV	0.78	0.77	0.44	0.40	0.49	0.51	0.56	0.63
	BQR	0.72	0.69	0.37	0.34	0.48	0.54	0.56	0.56
		10% VaR-ES score							
	AR-SV	142.1	139.6	91.1	89.6	88.8	87.0	86.7	85.0
base M	BMF-SV	127.1	98.4	68.9	65.5	64.7	63.5	77.8	76.2
	BQR	127.3	115.8	69.8	64.4	65.8	64.5	80.6	78.7
base M-F	BMF-SV	122.7	104.9	66.4	64.0	66.3	70.1	78.3	77.0
	BQR	127.0	112.6	65.9	62.5	64.6	67.9	79.8	78.1
	avg.	113.8	110.3	65.0	60.8	64.0	66.6	79.6	77.4
base M + small weekly	BMF-SV	111.7	105.4	65.4	62.7	60.5	59.6	76.2	73.1
	BQR	117.4	113.3	64.5	60.5	58.2	54.9	75.1	73.2
base M-F + small weekly	BMF-SV	110.2	105.8	63.7	61.3	61.6	66.1	75.8	74.0
	BQR	112.4	109.5	60.7	57.6	58.4	63.6	75.0	73.4
base M + large weekly	BMF-SV	113.0	109.4	63.2	62.7	61.3	58.1	76.8	73.5
	BQR	116.5	113.2	65.2	62.3	60.3	54.4	75.8	73.7
base M-F + large weekly	BMF-SV	111.1	108.0	61.6	61.2	60.7	65.2	75.2	72.6
	BQR	113.1	109.7	60.6	58.7	59.7	62.9	74.7	72.9

Notes: The weeks indicated in the columns refer to the weeks of forecast origins for the quarter (omitting even-numbered weeks to reduce the size of the table). In the top and middle panels, the top row gives the scores from the AR-SV benchmark, and other rows report the ratio of the score for the indicated variable set and model to the benchmark (lower is better). In the bottom panel, the top row gives the VaR-ES score from the AR-SV benchmark, and other rows report the the score for the indicated variable set and model less the benchmark score (higher is better)

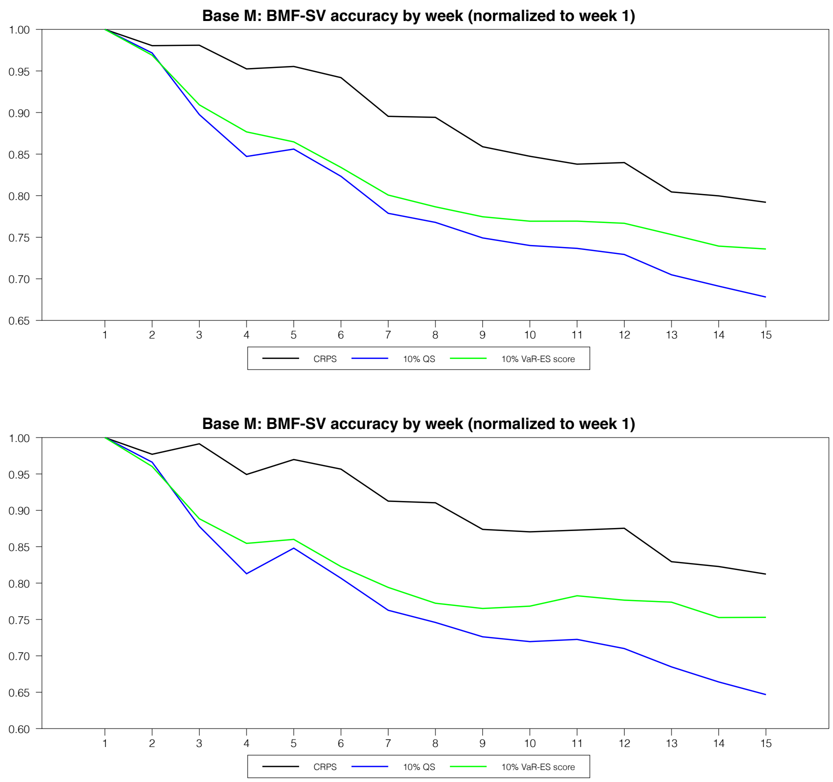


Figure 1: Results for CRPS, 10% QS, and 10% VaR-ES score across weeks 1 through 15 of forecast origins are indicated on the horizontal axis, indexed to 1.0 in week 1. The forecasts come from the BMF-SV model estimated with the base M variable set. The top and bottom panels provide results for the 1985:Q1-2019:Q4 and 2000:Q1-2019:Q4 samples, respectively.

