working paper

20 02R

Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions

Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino

FEDERAL RESERVE BANK OF CLEVELAND

ISSN: 2573-7953

Working Paper 20-02R September 2020*

Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions

Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino

A rapidly growing body of research has examined tail risks in macroeconomic outcomes. Most of this work has focused on the risks of significant declines in GDP, and it has relied on quantile regression methods to estimate tail risks. Although much of this work discusses asymmetries in conditional predictive distributions, the analysis often focuses on evidence of downside risk varying more than upside risk. We note that this pattern in risk estimates over time could obtain with conditional distributions that are symmetric but subject to simultaneous shifts in conditional means (down) and variances (up). Building on that insight, we examine the ability of Bayesian VARs with stochastic volatility to capture tail risks in macroeconomic forecast distributions and outcomes. We consider both a conventional stochastic volatility specification and a specification with a common factor in volatility that enters the VAR's conditional mean. Even though the one-step-ahead conditional predictive distributions from the conventional stochastic volatility specification are symmetric, the model estimates yield more time variation in downside risk as compared to upside risk. Results from the model that includes a volatility factor in the conditional mean and thereby allows for asymmetries in conditional distributions are very similar. Our paper also extends the recent literature by formally evaluating the accuracy of tail risk forecasts and assessing the performance of Bayesian quantile regression, as well as our Bayesian VARs, in this context. Overall, the BVAR models perform comparably to quantile regression for estimating and forecasting tail risks, complementing BVARs' established performance for forecasting and structural analysis.

JEL classification codes: C53, E17, E37, F47.
Keywords: forecasting, downside risk, asymmetries.

Suggested citation: 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. 20-02R. https://doi.org/10.26509/frbc-wp-202002r.

Andrea Carriero is at Queen Mary, University of London (a.carriero@qmul.ac.uk). Todd E. Clark is at the Federal Reserve Bank of Cleveland (todd.clark@clev.frb.org). Massimiliano Marcellino is at Bocconi University, IGIER, and CEPR (massimiliano.marcellino@unibocconi.it). The authors gratefully acknowledge research assistance from Jared Steinberg and John Zito and helpful comments from Andras Fulop, Domenico Giannone, Sune Karlsson, Elmar Mertens, James Mitchell, Simon Van Norden, Shaun Vahey, Saeed Zaman, Molin Zhong, and seminar participants at Indiana University, the International Association of Applied Econometrics, King's College, and the University of Strathclyde.

*First version Janaury 2020.

1 Introduction

Building on a longer tradition in finance of assessing tail risks in asset prices and returns, a rapidly growing body of research has examined tail risks in macroeconomic outcomes. 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 (2019a), 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). This work has emphasized the link of tail risks to output stemming from poor financial conditions.[1] Other work has considered tail risks to other variables, such as inflation (e.g., Ghysels, Iania, and Striaukas (2018)), unemployment (e.g., Galbraith and van Norden (2019) and Kiley (2018)), or used other methods, such as copula modeling (e.g., Smith and Vahey (2016) and Loaiza-Maya and Smith (2019)) or copula-based combinations of forecasts (e.g., Karagedikli, Vahey, and Wakerly (2019)) to quantify tail risks. Still other work (e.g., Loria, Matthes, and Zhang (2019)) has extended the analysis of Adrian, Boyarchenko, and Giannone (2019a) — henceforth, ABG — to better understand tail risks.[2] Earlier work of Manzan (2015) used quantile regression to assess the value of a large number of macroeconomic indicators in forecasting the complete distribution of some key variables.[3]

The interest in tail risks reflects an underlying perception or assumption of asymmetries in distributions of outcomes. Some form of asymmetry has long been incorporated in particular economic models (e.g., Morley and Piger (2012) assess the abilities of Markov switching and other nonlinear models to capture business cycle asymmetries in the output gap, and Alessandri and Mumtaz (2017) use threshold models to assess output forecasts in periods of financial distress). There has also been a body of research on asymmetries in the unemployment rate (e.g., Galbraith and van Norden (2019) and references therein). In addition, some recent research has considered simple evidence and concluded that GDP growth outcomes feature skewness (e.g., Kozeniauskas, Orlik and Veldkamp (2018) and Orlik and Veldkamp (2015)). Although some of this evidence of skewness in GDP growth has not established formal statistical significance, Jensen, et al. (2020) develop statistically significant evidence of increasingly negative business cycle asymmetry over

[1] For evidence on the euro area, see Figueres and Jarocinski (2020).

[2] Adams, et al. (2020) apply the ABG methodology to forecasts from the Survey of Professional Forecasters. To assess the joint distribution of economic and financial conditions, Adrian, Boyarchenko, and Giannone (2019b) develop a new approach that combines non-parametric and Monte Carlo methods.

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

time, which they link to the increased financial leverage of households and firms.[4]

At a practical level, monetary policymakers have commonly treated forecast distributions as being potentially asymmetric, at least at some points in time. The Bank of England's well-known fan charts for inflation are constructed with a two-piece normal distribution to reflect asymmetries as judged by the Monetary Policy Committee.[5] In the U.S., the Federal Open Market Committee's quarterly Summary of Economic Projections includes participants' subjective assessments of whether risks to each of GDP growth, unemployment, and inflation are "broadly balanced," "weighted to upside," or "weighted to downside." Assessments of asymmetric risks are commonly cited in policy discussions and decisions.

Although not always clearly distinguished in the recent literature, asymmetries could be present in either **conditional** predictive distributions or **unconditional** distributions. For example, the text of ABG sometimes refers to conditional distributions, as in "...recessions are associated with left-skewed distributions while, during expansions, the conditional distribution is closer to being symmetric" (p. 1264). Yet some of the features emphasized by ABG and others such as Adrian, et al. (2018), Caldara, Scotti, and Zhong (2020), and Adams, et al. (2020) could be associated with symmetric conditional distributions and asymmetric unconditional distributions. In particular, the pattern of downside risk varying more over time than upside risk (or, for other variables, upside risk varying more than downside risk) highlighted by ABG (e.g., p. 1264) and other studies could occur with conditional predictive distributions that are symmetric and subject to simultaneous mean (down) and variance (up) shifts. We appear to be the first in the literature to point this out.[6] For illustration, consider a very simple two-period example along the lines of what happens as the economy slows and then enters a recession. In the first period, the conditional one-step-ahead predictive distribution is a normal distribution with a mean of 0 and a standard deviation of 1; in the second period, the conditional one-step-ahead predictive distribution remains Gaussian but shifts left and widens, to have a mean of -2 and a standard deviation of 2. In this example, the 95 percent quantile of the predictive distribution changes relatively little, with values of 1.65 in period 1 and 1.29 in period 2. The 5 percent quantile drops significantly, from -1.65 in period 1 to -5.29 in

[4] Jensen, et al. (2020) also develop a dynamic stochastic general equilibrium model consistent with their empirical findings. Adrian, et al. (2020) extend an econometric formulation of a DSGE model to include endogenous risk. Salgado, Guvenen, and Bloom (2020) present evidence that skewness in firm-level panel data on growth of employment, sales, and productivity is pro-cyclical and develop a heterogenous agents with some of the same features.

