Nowcasting Tail Risks to Economic Activity with Many Indicators*

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: August 28, 2020

Abstract

This paper focuses on nowcasts of tail risk to GDP growth, with a potentially wide array of monthly and weekly information. We consider different models (Bayesian mixed frequency regressions with stochastic volatility, as well as classical and Bayesian quantile regressions) and also different methods for data reduction (either forecasts from models that incorporate data reduction or the combination of forecasts from smaller models). 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. The better-performing models or methods include the Bayesian regression model with stochastic volatility, Bayesian quantile regression, some approaches to data reduction that make use of factors, and forecast averaging. In contrast, simple quantile regression performs relatively poorly.

Keywords: forecasting, downside risk, pandemics, big data, 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 Knut Are Aastveit, Laurent Ferrara, Daniel Lewis, 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. Nowcasting is unique in that, to some degree, it involves "simply" adding up information in data releases for the current quarter. 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]

In 2020, the shutdown of significant portions of the economy to restrain the outbreak of the coronavirus has raised practical interest in high-frequency indicators of economic activity in the US and other economies. For example, in the US, it was clear by mid-March 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,

[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.

including — among others — initial 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, and Stock (2020) developed and began to publish regular updates of a weekly economic index formed as a principal component of 10 underlying series.

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. (2018), De Nicolo and Lucchetta (2017), and Giglio, Kelly, and Pruitt (2016) and extended to vector autoregressive models in Chavleishvili and Manganelli (2019).[2] Reichlin, Ricco, and Hasenzagl (2020) propose using leverage indicators to obtain earlier signals of economic vulnerabilities. Note also that, for output growth, nowcasting or forecasting tail risks has some precedent in the literature on forecasting recessions or just periods of negative growth (see, e.g., Aastveit, Ravazzolo, and van Dijk (2018)).

In this context, this paper assesses the ability of models to produce accurate nowcasts of tail risk to GDP with a potentially wide array of information. We consider accommodating large variable sets through a range of approaches including Bayesian shrinkage, data reduction (factor-based approaches, as well as approaches using Lasso penalties), and the combination of forecasts from smaller models. Our 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. Bayesian methods are used to estimate the model, which facilitates providing shrinkage on estimates of a model that can be quite large, conveniently generates predictive densities, and readily allows for stochastic volatility.

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

[2] Gonzalez-Rivera, Maldonado, and Ruiz (2019) use quantile regression-based methods to obtain a "growth-in-stress" index, based on stressed conditions of common factors in economic activity and their effects on growth, on a cross-country basis.

of information in the quarter, along with which indicators are most informative in that information flow. 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 risks to economic activity. Following precedents such as Adrian, Boyarchenko, and Giannone (2019) and Adrian, et al. (2018), we use the 5 percent quantile forecast as the measure of tail risk, which we evaluate with the quantile score (tick loss function). With some specifications, we also consider expected shortfall and jointly evaluate the quantile and shortfall forecasts. We consider tail risk nowcasts from not only Bayesian regressions but also several other quantile regression-based approaches making use of mixed frequency data: simple quantile regression, quantile regression with a Lasso penalty, and Bayesian quantile regression. (We have also considered quantile regression with mixed-data sampling (MIDAS) but provide these results in the appendix rather than the paper since other methods generally perform better.) We also consider models that make use of data reduction by forming factors of the data with principal components or the partial quantile regression approach of Giglio, Kelly, and Pruitt (2016).[3]

Our results show that, within some limits, more information helps the accuracy of tail risk forecasts. 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 real-time out-of-sample results, there is a benefit to adding a base 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 doesn't have much effect on forecast accuracy, either to help or harm.[4] Among the models or estimation approaches we consider, our regression with stochastic volatility and our Bayesian quantile regression perform reasonably consistently, offering solid gains in forecast accuracy (relative to a baseline quantile regression model with just a small set of macro indicators), with the greatest benefits accruing

[3] All of our nowcasting models are univariate, although based on a large information set, and therefore combined with direct forecasting. The use of multivariate models, such as vector autoregressions, combined with iterated forecasting could yield some additional gains in terms of forecasting accuracy if the models are correctly specified. Yet, the use of a large, mixed frequency information set would make estimation of a multivariate model much more complex, inefficient, and prone to mis-specification, therefore likely reducing nowcast and forecast accuracy. See Section 1.1 for some references and additional details.

[4] Admittedly, it is possible that, over time, as the time series samples of these weekly indicators grow and permit more precise model estimation, these indicators could become more helpful to forecast accuracy.

when financial indicators are included in the model. Some factor reduction methods, such as partial quantile regression, and forecast averaging also improve accuracy with some consistency. Some other methods do not perform well. In particular, tail risks obtained with simple quantile regression are typically less accurate than a Bayesian mixed frequency regression with stochastic volatility.

We conclude the paper with a report on recent — 2020:Q1 and 2020:Q2 — nowcasts from this set of specifications. In these results, the arrival of weekly information within the quarter typically lowers the tail risk nowcasts, although in 2020:Q1 this does not occur until some economic indicators are clearly showing sharp falls due to the economic shutdown that began in mid-March. For 2020:Q2, the 5 percent quantile nowcasts were sharply negative from the beginning of the forecast calendar we use, and in most cases, the nowcasts turned more negative several weeks later, in May. Since that downturn, most of the nowcasts put the 5 percent quantile at a historic -30 percent or worse. Although differences across models can be sizable, the nowcasts paint a broadly similar picture of the risks to GDP growth in these quarters. Given a variable set, the nowcasts from the Bayesian mixed frequency regressions with stochastic volatility and partial quantile regression models tend to be more similar to one another than to the nowcasts from the Bayesian quantile regression specification. Our takeaway from this illustration is similar to that of our historical forecast evaluation: additional information on the quarter as time moves forward in the quarter bears importantly on nowcasts of tail risk and their accuracy, and with sizable differences across models possible, it is helpful to consider a range of forecasts. Our starting points would be the mixed frequency regression with stochastic volatility, Bayesian quantile regression, and partial quantile regression, applied to our baseline set of macroeconomic and financial indicators. We would also consider forecasts from these same specifications but adding our small and large sets of weekly economic indicators, as well as averages of a broader set of nowcasts.

The paper proceeds as follows. The remainder of this section summarizes some other related nowcasting work. Sections 2 through 4 detail the data (including the release calendar setup), models, and forecast metrics, respectively. Section 5 provides our empirical results. Section 6 concludes. A supplemental appendix provides some additional empirical results.

1.1 Relationship to other nowcasting work

To place our proposed approach within the broader nowcasting literature, it is helpful to use the "partial model" (or single equation) methods and "full system" methods classification used by Banbura, et al. (2013). The former type of approach involves specifications focused on the low frequency variable, in which the high frequency explanatory variables are not modeled. In the latter approach, the low and high frequency variables are jointly modeled. Our proposed modeling approach falls in the partial models class.

Among partial model methods, bridge and MIDAS models are most commonly used. Bridge models, considered in such studies as Baffigi, Golinelli, and Parigi (2004), Diron (2008), and Bencivelli, Marcellino, and Moretti (2017), relate the period t value of the quarterly variable of interest, such as GDP growth, to the period t quarterly average of key monthly indicators. The period t average of each monthly indicator is obtained with data available within the quarter and forecasts for other months of the quarter (obtained typically from an autoregressive model for the monthly indicator). MIDAS-based models, developed in Ghysels, Santa-Clara, and Valkanov (2004) for financial applications and applied to macroeconomic forecasting by, e.g., Clements and Galvao (2008) and Guerin and Marcellino (2013), relate the period t value of the quarterly variable of interest to a constrained distributed lag of monthly or weekly or even daily data on the predictors of interest. The resulting model is then estimated by nonlinear least squares and used to forecast the variable of interest from constrained distributed lags of the available data. Foroni, Marcellino and Schumacher (2015) propose the use of unconstrained distributed lags of the high frequency indicators, a specification labeled unrestricted MIDAS, or U-MIDAS.

Full system methods for nowcasting include factor models and mixed frequency VARs. We refer to the surveys in Banbura et al. (2013) and Foroni, Ghysels, and Marcellino (2013) for details and references. Here we only mention a few studies closely related to our proposal. These include Aastveit, et al. (2014), which, in contrast to most of the nowcasting literature, focuses on density forecasts; Eraker, et al. (2015); Ghysels (2016); Schorfheide and Song (2015) and McCracken, Owyang, and Sekhposyan (2020), both of which develop mixed frequency Bayesian VARs; and Marcellino, Porqueddu, and Venditti (2016), which introduces a small scale factor model that allows for stochastic volatility in the common and idiosyncratic components and provides density forecasts.[5] Schorfheide and Song (2020) focuses on forecasting during the pandemic with a mixed frequency Bayesian VAR, using the original 11 monthly and quarterly indicators of Schorfheide and Song (2015).

In other related quantile forecasting work, some studies have considered tail risks to other variables, such as unemployment (e.g., Galbraith and van Norden (2019) and Kiley (2018)) and inflation (e.g., Ghysels, Iania, and Striaukas (2018)). Earlier work of Manzan (2015) used quan-

[5]Koop, Gefang, and Poon (2020) develop variational Bayes methods to estimate large Bayesian mixed frequency VARs with much greater computational efficiency.

tile regression to assess the value of a large number of macroeconomic indicators in forecasting the complete distribution of some key variables.[6] Cook and Doh (2019) apply quantile regression methods with a large set of predictors of growth, unemployment, and inflation, considering various approaches to dimension reduction and forecast combination. Lima, Meng, and Godeiro (2019) develop an approach for combining quantile forecasts to obtain point forecasts, in a setting with some mixed frequency data. Ferrara, Mogliani, and Sahuc (2019) use the quantile regression setup of Adrian, Boyarchenko, and Giannone (2019) to nowcast euro area GDP growth with an indicator of financial conditions updated on a daily basis, making use of quantile regression and Bayesian quantile regression with MIDAS. By comparison, our paper deploys a much richer set of economic and financial variables and model specifications, along with an alternative approach to accommodating mixed frequency data. Mazzi and Mitchell (2019) use quantile regression methods to form density nowcasts of euro area GDP growth. We build on their work by focusing on point and 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 risks to GDP growth and conclude that the evidence of such predictability is weak. In their assessment, none of their predictors yielded useful "...advance warnings of tail risks or indeed about any features of the GDP growth distribution other than the mean." 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. That said, the limited statistical significance in our results could be seen as consistent with the punchline of Plagborg-Moller, et al. (2020).

2 Data

As noted above, we focus on current-quarter forecasting of real GDP (or GNP for some of the sample) in real time. This section first explains the general design of the forecast calendar used in our analysis and then details the data used.

2.1 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 depart from and extend much of this literature by

[6] Other examples of studies of quantile forecasts in macroeconomics include Gaglianone and Lima (2012), Korobilis (2017), and Manzan and Zerom (2013, 2015).

considering 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, reflecting four weeks per month of the quarter and the first three weeks of the following quarter.[8]

We begin with the first full week of the quarter and end with the third week of the following quarter (the last week before GDP for the prior quarter is typically released). 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, stock prices, and the NFCI 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, the NFCI for week 1 of the quarter, and interest rates and stock prices for week 2 of the quarter.

Our starting point variable set largely corresponds to the small-model specification of CCM (in their results, the small model performed as well as models with additional leading indicators of the business cycle). In particular, we consider 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). In our baseline macro set, we add 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). We also consider financial indicators 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 prices as measured by 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.[9] Finally, in some additional specifications, we add to models

[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.

[8] In the interest of tractability of our real-time forecast evaluation of a range of models or methods, our calendar and forecasting design makes some simplifications that abstract from some complications that come with high frequency forecasting, relating to matters such as dealing with exact days on which series are published in real time and timing issues around weekly observations for overlapping starts and ends of months and quarters. Lewis, Mertens, and Stock (2020) use a notion of pseudo-weeks to deal with such timing issues in constructing their weekly index of economic activity.

[9] We conducted some limited checks replacing the NFCI with an indicator of financial stress in the US developed

some indicators of economic activity available at a weekly frequency and over time samples going back into at least the 1990s. For the most part, these series are those used in the weekly economic activity index of Lewis, Mertens, and Stock (2020). Our set of additional weekly indicators consists of continuing claims for unemployment insurance, consumer comfort from Bloomberg, raw steel production, electric utility output, loadings of railroad cars, total fuel sales, and Redbook same-store retail sales. Table 1 lists the variables and our calendar assumptions.

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 the percent change in the average weekly value of the index in one week compared to the average in the same week one quarter ago. In the case of the weekly indicators of steel production, utility output, car loadings, fuel sales, and Redbook retail sales, in light of the noisiness of the data and strong seasonality, we follow Ferrara and Simoni (2019) and Lewis, Mertens, and Stock (2020) and rely on year-over-year (52-week) growth rates. We leave as a subject for future research whether alternative treatments of seasonality in these series could be appropriate. We smooth the consumer comfort measure by using a four-week moving average of the weekly data. The next section will provide additional detail on the treatment of monthly and weekly data in our nowcasting models.

2.2 Details of data used

Quarterly real-time data on GDP or GNP are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM), in monthly vintages. For simplicity, hereafter "GDP" refers to the output series, even though the measures are based on GNP and a fixed weight deflator for much of the sample.

For the variables we use to nowcast GDP growth, for those 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

and published by the European Central Bank. The results we obtained with this indicator are similar to those obtained with the NFCI.

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 sources including the Federal Reserve Board's FAME database, the Federal Reserve Bank of St. Louis' FRED database, or Haver Analytics. Appendix Table A1 gives the source from which we obtained each series.

The full forecast evaluation period runs from 1985:Q1 through 2019:Q3 (using period t to refer to a forecast for period t), which involves real-time data vintages from January 1985 through February 2020. 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 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 our baseline macro models, estimation starts with 1970:Q2, the soonest possible given data availability and lags allowed in models. Adding financial indicators makes the estimation starting point 1971:Q2, due to the availability of the NFCI. 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 run from 2000:Q1 through 2019:Q3 and 2007:Q1 through 2019:Q3, respectively.

As discussed in such sources as Croushore (2006), Romer and Romer (2000), and Sims (2002), evaluating the accuracy of real-time forecasts requires a difficult decision on what to take as the actual data in calculating forecast errors. The GDP data available today for, say, 1985, represent the best available estimates of output in 1985. However, output as defined and measured today is quite different from output as defined and measured in 1970. For example, today we have available chain-weighted GDP; in the 1980s, output was measured with fixed-weight GNP. Forecasters in 1985 could not have foreseen such changes and the potential impact on measured output. Accordingly, we follow studies such as Clark (2011), Faust and Wright (2009), and Romer and Romer (2000) and use the second available estimates in the quarterly vintages of the RTDSM of GDP/GNP as actuals in evaluating forecast accuracy.

3 Models and Methods

This section details our proposed nowcasting models and methods. In general, to accommodate potentially large sets of indicators, we consider a range of approaches. (As noted earlier, we have also considered quantile regression with MIDAS but provide these results in the appendix rather than the paper since other methods generally perform better.) First, we use Bayesian shrinkage methods to estimate models with large sets of indicators. Second, we examine some factor-based approaches to dimension reduction, first forming common factors of available monthly and weekly indicators and then estimating regression models including the factors. Third, we consider forecast combination, by averaging forecasts across models, in two different ways. Under some estimation approaches (not featuring shrinkage) we estimate smaller models featuring one indicator at a time and then average forecasts from one-at-a-time models. In addition, we consider various other averages of other forecasts, such as an average across all forecasts. Note that, in averaging, we use equal weights, in light of evidence in the literature that, in practice, simple averages are typically hard to beat.[10]

To help the discussion flow, we first specify the general model and method forms in sections 3.1 through 3.6, and then in section 3.7, we detail the sets of indicators in the model. The appendix (included in the paper) provides the priors and algorithms used in Bayesian estimation.

3.1 General model forms: Bayesian mixed frequency model with stochastic volatility (BMF-SV)

At the outset, we should specify variable 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 $X_{w,t}$ (and therefore model) for each week $w = 1, \ldots, 15$ within the quarter, reflecting data availability. As also detailed below, the regressors of $X_{w,t}$ include a constant, past GDP growth, and selected available monthly and weekly variables. Let $\tilde{X}_{w,t}$ refer to the subset of $X_{w,t}$ that is the available monthly and weekly variables in week w, and let $Z_{w,t}$ refer to the subset of $X_{w,t}$ composed of the constant and lagged GDP growth (so $Z_{w,t} = (1, \text{GDP}_{t-p})'$), such that $X_{w,t} = \left(Z'_{w,t}, \tilde{X}'_{w,t}\right)'$. As detailed below, depending on timing and data availability, the GDP lag p is either 1 or 2.

[10]Studies including Giacomini and Komunjer (2005) and Taylor (2020) have developed methods for combining tail risk forecasts. In Taylor's (2020) results, a simple average sometimes performs as well as a more sophisticated combination. See Timmermann (2006) for a survey of evidence of the challenges of beating a simple average.

Among our models, we use CCM's Bayesian mixed frequency model with stochastic volatility (BMF-SV).[11] We consider nowcasting the quarterly growth rate of GDP in week w of the current quarter based on the following regression with stochastic volatility:

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

Following the approach pioneered in Cogley and Sargent (2005) and Primiceri (2005), the log of the conditional variance of the error term in equation (1) follows a random walk process. 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 results omitted in the interest of brevity, we have also considered a model with stochastic volatility that allows a link of financial conditions to not only the conditional mean of GDP growth but also its conditional variance. In this case, in a model we refer to as the BMF-GFSV specification, the stochastic volatility process has an AR(1) form augmented with an indicator of recent financial conditions, measured by the NFCI.[12] Studies such as Adrian, Boyarchenko, and Giannone (2019) and Carriero, Clark, and Marcellino (2020) have emphasized such a linkage as possibly helpful in forecasting macroeconomic tail risks. In our nowcast setting, this model yields results no better than those we report for the BMF-SV specification.

The specification of the regressor vector $X_{w,t}$ in the BMF-SV model 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, for each monthly variable, 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 period t as of week w of the quarter 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

[11] We also produced results for a homoskedastic version of the BMF regression specification. Stochastic volatility typically yields a significant improvement in in-sample model fit (marginal likelihood) and more accurate point and density nowcasts.

[12] As implemented in our nowcasting specification, the volatility process takes the form $\log(\lambda_{w,t}) = \delta_0 + \delta_1 \log(\lambda_{w,t-1}) + \delta_2 \overline{\text{NFCI}}_{w,t} + \nu_{w,t}$, where $\overline{\text{NFCI}}_{w,t}$ refers to the most recent 4-week average of the NFCI. The use of the 4-week average is meant to capture recent financial conditions in a simple way and with a modest amount of smoothing to mitigate weekly noise.

formed. As noted in section 2, we consider forecasts formed at 15 weeks associated with each quarter. At each of the 15 forecast origins we consider for each quarter t, the regressor set $X_{w,t}$ is specified to include the subset of variables available for t as of that week (details are given below in subsection 3.8). At these points in time, the availability of other indicators also varies. As a consequence, the model specification changes in each week of the quarter, reflecting and accommodating the ragged edge of the data, in line with a direct approach to forecasting.[13]

3.2 Quantile regression (QR) and quantile regression with Lasso penalty (QR-Lasso)

Our model set also includes quantile regression specifications patterned after the forecasting formulation developed in, among others, Adrian, Boyarchenko, and Giannone (2019). Out of concern for the variability of model estimates in small samples (especially small with respect to tail observations), we obtain our reported QR results by putting one predictor in the model at a time and averaging (simple equal weighting) the forecasts across indicators. In a different setting, Korobilis (2017) also uses a one-predictor-at-a-time approach, applying Bayesian model averaging to quantile regressions. We have produced out-of-sample nowcasts with models putting all elements of $X_{w,t}$ in a single regression at forecast origin w, but in larger models the coefficient estimates vary considerably over time, harming out-of-sample forecast accuracy (the appendix provides some of this evidence). Our averaging of one-at-a-time predictions works better.