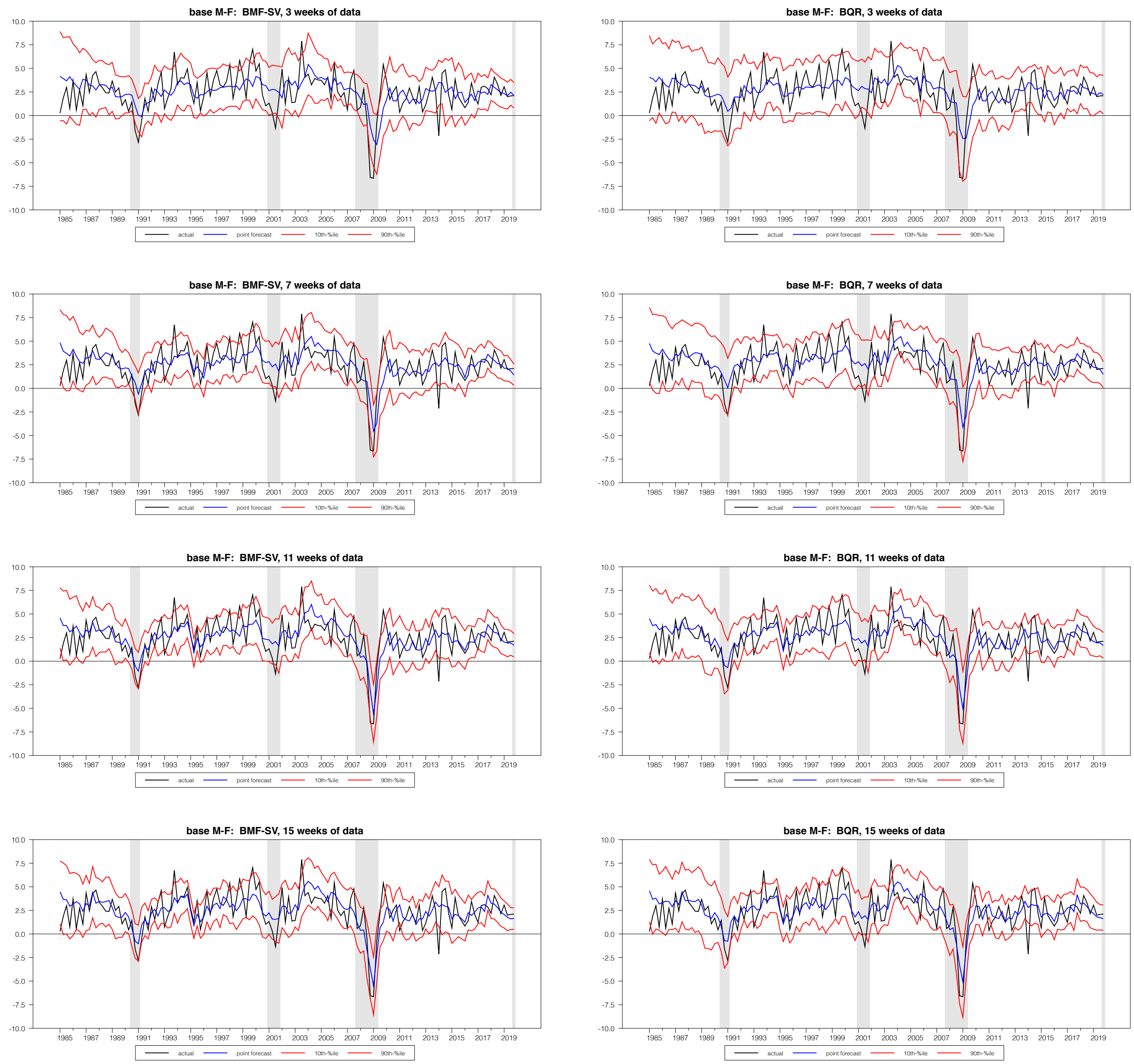
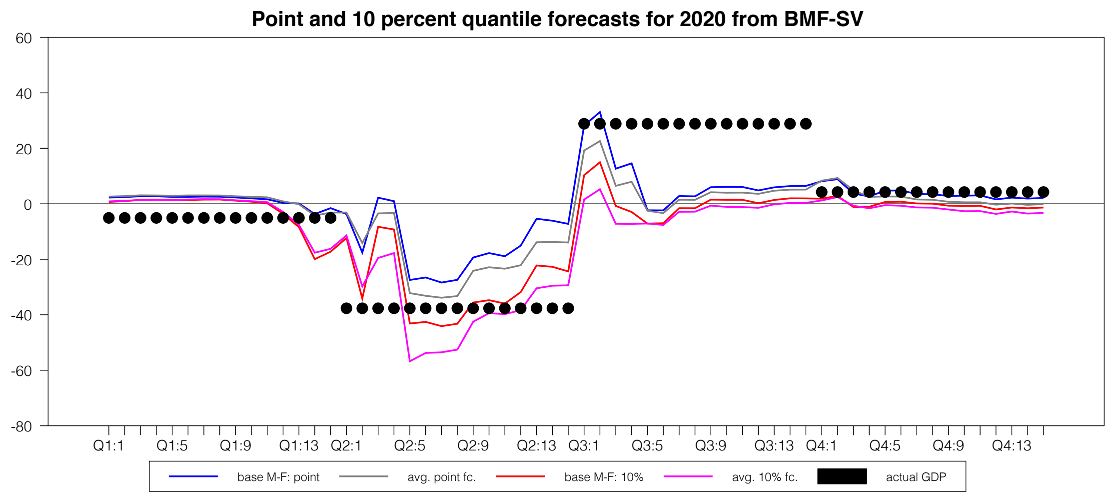