[5] Mitchell and Weale (2019) provide a recent summary of the Bank of England's approach and a time series of the skewness in the distributions.

[6] Caldara, Scotti, and Zhong (2020) emphasize the need for simultaneous mean and variance shifts — but in a context of obtaining asymmetric conditional distributions by allowing shocks to levels and volatilities to be correlated — to obtain asymmetric conditional distributions and more time variation in downside risk than upside risk.

period 2. Note that a change to both mean and variance is crucial to such asymmetries in changes in the quantiles; with just a mean change but not a variance change, the upper and lower quantiles move in lockstep.

Having drawn this distinction, in this paper we focus on capturing asymmetries in the time series behavior of measures of upside and downside risks that imply asymmetries in unconditional distributions but do not necessarily require asymmetries in conditional predictive distributions. In particular, we examine the ability of Bayesian VARs with stochastic volatility to capture tail risks in macroeconomic forecast distributions and outcomes.[7] Bayesian VARs are commonly used for point and density forecasting, are known to have a successful track record compared to structural models and survey-based forecasts (except that surveys have an advantage in nowcasting), and can be easily adapted to include a range of variables and produce a variety of forecast density measures. Bayesian VARs with stochastic volatility commonly improve on the point and density forecast accuracy of their homoskedastic counterparts (e.g., Clark (2011) and Clark and Ravazzolo (2015)). BVARs are also often used for structural analysis of the effects of various shocks. Hence, it would be very convenient for empirical macroeconomics if the same models could be also used to study tail risks.

To this point, the efficacy of BVARs for capturing tail risks has not yet been assessed, and therefore this paper pursues such an evaluation. At the one-step-ahead horizon, VARs with conventional stochastic volatility will generally yield conditional predictive distributions that are symmetric. Caldara, Scotti, and Zhong (2020) work through analytics for a VAR with stochastic volatility, abstracting from parameter uncertainty. At longer horizons, because of parameter uncertainty, the conditional predictive distributions may not be symmetric: Recall, for example, that the mean of the multi-step posterior forecast distribution is not in general equal to the forecast implied by the posterior mean of the coefficient vector.[8] But as noted above, asymmetries in conditional distributions are not necessary to obtain more time variation in downside risks than upside risks (or

[7] Carriero, Clark, and Marcellino (2020) examine the ability to nowcast tail risks to growth with a potentially wide array of information, using quantile regression and regressions with stochastic volatility with Bayesian shrinkage, data reduction (factor-based approaches), and forecast combination to manage large information sets. In the nowcast setting, they find that Bayesian regressions with stochastic volatility perform better than quantile regressions. Our results are also consistent with the more cautionary findings of Plagborg-Moller, et al. (2020). They 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." Using score-driven parametric models, Delle Monache, De Polis, and Petrella (2020) obtain a more favorable finding, in which predictive distributions become negatively skewed before and during recessions, preceded by tightening financial conditions.

[8] As a general matter, the multi-step predictive distribution is a complicated function of past data and estimates of parameters and latent volatility states; this distribution need not be symmetric.

vice versa). Rather, simultaneous shifts in means and variances — that is, negative comovement of volatility with the business cycle — are necessary. Historical estimates of stochastic volatility in VARs of macroeconomic data commonly display such comovement. For example, in full sample estimates of the five-variable model we describe below, the correlation between the estimated volatility of GDP growth and the level of GDP growth (using a four-quarter average for smoothing) is -0.3.

In addition, a VAR with stochastic volatility can be extended to capture or allow for asymmetries in conditional distributions at even the one-step-ahead horizon, as well as producing further asymmetry at longer horizons which is not stemming merely from parameter uncertainty. The necessary ingredient is a contemporaneous correlation between shocks to the levels and volatilities of variables. To include a model of this type, we rely on a Bayesian VAR with a common factor in volatility that enters the VAR's conditional mean, drawing on the model used in Carriero, Clark, and Marcellino (2018) to measure macroeconomic uncertainty and its effects. In this formulation, a shock to the volatility (aggregate uncertainty) factor yields simultaneous changes in the conditional mean and variance of the variables of the VAR. We will compare results from this more general model to a conventional BVAR with stochastic volatility. In a bivariate VAR setting, Caldara, Scotti, and Zhong (2020) take a different avenue to embedding the ingredient necessary for asymmetries in conditional distributions at the one-step-ahead horizon: their model has only lagged (not contemporaneous) volatility in the VAR's conditional mean but allows a correlation between the VAR's innovations and the volatility innovations.

Reflecting the combination of common practice in the VAR-based forecasting literature and the recent literature on macroeconomic tail risks, the BVAR models in our presented results include a small set of primary macroeconomic indicators and an indicator of financial conditions. Following ABG, in our baseline results we measure financial conditions with the Chicago Fed's national financial conditions index (we also report results with a measure of financial sector volatility from Giglio, Kelly, and Pruitt (2016)). In the presented results, in the interest of brevity and consistent with most of the recent literature on macroeconomic tail risks, we focus on risks to GDP growth; we provide selected other results for the unemployment rate, another key indicator of economic activity. Our basic results are robust to reducing the variable set to just a bivariate model in growth and the Chicago Fed's national financial conditions index (NFCI) and to expanding the model to make it medium-sized (along the lines of, e.g., the medium-sized forecasting models of Carriero, Clark, and Marcellino (2016)), with 15 variables.

Our paper also contributes to the recent macroeconomic tail risk literature by providing more formal evaluations of asymmetries and risk forecasts and the performance of alternative models than has much of the recent literature.[9] We begin by conducting formal tests of skewness and kurtosis in the data, BVAR residuals, and forecast errors using the inferential approach for time series developed in Bai and Ng (2005). In assessing the efficacy of the models in estimating tail risks, we not only provide basic comparisons of our BVAR-based forecast results to those obtained with the quantile regression-based methodology of ABG but also subject our BVAR-based and quantile regression-based models to comparisons of formal scoring of quantile and expected shortfall forecasts, using a conventional quantile scoring function and a recently developed joint scoring function for the quantile and its associated expected shortfall estimate. Our focus on forecasting with vector autoregressions and formal risk forecast evaluation, along with our distinction between asymmetries in unconditional distributions versus conditional predictive distributions, distinguishes our paper from a contemporary analysis by Caldara, Scotti, and Zhong (2020, hereafter CSZ). CSZ focus instead on assessing the mechanisms by which a small VAR (bivariate in their case) with stochastic volatility, particularly with an explicit correlation in shocks to levels and volatilities, can produce time-varying asymmetries in conditional predictive distributions, with downside risks to economic activity. More generally, our paper is one of the first in the recent tail risk literature to formally evaluate the accuracy of the tail risk forecasts. The recent work of Brownlees and Souza (2020) also engages in formal evaluation, in their case of tail risk forecasts for growth for a panel of countries using quantile regression methods and AR models augmented with GARCH.