More specifically, at the forecast origin of week w, let $\tilde{X}_{w,j,t}$ denote one of the (scalar) monthly or weekly indicators of $\tilde{X}_{w,t}$, and let $X_{w,j,t}$ denote the vector containing a constant, lagged GDP growth, and $\tilde{X}_{w,j,t}$, such that $X_{w,j,t} = \left(Z'_{w,t}, \tilde{X}'_{w,j,t}\right)'$. For each j, for a given quantile τ we estimate a quantile regression including as predictors the vector $X_{w,j,t}$, of the form:

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

in which the coefficient vector and innovation term are specific to quantile τ, as well as week w and j. The parameter vector $\beta_{\tau,w,j}$ is obtained with quantile regression:

$$\hat{\beta}_{\tau,w,j} = \underset{\beta_{\tau,w,j}}{\text{argmin}} \sum_{t=1}^{T} \left(\tau \cdot \mathbf{1}_{(y_t \geq X'_{w,j,t}\beta_{\tau,w,j})} |y_t - X'_{w,j,t}\beta_{\tau,w,j}| + (1-\tau) \cdot \mathbf{1}_{(y_t < X'_{w,j,t}\beta_{\tau,w,j})} |y_t - X'_{w,j,t}\beta_{\tau,w,j}| \right). \tag{3}$$

[13]We should stress that this approach does not involve bridge methods. Bridge methods require forecasting monthly or weekly observations of monthly or weekly variables for any months or weeks of quarter t for which data are not yet available. We do not use such forecasts. Rather, we only put on the right-hand side of the regression model the actual monthly and weekly observations that are available for the quarter, in the form of different quarterly variables associated with the different months and weeks of the quarter.

For each monthly and weekly indicator j of $X_{w,t}$, we form the quantile forecast as the predicted component $X'_{w,j,t}\hat{\beta}_{\tau,w,j}$. Our quantile regression forecast for quantile τ at forecast origin w is then a simple average of these individual nowcasts, from a total of J indicators in week w: $J^{-1}\sum_{j=1}^{J} X'_{w,j,t}\hat{\beta}_{\tau,w,j}$.

As one approach to dimension reduction, we also consider quantile regression with a Lasso penalty. In this case, we include in the regression the complete vector of weekly predictors $X_{w,t}$

$$y_t = X'_{w,t}\beta_{\tau,w} + \epsilon_{\tau,w,t}, \tag{4}$$

and the minimization problem underlying estimation includes the Lasso penalty:

$$\hat{\beta}_{\tau,w} = \underset{\beta_{\tau,w}}{\text{argmin}} \left\{ \sum_{t=1}^{T} \left(\tau \cdot \mathbf{1}_{(y_t \geq X'_{w,t}\beta_{\tau,w})} |y_t - X'_{w,t}\beta_{\tau,w}| + (1-\tau) \cdot \mathbf{1}_{(y_t < X'_{w,t}\beta_{\tau,w})} |y_t - X'_{w,t}\beta_{\tau,w}| \right) + \lambda \sum_{k=1}^{K} |\beta_k| \right\}. \tag{5}$$

For computational simplicity, we implement QR-Lasso with a fixed penalty parameter of $\lambda = 5$, using the quantreg package in R.[14]

3.3 Bayesian quantile regression (BQR) and Bayesian quantile regression with Lasso penalty (BQR-Lasso)

As just noted, with the large number of indicators associated with our mixed frequency approach, estimate imprecision can harm the forecast performance of simple quantile regression. Bayesian shrinkage may mitigate such imprecision and help the forecast performance of quantile models. Accordingly, we also consider models estimated with Bayesian quantile regression methods, including all predictors of $X_{w,t}$ in a single model rather than estimating separate models for each indicator. Yu and Moyeed (2001) established that quantile regression has a convenient mixture representation that enables Bayesian estimation. We use the Gibbs sampler of Khare and Hobert (2012), along with their mixture representation.

In our BQR formulation, for GDP growth in quarter t to be forecast as of week k of the quarter, we begin with a model

$$y_t = X'_{w,t}\beta_{\tau,w} + \sigma_{\tau,w}\epsilon_{\tau,w,t}, \tag{6}$$

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

$$y_t = X'_{w,t}\beta_{\tau,w} + \theta z_{\tau,w,t} + \kappa\sqrt{\sigma_{\tau,w} z_{\tau,w,t}}\, u_{\tau,w,t}, \tag{7}$$

[14] We experimented with a few other settings for λ, obtaining either similar results (for $\lambda = 3$) or worse results (for still smaller settings).

where θ and κ are fixed parameters as functions of the quantile κ and $u_{\tau,w,t}$ is i.i.d. standard normal. The quantile forecast is the predicted component $X'_{w,t}\hat{\beta}_{\tau,w}$. We compute the forecast at the single point of the posterior mean of the coefficient vector $\beta_{\tau,w}$.

In addition to a Bayesian shrinkage approach to QR, we consider a combination of shrinkage with a Lasso penalty that is meant to capture a form of variable selection. In the Bayesian setting, the Lasso penalty (in the loss function minimized to obtain estimates in the frequentist setting) on the sum of the absolute value of coefficients is tantamount to adjusting BQR to feature a Laplace prior on the regression coefficients. The Laplace distribution has a sharp peak at 0, which captures the Lasso idea of setting some coefficients to 0. The resulting model has a hierarchical form, for which Li, Xi, and Lin (2010) develop a Gibbs sampler. We follow their approach with our BQR-Lasso model. Mazzi and Mitchell (2019) use the model to examine density nowcasts.

3.4 Factor-based specifications: BMF-factor-SV, QR-factor, and BQR-factor

As another possible approach to nowcasting tail risks with large sets of regressors, we consider several specifications that rely on dimension reduction, in the form of common factors in the monthly and weekly indicators used in a given specification. Specifically, at each forecast origin, consider the vector of variables $X_{w,t}$ included in the BMF-SV specification above, associated with week w. We estimate a vector of common factors $f_{w,t}$ as the first three principal components of the monthly and weekly indicators of $\tilde{X}_{w,t}$. We then estimate forecasting models relating GDP growth to $Z_{w,t}$ (a constant and lagged GDP growth) and the factor vector $f_{w,t}$. We only apply factor-based approaches with our larger variable sets (detailed below) that, in addition to the baseline macroeconomic indicators, include sets of financial indicators or sets of weekly activity indicators.

We consider this factor-based data reduction approach with three different specifications. One, the BMF-factor-SV model, takes the form of our BMF-SV model but replaces the full predictor vector $X_{w,t}$ with $Z_{w,t}$ and $f_{w,t}$:

$$y_t = Z'_{w,t}\beta_{w,Z} + f'_{w,t}\beta_{w,f} + v_{w,t}, \tag{8}$$

with stochastic volatility in the innovation process. Our QR-factor and BQR-factor specifications rely on quantile regression with $Z_{w,t}$ and $f_{w,t}$ as predictors in a regression model of the form:

$$y_t = Z'_{w,t}\beta_{\tau,w,Z} + f'_{w,t}\beta_{\tau,w,f} + \sigma_{\tau,w}\epsilon_{\tau,w,t}. \tag{9}$$

Giglio, Kelly, and Pruitt (2016) refer to this type of approach as principal components quantile regression, which they apply to a panel of systemic risk measures for forecasting indicators of economic activity.

3.5 Mixed frequency partial quantile regression (PQR)

As another approach to data reduction with large sets of regressors, 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.[15]

At each forecast origin, consider the vector of variables $X_{w,t}$ included in the BMF-SV specification 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 $\tilde{X}_{w,t}$.[16] We then estimate the mixed frequency quantile regression

$$y_t = Z'_{w,t}\beta_{\tau,w,Z} + f_{\tau,w,t}\beta_{\tau,w,f} + \epsilon_{\tau,w,t} \tag{10}$$

and form the associated PQR nowcast for quantile τ with the resulting coefficient estimates.

3.6 Forecast averages

As indicated above, with some models (such as QR), we form forecasts by estimating specifications with a single weekly or monthly indicator and then averaging across the smaller models' projections. We also consider some combinations of larger sets of forecasts as another approach to exploiting large sets of information and models. Combination has been shown to be helpful to forecast accuracy in a range of settings (see, e.g., the survey of Timmermann (2006)). In our implementation, partly in light of the known difficulty of beating simple averages when the forecast set can be large, we only consider simple averages. Note that each average uses only those forecasts of the indicated type available over the evaluation sample being used. More forecasts are available for the shorter samples than for the full 1985-2019 sample. Our average forecasts consist of the following, using data sets defined in the next section:

- avg. all: simple average of all forecasts;
- avg. base M-F: simple average of all forecasts using the base M-F variable set;
- avg. base M-F + small weekly: simple average of all forecasts using the base M-F + small weekly variable set;

[15] As an alternative to the PQR approach that targets the prediction of GDP growth, one might instead construct quantile factors of economic and financial conditions as in Chen, Dolado, and Gonzalo (2019) and use those factors to predict tail risk to GDP growth.

[16] In the first stage quantile regression used to obtain the factor, we include a constant, lagged GDP growth, and one of the components of $\tilde{X}_{w,t}$, on a one-at-a-time basis.

- avg. base M-F + large weekly: simple average of all forecasts using the base M-F + large weekly variable set;
- avg. BMF-SV: simple average of all forecasts from the BMF-SV specification;
- avg. BQR: simple average of all forecasts from the BQR specification; and
- avg. BQR-Lasso: simple average of all forecasts from the BQR-Lasso specification.

Among the averages that focus on a single model specification and average forecasts obtained with different variable sets, we limit our attention to the few models that do not already include some averaging of forecasts or data reduction via factors. Based on this rationale, we include an average of BQR forecasts but exclude a similar average of QR forecasts, since the latter are obtained by averaging across one-at-a-time models.

3.7 Indicators used

We report below results for a total of five different variable combinations, each (with some exceptions noted below) applied with the BMF-SV, QR, QR-Lasso, BQR, BQR-Lasso, BMF-factor-SV, QR-factor, BQR-factor, and PQR models or methods, along with some forecast averages.

- Our starting point is the *base M* ("M" for macro) set of six macroeconomic activity indicators consisting of the small variable set of CCM plus initial claims for unemployment insurance. In this case, we have six monthly variables (payroll employment, ISM manufacturing index, industrial production, real retail sales, housing starts, and initial claims) and one weekly variable (initial claims).
- We also consider a *base M + NFCI* set that adds to the base M variable set a single summary indicator of financial conditions, the NFCI, available both monthly and weekly.
- Our third variable set — *base M-F* — adds to the base M variable set a set of monthly and weekly financial indicators ("F" for financial) consisting of the NFCI, S&P 500 stock returns, the Treasury term spread, and the Baa-Treasury credit spread.
- The variable set *base M-F + small weekly* adds to the base M-F variable set the small set of weekly indicators (continuing claims for unemployment insurance, consumer comfort, steel production, and electric utility output).

- The variable set *base M-F + large weekly* adds to the base M-F variable set the large set of weekly indicators (continuing claims for unemployment insurance, consumer comfort, steel production, electric utility output, car loadings, fuel sales, and Redbook sales). Again, with a short sample available, we only evaluate forecasts starting in 2007 and omit some models or methods from consideration.

In results omitted in the interest of brevity, we also considered variants of the variable sets base M-F + small weekly and base M-F + large weekly that omit the financial indicators. These results indicate that adding weekly indicators of economic activity to the base M variable set does not improve forecast accuracy. It is more helpful to add financial indicators (i.e., use the base M-F variable set) than it is to add the small or large sets of weekly indicators.

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 didn't find any payoff to longer lags. However, in most of our variable-model combinations, our general approach easily allows the use of values from previous quarters (while this makes the models even larger, Bayesian shrinkage helps limit the effects of parameter estimation error on forecast accuracy).

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 macro indicators, the second row covering the base 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. Of the other variable sets, models for the *base M + NFCI* set include the first row indicators plus the NFCI indicators shown in the second row (for a given week). 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.

We now offer some additional explanations of model details:

- The dependent variable of the model is GDP growth in quarter t. Subscripts of $t+1$, t, $t-1$, and $t-2$ refer to the next, 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 $w1$, $w2$, $\ldots$, $w12$ 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 included 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 the quarter. 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), large enough that, with the short sample available for estimation and forecast evaluation, we omit results for specifications that do not feature any variable reduction (BMF-SV, BQR, and BQR-Lasso). With models of these medium to large sizes, under simple OLS or QR estimation, parameter estimation error may be expected to have large adverse effects on forecast accuracy. Some results we have produced (most of these in the appendix) bear this out: For example, QR models featuring large variable sets together (rather than on a one-at-a-time basis) fare poorly in out-of-sample forecast accuracy. Our Bayesian approaches to estimation incorporate shrinkage to help limit the effects of parameter estimation error on forecast accuracy, as can some of the other dimension reduction approaches we consider.

4 Forecast Metrics

In assessing the efficacy of the models described in the previous section, we consider a range of forecast metrics, with a focus on the 5 percent tail. Note that the tail quantile forecast corresponds to the value at risk (VaR) forecast, and Adrian, et al. (2018) coined the term "growth at risk" for the 5 percent quantile of GDP growth forecasts (De Nicolo and Lucchetta (2017) coined similar terms for industrial production and employment). We have verified that our main results on lower tail forecasts (QS, FZG score, and coverage, detailed below) are robust to using 10 percent and 15 percent quantiles in lieu of the baseline 5 percent quantile; the appendix provides 10 percent and 15 percent quantile results. We also provide some comparisons that include an upper quantile of 95 percent.

In some limited results, to establish some patterns with different information sets using a familiar metric, we refer to point forecasts. In the assessment of point forecasts, we define them as the mean of the predictive distributions for the BMF-SV and related models and the prediction obtained at the quantile $\tau = 0.5$ for the quantile-based regression models. We evaluate the point forecasts with the root mean squared error (RMSE).

To assess the efficacy of the models in quantifying tail risks, for all models or methods we consider two basic measures of the accuracy of the lower tail quantile estimate. For the BMF-SV and related models, the quantile is simply estimated as the associated percentile of the simulated predictive distribution. For the quantile regression and related models, we use the prediction obtained from the quantile regression at the quantiles $\tau = 0.05$ and 0.10. Applied to these quantile estimates, the first accuracy measure 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

$$\mathrm{QS}_t = (y_t - Q_{\tau,t})(\tau - \mathbf{1}_{(y_t <= Q_{\tau,t})}), \tag{11}$$

where y_t is the actual outcome for GDP growth, $Q_{\tau,t}$ is the forecast quantile at quantile $\tau = 0.05$ or 0.10, and the indicator function $\mathbf{1}_{(y_t <= Q_{\tau,t})}$ has a value of 1 if the outcome is at or below the forecast quantile and 0 otherwise. Although much of the recent literature has not included formal statistical evaluations of quantile accuracy, Manzan (2015) relied on the quantile score. The second accuracy measure is a simple coverage measure for the interval forecast: the percentage of outcomes falling below the 5 and 10 percent quantiles of the forecast distribution.

For some models, we consider another measure of the accuracy of lower tail nowcasts. Conceptually, VaR has a number of disadvantages, and in research and practice (e.g., Basel standards on financial risk management), some prefer to focus on expected shortfall (ES). In our context, at the 5 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.[17] Accordingly, in some cases we report results for evaluating shortfall nowcasts using the joint value at risk-expected shortfall (VaR-ES) score from Fissler, Ziegel, and Gneiting (2015). The joint VaR-ES score is computed as

$$\begin{aligned} \mathrm{S}_t &= Q_{\tau,t} \cdot (\mathbf{1}_{(y_t <= Q_{\tau,t})} - \tau) - y_t \cdot \mathbf{1}_{(y_t <= Q_{\tau,t})} \\ &+ \frac{e^{\mathrm{ES}_{\tau,t}}}{1 + e^{\mathrm{ES}_{\tau,t}}} \left(\mathrm{ES}_{\tau,t} - Q_{\tau,t} + \tau^{-1} \left(Q_{\tau,t} - y_t \right) \mathbf{1}_{(y_t <= Q_{\tau,t})} \right) + \ln \frac{2}{1 + e^{\mathrm{ES}_{\tau,t}}}, \end{aligned} \tag{12}$$

where $\tau = 0.05$ and $\mathrm{ES}_{\tau,t}$ denotes the expected shortfall nowcast at quantile τ.

Out of consideration of computational simplicity and performance in the simpler QS results, we only report VaR-ES scores for the BMF-SV and related models, as well as the base M: QR specification we will take as a baseline. With the BMF-SV and related specifications, it is easy to estimate — using draws of the posterior forecast distribution — the expected shortfall as the

[17]See Taylor (2019) for another application using the Fissler, Ziegel, and Gneiting (2015) score and a useful discussion of challenges in evaluating shortfall by itself.

mean of forecast draws in the 5 percent tail of the predictive distributions. In contrast, for the methods relying on quantile regression, obtaining a shortfall estimate would involve a second-step estimation of a predictive distribution from different quantile estimates (see, e.g., the second step used in Adrian, Boyarchenko, and Giannone (2019)). In light of the considerable number of forecasts we examine and the performance of quantile regression-based methods vis-a-vis BMF-SV specifications, we only make these computations for the QR specification estimates with the base M variable set and otherwise omit these QR-based approaches from our VaR-ES score results.

To gauge statistical significance, we estimate Diebold and Mariano (1995)–West (1996) t-tests for equality of the average loss (with loss defined as squared error, quantile score, or VaR-ES score). We also compute t-tests for the empirical coverage rate equaling the nominal rate of 5 or 10 percent. In the tables, differences in accuracy or departures from nominal coverage that are statistically different from zero are denoted by one, two, or three asterisks, corresponding to significance levels of 10 percent, 5 percent, and 1 percent, respectively. The underlying p-values are based on t-statistics computed with a serial correlation-robust variance, using the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992). For the equal MSE, QS, and VaR-ES score tests, we conduct them on a one-sided basis, such that the alternative hypothesis is that the indicated forecast is more accurate than the benchmark.

5 Empirical Results

This section presents our results on the accuracy of out-of-sample nowcasts of GDP growth. Again, with the tail risk results presented herein, we focus on the 5 percent quantile; our general results also apply to the 10 percent and 15 percent quantiles, and the appendix reports these estimates. In light of the interest in tail risks and recent events, we also report some forecast accuracy results for just the periods of NBER-dated recessions (using their quarterly dating, in line with our forecasting of quarterly GDP growth). These results address forecast accuracy conditional on being in a state of recession (taking as given the ex post dating by the NBER). In these results, though, in light of the small samples of observations occurring in recessions, we abstract from tests of statistical significance. The section concludes with recent examples of nowcasts, using 2020:Q1 nowcasts produced with data as of late April 2020 and 2020:Q2 nowcasts produced with data as of late July 2020.