Figure 2: Out-of-sample forecasts from the base M-F variable set and BMF-SV and BQR models, selected weeks indicated in panel headers. Each panel reports actual GDP growth (black line), the point forecast (blue line), and 10%-90% forecast quantiles (red lines). Shaded regions denote NBER recessions.



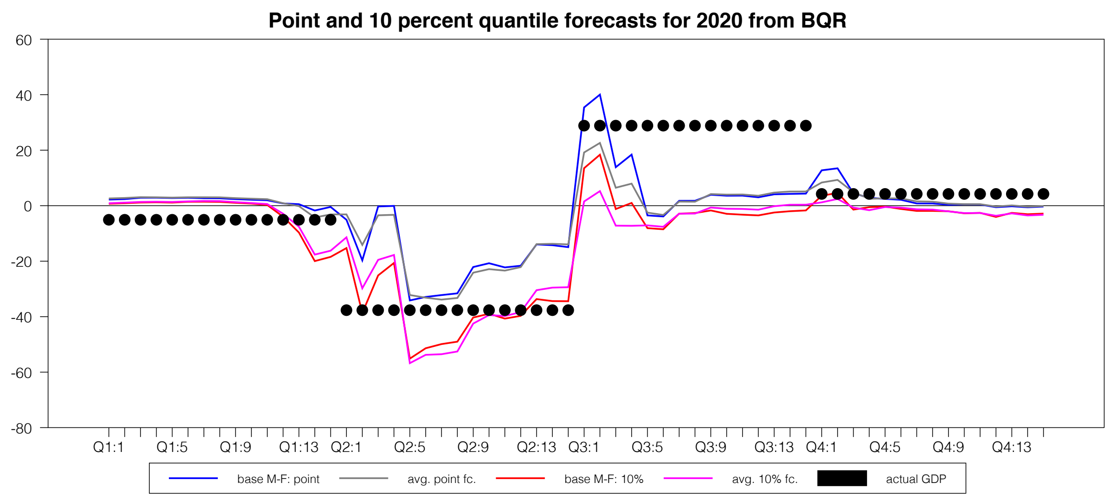


Figure 3: Real-time forecasts of GDP growth in 2020:Q1-2020:Q4 from base M-F variable set with BMF-SV and BQR models and simple average. Weeks of forecast origin are indicated on the horizontal axis, in the form Qi:j for week j of quarter i.