Our analysis yields the following main results. First, the formal statistical evidence of skewness in output growth is generally weak (prior work has typically considered skewness estimates without assessing statistical significance); the skewness statistics are often large, but not often statistically significant. There is more evidence of asymmetries in the unemployment rate and our indicator of financial conditions; relationships between these variables and GDP growth may lead to asymmetries in the predictive distributions of GDP. Of course, some of the considerable policy and research interest in downside risks to output focuses on asymmetries in conditional distributions that emerge at certain points in time and are not necessarily constant, whereas simple checks of skewness abstract from time variation and can be seen as applying to marginal distributions.

Second, although quantile regression-based approaches can be a useful tool for quantifying

[9] Although other work in the tail risk literature, such as Adrian, Boyarchenko, and Giannone (2019a) and Caldara, Scotti, and Zhong (2020) consider forecast accuracy, their measures are limited to predictive scores and probability integral transforms, which get at general density accuracy and calibration, rather than tail forecast accuracy specifically.

tail risks in macroeconomics, they can come with some challenges in data samples of the size typically available in macroeconomics. In addition to presenting some quantile regression results with fragility with respect to the choice of financial indicator, we document examples of problems of crossing of the quantile estimates as well as considerable variability in the coefficient estimates of quantile regressions. To potentially alleviate these problems, we also considered Bayesian quantile regression. In our limited checks, making use of some Bayesian shrinkage seemed to mitigate these problems but produced tail risk forecasts very similar to those obtained with simple quantile regression. In the quantile literature, the challenges of estimating extreme quantiles with small samples of observations are well known, resulting in coefficient bias and complicating inference (see, e.g., Chernozhukov, Fernandez-Val, and Kaji (2017)).

Our third main result is that, with our BVAR specifications featuring time-varying volatility, we are able to capture time variation in tail risks to output growth — with downside risks more variable than upside risks — like that emphasized in ABG and captured in their case by quantile regression. Put another way, our estimates indicate that familiar BVARs with time-varying volatility — which are known to be broadly successful in macroeconomic point and density forecasting and to be useful for structural analysis — can perform as well as quantile regression for the purposes of capturing downside risks to output growth.[10] In the case of the BVARs, the time variation in downside tail risks as compared to upside risks is driven by simultaneous shifts in the means and variances of conditional predictive distributions without requiring asymmetries in the conditional distributions. As noted above, changes in conditional variances like those captured by stochastic volatility are crucial to this result. We should emphasize that, in these comparisons, we extend prior work by including comparisons of formal scoring of quantile and expected shortfall forecasts. In this formal scoring, there is little to distinguish our BVAR performance from quantile regression performance. Finally, in the BVAR model set considered, we obtain these results for both (i) the conventional BVAR with stochastic volatility on which we focus and (ii) the model featuring a factor structure to volatility in which the volatility factor is a function of past economic and financial conditions and appears in the VAR's conditional mean. As noted above, this model, unlike the BVAR with conventional stochastic volatility, is capable of producing or capturing asymmetries in one-step-ahead predictive distributions.

The paper proceeds as follows. Sections 2 and 3 describe the models and data, respectively. Section 4 explains the forecast metrics. Section 5 reports the empirical results. Section 6 concludes.

[10]Brownlees and Souza (2020) conclude that AR models with GARCH outperform quantile regression in a panel of OECD countries.

A supplemental appendix provides some additional results.

2 Models

We present estimates and forecasts from three different models: a Bayesian VAR with stochastic volatility (BVAR-SV); a Bayesian VAR with a common factor in volatility that enters the VAR's conditional mean (BVAR-SVF-M), as in Carriero, Clark, and Marcellino (2018); and quantile regression as in ABG. Although not essential, we estimate both the BVAR-SV and BVAR-SVF-M models with standardized data (so both models exposited below exclude an intercept). We then renormalize the sample variances and add back in the means to obtain forecasts in the same units as the not-standardized data. In the out-of-sample analysis, we compute relevant means and variances with each data vintage, so that future information is not used and real-time information timing is preserved. As we will detail below, we consider models with different number of variables, but they all include a financial indicator — the NFCI as a measure of broader financial conditions in the baseline results, a measure of financial sector volatility from Giglio, Kelly, and Pruitt (2016) in one of our robustness checks, or both the NFCI and the credit spread of Gilchrist and Zakrajsek (2012) in another robustness check.

2.1 BVAR-SV Model

The conventional BVAR with stochastic volatility, referred to as a **BVAR-SV** specification, takes the following form, for the $n \times 1$ data vector y_t:

$$\begin{aligned} y_t &= \sum_{i=1}^{p} \Pi_i y_{t-i} + v_t \\ v_t &= A^{-1}\Lambda_t^{0.5}\epsilon_t,\ \epsilon_t \sim N(0, I_n),\ \Lambda_t \equiv \text{diag}(\lambda_{1,t}, \ldots, \lambda_{n,t}) \\ \ln(\lambda_{i,t}) &= \gamma_{0,i} + \gamma_{1,i}\ln(\lambda_{i,t-1}) + \nu_{i,t},\ i = 1, \ldots, n \\ \nu_t &\equiv (\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{n,t})' \sim\ N(0, \Phi), \end{aligned} \tag{1}$$

where A is a lower triangular matrix with ones on the diagonal and non-zero coefficients below the diagonal, and the diagonal matrix Λ_t contains the time-varying variances of conditionally Gaussian shocks.[11] This model implies that the reduced-form variance-covariance matrix of innovations to the VAR is $\text{var}(v_t) \equiv \Sigma_t = A^{-1}\Lambda_t A^{-1\prime}$. Note that, as in Primiceri's (2005) implementation, innovations

[11] In unreported results, we also considered making A time-varying as in Primiceri (2005) and obtained very similar estimates over the full sample. Other forecast studies (e.g., Clark and Ravazzolo (2015)) have obtained a similar finding of little payoff to making A time-varying. So, in this paper's results, to reduce computational requirements, we make A constant over time.

to log volatility are allowed to be correlated across variables; Φ is not restricted to be diagonal. For notational simplicity, let Π denote the collection of the VAR's coefficients. In implementation, we include four lags in the VAR.