Although our focus is on conventional out-of-sample forecasts, we have produced and examined results on in-sample forecasts, and some of the results are in the appendix.[18] Our rationale for considering — only briefly here in the text — in-sample results is that, with some of the weekly indicators having data starting only in the 1990s, only a short out-of-sample evaluation period is possible, which makes it more difficult to establish meaningful differences in accuracy. In addition, for assessing tail risks, the most relevant periods are probably recessions, and these only occur periodically (roughly once every 10 years since the early 1980s). The in-sample evidence yields the following broad results. First, using the full sample of data and model fit measured by the marginal likelihood of the BMF-SV specification, model fit typically improves with additional data — both more observations on the quarter as the weekly origin moves forward and more indicators in a larger variable set compared to a smaller. Second, in-sample prediction accuracy also typically improves with the addition of more variables. Third, the gap between in-sample and out-of-sample performance can be wide, particularly with large models estimated without some form of shrinkage. In forecast accuracy, some models perform relatively much better on an in-sample basis than on an out-of-sample basis. our benchmark model is much more readily beaten in in-sample predictions than in real-time out-of-sample nowcasts. On an in-sample basis, quantile regression and can be estimated with full sets of predictors rather than one predictor at a time and yield predictions of high accuracy. But in the shorter estimation samples of out-of-sample forecast comparisons, parameter estimates can become highly variable over time and make these methods much less accurate. (And it is for that reason that, in the out-of-sample results we present below, we estimate these models with one indicator at a time and average the resulting nowcasts.)

5.1 Out-of-sample nowcast accuracy

As a starting point, we assess how increasing the basic information set over the weeks of the quarter affects forecast accuracy. Due to the familiarity of point forecast accuracy, for comparison we include RMSE measures of point forecasts along with measures of tail risk forecast accuracy. Figure 1 reports RMSEs, 5 percent quantile scores, and 5 percent VaR-ES scores from the BMF-SV model estimated with the base M variable set, for samples of 1985:Q1-2019:Q3 (upper panel) and 2000:Q1-2019:Q3 (lower panel). To facilitate comparisons, 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 the 1985:Q1-2019:Q3 sample, both RMSE and 5 percent QS fairly steadily improve by the week; additional information on the quarter materially improves the accuracy of both point and tail risk

[18] We compute in-sample forecast results just as we do for the out-of-sample case, with the differences that the parameter estimates used are obtained for the full sample rather than a recursive window, and we abstract from real-time data in the in-sample results.

forecasts. From week 1 to week 15, the RMSE and QS for the model improve by 25 to 30 percent. That said, even at week 15, shortly before the initial estimate of GDP is released, the forecast uncertainty is considerable, with the level of the RMSE at 1.44 (not shown). The VaR-ES score also improves considerably as more information becomes available by the week, but not uniformly: By this measure, tail forecast accuracy improves substantially from week 1 through about week 9 and then levels off or even deteriorates some. For the shorter sample of 2000:Q1-2019:Q3, the patterns are largely the same.

To compare real-time accuracy across models and variable sets, Figures 2 and 3 and Table 3 provide QS and coverage results using the variable sets available for the forecast evaluation sample of 1985:Q1 through 2019:Q3. To facilitate comparisons of quantile score accuracy, we use quantile regression with the base M set of indicators as the benchmark specification and for other specifications report results relative to this benchmark. In tail risk forecast accuracy as measured by the QS, the benchmark base M: QR specification is easily beaten by a number of other specifications (the appendix provides RMSE results for point forecasts in which the same is true). Using just the base M variable set, a few of the other models or methods considered — including QR-Lasso, BQR, and BMF-SV — consistently and significantly improve the QS. For example, with the base M variable set, the BMF-SV specification yields scores as much as 35 percent better than the QR benchmark. For most models, extending the variable set to include the NFCI slightly to modestly improves tail forecast accuracy. Further extending the variable set to the base M-F set does not yield much further improvement in accuracy relative to the base M + NFCI set (although QR performance is better with just the NFCI than the full set of financial indicators added), but it does not harm accuracy, either. With the base M-F variable set, almost all of the models significantly improve on the accuracy of the base M: QR baseline, with the better-performing choices (such as BMF-SV) lowering the QS by 30 to 40 percent across weeks. In general, then, the results indicate that, along some dimensions, tail risk forecast accuracy improves with more data — both with the addition of more timely indicators across weeks (with the benchmark scores falling across weeks, as noted above and reported in the top row of the table) and with the addition of financial variables to the base M variable set.

Considering alternative models for nowcasting tail risks with potentially large variable sets, the QR method stands out for not faring very well. Quantile regression approaches that make use of shrinkage or dimension reduction through factors fare better. As summarized visually in Figure 2, with the base M-F variable set, BMF-SV, BQR, and PQR all perform similarly (and BQR-factor does too, although with a little more variability across weeks), as does the BMF-factor-SV approach covered in Table 3 but not in the figure. These better approaches materially improve on the base M: QR results. Bayesian shrinkage and dimension reduction through factors both seem helpful to tail risk forecasting. The alternative shrinkage approach of combining forecasts — either averaging all available or averaging all of those using the base M-F variable set — can achieve comparable gains in accuracy of tail risk forecasts. Despite potentially small samples of tail outcomes (for example, with a total of 139 observations in the 1985-2019 evaluation sample, only about 7 should fall in the 5 percent tail of focus), a number of these approaches improve significantly on the quantile regression baseline, and as noted above, we obtain similar results at quantiles of 10 percent and 15 percent.

Regarding the empirical coverage of the 5 percent quantile forecast, the only specifications that consistently (across weeks and variable sets) deliver accurate coverage are those with stochastic volatility, the BMF-SV and BMF-factor-SV models. As a quick summary of this evidence, consider the asterisks reported in the lower panel of Table 3. Hardly any appear for these model specifications. Other models or methods are more challenged to yield accurate coverage in the 5 percent tail, most typically with coverage rates that are too low, reflecting a quantile estimate that is too low. For example, coverage rates with QR are too low with each variable set considered for the 1985:Q1 through 2019:Q sample. Coverage is also routinely too low with the base M-F: PQR specification. With the base M-F variable set, some other methods yield accurate coverage: these include the QR-factor and BQR-factor specifications (as well as QR-Lasso, to a lesser extent).

Figure 4 and Tables 4 and 5 provide QS and coverage results using the variable sets available for the forecast evaluation sample of 2000:Q1 through 2019:Q3. With the shorter sample, we are able to include comparisons to nowcasts that make use of an additional variable set — base M-F + small weekly. The patterns in results are similar to those for the full sample. With just the base M variable set, a few models — including QR-Lasso, BQR, and BMF-SV — significantly improve on the quantile score accuracy of the baseline model. Relative to the benchmark variable set and model (base M: QR), adding financial indicators to the base variable set yields significant gains in accuracy as measured by the quantile score. For example, for many weeks, the QS for the base M-F: BMV-SV specification is roughly 25-30 percent lower than the benchmark. With financial indicators in the models, adding the small set of weekly indicators doesn't much help or harm accuracy: For the better performing approaches, QS ratios are broadly similar across the base M-F and base M-F + small weekly variable sets. Focusing on alternative models or methods, it remains the case that QR does not perform well. Several approaches, including BQR, BQR-factor, PQR, and BMF-SV, perform comparably, once again indicating that Bayesian shrinkage and dimension reduction through factors can be helpful to tail risk forecasting. The alternative approach of combining forecasts — averaging all forecasts, averaging just the base M-F forecasts, and averaging the base M-F + small weekly forecasts — achieves comparable gains.

With the shorter sample of 2000:Q1-2019:Q3, empirical coverage rates are modestly better, in the sense that fewer forecasts have coverage rates significantly different from 5 percent. For some forecast methods, such as BQR, the 5 percent quantile forecast is systematically low in the 1985-2000 period and not as low relative to the data in the remainder of the sample (see Figure 7). The bands for the 5 percent and 95 percent quantiles are noticeably narrower early in the sample with the BMF-SV specification (see Figure 6); the coverage challenges of other models may be due to the treatment of innovation variances as constant, when in fact volatility fell sharply with the Great Moderation of the mid-1980s.

Figure 5 and Table 6 provide QS results using the variable sets available for the forecast evaluation sample of 2007:Q1 through 2019:Q3. This sample allows us to make use of the large weekly variable set (but due to the short data sample available for model estimation in the early years of the evaluation period, we omit results for the shrinkage-based approaches that allow all available indicators to enter the model). However, the short evaluation sample makes it more difficult to achieve statistical significance in accuracy differences. The broad patterns in tail forecast accuracy over this sample are quite similar to those over longer samples. Again, tail forecast accuracy generally improves as additional observations on the quarter become available across weeks. In addition, extending the base M variable set to include financial indicators improves accuracy. With financial indicators in the model, adding weekly activity indicators does not harm accuracy (however, this is more clearly the case for the small weekly set than for the large weekly set). As to model or approach choice, it remains the case that QR does not fare well; Bayesian shrinkage or variable reduction typically yields more accurate forecasts. Over this shorter sample, some of the forecast averages — such as averaging all forecasts or averaging the base M-F or base M-F + small weekly projections — may be seen as having advantages of consistency and relative accuracy.

As another assessment of the accuracy of tail risk forecasts, Table 7 reports VaR-ES scores from the BMF-SV and BMF-factor-SV specifications, relative to the benchmark base M: QR specification, for each evaluation sample (as noted above, we consider just these models in part because the expected shortfall can easily be computed from the posterior sampler). The top row of each panel shows that, as additional information becomes available across weeks of the quarter, the quantile and shortfall forecasts improve and the VaR-ES scores fall. With the base M variable set, a number of the specifications with stochastic volatility significantly improve (more readily in the 1985-2019 and 2000-2019 samples than the 2007-2019 sample) on the score accuracy of the QR benchmark. However, in contrast to the QS results, changing the variable set to include more indicators does not consistently improve the tail risk forecasts. Consider, for example, results with the BMF-SV model. Over the 1985:Q1-2019:Q3 sample (top panel), adding the NFCI or the full set of financial indicators to the base M variable set doesn't add much to the forecast gains achieved by the model with the base M variable set. Overall, by this measure of tail risk accuracy, adding information with the weekly flow of information helps, but there isn't much evidence that adding other variables does the same.

In light of the interest in downside tail risks and the interest in quickly detecting the extent of the downturn following the recent outbreak of the pandemic, we now consider the accuracy of tail risk forecasts during past NBER recessions.[19] To be clear, these quantile scores are not any different from the quantile scores for the 5 percent quantile forecasts. Rather, we are just computing their averages in the subset of quarters that fall in recessions rather than for the full sample. It is in these quarters that we might expect GDP growth to be close to or below its 5 percent quantile forecast, although growth may be close to or below its tail quantile even in periods of expansions. Table 8 provides QS and selected VaR-ES score (5 percent) results for 1985:Q1-2019:Q3, the longest sample available among our variable sets, in order to maximize the coverage of recessions. Still, the sample only contains three recessions and a total of 11 quarterly observations of GDP growth, which likely factors into some of the considerable variation across models and weeks evident in the results. The top rows of the table's two panels show that, with recessions, as with the full sample, nowcast accuracy improves substantially as more information becomes available across the weeks of the quarter. For the most part, using the larger set of base M-F variables improves on the forecasts of the base M variable set, but not always. For example, the QS and VaR-ES scores are lower in the base M-F: BMF-SV estimates than in the base M: BMF-SV estimates. But to take a contrary example, for some weeks early in the quarter, before many monthly observations on the quarter are available, the QS scores are lower for base M: QR than for base M-F: QR. A number of models or approaches perform fairly well, including the shrinkage-based BMF-SV specification, the factor-based QR-factor, BQR-factor, and PQR approaches, and some of the forecast averages, particularly the average across all base M-F forecasts.

[19] We define the periods of recessions using the NBER's quarterly dating of business cycle peaks and troughs.

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. We first show in Figures 6, 7, and 8 forecasts from the BMF-SV, BQR, and PQR specifications estimated with the base M-F variable set. These charts include the point, 5 percent quantile, and 95 percent quantile forecasts produced at a set of selected forecast origins within the quarter, along with the actual GDP growth outcome. Figure 9 shows just 5 percent quantile forecasts, at weeks 7, 11, and 15, for a slightly wider array of specifications. Figure 10 reports 5 percent expected shortfall forecasts for BMF-SV specifications with different variable sets.

As indicated in Figure 6, with the BMF-SV model and the base M-F variable set, the 5 percent and 95 percent forecast quantiles tend to move together. In nowcasting, with conditioning on some information for the quarter being forecast, we don't 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 (2020). Consistent with the results already documented, the visual evidence suggests that the point forecasts get a little more accurate as more data become available across the weeks of the quarter, especially during recessions. In addition, during recessions, the 5 percent forecast quantile declines with more weeks of data becoming available in the quarter, as does the 95 percent quantile. In weeks of the first half of the quarter, the tail forecast ends up being close to actual GDP growth in downturns, but in later weeks, the outcome is not as bad as the 5 percent quantile forecast.

As evident from Figures 7 and 8, with the BQR and PQR specifications, as with the BMF-SV model, with some conditioning on information within the quarter for nowcasting, there doesn't appear to be much asymmetry in movements of downside tail risks as compared to upside risks — notwithstanding the use of quantile regression-based methods. Early in the sample, the BQR and PQR bands are quite wide, consistent with the coverage issues noted above; the narrower bands produced by the BMF-SV specification reflect its incorporation of time-varying volatility and the influences of the Great Moderation. The tendency of the PQR model estimated with the base M-F variable set to produce wide bands persists into the later portion of the sample.

Figure 9's selected comparisons make clear that the addition of data across weeks of the quarter can produce sizable changes in tail quantile forecasts. With the information available as of week 7 of the quarter, the base M-F: BMF-SV and base M-F: PQR nowcasts move together fairly closely in the early years of the sample (top panel). However, their comovement in the early years of the sample (before the 1990-1991 recession) is noticeably weaker with information as of week 11 of the quarter (middle panel). The figure also indicates that — for the estimates reported — the 5 percent quantile forecasts become more similar in the 1990-1991 and 2007-2009 recessions than in other periods, including the shallower recession of 2001 and periods of economic expansions. This pattern, though, does not appear to be a general pattern in the wider array of our nowcasts. We have checked the dispersion of nowcasts across the models and variable sets considered, and by this broader quantification the dispersion of 5 percent quantile nowcasts tends to go up rather than down during recessions.

The expected shortfall forecasts in Figure 10 display more similarity to one another than do the quantile forecasts, but that is partly because the shortfall estimates are all based on BMF-SV specifications. In recessions, the shortfall estimates tend to become more negative as additional weeks of data become available. The shortfall nowcasts are similar over time for the base M, base M + NFCI, and base M-F variable sets; from the late 1990s until the mid-2000s, estimates using the base M-F + small weekly variable set can differ notably from the other estimates, but then become more similar. Some of the differential could have to do with the shorter time sample available for estimation with the base M-F + small weekly variable set compared to the others.

5.2 Current example: Forecasts for 2020:Q1 and 2020:Q2

We conclude our analysis with real-time examples of nowcasting tail risks to GDP growth in the early stages of a pandemic-driven recession in the US. In particular, with selected variable sets and models, we report forecasts for growth in 2020:Q1 produced across the weeks of the quarter using data available in late April 2020 (about the time a first draft of this paper was prepared), which is weeks 1 through 15 in our setup.[20] Note that, shortly after the 2020:Q1 forecasts were produced, the Bureau of Economic Analysis published the first estimate of GDP in 2020:Q1, and that outcome (as an annualized log growth rate) was -4.9 percent, subsequently revised to -5.2 percent in the second estimate. We also report forecasts for growth in 2020:Q2 produced across the weeks of the quarter using data available in late July 2020 (at the time we began to prepare a third draft of this paper), spanning weeks 1 through 15 in our setup. Shortly after the 2020:Q2 forecasts were produced, the Bureau of Economic Analysis published the first estimate of GDP in 2020:Q2, and that outcome (as an annualized log growth rate) was -39.9 percent.

[20] In the 2020:Q1 exercise, for simplicity we use the third estimate of GDP growth in 2019:Q4 and the current vintage time series of monthly and weekly indicators available in late April and abstract from updates of preliminary data that occurred over the course of 2020:Q1. In the 2020:Q2 exercise, we use the third estimate of GDP growth in 2020:Q1 and the current vintage time series of monthly and weekly indicators available in late July and abstract from updates of preliminary data that occurred over the course of 2020:Q2.

To limit the volume of results, we use just a few variable sets and models, selected on the basis of our assessment of the more successful historical approaches. In particular, we report results for the BMF-SV, BQR, and PQR models estimated with the base M-F, base M-F + small weekly, and base M-F + large weekly variable sets.

As indicated in the nowcasts in the left column of Figure 11, early in 2020:Q1, before the pandemic spread and shutdowns began (in mid- to late March in most of the US), almost all of the BMF-SV, BQR, and PQR nowcasts of the 5 percent quantile were above 0. For example, in week 9, the reported nowcasts ranged from 0.2 percent to 1.5 percent. 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. The nowcasts suddenly turned sharply negative in week 13, and most, although not all, turned more negative in week 14. Except for the base M-F: BQR specification, all of the week 14 nowcasts of the 5 percent quantile are below the second estimate of actual GDP growth in the quarter. For 2020:Q2, the 5 percent quantile nowcasts were sharply negative from the beginning (weeks overlapping with the last couple of weeks of nowcasts for 2020:Q1), with the week 1 projections ranging from -7.3 percent to -19.2 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 few weeks, some of the nowcasts deteriorated further and others remained broadly stable, before some became modestly less negative over the last several weeks. At the last weekly forecast origin, all but one of the nowcasts projected a historic tail risk to GDP growth, with a decline of at least 20 percent. As to patterns across models and variable sets, the nowcasts are relatively similar with the base M-F + small weekly and base M-F + large weekly variable sets; for a given model, these nowcasts are more similar to one another than to the nowcast obtained with just the base M-F variable set. Although differences across models can be sizable, the nowcasts paint a broadly similar picture of the risks to GDP growth in these quarters. Given a variable set, the nowcasts from the BMF-SV and PQR models tend to be more similar to one another than to the nowcasts from the BQR specification (albeit more so for 2020:Q2 than 2020:Q1). Our takeaway from this illustration is similar to that of our historical forecast evaluation: additional information on the quarter as time moves forward in the quarter bears importantly on nowcasts and their accuracy, and with sizable differences across models possible, it is likely helpful in practice to consider a range of forecasts.

6 Conclusions

This paper focuses on nowcasts of tail risk to GDP growth, with a potentially wide array of monthly and weekly information. We consider different models (Bayesian mixed frequency regressions with stochastic volatility, as well as classical and Bayesian quantile regressions) and also different methods for data reduction (either forecasts from models that incorporate data reduction through factors or the combination of forecasts from smaller models).

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, with monthly indicators more important than weekly indicators. In addition, extending the base macro variable set to include financial indicators improves accuracy. Adding just the small weekly or large weekly 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, our regression with stochastic volatility and our Bayesian quantile regression perform reasonably consistently, offering solid gains in forecast accuracy (relative to a baseline quantile regression model with just a small set of macro indicators), with benefits maximized when financial indicators are included in the model. Some factor reduction methods, such as partial quantile regression, and forecast averaging also improve accuracy with some consistency. Simple quantile regression is consistently less accurate than the benchmark nowcast.