Regarding the priors for the BVAR-SV model, we use settings like those common in the forecasting literature. For the VAR coefficients contained in Π, we use a Minnesota-type prior. With the variables of interest transformed for stationarity, we set the prior mean of all the VAR coefficients to 0. We make the prior variance-covariance matrix $\underline{\Omega}_{\Pi}$ diagonal. For lag l of variable j in equation i, the prior variance is $\frac{\theta_1^2}{l^2}$ for $i = j$ and $\frac{\theta_1^2 \theta_2^2}{l^2} \frac{\sigma_i^2}{\sigma_j^2}$ otherwise. In line with common settings, we set overall shrinkage $\theta_1 = 0.2$ and cross-variable shrinkage $\theta_2 = 0.5$. Consistent with common settings, the scale parameters σ_i^2 take the values of residual variances from AR(p) models fit over the estimation sample.

For each row a_j of the matrix A, we follow Cogley and Sargent (2005) and make the prior fairly uninformative, with prior means of 0 and variances of 10 for all coefficients. The variance of 10 is large enough for this prior to be considered uninformative. For the coefficients $(\gamma_{i,0}, \gamma_{i,1})$ (intercept, slope) of the log volatility process of equation i, $i = 1, \ldots, n$, the prior mean is $(0.1 \times \ln \sigma_i^2,\ 0.9)$, where σ_i^2 is the residual variance of an AR(p) model over the estimation sample; this prior implies that the mean level of volatility is $\ln \sigma_i^2$. The prior standard deviations (assuming 0 covariance) are $(2^{0.5}, 0.2)$. For the variance matrix Φ of innovations to log volatility, we use an inverse Wishart prior with mean of $0.03 \times I_n$ and n degrees of freedom, with $n = 10$. For the period 0 values of $\ln \lambda_t$, we set the prior mean and variance at $\ln \sigma_i^2$ and 2.0, respectively.

We estimate the model with a conventional Gibbs sampler, detailed in such sources as Clark and Ravazzolo (2015).[12] Volatility is sampled with a Gibbs step based on Kim, Shephard, and Chib (1998). Estimates derived from the BVAR-SV model are based on samples of 5,000 retained draws, obtained by sampling a total of 30,000 draws, discarding the first 5,000, and retaining every 5th draw of the post-burn sample. Our forecast results are obtained using draws from the predictive distribution (see, e.g., Cogley, Morozov, and Sargent (2005) for a detailed explanation).

For the reasons analyzed in CSZ (abstracting from parameter uncertainty), the conditional one-step-ahead predictive distributions obtained from the BVAR-SV model will be symmetric. At longer horizons, symmetry may not apply because, as noted above, the predictive distribution is a complicated function of past data and estimates of parameters and latent volatility states. That said, applying at each forecast origin the conditional symmetry test of Bai and Ng (2001)

[12]To speed computation with the 15-variable model considered in robustness checks, we include in the algorithm the triangularization approach developed in Carriero, Clark, and Marcellino (2019).

— a test that is based on the empirical distribution — to the forecast draws from the predictive densities of our estimated BVAR-SV yields results broadly consistent with symmetry. In the in-sample forecasts of GDP growth detailed below, with a 5 percent critical value the Bai-Ng test rejects conditional symmetry in only 4 percent of forecast origins at the one-step-ahead horizon and 5 percent at the four-steps-ahead horizon. In the out-of-sample forecasts of GDP growth, the corresponding rejection rates are 8 (one-step) and 6 percent (four-step). Although the tests are unlikely to be independent across the forecast origins, these rejection rates are close to the test's nominal significance level. Accordingly, we take these results as consistent with symmetry of the conditional predictive distributions from the BVAR-SV specification, with the caveat that this pattern reflects both the data and the model, and the model embeds conditional symmetry at the 1-step horizon. However, in the results below we will obtain more time variation in downside risks than upside (for output growth) because the BVAR-SV model is able to capture simultaneous shifts in the means and variances of the conditional distributions. In the estimates, the simultaneous downward shift of the conditional mean and upward shift of the conditional variance occurs over short periods as the model captures patterns of correlated shocks to levels and volatilities.

2.2 BVAR-SVF-M Model

Following Carriero, Clark, and Marcellino (2018), the **BVAR-SVF-M** specification incorporates a factor structure of volatility in a VAR with stochastic volatility, links the (unobservable) factor in volatility to the last quarter's levels of the VAR's variables, and allows the volatility factor to enter the VAR's conditional mean.[13] Note that the volatility factor, being common to all the volatilities, can be interpreted as a measure of uncertainty, along the lines of studies such as Jurado, Ludvigson, and Ng (2015). This model takes the form:

$$\begin{aligned}
y_t &= \sum_{i=1}^{p} \Pi_i y_{t-i} + \sum_{i=0}^{p_m} \Pi_{m,i} \ln m_{t-i} + v_t \\
v_t &= A^{-1} \Lambda_t^{0.5} \epsilon_t, \ \epsilon_t \sim N(0, I_n), \ \Lambda_t \equiv \text{diag}(\lambda_{1,t}, \ldots, \lambda_{n,t}) \\
\ln \lambda_{i,t} &= \beta_{m,i} \ln m_t + \ln h_{i,t}, \ i = 1, \ldots, n \qquad (2) \\
\ln m_t &= \sum_{i=1}^{p_m} \delta_{m,i} \ln m_{t-i} + \delta_y' y_{t-1} + u_{m,t}, \ u_{m,t} \sim iid\ N(0, \phi_m) \\
\ln h_{i,t} &= \gamma_{i,0} + \gamma_{i,1} \ln h_{i,t-1} + e_{i,t}, \ e_{i,t} \sim iid\ N(0, \phi_i), \ i = 1, \ldots, n.
\end{aligned}$$

[13] In Carriero, Clark, and Marcellino (2016), we first considered a factor structure of volatility for the purpose of making large models with stochastic volatility computationally feasible, in a model simpler than that of Carriero, Clark, and Marcellino (2018) and this paper.

For each variable i, its log-volatility follows a linear factor model with a common uncertainty factor $\ln m_t$ that captures (unobservable) aggregate uncertainty. This factor follows an AR(p_m) process augmented to include the previous period's macroeconomic and financial conditions as captured by y_{t-1}. This factor also appears in the VAR's conditional mean, contemporaneously and with lags. The idiosyncratic component $\ln h_{i,t}$ — which captures time variation in each variable's volatility unique to that variable — follows an AR(1) process.