To conclude, based on the many results already presented, what would we recommend for tail risk nowcasts? Our starting points would be the mixed frequency regression with stochastic volatility, Bayesian quantile regression, and partial quantile regression, applied to our baseline set of macroeconomic and financial indicators. As a practical matter, we would also consider forecasts from these same specifications but adding our small and large sets of weekly economic indicators, as well as averages of a broader set of nowcasts.

References

[1] Aastveit, Knut Are, Claudia Foroni, and Francesco Ravazzolo (2017), "Density Forecasts With Midas Models," *Journal of Applied Econometrics* 32, 783-801. `https://doi.org/10.1002/jae.2545`.

[2] Aastveit, Knut Are, Karsten R. Gerdrup, Anne Sofie Jore, and Leif Anders Thorsrud (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] Aastveit, Knut Are, Francesco Ravazzolo, and Herman K. van Dijk (2018), "Combined Density Nowcasting in an Uncertain Economic Environment," *Journal of Business and Economic Statistics* 36, 131-145. `https://doi.org/10.1080/07350015.2015.1137760`.

[4] Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019), "Vulnerable Growth," *American Economic Review* 109, 1263-1289. `https://doi.org/10.1257/aer.20161923`.

[5] Adrian, Tobias, Federico Grinberg, Nellie Liang, and Sheheryar Malik (2018), "The Term Structure of Growth-at-Risk," IMF Working Paper WP/18/180. `https://doi.org/10.5089/9781484372364.001`.

[6] Andreou, Elena, Eric Ghysels, and Andros Kourtellos (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] Andrews, Donald W.K., and J. Christopher Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica* 60, 953-966. `https://doi.org/10.2307/2951574`.

[8] Aruoba, S. Boragan, Francis X. Diebold, and Chiari Scotti (2009), "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics* 27, 417-427. `https://doi.org/10.1198/jbes.2009.07205`.

[9] Baffigi, A., Golinelli, R., and Parigi, G. (2004), "Bridge Models to Forecast the Euro Area GDP," *International Journal of Forecasting* 20, 447-460. `https://doi.org/10.1016/S0169-2070(03)00067-0`.

[10] Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2011), "Nowcasting," in *Oxford Handbook on Economic Forecasting*, Clements, M.P., and Hendry D.F. (eds). Oxford University Press: Oxford. `https://doi.org/10.1093/oxfordhb/9780195398649.013.0008`.

[11] Banbura, Marta, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin (2013), "Nowcasting and the Real Time Data Flow," in *Handbook of Economic Forecasting, Volume 2*, Elliott, G., and Timmermann, A. (eds). Amsterdam: North Holland. `https://doi.org/10.1016/B978-0-444-53683-9.00004-9`.

[12] Bencivelli, Lorenzo, and Massimiliano Marcellino, and Gianluca Moretti (2017), "Forecasting Economic Activity by Bayesian Bridge Model Averaging," *Empirical Economics* 53, 21-40. `https://doi.org/10.1007/s00181-016-1199-9`.

[13] Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino (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`.

[14] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2020), "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`.

[15] Chavleishvili, Sulkhan, and Simone Manganelli (2019), "Forecasting and Stress Testing with Quantile Vector Autoregression," ECB Working Paper No. 2330.

[16] Chen, Liang, Juan J. Dolado, and Jesus Gonzalo (2019), "Quantile Factor Models," manuscript.

[17] Clark, Todd 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`.

[18] Clark, Todd E., and Francesco Ravazzolo (2015), "Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics* 30, 551-575. `https://doi.org/10.1002/jae.2379`.

[19] Clements, M.P., and Galvao, A.B. (2008), "Macroeconomic Forecasting With Mixed-Frequency Data: Forecasting Output Growth in the United States," *Journal of Business and Economic Statistics* 26, 546-554. `https://doi.org/10.1198/073500108000000015`.

[20] Cogley Timothy, and Thomas J. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post-World War II U.S.," *Review of Economic Dynamics* 8, 262-302. https://doi.org/10.1016/j.red.2004.10.009.

[21] Cook, Thomas, and Taeyoung Doh (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.

[22] Croushore, Dean (2006), "Forecasting with Real-Time Macroeconomic Data," in *Handbook of Economic Forecasting*, Elliott, G., Granger, C., and Timmermann, A. (eds.). Amsterdam: North Holland. https://doi.org/10.1016/S1574-0706(05)01017-7.

[23] D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), "Macroeconomic Forecasting and Structural Change," *Journal of Applied Econometrics* 28, 82-101. https://doi.org/10.1002/jae.1257.

[24] De Nicolo, Gianni, and Marcella Lucchetta (2017), "Forecasting Tail Risks," *Journal of Applied Econometrics* 32, 159-170. https://doi.org/10.1002/jae.2509.

[25] Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263. https://doi.org/10.2307/1392185.

[26] Diron, Marie (2008), "Short-Term Forecasts of Euro Area Real GDP Growth: An Assessment of Real-Time Performance Based on Vintage Data," *Journal of Forecasting* 27, 371-390. https://doi.org/10.1002/for.1067.

[27] Eraker, Bjorn, Ching Wai (Jeremy) Chiu, Andrew T. Foerster, Tae Bong Kim, and Hernan D. Seoane (2015), "Bayesian Mixed Frequency VARs," *Journal of Financial Econometrics* 13, 698-721. https://doi.org/10.1093/jjfinec/nbu027.

[28] Faust, Jon, and Jonathan H. Wright (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.

[29] Faust, Jon, and Jonathan H. Wright (2013), "Forecasting Inflation," in *Handbook of Economic Forecasting, Volume 2*, Elliott, G., and Timmermann, A. (eds). Amsterdam: North Holland. https://doi.org/10.1016/B978-0-444-53683-9.00001-3.

[30] Ferrara, Laurent, M. Mogliani, and J.G. Sahuc (2019), "Real-time High Frequency Monitoring of Growth-at-Risk," mimeo.

[31] Ferrara, Laurent and Anna Simoni (2019), "When Are Google Data Useful to Nowcast GDP? An Approach Via Pre-Selection and Shrinkage," Banque de France Working Paper No. 717.

[32] Fissler, Tobias, and Johanna F. Ziegel (2016), "Higher Order Elicitability and Osband's Principle," *Annals of Statistics* 44, 1680-1707. https://doi.org/10.1214/16-AOS1439.

[33] Fissler, Tobias, Johanna Ziegel, and Tilmann Gneiting (2015), "Expected Shortfall is Jointly Elicitable with Value at Risk – Implications for Backtesting," *Risk*, available at: https://www.risk.net/node/2439862.

[34] Foroni, Claudia, Eric Ghysels, and Massimiliano Marcellino (2013), "Mixed-Frequency Vector Autoregressive Models," in *Advances in Econometrics: VAR Models in Macroeconomics, Financial Econometrics, and Forecasting — New Developments and Applications*, Fomby, T., Kilian, L. and Murphy, A. (eds.). https://doi.org/10.1108/S0731-9053(2013)0000031007.

[35] Foroni, Claudia, Massimiliano Marcellino, and Christian Schumacher (2015), "Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials," *Journal of the Royal Statistical Society* 178, 57-82. https://doi.org/10.1111/rssa.12043.

[36] Gaglianone, Wagner P., and Luiz Renato Lima (2012), "Constructing Density Forecasts from Quantile Regressions," *Journal of Money, Credit and Banking* 44, 1589-1607. https://doi.org/10.1111/j.1538-4616.2012.00545.x.

[37] Galbraith, John W., and Simon van Norden (2019), "Asymmetry in Unemployment Rate Forecast Errors," *International Journal of Forecasting* 35, 1613-1626. https://doi.org/10.1016/j.ijforecast.2018.11.006.

[38] Ghysels, Eric, Leonardo Iania, and Jonas Striaukas (2018), "Quantile-based Inflation Risk Models," Research Working Paper 349, National Bank of Belgium.

[39] Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov (2004), "The MIDAS Touch: Mixed Data Sampling Regression Models," mimeo, University of North Carolina.

[40] Ghysels, Eric (2016), "Macroeconomics and the Reality of Mixed Frequency Data," *Journal of Econometrics* 193, 294-314. https://doi.org/10.1016/j.jeconom.2016.04.008.

[41] Giacomini, Raffaella, and Ivana Komunjer (2005), "Evaluation and Combination of Conditional Quantile Forecasts," *Journal of Business and Economic Statistics* 23, 416-431. https://doi.org/10.1198/073500105000000018.

[42] Giannone, Domenico, Lucrezia Reichlin, and David Small (2008), "Nowcasting: The Real-Time Informational Content of Macroeconomic Data," *Journal of Monetary Economics* 55, 665-676. https://doi.org/10.1016/j.jmoneco.2008.05.010.

[43] Giglio, Stefano, Bryan Kelly, and Seth Pruitt (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.

[44] Gonzalez-Rivera, Gloria, Javier Maldonado, and Esther Ruiz (2019), "Growth in Stress," *International Journal of Forecasting* 35, 948-966. https://doi.org/10.1016/j.ijforecast.2019.04.006.

[45] Guerin, Pierre, and Massimiliano Marcellino (2013), "Markov Switching MIDAS Models," *Journal of Business and Economic Statistics* 31, 45-56. https://doi.org/10.1080/07350015.2012.727721.

[46] Khare, Kshitij, and James P. Hobert, (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.

[47] Kiley, Michael T. (2018), "Unemployment Risk," Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2018-067. https://doi.org/10.17016/FEDS.2018.067.

[48] Gefang, Deborah, Gary Koop, and Aubrey Poon (2020), "Computationally Efficient Inference in Large Bayesian Mixed Frequency VARs," *Economics Letters* 191, 109-120. https://doi.org/10.1016/j.econlet.2020.109120.

[49] Korobilis, Dimitris (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.

[50] Kozicki, Sharon, and Barak Hoffman (2004), "Rounding Error: A Distorting Influence on Index Data," *Journal of Money, Credit and Banking* 36, 319-338. https://doi.org/10.1353/mcb.2004.0056.

[51] Lewis, Daniel, Karel Mertens and James Stock (2020), "U.S. Economic Activity during the Early Weeks of the SARS-Cov-2 Outbreak," Federal Reserve Bank of Dallas Working Paper 2011. https://doi.org/10.24149/wp2011.

[52] Li, Qing, Ruibin Xi, and Nan Lin (2010), "Bayesian Regularized Quantile Regression," *Bayesian Analysis* 5, 533-556. https://doi.org/10.1214/10-BA521.

[53] Lima, Luiz Renato, Fanning Meng, and Lucas Godeiro (2019), "Quantile Forecasting with Mixed-Frequency Data," *International Journal of Forecasting*, forthcoming. https://doi.org/10.1016/j.ijforecast.2018.09.011.

[54] Manzan, Sebastiano (2015), "Forecasting the Distribution of Economic Variables in a Data-Rich Environment," *Journal of Business and Economic Statistics* 33, 144-164. https://doi.org/10.1080/07350015.2014.937436.

[55] Manzan, Sebastiano, and Dawit Zerom (2013), "Are Macroeconomic Variables Useful for Forecasting the Distribution of U.S. Inflation?" *International Journal of Forecasting* 29, 469-478. https://doi.org/10.1016/j.ijforecast.2013.01.005.

[56] Manzan, Sebastiano, and Dawit Zerom (2015), "Asymmetric Quantile Persistence and Predictability: The Case of US Inflation," *Oxford Bulletin of Economics and Statistics* 77, 297-318. https://doi.org/10.1111/obes.12065.

[57] Marcellino, Massimiliano, Mario Porqueddu, and Fabrizio Venditti (2016), "Short-Term GDP Forecasting with a Mixed-Frequency Dynamic Factor Model with Stochastic Volatility," *Journal of Business and Economic Statistics* 34, 118-127. https://doi.org/10.1080/07350015.2015.1006773.

[58] Mazzi, Gian Luigi, and James Mitchell (2019), "Nowcasting Euro Area GDP Growth Using Quantile Regression," manuscript.

[59] McCracken, Michael W., Michael T. Owyang, and Tatevik Sekhposyan (2020), "Real-Time Forecasting with a Large, Mixed Frequency, Bayesian VAR," manuscript.

[60] Plagborg-Moller, Mikkel, Lucrezia Reichlin, Giovanni Ricco, and Thomas Hasenzagl (2020), "When Is Growth at Risk?" *Brookings Papers on Economic Activity*, forthcoming.

[61] Primiceri, Giorgio E. (2005), "Time-Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies* 72, 821-852. https://doi.org/10.1111/j.1467-937X.2005.00353.x.

[62] Reichlin, Lucrezia, Giovanni Ricco, and Thomas Hasenzagl (2020), "Financial Variables as Predictors of Real Growth Vulnerability," Deutsche Bundesbank Discussion Paper No. 05/2020.

[63] Romer, Christina D., and David H. Romer (2000), "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review* 90, 429-457. https://doi.org/10.1257/aer.90.3.429.

[64] Schorfheide, Frank, and Dongho Song (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.

[65] Schorfheide, Frank, and Dongho Song (2020), "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic," manuscript, University of Pennsylvania.

[66] Sims, Christopher A. (2002), "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity* 2, 1-40. https://doi.org/10.1353/eca.2003.0009.

[67] Taylor, James 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.

[68] Taylor, James 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.

[69] Timmermann, Allan (2006), "Forecast Combinations," in *Handbook of Economic Forecasting*, Elliott, G., Granger, C., and Timmermann, A. (eds.). Amsterdam: North Holland. https://doi.org/10.1016/S1574-0706(05)01004-9

[70] Wallis, Kenneth F. (1986), "Forecasting with an Econometric Model: The 'Ragged Edge' Problem," *Journal of Forecasting* 5, 1-13. https://doi.org/10.1002/for.3980050102.

[71] West, Kenneth D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica* 64, 1067-1084. `https://doi.org/10.2307/2171956`.

[72] Yu, Keming, and Rana A. Moyeed (2001), "Bayesian Quantile Regression," *Statistics and Probability Letters* 54, 437-447. `https://doi.org/10.1016/S0167-7152(01)00124-9`.

A Appendix

This appendix provides information on the Bayesian priors and algorithms used with some of our models.

A.1 Priors for models estimated with Bayesian methods

In the case of the BMF-SV, BQR, and BQR-Lasso models, Bayesian estimation methods necessitate priors. For the BMF-SV models with stochastic volatility (the same approach is used for the BMF-factor-SV models), we use independent priors for the coefficients (normal distribution) and volatility components (details below). Since the form of the prior is not dependent on m, in spelling out the prior we drop the index m from the model parameters for notational simplicity. 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 (following Khare and Hobert (2012)). The BQR-Lasso specification has many similarities to the BQR case, but with a hierarchical structure and a Laplace (rather than Normal) prior on the regression coefficients, with independence across coefficients.

With BMF-SV and BQR specifications, the normal priors on the coefficient vector β 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 shrinkage increases with the lag (with the quarter, not with the month within the quarter), and in the sense that we take account of the relative scales of variables. The shrinkage is controlled by three hyperparameters (in all cases, a smaller number means more shrinkage): λ_1, which controls the overall rate of shrinkage; λ_2, which controls the rate of shrinkage on variables other than lags of the dependent variable; and λ_3, which determines the rate of shrinkage associated with longer lags of GDP growth (it is not applied with monthly variables).

At each forecast origin, the prior standard deviation associated with the coefficient on the monthly or weekly variable $X_{w,j,t}$ of $X_{w,t}$ is specified as follows:

$$\mathrm{sd}_{j,t} = \lambda_1 \lambda_2 \frac{\sigma_{GDP}}{\sigma_j}. \tag{13}$$

For coefficients on lag l of GDP, the prior standard deviation is

$$\mathrm{sd}_l = \frac{\lambda_1}{l^{\lambda_3}}. \tag{14}$$

Finally, for the intercept, the prior is uninformative:

$$\mathrm{sd}_{int} = 1000 \sigma_{GDP}. \tag{15}$$

In setting these components of the prior, for σ_{GDP} and σ_j we use standard deviations from AR(4) models for GDP growth and $X_{w,j,t}$ estimated with the available sample of data as of the forecast origin. In all of our results, we follow CCM and fix the hyperparameters at values that may be considered very common in Minnesota-type priors and forecasting: $\lambda_1 = 0.2$, $\lambda_2 = 0.2$, and $\lambda_3 = 1$.

In the prior for the volatility-related components of the model, our approach is similar to that used in such studies as Clark (2011), Cogley and Sargent (2005), and Primiceri (2005). For the prior on ϕ, we use a mean of 0.035 and 5 degrees of freedom. For the period 0 value of volatility, we use a prior of

$$\underline{\mu}_\lambda = \log \hat{\lambda}_{0,OLS}, \ \underline{\Omega}_\lambda = 4. \tag{16}$$

To obtain $\log \hat{\lambda}_{0,OLS}$, we use a training sample of 40 observations preceding the estimation sample to fit an AR(4) model to GDP growth.

For the scale parameter $\sigma_{\tau,w}$ of the BQR and BQR-Lasso models, we use an inverse Gamma prior with 5 degrees of freedom and, for simplicity, with the mean set at the standard deviation of the residuals from regressing GDP growth on the variables of the model over the sample. In the Gamma prior on the parameter η^2 that governs the regularization rate of the BQR-Lasso model, we set the scale parameter at 2 and set the shape parameter to make the prior mean equal 5.

A.2 Estimation algorithms

We estimate the BMF-SV models with a Metropolis-within-Gibbs algorithm, used in such studies as Clark (2011) and CCM. The posterior mean and variance of the coefficient vector are given by

$$\bar{\mu}_\beta = \bar{\Omega}_\beta \left\{ \sum_{t=1}^{T} \lambda_t^{-1} X_{w,t} y_t + \underline{\Omega}_\beta^{-1} \underline{\mu}_\beta \right\} \tag{17}$$

$$\bar{\Omega}_\beta^{-1} = \underline{\Omega}_\beta^{-1} + \sum_{t=1}^{T} \lambda_t^{-1} X_{w,t} X'_{w,t}, \tag{18}$$

where we again omit the w index from the parameters for notational simplicity. For the BMF-SV model and its variants, we obtain forecasts from the posterior predictive distribution. The point forecast is the posterior mean forecast, and we compute the quantiles of interest from the quantiles of forecast draws.

We estimate the Bayesian quantile regression with the three-step Gibbs sampling approach 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, with posterior mean and variance that can be expressed in the same basic form indicated above for the BMF-SV case.

Finally, we estimate BQR-Lasso models with the Gibbs sampler of Li, Xi, and Lin (2010), which shares a number of the aspects of the BQR algorithm. The first step samples the mixture state time series z from an inverse Gaussian distribution. The second draws variance scale parameters (denoted s_k for each parameter k in the notation of Li, Xi, and Lin (2010)) associated with each regression coefficient from an inverse Gaussian distribution. In the third step, each individual element of the regression parameter vector $\beta_{\tau,w}$ is drawn from its Normal conditional posterior, with posterior mean and variance of the same basic form as that of the BMF-SV and BQR cases. The fourth step draws the scale parameter $\sigma_{\tau,w}$ from its inverse Gamma conditional posterior, and the fifth draws from a Gamma distribution the parameter η^2 that governs the Lasso regularization rate.