With a common factor in volatility that enters the VAR's conditional mean, the BVAR-SVF-M model effectively embeds a contemporaneous correlation between shocks to the levels and volatilities of variables. In particular, a shock to the uncertainty factor m_t yields simultaneous changes in the conditional mean and variance of y_t. As a result, the BVAR-SVF-M specification can produce or capture asymmetries in the conditional predictive distribution, including at the 1-step ahead horizon.[14] In a bivariate VAR with stochastic volatility, Caldara, Scotti, and Zhong (2020) include lagged volatility in the VAR's conditional mean and allow correlated shocks to levels and volatilities. Our respective models can be seen as different formulations capturing the same basic feature.

We use priors for the BVAR-SVF-M aligned with those of the BVAR-SV model (indicated above). Regarding the unique components of the BVAR-SVF-M model, for the coefficients Π_m on uncertainty in the VAR's conditional mean, we set the prior means at small values to imply adverse effects of uncertainty on growth and unemployment and 0 for other variables, and in equation i we set the prior variances at $4\sigma_i^2$. In the case of the loading $\beta_{m,i,}$, $i = 1, \ldots, n$, on the uncertainty factor $\ln m_t$, we use a prior mean of 1 and a standard deviation of 0.5. The prior is meant to be consistent with average volatility approximating aggregate uncertainty. For the coefficients of the process of the factor, we use priors consistent with some persistence in volatility. For the coefficients on lags 1 and 2 of $\ln m_t$, we use means of 0.9 and 0.0, respectively, with standard deviations of 0.2. For the coefficient on each variable of y_{t-1}, we use a mean of 0 and standard deviation of 0.4. For the period 0 value of $\ln m_t$, we set the mean at 0 and in each draw use the variances implied by the AR representations of the factors and the draws of the coefficients and error variance matrix. For the idiosyncratic volatility components, the prior means on the intercepts and slope coefficients are $\ln(0.75 \times \sigma_i^2)$ and 0, respectively, where σ_i^2 is the residual variance of an AR(p) model over the estimation sample. For the variance of innovations to the log idiosyncratic volatilities, we use a mean of 0.03 and 10 degrees of freedom.

We close the discussion of the BVAR-SVF-M model with a few other specification or imple-

[14] With a simple version of the model, it is straightforward to work through algebra like that of Caldara, Scotti, and Zhong (2020) to verify that 1-step ahead prediction errors can feature skewness.

mentation details. First, the uncertainty shock $u_{m,t}$ is independent of the conditional errors ϵ_t as well as the elements of $\nu_t = (e_{1,t}, \dots, e_{n,t})'$, which are distributed independently of one another as i.i.d. $N(0, \Phi_\nu)$, with $\Phi_\nu = diag(\phi_1, \dots, \phi_n)$. Second, for identification, we follow common practice in the dynamic factor model literature and assume $\ln m_t$ to have a zero unconditional mean, fix the factor's innovation variance ϕ_m at 0.03, and use an accept-reject step to force GDP volatility's factor loading to be positive. Third, in our implementation, we set the model's lag orders at $p = 4$ and $p_m = 2$. Finally, we estimate the model with a Gibbs sampler as detailed in Carriero, Clark, and Marcellino (2018). The algorithm is similar to that used for the BVAR-SV model, except that the volatility state is estimated with a particle Gibbs step instead of a Gibbs step. Estimates derived from the BVAR-SVF-M model are based on samples of 5,000 retained draws, with the same burn and skip specifications as for the BVAR-SV model, and forecast results are obtained using draws from the predictive distribution.

For the reasons described above, conceptually the BVAR-SVF-M model is capable of producing asymmetries in conditional predictive distributions at any forecast horizon, with the asymmetries stemming from the correlation between the shocks in the model (driven by innovations to the volatility factor that simultaneously affect the conditional mean and conditional variance of the VAR) as well as from parameter uncertainty. Applying at each forecast origin the conditional symmetry test of Bai and Ng (2001) — a test that is based on the empirical distribution — to the forecast draws from the predictive densities of our estimated BVAR-SVF-M model yields evidence of such asymmetries for some variables, although less so for GDP growth. In the in-sample forecasts of GDP growth detailed below, with a 5 percent critical value the Bai-Ng test rejects conditional symmetry in only 3 percent of forecast origins at the one-step-ahead horizon but 27 percent at the four-steps-ahead horizon (corresponding figures for the unemployment rate are 17 percent and 62 percent, respectively). The rejection rates for the federal funds rate and NFCI exceed 80 percent at both horizons. In the out-of-sample forecasts at the one-step-ahead horizon (rates are higher at the four-steps-ahead horizon), the rejection rate across the 1985-2018 sample of forecast origins is only 4 percent for GDP growth but 56 and 87 percent for the federal funds rate and NFCI, respectively. More generally, as noted earlier, even with symmetric distributions for GDP growth, the model is capable of capturing more time variation in downside risks than upside risks.

2.3 Quantile Regressions

To assess the efficacy of the BVAR-SV and BVAR-SVF-M specifications, we include comparisons to results obtained with the quantile regression approach of ABG. More specifically, in our quantile regression analysis, for a given quantile τ we estimate a regression model using a direct multi-step form as in ABG:

$$y_{t+h}^{(h)} = x_t'\beta_\tau + \epsilon_{\tau,t+h}, \tag{3}$$

where h is the forecast horizon and the coefficient vector and innovation term are specific to quantile τ. The vector of predictors x_t includes a constant, y_t, and either NFCI_t in the baseline results or a financial volatility indicator in the robustness check. In our results for GDP growth, we consider forecast horizons of both one and four quarters, y_t is annualized quarterly GDP growth computed as 400 times the log change, and $y_{t+h}^{(h)} \equiv h^{-1}\sum_{i=1}^{h} y_{t+i}$. In results for the unemployment rate, our specification is patterned along the lines of Kiley's (2018), to predict the change in unemployment rather than the level. We use just a forecast horizon of four quarters, y_t is the unemployment rate, and $y_{t+4}^{(4)} \equiv y_{t+4} - y_t$.