The last aspect of estimation to mention is that our forecasts are produced by estimating the forecasting models with a recursive scheme: the estimation sample expands as forecasting moves forward in time. A rolling scheme, under which the size of the estimation sample remains fixed over time but the first observation moves forward in time, is in general less efficient but can be more robust in the presence of changes in regression parameters and (for density-related forecasts) error variances. However, in the nowcast (point and density) comparisons of CCM, recursive scheme forecasts were more accurate than rolling scheme forecasts.

Table 1: **Variables used**

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

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 available and used. The final column gives the week in which each indicator is commonly reported and which determines which variables enter our models at each forecast origin (our dating is based on end-of-week availability). As examples, GDP for quarter $t-1$ is typically reported in the last (fourth) week of month 1 of quarter t, employment for month $t-1$ is normally published in week 1 of month t, the NFCI for week $t-1$ is reported in week t, and Treasury yields and stock prices for week t are published in (at the end of) week t.

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}, $\text{emp}_{t-1}^{(m1,m2,m3)}$, $\text{ISM}_{t-1}^{(m1,m2,m3)}$, $\text{retail}_{t-1}^{(m1,m2,m3)}$, $\text{IP}_{t-1}^{(m1,m2)}$, $\text{starts}_{t-1}^{(m1,m2)}$, $\text{claims}_{t-1}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(w1)}$ $\text{NFCI}_{t-1}^{(m1,m2,m3)}$, $\text{SP}_{t-1}^{(m1,m2,m3)}$, $\text{TS}_{t-1}^{(m1,m2,m3)}$, $\text{CS}_{t-1}^{(m1,m2,m3)}$, $\text{NFCI}_{t}^{(w1)}$, $\text{SP}_{t}^{(w1+w2)}$, $\text{TS}_{t}^{(w1+w2)}$, $\text{CS}_{t}^{(w1+w2)}$ $\text{cclaims}_{t-1}^{(m1,m2,m3)}$, $\text{cclaims}_{t}^{(w1)}$, $\text{sment}_{t}^{(w1)}$, $\text{steel}_{t}^{(w1)}$, $\text{util}_{t}^{(w1)}$
3 (qrtr. t)	GDP_{t-2}, $\text{emp}_{t-1}^{(m1,m2,m3)}$, $\text{ISM}_{t-1}^{(m1,m2,m3)}$, $\text{retail}_{t-1}^{(m1,m2,m3)}$, $\text{IP}_{t-1}^{(m1,m2,m3)}$, $\text{starts}_{t-1}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(w1+w2)}$ $\text{NFCI}_{t}^{(w1+w2)}$, $\text{SP}_{t}^{(w1+w2+w3)}$, $\text{TS}_{t}^{(w1+w2+w3)}$, $\text{CS}_{t}^{(w1+w2+w3)}$ $\text{cclaims}_{t}^{(w1+w2)}$, $\text{sment}_{t}^{(w2)}$, $\text{steel}_{t}^{(w2)}$, $\text{util}_{t}^{(w2)}$
4 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t-1}^{(m1,m2,m3)}$, $\text{ISM}_{t-1}^{(m1,m2,m3)}$, $\text{retail}_{t-1}^{(m1,m2,m3)}$, $\text{IP}_{t-1}^{(m1,m2,m3)}$, $\text{starts}_{t-1}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(w1+w2+w3)}$ $\text{NFCI}_{t}^{(w1+w2+w3)}$, $\text{SP}_{t}^{(w1+w2+w3+w4)}$, $\text{TS}_{t}^{(w1+w2+w3+w4)}$, $\text{CS}_{t}^{(w1+w2+w3+w4)}$ $\text{cclaims}_{t}^{(w1+w2+w3)}$, $\text{sment}_{t}^{(w3)}$, $\text{steel}_{t}^{(w3)}$, $\text{util}_{t}^{(w3)}$
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}, $\text{emp}_{t}^{(m1)}$, $\text{ISM}_{t}^{(m1)}$, $\text{retail}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,w5)}$ $\text{NFCI}_{t}^{(m1,w5)}$, $\text{SP}_{t}^{(m1,w5+w6)}$, $\text{TS}_{t}^{(m1,w5+w6)}$, $\text{CS}_{t}^{(m1,w5+w6)}$ $\text{cclaims}_{t}^{(m1,w5)}$, $\text{sment}_{t}^{(w5)}$, $\text{steel}_{t}^{(w5)}$, $\text{util}_{t}^{(w5)}$
7 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1)}$, $\text{ISM}_{t}^{(m1)}$, $\text{retail}_{t}^{(m1)}$, $\text{IP}_{t}^{(m1)}$, $\text{starts}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,w5+w6)}$ $\text{NFCI}_{t}^{(m1,w5+w6)}$, $\text{SP}_{t}^{(m1,w5+w6+w7)}$, $\text{TS}_{t}^{(m1,w5+w6+w7)}$, $\text{CS}_{t}^{(m1,w5+w6+w7)}$ $\text{cclaims}_{t}^{(m1,w5+w6)}$, $\text{sment}_{t}^{(w6)}$, $\text{steel}_{t}^{(w6)}$, $\text{util}_{t}^{(w6)}$
8 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1)}$, $\text{ISM}_{t}^{(m1)}$, $\text{retail}_{t}^{(m1)}$, $\text{IP}_{t}^{(m1)}$, $\text{starts}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,w5+w6+w7)}$ $\text{NFCI}_{t}^{(m1,w5+w6+w7)}$, $\text{SP}_{t}^{(m1,w5+w6+w7+w8)}$, $\text{TS}_{t}^{(m1,w5+w6+w7+w8)}$, $\text{CS}_{t}^{(m1,w5+w6+w7+w8)}$ $\text{cclaims}_{t}^{(m1,w5+w6+w7)}$, $\text{sment}_{t}^{(w7)}$, $\text{steel}_{t}^{(w7)}$, $\text{util}_{t}^{(w7)}$
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}, $\text{emp}_{t}^{(m1,m2)}$, $\text{ISM}_{t}^{(m1,m2)}$, $\text{retail}_{t}^{(m1,m2)}$, $\text{IP}_{t}^{(m1)}$, $\text{starts}_{t}^{(m1)}$, $\text{claims}_{t}^{(m1,m2,w9)}$ $\text{NFCI}_{t}^{(m1,m2,w9)}$, $\text{SP}_{t}^{(m1,m2,w9+w10)}$, $\text{TS}_{t}^{(m1,m2,w9+w10)}$, $\text{CS}_{t}^{(m1,m2,w9+w10)}$ $\text{cclaims}_{t}^{(m1,m2,w9)}$, $\text{sment}_{t}^{(w9)}$, $\text{steel}_{t}^{(w9)}$, $\text{util}_{t}^{(w9)}$
11 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2)}$, $\text{ISM}_{t}^{(m1,m2)}$, $\text{retail}_{t}^{(m1,m2)}$, $\text{IP}_{t}^{(m1,m2)}$, $\text{starts}_{t}^{(m1,m2)}$, $\text{claims}_{t}^{(m1,m2,w9+w10)}$ $\text{NFCI}_{t}^{(m1,m2,w9+w10)}$, $\text{SP}_{t}^{(m1,m2,w9+w10+w11)}$, $\text{TS}_{t}^{(m1,m2,w9+w10+w11)}$, $\text{CS}_{t}^{(m1,m2,w9+w10+w11)}$ $\text{cclaims}_{t}^{(m1,m2,w9+w10)}$, $\text{sment}_{t}^{(w10)}$, $\text{steel}_{t}^{(w10)}$, $\text{util}_{t}^{(w10)}$
12 (qrtr. t)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2)}$, $\text{ISM}_{t}^{(m1,m2)}$, $\text{retail}_{t}^{(m1,m2)}$, $\text{IP}_{t}^{(m1,m2)}$, $\text{starts}_{t}^{(m1,m2)}$, $\text{claims}_{t}^{(m1,m2,w9+w10+w11)}$ $\text{NFCI}_{t}^{(m1,m2,w9+w10+w11)}$, $\text{SP}_{t}^{(m1,m2,w9+w10+w11+w12)}$, $\text{TS}_{t}^{(m1,m2,w9+w10+w11+w12)}$, $\text{CS}_{t}^{(m1,m2,w9+w10+w11+w12)}$ $\text{cclaims}_{t}^{(m1,m2,w9+w10+w11)}$, $\text{sment}_{t}^{(w11)}$, $\text{steel}_{t}^{(w11)}$, $\text{util}_{t}^{(w11)}$
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}, $\text{emp}_{t}^{(m1,m2,m3)}$, $\text{ISM}_{t}^{(m1,m2,m3)}$, $\text{retail}_{t}^{(m1,m2,m3)}$, $\text{IP}_{t}^{(m1,m2)}$, $\text{starts}_{t}^{(m1,m2)}$, $\text{claims}_{t}^{(m1,m2,m3,w13)}$ $\text{NFCI}_{t}^{(m1,m2,m3,w13)}$, $\text{SP}_{t}^{(m1,m2,m3,w13+w14)}$, $\text{TS}_{t}^{(m1,m2,m3,w13+w14)}$, $\text{CS}_{t}^{(m1,m2,m3,w13+w14)}$ $\text{cclaims}_{t}^{(m1,m2,m3,w13)}$, $\text{sment}_{t}^{(w13)}$, $\text{steel}_{t}^{(w13)}$, $\text{util}_{t}^{(w13)}$
15 (qrtr. $t+1$)	GDP_{t-1}, $\text{emp}_{t}^{(m1,m2,m3)}$, $\text{ISM}_{t}^{(m1,m2,m3)}$, $\text{retail}_{t}^{(m1,m2,m3)}$, $\text{IP}_{t}^{(m1,m2,m3)}$, $\text{starts}_{t}^{(m1,m2,m3)}$, $\text{claims}_{t}^{(m1,m2,m3,w13+w14)}$ $\text{NFCI}_{t}^{(m1,m2,m3,w13+w14)}$, $\text{SP}_{t}^{(m1,m2,m3,w13+w14+w15)}$, $\text{TS}_{t}^{(m1,m2,m3,w13+w14+w15)}$, $\text{CS}_{t}^{(m1,m2,m3,w13+w14+w15)}$ $\text{cclaims}_{t}^{(m1,m2,m3,w13+w14)}$, $\text{sment}_{t}^{(w14)}$, $\text{steel}_{t}^{(w14)}$, $\text{util}_{t}^{(w14)}$

Notes: For each week indicated in the first column, the table has three rows of entries, with the first listing the relevant base macro indicators, the second row covering the base 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, 1985:Q1-2019:Q3**

variable and model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
	5% quantile score							
base M: QR	0.31	0.30	0.25	0.27	0.26	0.26	0.26	0.26
base M: QR-Lasso	0.86 **	0.88 **	0.98	0.74 ***	0.76 ***	0.72 ***	0.71 ***	0.72 ***
base M: BQR	0.97	0.91 **	1.10	0.90 ***	0.90 ***	0.86 ***	0.86 ***	0.85 ***
base M: BQR-Lasso	0.73 ***	0.80 ***	0.99	0.89	0.93	0.94	0.92	0.91
base M: BMF-SV	0.78 ***	0.72 ***	0.83	0.72 ***	0.72 ***	0.71 ***	0.68 ***	0.65 ***
base M + NFCI: QR	0.93 ***	0.98	0.90 ***	0.89 ***	0.92 ***	0.91 ***	0.92 ***	0.92 ***
base M + NFCI: QR-Lasso	0.64 ***	0.73 ***	0.83 ***	0.71 ***	0.69 ***	0.67 ***	0.66 ***	0.68 ***
base M + NFCI: BQR	0.73 ***	0.74 ***	0.85 ***	0.72 ***	0.73 ***	0.70 ***	0.67 ***	0.64 ***
base M + NFCI: BQR-Lasso	0.65 ***	0.66 ***	0.71 ***	0.80 *	0.82	0.86	0.87	0.79 *
base M + NFCI: BMF-SV	0.69 ***	0.67 ***	0.77 **	0.67 ***	0.69 ***	0.69 ***	0.67 ***	0.64 ***
base M-F: QR	0.98	1.00	1.08	0.95 **	1.00	0.98 *	1.01	0.98 *
base M-F: QR-Lasso	0.81 *	0.77 **	0.99	0.83	0.82	0.87	0.81 *	0.75 **
base M-F: QR-factor	0.74 ***	0.76 ***	0.85 *	0.62 ***	0.73 ***	0.64 ***	0.74 ***	0.69 ***
base M-F: BQR	0.71 ***	0.67 ***	0.76 ***	0.62 ***	0.68 ***	0.65 ***	0.65 ***	0.66 ***
base M-F: BQR-Lasso	0.69 ***	0.63 ***	0.76 **	0.69 ***	0.67 ***	0.76 **	0.78 **	0.74 **
base M-F: BQR-factor	0.67 ***	0.71 ***	0.77 ***	0.62 ***	0.69 ***	0.63 ***	0.72 ***	0.68 ***
base M-F: PQR	0.66 ***	0.70 ***	0.74 ***	0.61 ***	0.70 ***	0.65 ***	0.70 ***	0.67 ***
base M-F: BMF-SV	0.69 ***	0.66 ***	0.76 ***	0.63 ***	0.68 ***	0.70 ***	0.71 ***	0.71 ***
base M-F: BMF-factor-SV	0.69 ***	0.61 ***	0.78 ***	0.68 ***	0.68 ***	0.64 ***	0.66 ***	0.64 ***
avg. all	0.69 ***	0.70 ***	0.79 ***	0.68 ***	0.71 ***	0.69 ***	0.69 ***	0.67 ***
avg. base M-F	0.65 ***	0.62 ***	0.71 ***	0.61 ***	0.64 ***	0.62 ***	0.65 ***	0.64 ***
avg. BQR	0.77 ***	0.76 ***	0.89 ***	0.74 ***	0.76 ***	0.71 ***	0.70 ***	0.67 ***
avg. BQR-Lasso	0.62 ***	0.62 ***	0.75 ***	0.76 ***	0.78 **	0.80 *	0.83	0.77 **
avg. BMF-SV	0.70 ***	0.68 ***	0.77 **	0.67 ***	0.69 ***	0.69 ***	0.67 ***	0.66 ***
	5% coverage							
base M: QR	0.02 *	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***
base M: QR-Lasso	0.02 **	0.02 ***	0.01 ***	0.01 ***	0.01 ***	0.02 **	0.02 **	0.02 **
base M: BQR	0.01 **	0.01 ***	0.02 *	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***
base M: BQR-Lasso	0.02 **	0.02 **	0.03	0.04	0.02 **	0.04	0.02 **	0.03
base M: BMF-SV	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03 *
base M + NFCI: QR	0.02 *	0.02 *	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***
base M + NFCI: QR-Lasso	0.04	0.07	0.04	0.05	0.02 **	0.06	0.04	0.04
base M + NFCI: BQR	0.03	0.01 ***	0.02 *	0.02 **	0.02 **	0.03	0.03	0.04
base M + NFCI: BQR-Lasso	0.07	0.05	0.04	0.06	0.06	0.05	0.07	0.08
base M + NFCI: BMF-SV	0.04	0.05	0.04	0.04	0.04	0.04	0.04	0.04
base M-F: QR	0.02 *	0.02 *	0.02 *	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***
base M-F: QR-Lasso	0.11 *	0.07	0.06	0.09	0.08	0.09	0.07	0.09
base M-F: QR-factor	0.07	0.06	0.07	0.04	0.08	0.06	0.05	0.05
base M-F: BQR	0.04	0.01 ***	0.03	0.03	0.04	0.03	0.03	0.04
base M-F: BQR-Lasso	0.09 **	0.08	0.04	0.09 *	0.07	0.09	0.09 *	0.07
base M-F: BQR-factor	0.05	0.04	0.06	0.04	0.06	0.05	0.06	0.04
base M-F: PQR	0.04	0.01 ***	0.05	0.04	0.01 ***	0.01 ***	0.01 ***	0.01 ***
base M-F: BMF-SV	0.08	0.05	0.06	0.06	0.04	0.04	0.04	0.04
base M-F: BMF-factor-SV	0.04	0.04	0.04	0.03 *	0.04	0.04	0.03 *	0.03 *
avg. all	0.02 **	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.02 **	0.02 **	0.02 **
avg. base M-F	0.04	0.02 **	0.04	0.03 *	0.01 ***	0.02 **	0.02 **	0.03 *
avg. BQR	0.03	0.01 ***	0.02 *	0.02 **	0.02 **	0.02 **	0.02 **	0.01 ***
avg. BQR-Lasso	0.05	0.02 **	0.02 **	0.04	0.04	0.04	0.04	0.05
avg. BMF-SV	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04

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 panel, the top row gives the 5% 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). The lower panel reports empirical coverage rates for 5% quantile forecasts (percentage of outcomes at or below the quantile). 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. Statistical significance of departures of empirical coverage from the nominal 5% is also indicated by *** (1%), ** (5%), or * (10%), obtained with two-sided t-tests.