In all cases, the parameter vector β_τ is obtained with quantile regression:

$$\hat{\beta}_\tau = \underset{\beta_\tau}{\text{argmin}} \sum_{t=1}^{T-h} \left(\tau \cdot \mathbf{1}_{(y_{t+h}^{(h)} \geq x_t'\beta_\tau)} |y_{t+h}^{(h)} - x_t'\beta_\tau| + (1-\tau) \cdot \mathbf{1}_{(y_{t+h}^{(h)} < x_t'\beta_\tau)} |y_{t+h}^{(h)} - x_t'\beta_\tau| \right). \tag{4}$$

We estimate the model for quantiles of $\tau = 0.05$, 0.25, 0.75, and 0.95, as well as $\tau = 0.5$. Following ABG, to obtain some of the forecast measures considered — not all, as detailed in the next section — at each point in time we use the estimates of the first four quantiles in a second step that consists of fitting the skewed t-distribution developed by Azzalini and Capitanio (2003). This second step serves to smooth the estimated quantile function and provide a complete probability density function needed for some of the forecast comparisons.[15]

3 Data

Reflecting our VAR specifications, we focus on a data set of five variables, at a quarterly frequency. In the baseline case, the variables consist of GDP growth (annualized, as $400\Delta \ln \text{GDP}$), the unemployment rate, inflation in the GDP price index (annualized, as $\Delta \ln P$), the federal funds rate, and the Chicago Fed's NFCI. The first four variables are very commonly used in small VARs in the forecasting literature (see, e.g., Clark and Ravazzolo (2015) and Carriero, Clark, and Marcellino (2016)). In our baseline results, we use the NFCI to measure financial conditions following prior research that has found it to be related to recessions or business cycle asymmetries more generally (e.g., ABG). Although we focus on the five-variable model, in robustness checks with quantile

[15]Other studies in the forecasting literature with quantile models, such as Gaglianone and Lima (2012) and Korobilis (2017), have also used two-step approaches that involve fitting a density to the quantile estimates. In this step, for comparability we follow exactly the procedure of ABG, using the Matlab programs accompanying their paper.

regression and the BVAR-SV model, we obtained qualitatively similar results with just a bivariate model in GDP growth and the NFCI and with a 15-variable model consisting of the 14-variable set of Carriero, Clark, and Marcellino (2016) augmented by the NFCI. In another set of robustness checks we do report, we replace the NFCI with a measure of financial market volatility considered in Giglio, Kelly, and Pruitt (2016). In particular, we use their turbulence variable, computed from asset returns for the 20 largest financial institutions each year, to measure the distance (as proposed by Kritzman and Li (2010)) between recent and historical covariation. Note that using this measure allows us to consider robustness not only to the use of a different financial indicator but also to a longer sample, due to the earlier availability of the turbulence variable.

In the real-time forecast analysis, output is measured as GDP or GNP, depending on data vintage. Inflation is measured with the GDP or GNP deflator or price index. For simplicity, hereafter "GDP" and "GDP price index" refer to the output or price series, even though the measures are based on GNP and a fixed weight deflator for some of the sample. Real-time data on GDP, the unemployment rate, and the GDP price index are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM). But in full-sample estimates described below, we use current-vintage data on GDP, the GDP deflator, and unemployment rate obtained from the FAME database of the Federal Reserve Board of Governors.

In the case of interest rates, for which real-time revisions are non-existent, we abstract from real-time aspects of the data and use current vintage data obtained from the FAME database. In the case of the NFCI, its construction from a factor model means it will be subject to revision over time, but in the absence of a long history of vintages of real-time data on it, we abstract from the real-time aspect of the NFCI and use a current vintage series obtained from the website of the Federal Reserve Bank of Chicago. In the case of the turbulence variable, we took the series from the data files of Giglio, Kelly, and Pruitt (2016).

In our main results, our analysis of real-time forecasts uses real-time data vintages from 1985:Q1 through 2018:Q3. As described in Croushore and Stark (2001), the vintages of the RTDSM are dated to reflect the information available around the middle of each quarter. For each forecast origin t starting with 1985:Q1, we use the real-time data vintage t containing data through $t-1$ to estimate the forecast models and construct forecasts for periods t and beyond. Note that this timing means that, in our main results, the last data vintage of 2018:Q3 contains data ending in 2018:Q2. The out-of-sample forecast evaluation uses a sample of forecasts produced starting in 1985:Q1 and ending in 2018:Q2. In our results using the turbulence variable, we use vintages starting with 1972:Q1 and ending with 2012:Q1 (containing data through 2011:Q4) and evaluate out-of-sample forecasts produced starting in 1972:Q1 and ending in 2011:Q4. We report results for forecasts at horizons of one and four quarters ahead. In keeping with ABG, the results for the four-steps-ahead horizon are for the four-quarter average growth rate of GDP (equivalent to $100 \ln(\text{GDP}_{t+4}/\text{GDP}_t)$).

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 years ago. For example, in the mid-1990s, the measure of national output switched from fixed-weight GNP to chain-weighted GDP. Forecasters in 1985 could not have foreseen such changes and the potential impact on measured output. Accordingly, following studies such as Clark (2011), Faust and Wright (2009), and Romer and Romer (2000) that have used early estimates, we use the first available estimates of the real-time measured variables as actuals in evaluating forecast accuracy. In light of the sometimes sizable revisions to GDP growth, we have verified that our results on the accuracy of tail risk forecasts are robust to instead measuring growth outcomes with the second available estimate or the final vintage (2018:Q3 in our case) estimate.[16] For interest rates and the NFCI, the real-time data correspond to the final vintage data.

Finally, in our main results all models are estimated with data on GDP growth, other macro variables, and the NFCI that start in 1971:Q1, reflecting the starting point of the NFCI. The starting point of the estimation sample is then 1971:Q1 + (the lag order of the model as indicated in the previous section) + (the forecast horizon less 1).[17] In our robustness checks with the turbulence indicator, the underlying data on GDP growth and the other variables included are available back to 1959:Q2. In this case, the starting point of the estimation sample is then 1959:Q2 + (the lag order of the model as indicated in the previous section) + (the forecast horizon less 1).[18]

[16] More specifically, in evaluating 5 percent tail forecasts in terms of the coverage, quantile score, VaR-ES score performance of quantile regression and our BVAR models with time-varying volatility, we obtain qualitatively similar results using actual GDP growth measured with the first available, second available, and 2018:Q3 vintage estimates.

[17] The adjustment for the forecast horizon exceeding 1 aligns the estimation sample of the BVARs with that available for the direct multi-step estimation of the quantile regression.

[18] In real-time out-of-sample forecasting with the BVAR, there are a few vintages in which the GDP data start a few quarters later than normal, and in these cases the estimation sample is shortened accordingly.

4 Forecast Metrics

In assessing the efficacy of the models described in the previous section, we consider a range of forecast metrics. In the paper, we primarily provide results using lower and upper quantiles of 5 and 95 percent, respectively (some presented results will pertain to other quantiles). Using lower and upper quantiles of 10 and 90 percent yields very similar results, provided in the supplemental appendix.