Table 4: **Out-of-sample forecast accuracy: 5% quantile score, 2000:Q1-2019:Q3**

variable and model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
base M: QR	0.31	0.29	0.22	0.25	0.24	0.24	0.23	0.24
base M: QR-Lasso	0.88	0.85 *	1.12	0.73 **	0.78 **	0.75 ***	0.76 ***	0.80 **
base M: BQR	0.96	0.90 *	1.12	0.89 **	0.89 ***	0.87 ***	0.88 ***	0.88 **
base M: BQR-Lasso	0.75 **	0.78 **	0.95	1.07	1.15	1.22	1.19	1.18
base M: BMF-SV	0.88	0.80 **	1.07	0.84 *	0.84 *	0.83 **	0.78 ***	0.77 **
base M + NFCI: QR	0.94 **	0.99	0.93 ***	0.90 ***	0.92 ***	0.92 ***	0.93 ***	0.93 ***
base M + NFCI: QR-Lasso	0.77 *	0.78 **	0.95	0.77 **	0.75 **	0.75 **	0.73 **	0.81
base M + NFCI: BQR	0.79 **	0.75 **	0.86 *	0.81 **	0.81 **	0.80 **	0.76 ***	0.75 **
base M + NFCI: BQR-Lasso	0.71 **	0.66 ***	0.83 *	1.01	1.08	1.12	1.20	1.09
base M + NFCI: BMF-SV	0.81 **	0.78 **	0.98	0.80 ***	0.81 **	0.80 ***	0.78 ***	0.76 ***
base M-F: QR	1.03	1.02	1.15	0.96	0.99	0.97	1.01	0.99
base M-F: QR-Lasso	0.86	0.90	1.23	0.97	0.96	1.11	1.02	0.94
base M-F: QR-factor	0.73 *	0.63 ***	0.92	0.65 **	0.78	0.70 **	0.82	0.76
base M-F: BQR	0.80 *	0.75 **	0.83	0.68 ***	0.75 **	0.71 **	0.75 **	0.79 *
base M-F: BQR-Lasso	0.65 **	0.67 **	0.77 *	0.78	0.77 **	0.91	1.02	0.94
base M-F: BQR-factor	0.72 **	0.63 ***	0.89	0.65 **	0.77 *	0.72 *	0.82	0.77
base M-F: PQR	0.69 **	0.78 ***	0.81 **	0.67 ***	0.76 **	0.71 ***	0.75 **	0.75 **
base M-F: BMF-SV	0.77 **	0.76 **	0.88	0.69 ***	0.71 ***	0.71 ***	0.73 ***	0.72 ***
base M-F: BMF-factor-SV	0.76 **	0.65 ***	0.84 *	0.69 **	0.73 **	0.74 **	0.77 *	0.76 *
base M-F + small weekly: QR	1.01	1.03	1.07	0.96	0.89	0.87 *	0.90	0.89
base M-F + small weekly: QR-Lasso	0.91	1.00	1.19	1.10	1.18	1.30	1.11	1.22
base M-F + small weekly: QR-factor	0.96	0.94	1.14	0.92	0.92	1.12	1.00	0.99
base M-F + small weekly: BQR	0.78 **	0.87	0.91	0.71 ***	0.71 ***	0.72 ***	0.75 **	0.77 *
base M-F + small weekly: BQR-Lasso	0.86	1.11	1.31	0.90	0.90	1.07	1.22	0.93
base M-F + small weekly: BQR-factor	0.76 **	0.79 *	0.78 **	0.72 **	0.76 **	0.75 *	0.79 *	0.80
base M-F + small weekly: PQR	0.75 **	0.71 ***	0.90 *	0.71 ***	0.78 **	0.80 **	0.85	0.83 *
base M-F + small weekly: BMF-SV	0.78 **	0.76 **	0.87	0.78 *	0.79	0.76 *	0.78 *	0.78 *
base M-F + small weekly: BMF-factor-SV	0.72 ***	0.80	0.79 **	0.72 **	0.76 **	0.74 **	0.77 *	0.81
avg. all	0.73 ***	0.73 ***	0.86 **	0.72 ***	0.74 ***	0.74 ***	0.73 ***	0.71 ***
avg. base M-F	0.70 ***	0.66 ***	0.78 **	0.64 ***	0.69 ***	0.67 ***	0.71 **	0.71 **
avg. base M-F + small weekly	0.72 ***	0.83	0.85 *	0.70 ***	0.69 ***	0.72 **	0.71 **	0.71 **
avg. BQR	0.81 ***	0.78 ***	0.90	0.76 ***	0.76 ***	0.72 ***	0.71 ***	0.71 ***
avg. BQR-Lasso	0.65 **	0.69 **	0.83 **	0.85	0.84	0.98	0.96	0.93
avg. BMF-SV	0.78 **	0.77 **	0.95	0.76 ***	0.76 **	0.75 ***	0.73 ***	0.69 ***

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). The top row gives the 5% 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). 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: 5% coverage, 2000:Q1-2019:Q3**

variable and model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
base M: QR	0.04	0.01 ***	0.03	0.01 ***	0.01 ***	0.01 ***	0.03	0.03
base M: QR-Lasso	0.04	0.01 ***	0.03	0.03	0.03	0.04	0.04	0.04
base M: BQR	0.03	0.01 ***	0.04	0.01 ***	0.03	0.03	0.03	0.03
base M: BQR-Lasso	0.03	0.01 ***	0.05	0.05	0.04	0.06	0.04	0.05
base M: BMF-SV	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.04
base M + NFCI: QR	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03
base M + NFCI: QR-Lasso	0.06	0.08	0.06	0.08	0.04	0.10 *	0.05	0.06
base M + NFCI: BQR	0.05	0.03	0.04	0.04	0.04	0.05	0.05	0.06
base M + NFCI: BQR-Lasso	0.08	0.06	0.06	0.06	0.08	0.08	0.10 *	0.11 **
base M + NFCI: BMF-SV	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05
base M-F: QR	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03
base M-F: QR-Lasso	0.11	0.10	0.10	0.15 **	0.14 *	0.15 **	0.13 **	0.15 **
base M-F: QR-factor	0.05	0.04	0.06	0.05	0.08	0.08	0.06	0.06
base M-F: BQR	0.06	0.03	0.05	0.04	0.05	0.04	0.04	0.04
base M-F: BQR-Lasso	0.10 *	0.08	0.04	0.10 *	0.09	0.10 *	0.13 **	0.08
base M-F: BQR-factor	0.06	0.01 ***	0.06	0.04	0.08	0.06	0.06	0.06
base M-F: PQR	0.04	0.01 ***	0.03	0.04	0.03	0.03	0.03	0.01 ***
base M-F: BMF-SV	0.08	0.06	0.06	0.05	0.03	0.04	0.04	0.05
base M-F: BMF-factor-SV	0.04	0.04	0.04	0.03	0.05	0.04	0.03	0.04
base M-F + small weekly: QR	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03
base M-F + small weekly: QR-Lasso	0.10	0.08	0.05	0.10	0.16 **	0.13 **	0.11	0.18 **
base M-F + small weekly: QR-factor	0.09	0.09	0.09	0.10	0.08	0.10 *	0.09	0.09
base M-F + small weekly: BQR	0.05	0.06	0.06	0.05	0.05	0.06	0.06	0.05
base M-F + small weekly: BQR-Lasso	0.13 **	0.13 *	0.18 ***	0.16 ***	0.14 ***	0.18 ***	0.19 ***	0.16 ***
base M-F + small weekly: BQR-factor	0.04	0.04	0.04	0.05	0.05	0.06	0.08	0.08
base M-F + small weekly: PQR	0.04	0.03	0.04	0.05	0.03	0.05	0.06	0.05
base M-F + small weekly: BMF-SV	0.08	0.08	0.05	0.08	0.11 **	0.09	0.11 *	0.15 **
base M-F + small weekly: BMF-factor-SV	0.03	0.05	0.04	0.04	0.04	0.04	0.04	0.05
avg. all	0.04	0.03	0.04	0.04	0.03	0.04	0.04	0.04
avg. base M-F	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04
avg. base M-F + small weekly	0.04	0.05	0.04	0.04	0.06	0.06	0.08	0.05
avg. BQR	0.05	0.04	0.04	0.04	0.04	0.05	0.04	0.05
avg. BQR-Lasso	0.05	0.05	0.06	0.06	0.06	0.08	0.06	0.08
avg. BMF-SV	0.05	0.06	0.05	0.04	0.05	0.05	0.05	0.04

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). The table reports empirical coverage rates for 5% quantile forecasts (percentage of outcomes at or below the quantile). Statistical significance of departures of empirical coverage from the nominal 5% is also indicated by *** (1%), ** (5%), or * (10%), obtained with two-sided t-tests.

Table 6: **Out-of-sample forecast accuracy: 5% quantile score, 2007:Q1-2019:Q3**

variable and model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
base M: QR	0.33	0.29	0.21	0.25	0.24	0.24	0.23	0.23
base M: QR-Lasso	0.87	0.87	1.31	0.73 *	0.81 *	0.78 **	0.82	0.88
base M: BQR	0.99	0.93	1.15	0.94	0.93	0.91 ***	0.92 *	0.92
base M: BQR-Lasso	0.69 *	0.72 *	0.98	1.13	1.22	1.15	1.21	1.24
base M: BMF-SV	1.02	0.88	1.28	0.92	0.93	0.87	0.86	0.84
base M + NFCI: QR	0.98	1.03	0.99	0.91 ***	0.94 ***	0.93 ***	0.95 *	0.95
base M + NFCI: QR-Lasso	0.73 **	0.82	1.15	0.84	0.80	0.73 **	0.77	0.85
base M + NFCI: BQR	0.85	0.77 **	0.98	0.85 **	0.87 *	0.82 *	0.80 *	0.79 *
base M + NFCI: BQR-Lasso	0.72 *	0.68 *	0.94	1.04	1.12	1.11	1.25	1.14
base M + NFCI: BMF-SV	0.89	0.86	1.17	0.84 **	0.87	0.84 *	0.84	0.81
base M-F: QR	1.11	1.07	1.27	0.95	0.99	0.97	1.02	1.01
base M-F: QR-Lasso	0.73 **	0.88	1.56	1.16	1.14	1.22	1.18	1.02
base M-F: QR-factor	0.81	0.67 **	1.12	0.72	0.86	0.80	0.88	0.89
base M-F: BQR	0.84	0.78	1.00	0.69 **	0.78 *	0.74 *	0.78	0.87
base M-F: BQR-Lasso	0.69	0.70	0.92	0.80	0.76 **	0.84	1.01	0.90
base M-F: BQR-factor	0.79	0.67 **	1.07	0.71	0.86	0.80	0.86	0.89
base M-F: PQR	0.72 *	0.82 *	0.87	0.70 *	0.78 *	0.78 *	0.82	0.83
base M-F: BMF-SV	0.83	0.83	1.01	0.69 **	0.75 *	0.75 *	0.78 *	0.79
base M-F: BMF-factor-SV	0.81	0.66 **	0.94	0.74	0.80	0.80	0.84	0.88
base M-F + small weekly: QR	1.21	1.27	1.34	1.11	1.01	0.98	1.03	1.02
base M-F + small weekly: QR-Lasso	0.92	1.06	1.55	1.16	1.40	1.42	1.26	1.28
base M-F + small weekly: QR-factor	1.14	1.06	1.33	0.99	1.06	1.33	1.18	1.20
base M-F + small weekly: BQR	0.90	1.02	1.02	0.67 **	0.72 *	0.75 *	0.83	0.89
base M-F + small weekly: BQR-Lasso	0.95	1.34	1.64	0.88	0.91	1.06	1.01	0.95
base M-F + small weekly: BQR-factor	0.84	0.93	0.81	0.72 *	0.81	0.80	0.87	0.91
base M-F + small weekly: PQR	0.80 *	0.76 ***	0.99	0.71 **	0.81 *	0.86	0.91	0.93
base M-F + small weekly: BMF-SV	0.85	0.83	1.00	0.73 *	0.83	0.76	0.82	0.80
base M-F + small weekly: BMF-factor-SV	0.78 *	0.89	0.85	0.72	0.78	0.77	0.84	0.90
base M-F + large weekly: QR	1.29	1.41	1.51	1.24	1.11	1.08	1.15	1.15
base M-F + large weekly: QR-Lasso	1.33	1.51	2.08	1.77	1.41	1.45	1.17	1.56
base M-F + large weekly: QR-factor	0.92	1.40	0.99	0.92	0.97	0.93	1.18	1.36
base M-F + large weekly: BQR-factor	0.86	0.87	0.97	0.81	0.88	0.86	0.90	0.97
base M-F + large weekly: PQR	1.00	1.08	1.29	0.91	0.93	0.95	1.04	1.02
base M-F + large weekly: BMF-factor-SV	0.78 **	0.85	0.88	0.69 *	0.85	0.79	0.85	0.88
avg. all	0.79 *	0.84	0.99	0.76 **	0.79 *	0.78 *	0.79	0.81
avg. base M-F	0.74 *	0.69 **	0.86	0.66 **	0.73 *	0.71 *	0.76	0.79
avg. base M-F + small weekly	0.82	0.99	0.99	0.73 **	0.72 *	0.76 *	0.76	0.79
avg. base M-F + large weekly	0.91	1.11	1.18	0.79 *	0.77 *	0.77 *	0.84	0.99
avg. BQR	0.89	0.85	1.01	0.78 ***	0.81 **	0.76 **	0.78 *	0.80
avg. BQR-Lasso	0.67 *	0.73	0.97	0.86	0.83	0.92	0.97	0.96
avg. BMF-SV	0.87	0.85	1.11	0.80 **	0.82 *	0.78 *	0.79 *	0.76 *

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). The top row gives the 5% 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). Statistical significance of differences in MSEs and 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 7: **Out-of-sample forecast accuracy, 5% VaR-ES scores**

variable and model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
	1985:Q1-2019:Q3							
base M: QR	0.86	0.84	0.65	0.74	0.73	0.74	0.74	0.76
base M: BMF-SV	0.26 ***	0.35 ***	0.28 ***	0.47 ***	0.49 ***	0.38 ***	0.46 ***	0.49 ***
base M + NFCI: BMF-SV	0.42 ***	0.42 ***	0.36 ***	0.49 ***	0.44 ***	0.36 ***	0.42 ***	0.45 ***
base M-F: BMF-SV	0.30 **	0.38 ***	0.28 ***	0.40 ***	0.25 *	0.11	0.11	-0.00
base M-F: BMF-factor-SV	0.23 *	0.47 ***	0.05	0.14	0.29 *	0.45 ***	0.40 **	0.39 *
	2000:Q1-2019:Q3							
base M: QR	0.86	0.83	0.57	0.69	0.69	0.70	0.70	0.71
base M: BMF-SV	0.23 *	0.32 ***	0.11	0.44 ***	0.43 ***	0.32 **	0.39 ***	0.34 **
base M + NFCI: BMF-SV	0.34 **	0.31 **	0.17	0.45 ***	0.42 ***	0.36 **	0.35 ***	0.35 **
base M-F: BMF-SV	0.27	0.33 **	0.14	0.49 ***	0.46 ***	0.42 ***	0.39 **	0.40 **
base M-F: BMF-factor-SV	0.24	0.43 **	0.04	0.33	0.29	0.32	0.21	0.16
base M-F + small weekly: BMF-SV	0.20	0.48 **	0.13	0.20	0.09	0.11	0.05	-0.04
base M-F + small weekly: BMF-factor-SV	0.36 *	0.45 ***	0.16	0.31	0.24	0.28	0.21	0.02
	2007:Q1-2019:Q3							
base M: QR	0.90	0.87	0.53	0.73	0.72	0.74	0.73	0.75
base M: BMF-SV	0.00	0.16	0.02	0.30 ***	0.27 **	0.29 ***	0.27 **	0.20
base M + NFCI: BMF-SV	0.13	0.09	0.04	0.34 ***	0.30 ***	0.29 ***	0.22 *	0.19
base M-F: BMF-SV	0.00	0.15	-0.01	0.32 **	0.23	0.24 *	0.19	0.15
base M-F: BMF-factor-SV	0.02	0.25	-0.25	0.04	-0.02	0.06	-0.10	-0.24
base M-F + small weekly: BMF-SV	-0.10	0.28	-0.13	0.24	-0.17	0.02	-0.17	-0.09
base M-F + small weekly: BMF-factor-SV	0.17	0.34 *	-0.01	0.12	0.04	0.09	-0.07	-0.33
base M-F + large weekly: BMF-factor-SV	0.27	0.32	0.08	0.28	-0.13	0.06	-0.20	-0.36

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). The first row gives the 5% VaR-ES from the benchmark model and variable set, and other rows report the difference in score for the indicated variable set and model relative to the benchmark (higher is better). 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 8: **Out-of-sample forecast accuracy during recessions: 5% QS and VaR-ES score, 1985:Q1-2019:Q3**

variable and model	week 1	week 3	week 5	week 7	week 9	week 11	week 13	week 15
	5% quantile score							
base M: QR	0.52	0.33	0.28	0.36	0.31	0.30	0.27	0.24
base M: QR-Lasso	0.44	0.49	1.31	0.58	0.71	0.73	0.68	1.05
base M: BQR	0.65	0.60	1.28	1.12	1.10	0.95	0.98	0.97
base M: BQR-Lasso	0.49	0.47	0.98	2.13	2.39	2.10	2.55	2.60
base M: BMF-SV	1.37	1.24	1.96	1.37	1.32	1.20	1.00	0.89
base M + NFCI: QR	1.03	1.19	0.92	0.88	0.89	0.86	0.88	0.88
base M + NFCI: QR-Lasso	1.25	1.07	1.13	0.85	0.80	0.84	0.52	0.69
base M + NFCI: BQR	1.02	0.80	0.90	1.07	1.10	1.11	1.01	0.92
base M + NFCI: BQR-Lasso	0.77	0.65	0.91	1.90	2.13	2.28	2.54	1.98
base M + NFCI: BMF-SV	1.29	1.28	1.78	1.19	1.19	1.09	0.93	0.87
base M-F: QR	1.40	1.36	1.37	0.78	0.82	0.83	0.95	0.97
base M-F: QR-Lasso	1.42	2.11	2.82	2.15	1.91	2.56	2.25	1.93
base M-F: QR-factor	0.59	0.42	0.49	0.30	0.36	0.41	0.87	0.70
base M-F: BQR	1.04	0.83	0.70	0.58	0.75	0.66	0.88	0.94
base M-F: BQR-Lasso	0.31	0.42	1.12	0.45	0.60	0.58	1.20	0.69
base M-F: BQR-factor	0.62	0.44	0.42	0.29	0.38	0.52	0.88	0.67
base M-F: PQR	0.48	0.63	0.76	0.33	0.53	0.48	0.63	0.72
base M-F: BMF-SV	1.04	1.09	1.32	0.66	0.57	0.57	0.62	0.64
base M-F: BMF-factor-SV	1.02	0.54	0.61	0.37	0.43	0.53	0.67	0.63
avg. all	0.64	0.53	0.76	0.74	0.82	0.74	0.78	0.67
avg. base M-F	0.68	0.41	0.44	0.37	0.43	0.41	0.58	0.59
avg. BQR	0.80	0.62	0.96	0.87	0.89	0.75	0.77	0.64
avg. BQR-Lasso	0.39	0.37	0.60	1.43	1.66	1.57	2.01	1.39
avg. BMF-SV	1.23	1.18	1.67	1.07	1.00	0.90	0.82	0.75
	5% VaR-ESscore							
base M: QR	1.28	1.09	1.03	1.11	1.06	1.05	1.02	0.99
base M: BMF-SV	-0.38	-0.13	-0.31	-0.64	-0.27	-0.16	0.01	0.07
base M + NFCI: BMF-SV	-0.53	-0.26	-0.27	-0.59	-0.41	-0.24	-0.05	-0.00
base M-F: BMF-SV	-0.34	-0.12	-0.18	-0.31	0.02	0.10	0.06	0.11
base M-F: BMF-factor-SV	-0.38	0.13	0.11	0.27	0.22	0.15	0.04	0.14

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 upper panel, the first row gives the 5% 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). In the lower panel, the first row gives the 5% VaR-ES from the benchmark model and variable set, and other rows report the difference in score for the indicated variable set and model relative to the benchmark (higher is better).

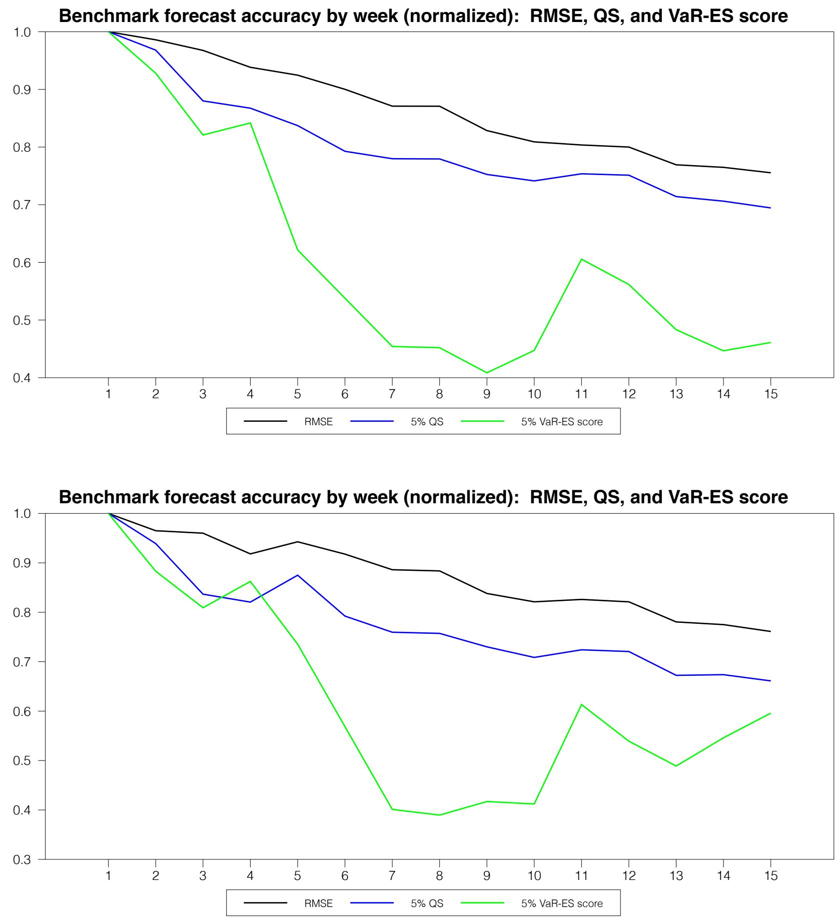

Figure 1: Out-of-sample forecast accuracy of base M: BMF-SV forecasts: levels of RMSE, 5% QS, and 5% VaR-ES score across weeks 1 through 15 of forecast origins are indicated on the horizontal axis. The benchmark forecasts come from the BMF-SV model estimated with the base macro variable set. The top and bottom panels provide results for the 1985:Q1-2019:Q3 and 2000:Q1-2019:Q3 samples, respectively.