Although not a focus of the paper, as a basic check we consider point forecasts, defined as the median of the predictive distributions for the BVAR-SV and BVAR-SVF-M models and the prediction obtained from the quantile regression at the quantile $\tau = 0.5$. We evaluate the point forecasts with the root mean squared error (RMSE). Although the quantile methods advocated in ABG are not intended to produce good point forecasts, the point forecasts provide a basic check of the model; at a minimum, in practice it is useful to know if a model that might be useful for assessing downside risks is also successful at capturing the center of the distribution.

We also examine overall density forecast accuracy, measured with the average log predictive score. As with RMSE and point forecasts, we treat log score performance as a basic check and not the focal point of the comparison. To compute the score with the BVAR-SV and BVAR-SVF-M models, we use a kernel-smoothed estimate of the density of the draws from the predictive distribution. With quantile regression, we compute the predictive score as in ABG, using the second-step, smoothed estimate of the quantile function.

To assess the efficacy of the models in quantifying tail risks, we consider two basic measures of the accuracy of the lower tail quantile estimate, at the 5 percent quantile. For the BVAR-SV and BVAR-SVF-M models, the quantile is simply estimated as the associated percentile of the simulated predictive distribution. For the quantile regression, we use the prediction obtained from the quantile regression at the quantile $\tau = 0.05$. Applied to these quantile estimates, the first accuracy measure is a simple coverage measure for the interval forecast: the percentage of outcomes falling below the 5 percent quantile of the forecast distribution. The second is the quantile score, commonly associated with the tick loss function (see, e.g., Giacomini and Komunjer (2005)). The quantile score is computed as

$$\text{QS}_{\tau,t+h} = (y_{t+h} - Q_{\tau,t+h})(\tau - \mathbf{1}_{(y_{t+h} <= Q_{\tau,t+h})}), \tag{5}$$

where y_{t+h} is the actual outcome for GDP growth, $Q_{\tau,t+h}$ is the forecast quantile at quantile $\tau = 0.05$, and the indicator function $\mathbf{1}_{(y_{t+h} <= Q_{\tau,t+h})}$ 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. Note also that, in coverage and quantile score comparisons, the results from quantile regression are based on the direct quantile estimate; the smoothing step of ABG does not factor into these comparisons.

In comparing the models, we also rely on estimates of expected shortfall (ES) and long-rise measures as in ABG. The shortfall is the conditional expectation (mean or average) of GDP growth rates in the 5 percent tail of the predictive distribution, and the long-rise is the conditional expectation of GDP growth rates in the 95 percent tail of the predictive distribution (see sources such as ABG for explicit formulae). The 5 percent quantile corresponds to the Value at Risk (VaR) — the GDP growth rate that would occur with 5 percent probability; the expected shortfall provides a measure of the average growth rate that would be observed if growth were in that tail of the distribution. With the BVAR-SV and BVAR-SVF-M models, we estimate the expected shortfall and long-rise as the means of forecast draws in, respectively, the 5 percent and 95 percent tails of the predictive distributions. For the quantile regression, we estimate the shortfall and long-rise as in ABG, using the second-step, smoothed estimate of the quantile function to obtain the complete density function and, in turn, the shortfall and long-rise.

We evaluate the shortfall forecasts using the joint value at risk-expected shortfall (VaR-ES) score from Fissler, Ziegel, and Gneiting (2015). 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), whereas value at risk and expected shortfall can be jointly elicited.[19] The joint VaR-ES score is computed as

$$\begin{aligned} \mathrm{S}_{\tau,t+h} &= Q_{\tau,t+h} \cdot (\mathbf{1}_{(y_{t+h} <= Q_{\tau,t+h})} - \tau) - y_{t+h} \cdot \mathbf{1}_{(y_{t+h} <= Q_{\tau,t+h})} \\ &+ \frac{e^{\mathrm{ES}_{\tau,t+h}}}{1+e^{\mathrm{ES}_{\tau,t+h}}} \left(\mathrm{ES}_{\tau,t+h} - Q_{\tau,t+h} + \tau^{-1} \left(Q_{\tau,t+h} - y_{t+h} \right) \mathbf{1}_{(y_{t+h} <= Q_{\tau,t+h})} \right) + \ln \frac{2}{1+e^{\mathrm{ES}_{\tau,t+h}}}, \end{aligned} \tag{6}$$

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

We also evaluate the QR, BVAR-SV, and BVAR-SVF-M forecasts with two additional metrics, the (left tail) quantile-weighted continuous ranked probability score (qwCRPS) and dynamic quantile (DQ) tests. Gneiting and Ranjan (2011) develop the qwCRPS as a proper scoring function of the entire predictive density with quantile weighting to emphasize the left tail. The qwCRPS is computed as a weighted sum of quantile scores at a range of quantiles, with more weight given to

[19]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.

the left tail than the right tail:

$$\text{qwCRPS}_{t+h} = \frac{2}{J-1} \sum_{j=1}^{J-1} v(\tau_j)\text{QS}_{\tau_j,t+h} \tag{7}$$

with $\tau_j = j/J$, and their proposed weighting function for interest in the left tail is $v(\tau_j) = (1-\tau_j)^2$. We set $J = 20$ and compute this score using 19 quantiles, of $0.05, 0.10, 0.15, \ldots, 0.90, 0.95$.

Engle and Manganelli (2004) develop the DQ test to assess whether quantile forecasts meet basic requirements of unbiasedness and, at the one-step-ahead horizon, independence of hits and independence of the quantile estimates — a test that can be thought of as being analogous to the familiar rationality test applied to point forecasts. Our implementation of the DQ test is patterned after that of Brownlees and Souza (2020): We regress the hit rate $\mathbf{1}_{(y_{t+h} <= Q_{\tau,t+h})}$ on a constant and the lagged hit rate of periods t and $t-1$ (we also compute tests instead using the period t and $t-1$ values of the NFCI) and compute a Wald test of the null of 0 coefficients on the lagged hit rates. In light of the small samples of "hit" observations and considerations of brevity, we provide these results in the supplemental appendix, using the 10 percent quantile (see Table A3). At the one-step-ahead horizon, the tests never reject the null of 0 coefficients; although a few rejections occur at the four-steps-ahead horizon, inference may be less reliable at this horizon case due to the small sample and challenges of autocorrelation-robust inference. Nieto and Ruiz (2016) provide an overview of research that has highlighted some of the small-sample size and power challenges with the DQ test and others used in evaluation of VaR forecasts.