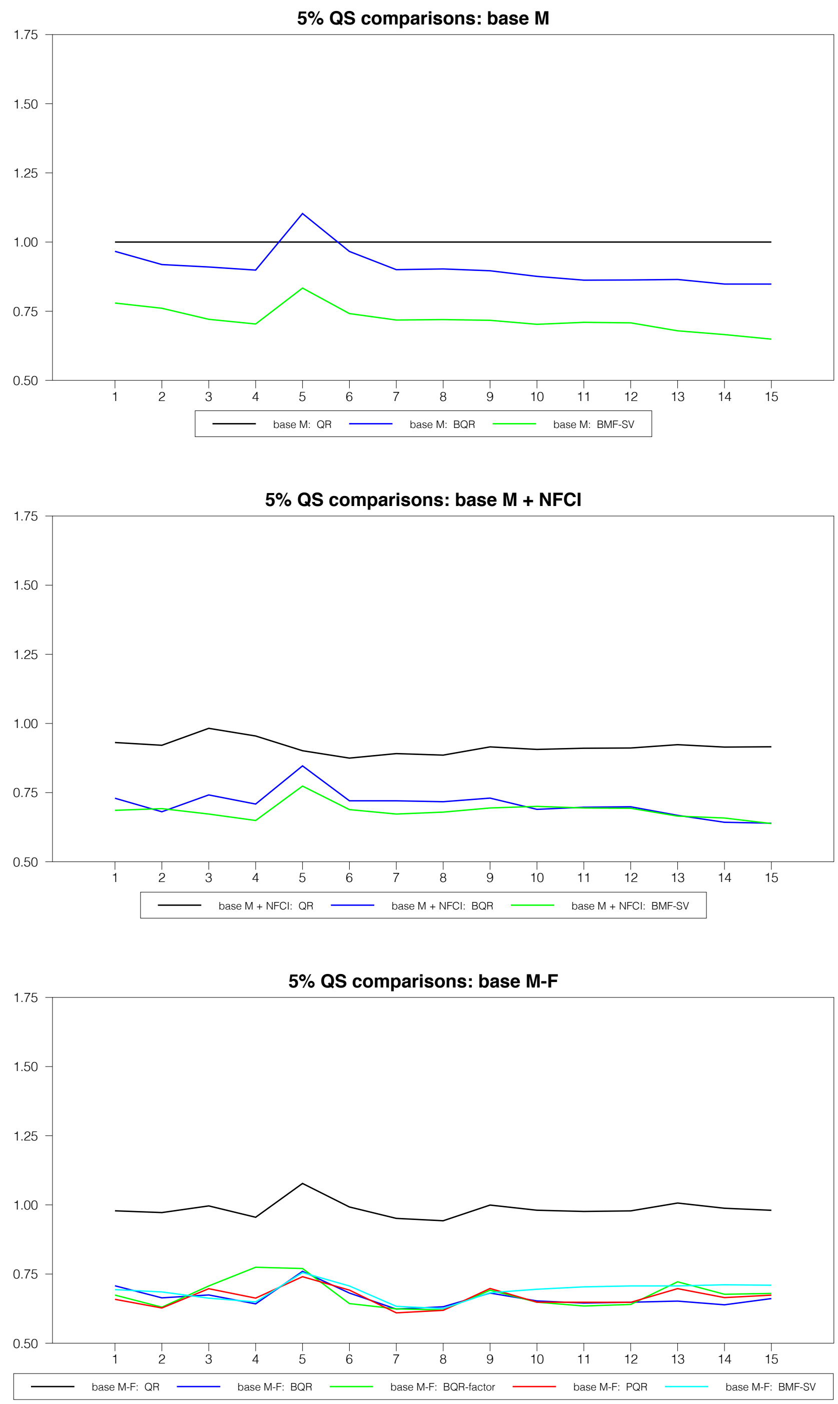

Figure 2: Out-of-sample forecast accuracy, 1985:Q1-2019:Q3: comparisons of 5% QS across variable sets (indicated in panel header) and models (indicated in key label). Scores are reported as relative to the base M: QR specification, so lower numbers represent more accurate forecasts. The weeks 1 through 15 of forecast origins are indicated on the horizontal axis.

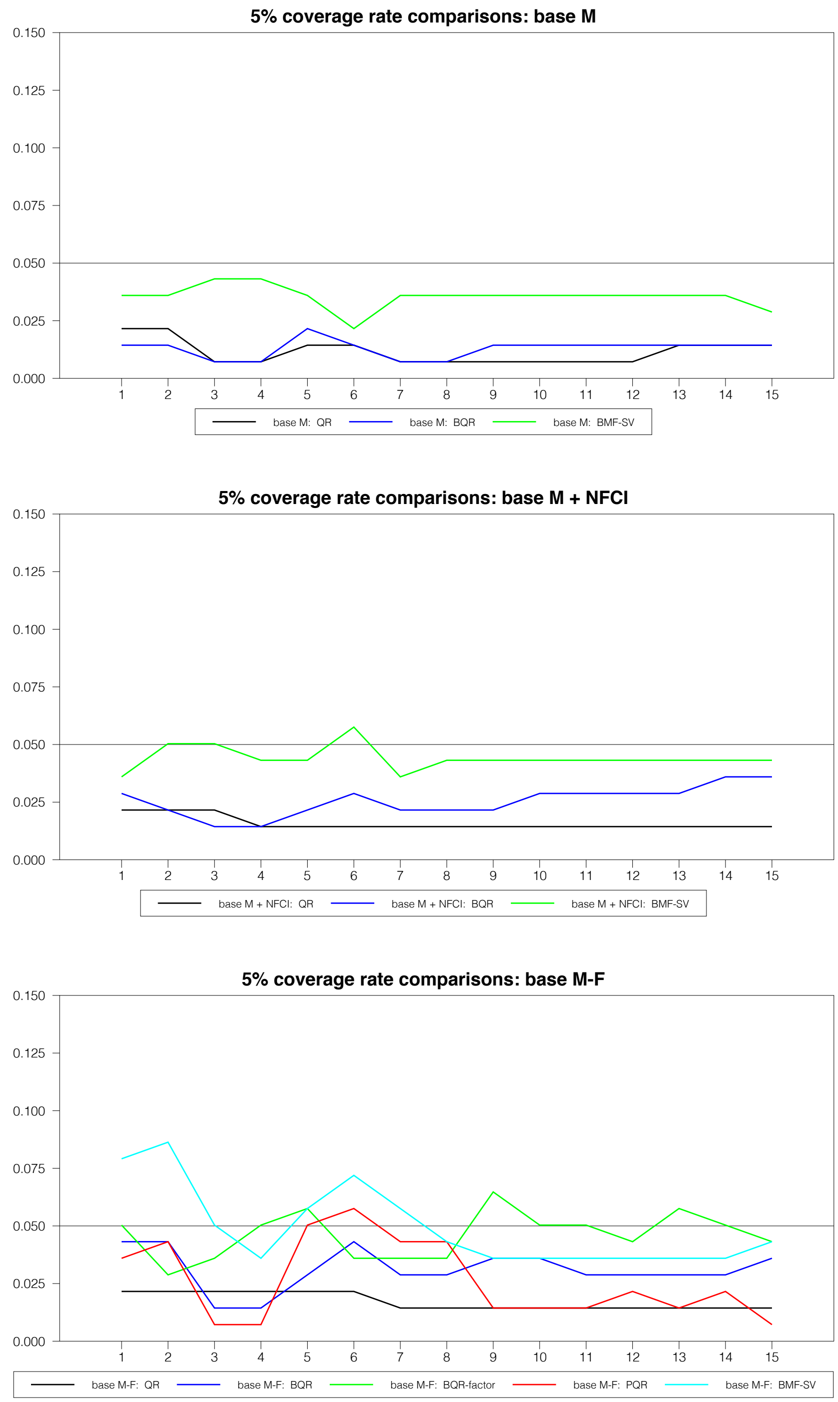

Figure 3: Out-of-sample forecast accuracy, 1985:Q1-2019:Q3: comparisons of 5% coverage rates across variable sets (indicated in panel header) and models (indicated in key label). The black horizontal line at 0.05 denotes the nominal coverage rate. The weeks 1 through 15 of forecast origins are indicated on the horizontal axis.

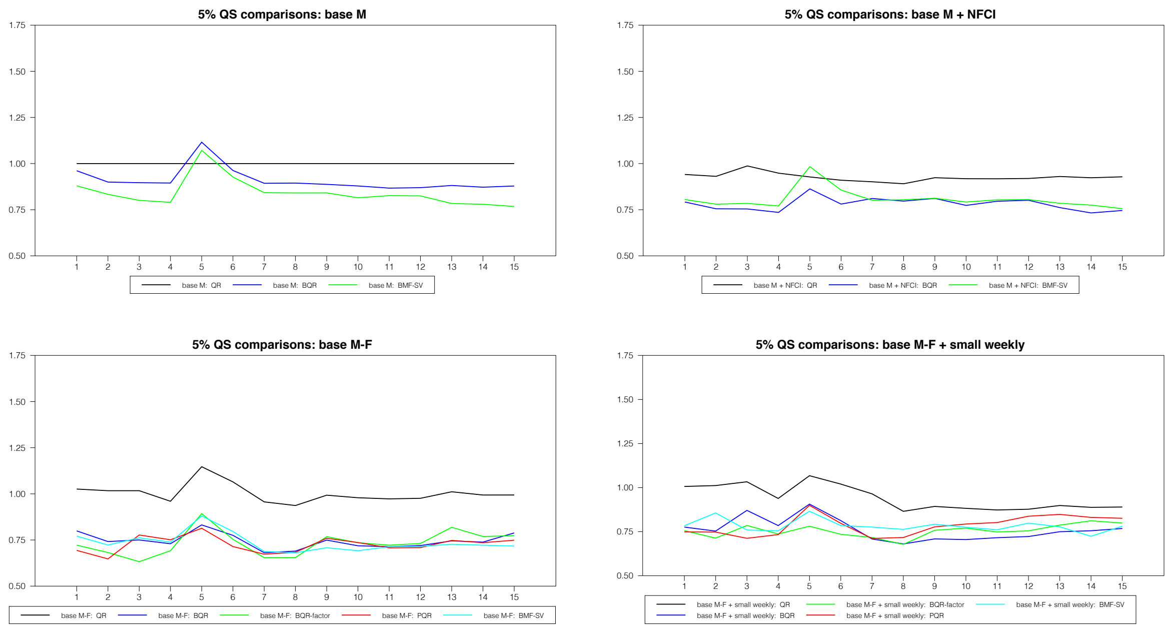

Figure 4: Out-of-sample forecast accuracy, 2000:Q1-2019:Q3: comparisons of 5% QS across variable sets (indicated in panel header) and models (indicated in key label). Scores are reported as relative to the base M: QR specification, so lower numbers represent more accurate forecasts. The weeks 1 through 15 of forecast origins are indicated on the horizontal axis.

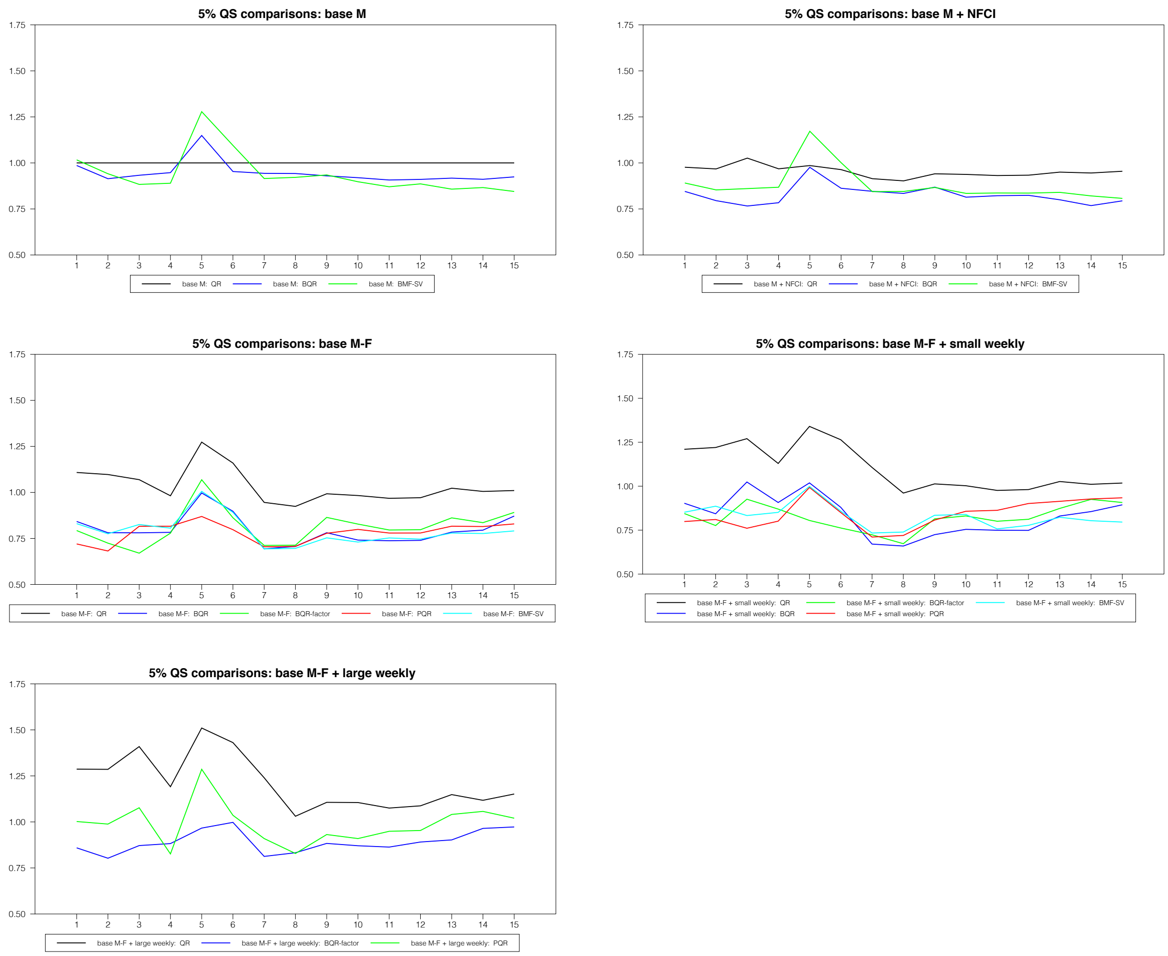

Figure 5: Out-of-sample forecast accuracy, 2007:Q1-2019:Q3: comparisons of 5% QS across variable sets (indicated in panel header) and models (indicated in key label). Scores are reported as relative to the base M: QR specification with monthly macroeconomic indicators, so lower numbers represent more accurate forecasts. The weeks 1 through 15 of forecast origins are indicated on the horizontal axis.

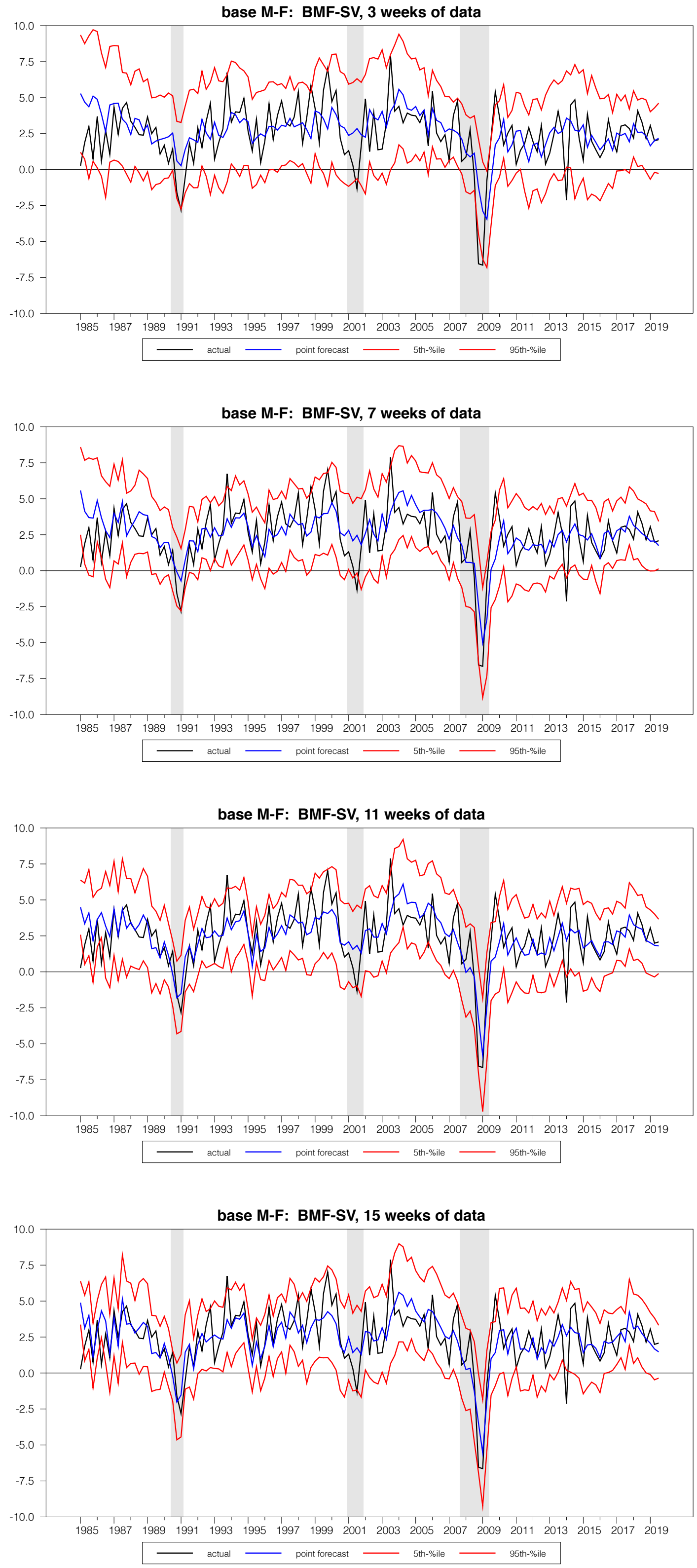

Figure 6: Out-of-sample forecasts from the base M-F variable set and BMF-SV model, selected weeks indicated in panel headers. Each panel reports actual GDP growth (black line), the point forecast (blue line), and 5%-95% forecast quantiles (red lines). Shaded regions denote NBER recessions.

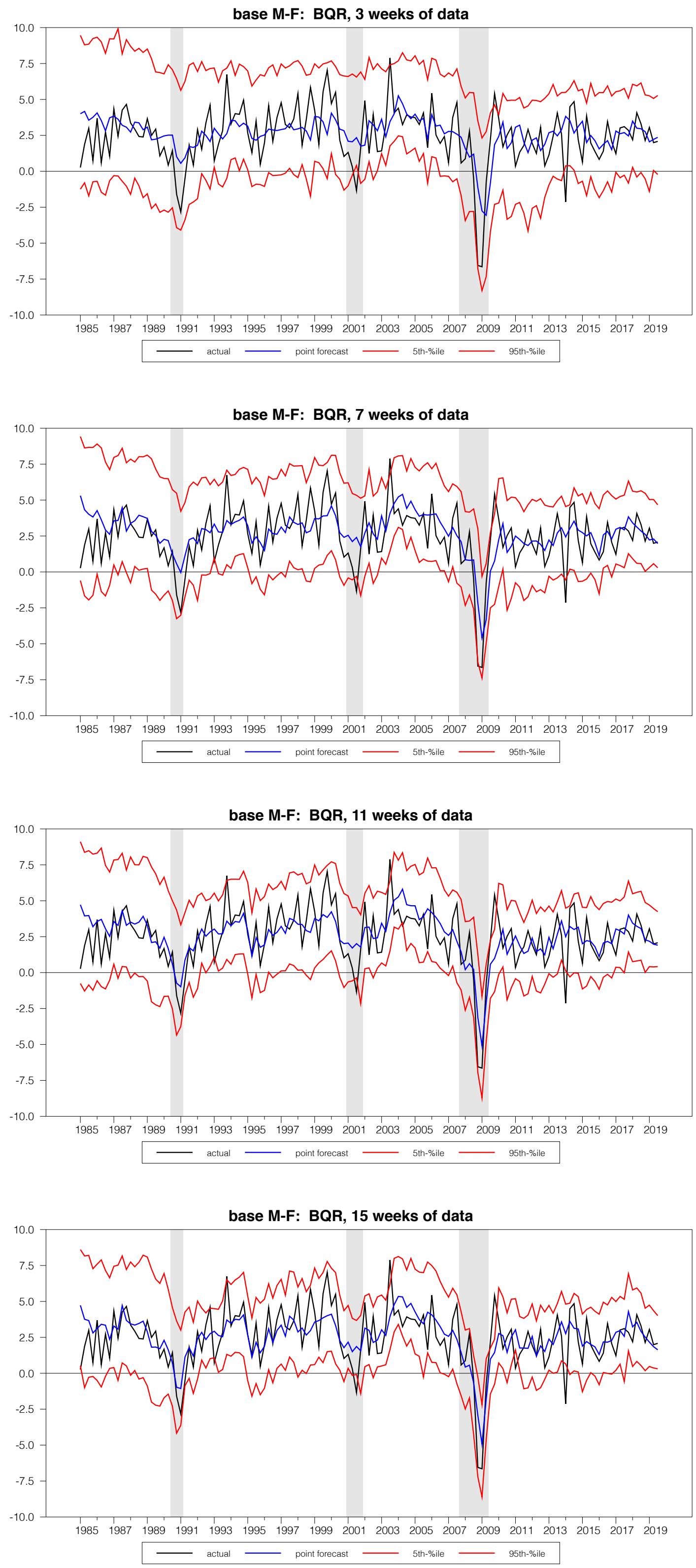

Figure 7: Out-of-sample forecasts from the base M-F variable set and BQR model, selected weeks indicated in panel headers. Each panel reports actual GDP growth (black line), the point forecast (blue line), and 5%-95% forecast quantiles (red lines). Shaded regions denote NBER recessions.

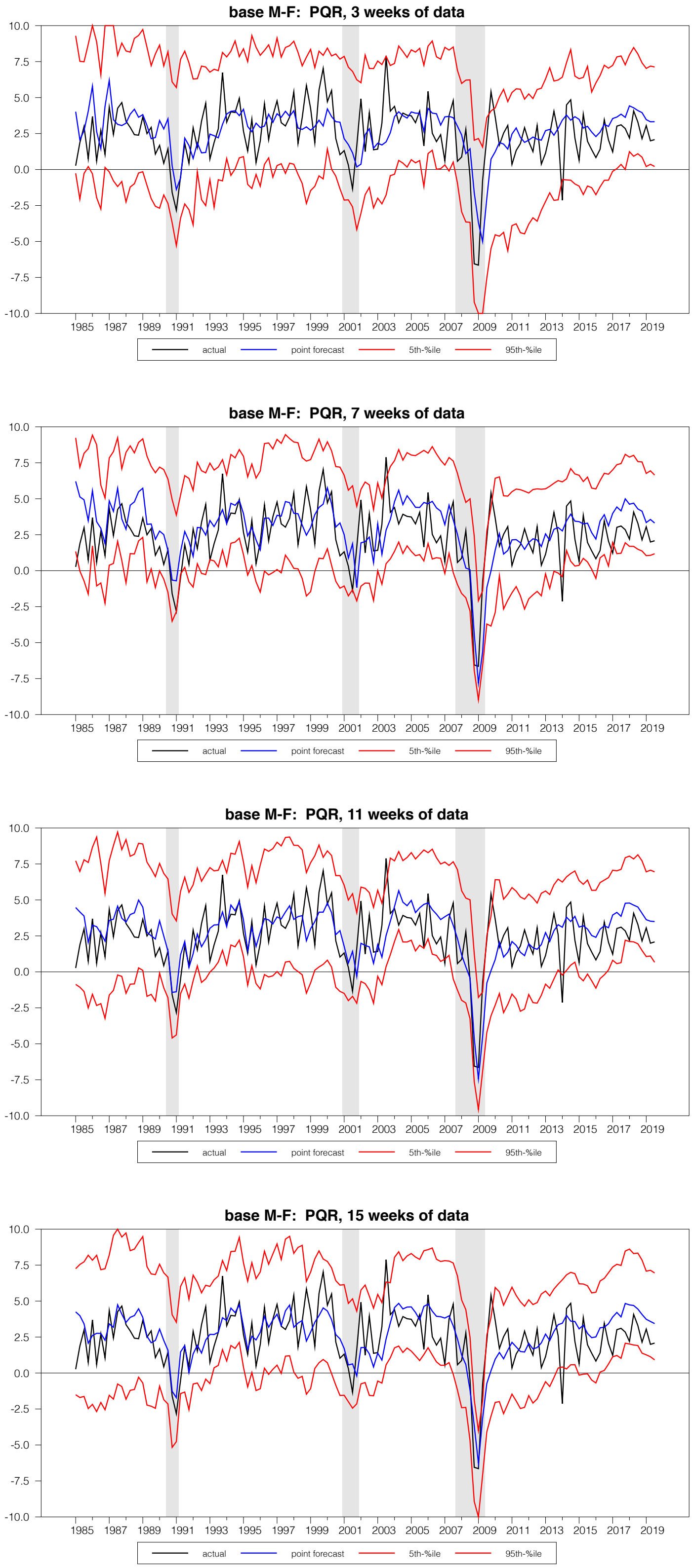

Figure 8: Out-of-sample forecasts from the base M-F variable set and PQR model, selected weeks indicated in panel headers. Each panel reports actual GDP growth (black line), the point forecast (blue line), and 5%-95% forecast quantiles (red lines). Shaded regions denote NBER recessions.

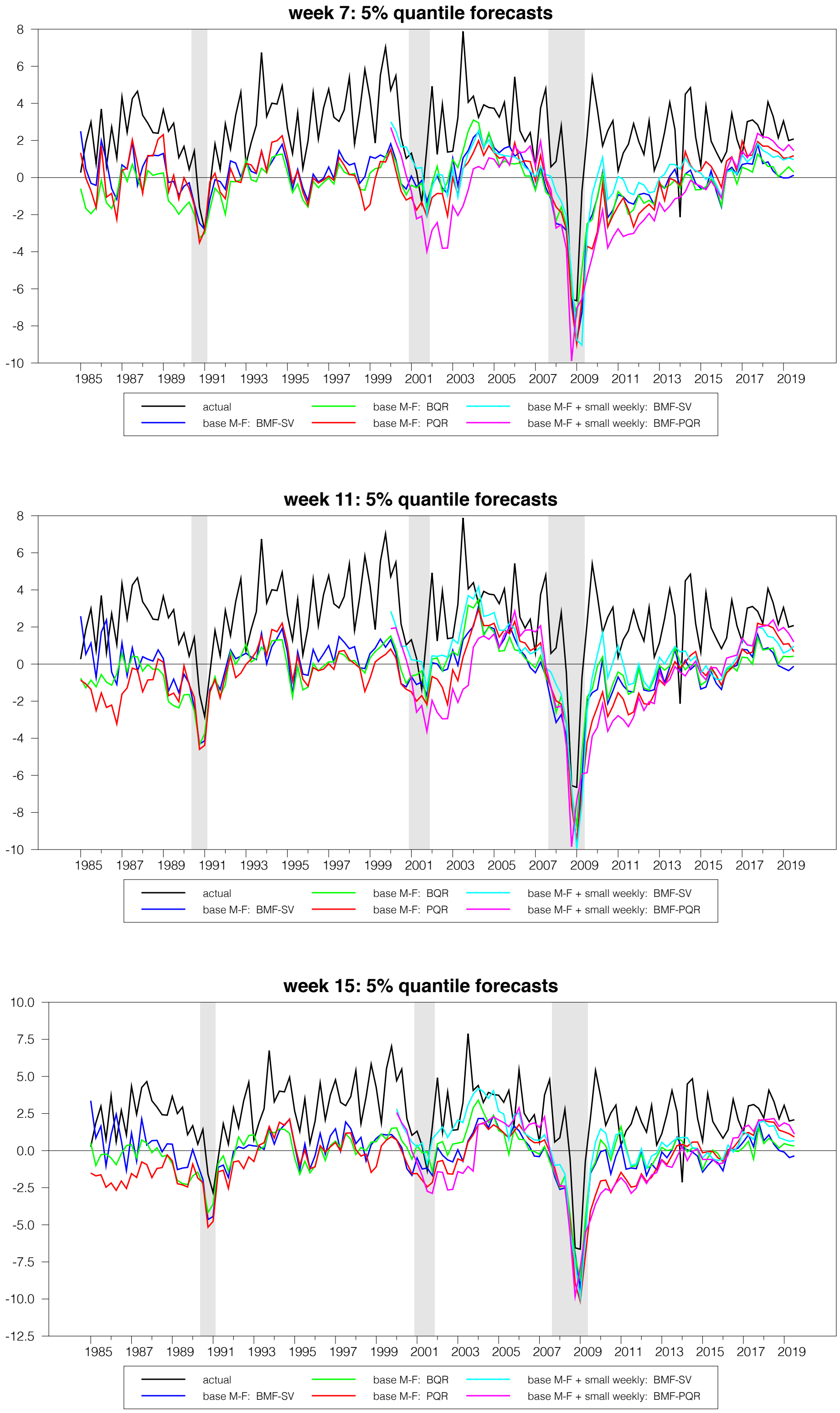

Figure 9: Comparisons of 5 percent quantile forecasts across selected variable-model combinations (indicated in key labels for each chart), for selected weeks of the quarter, indicated in panel headers. Shaded regions denote NBER recessions.

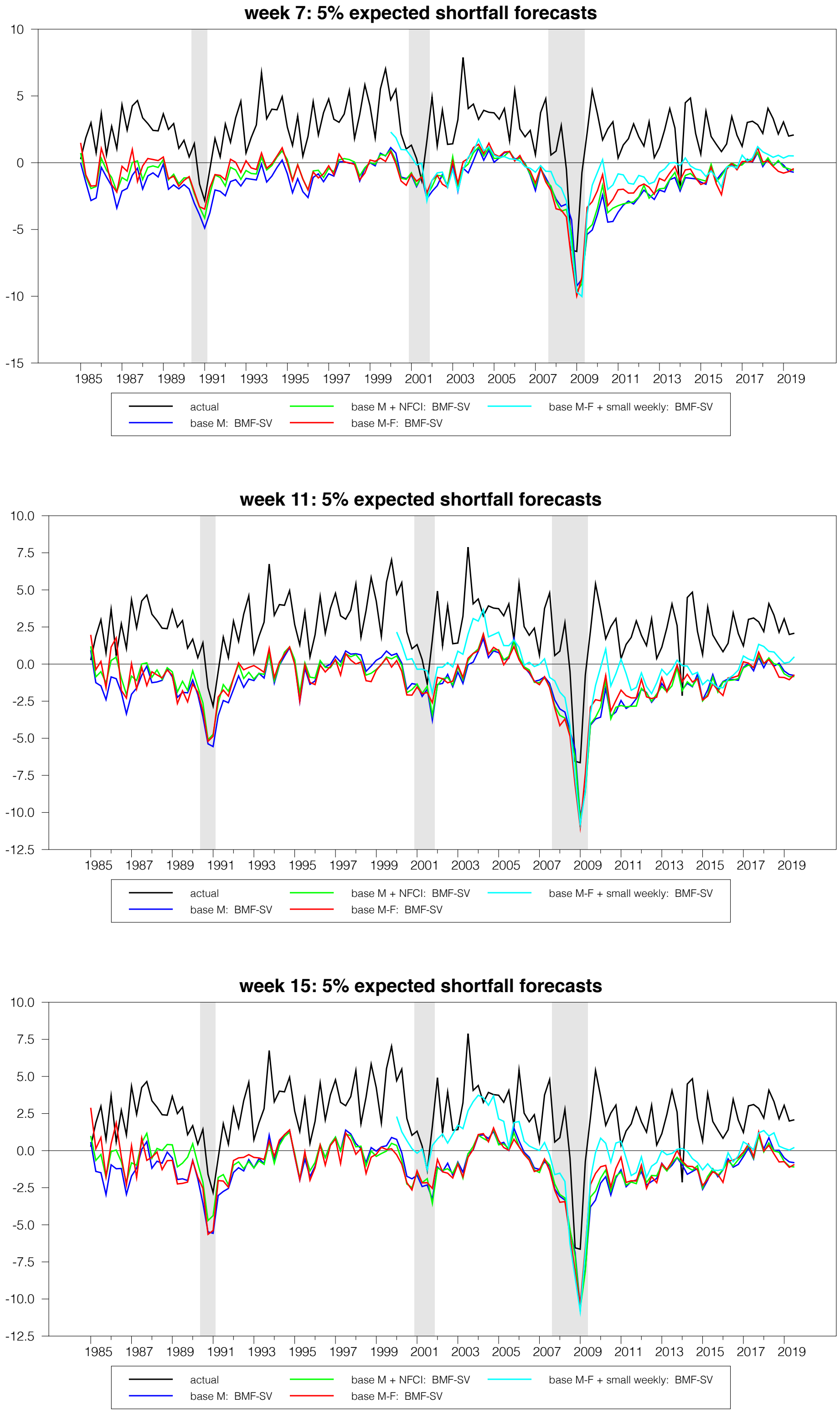

Figure 10: Comparisons of 5 percent expected shortfall forecasts across selected variable-model combinations (indicated in key labels for each chart), for selected weeks of the quarter, indicated in panel headers. Shaded regions denote NBER recessions.

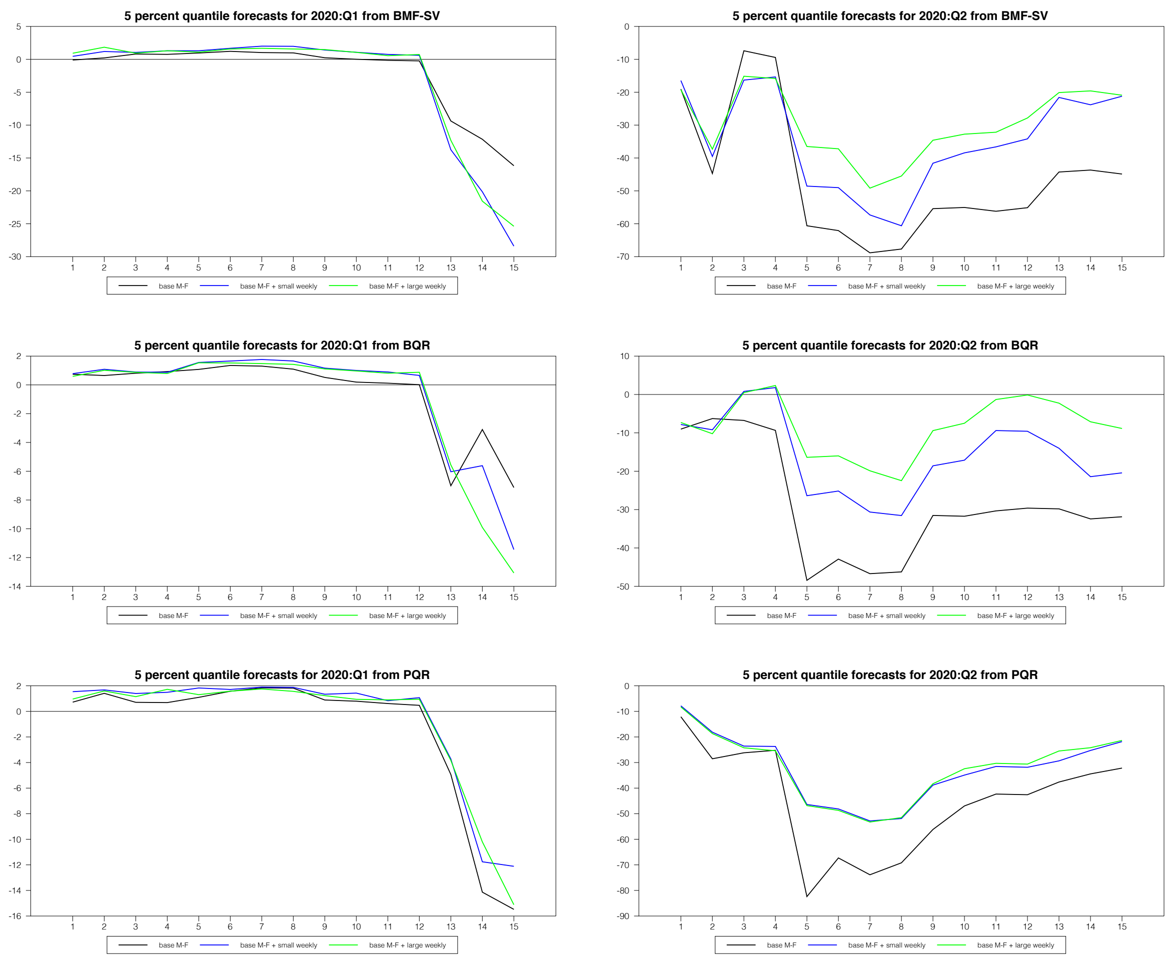

Figure 11: Forecasts of GDP growth in 2020:Q1 (left column) and 2020:Q2 (right column) from selected variable sets and models, for weeks in which data were available up through late April 2020 for 2020:Q1 and late July 2020 for 2020:Q2. Weeks of forecast origin are indicated on the horizontal axis.