The remaining sections of the paper present results using these forecast metrics. Although our focus is on conventional out-of-sample forecasts, we also provide some results on in-sample forecasts. We do so in part because, due to the NFCI data being available back to only 1971, the overall sample is too short to allow out-of-sample evaluation over the recessions of the 1970s and early 1980s. In addition, with many of the results reported in ABG being in-sample rather than out-of-sample, providing in-sample results in our paper facilitates comparison of our BVAR-based models to quantile regression-based results. 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.[20]

[20] Regarding the treatment of the latent volatility states in the in-sample forecasts, we construct the forecasts so as to reflect some of the uncertainty around the path of volatility over each forecast horizon. Specifically, at each forecast origin t in the sample, for each MCMC draw we simulate a path of log volatility from t through $t+H-1$ periods ahead, starting from the smoothed estimate of log volatility in period $t-1$. We then compute the implied Σ_{t+h} and draw shocks to y for period $t+h$ with variance Σ_{t+h}. We feed in the shocks and iterate the VAR forward starting from the data y_{t-1} to obtain draws of forecasts for periods t through $t+H-1$.

5 Empirical Results

This section begins with tests of skewness and kurtosis in the raw data, BVAR-SV residuals, and BVAR-SV out-of-sample forecast errors (results with the BVAR-SVF-M specification are qualitatively similar and omitted for the sake of brevity). It next reviews some of the practical challenges with using quantile regression to assess tail risks in short macroeconomic time series. The section then examines — for GDP growth — estimates of expected shortfall and long-rise in both in-sample and out-of-sample forecasts to compare the abilities of the models under consideration to capture downside risks. The section next provides an analysis of in-sample and out-of-sample forecast accuracy for GDP growth. The final two subsections consider robustness, first with respect to replacing the NFCI with a measure of financial volatility and second with respect to forecasting the unemployment rate rather than GDP growth.

5.1 Skewness and Kurtosis Properties

As noted in the introduction, asymmetries in the distributions of outcomes and predictive distributions might be expected to be associated with skewness in the data, residuals, and forecast errors (subject to the caveat that simple checks of skewness abstract from time variation and can be seen as applying to marginal distributions, which might show little skewness on average even if conditional distributions are sometimes asymmetric). Although some papers in the literature present skewness statistics, they normally do so without formal inference of statistical significance (e.g., Kozeniauskas, Orlik, and Veldkamp (2018) and Orlik and Veldkamp (2015)). To assess the significance of skewness, we use the formal time series tests of Bai and Ng (2005), using HAC variances computed with the pre-whitened quadratic spectral kernel, as developed in Andrews and Monahan (1992). We also include the Bai and Ng (2001) test of conditional symmetry, computed as described in their paper.[21] In part because some studies consider overall tests for normality that cover both skewness and kurtosis (e.g., Jensen, et al. (2020)), we also include test results for kurtosis and normality. To provide a graphical illustration of possible asymmetries, we also follow some recent studies (e.g., Galbraith and van Norden (2019)) in providing quantile-quantile (Q-Q) plots, in our case for the BVAR-SV residuals and forecast errors. These plots compare the empirical quantiles of the residuals or forecast errors with quantiles of the normal distribution. To facilitate this comparison, we first compute the quantiles of the standard normal distribution and

[21] In applying the conditional symmetry test to the data, we use the residuals from AR(4) models estimated for each series. In applying the test to VAR residuals or forecast errors, we use the residuals or errors in question without further transformation.

then adjust these quantiles to reflect the same mean and variance as each time series in question.

In these estimates, the sample is 1972:Q1-2018:Q2 for the data and BVAR-SV residuals and 1985:Q1-2018:Q2 for the out-of-sample forecast errors. The residuals used in these estimates are the posterior medians of residuals across simulation draws. We also consider normalized residuals, which are the posterior medians of draws of residuals divided by the standard deviation, where the standard deviation is the square root of the corresponding diagonal element of Σ_t for each draw.[22] In the case of the out-of-sample forecast errors from the BVAR-SV specification, the forecast errors are computed using point forecasts defined as posterior medians, at forecast horizons of one and four quarters.

In Table 1's results for the raw data (first panel) and BVAR-SV residuals (second panel), the skewness statistics are often large, but not often statistically significant. For example, GDP growth has a skewness statistic of -0.364, but the estimate is sufficiently imprecise that the Bai and Ng (2005) test statistic for skewness is not close to rejection. In the raw data and BVAR-SV residuals, there is more evidence of kurtosis than skewness, with rejections of no-kurtosis for several of the variables.[23] In contrast, when the BVAR residuals are normalized by their time-varying volatilities, the evidence of kurtosis declines, whereas the evidence of skewness increases, with rejections of no-skewness and conditional symmetry for the unemployment rate and NFCI.[24] As indicated in the bottom two panels of the table, the evidence of skewness and kurtosis is modestly weaker in the shorter sample of out-of-sample forecast errors from the BVAR-SV model (based on real-time data, as noted above). In this case, the Bai-Ng tests yield little evidence of significant skewness or kurtosis, at any horizon, although the test for conditional symmetry provides some evidence for asymmetries.

As another, more informal and visual assessment of skewness and kurtosis in the BVAR-SV residuals, Figure 1 provides Q-Q plots for each variable's residual, with the raw residuals in the

[22] More specifically, in the case of the BVAR-SV residuals results, for the time series of each variable i, we compute skewness statistics for the posterior median of the draws of $v_{i,t}$. In the normalized residuals results, for each draw j, we compute $v_{i,t}^{(j)}/\sqrt{\sigma_{i,t}^{(j)}}$, where $\sigma_{i,t}$ denotes the i-th diagonal element of Σ_t (the smoothed estimate from draw j) and its median across draws. We then apply the tests to this time series.

[23] The large kurtosis statistic for the federal funds rate residual may be associated with the very high volatility that occurred with the monetary policy regime changes in the 1979-82 period, which likely produces a large kurtosis point estimate at the same time it inflates the imprecision of the fourth moment. The kurtosis statistic for the funds rate residual is much lower when the sample is shortened to start in 1985.

[24] In results detailed in the supplemental appendix, when we shorten the sample to 1985-2018 (without changing the sample of model estimates used to obtain the residuals), it remains the case that the normalized residuals show significant asymmetries for the unemployment rate and NFCI, although the evidence of significant skewness or kurtosis in the data and raw residuals is a little weaker than in the 1972-2018 sample. In the shorter sample, our estimate of skewness in GDP growth data is similar to the estimate of Jensen, et al. (2020); whereas their bootstrap approach to inference implies that their estimate is statistically significant, our approach based on the normal-based inference of Bai and Ng (2005) implies that our estimate is not significant.

upper panel and the volatility-normalized residuals in the lower panel, and Figure 2 provides Q-Q plots for real-time out of sample (OOS) forecast errors from the BVAR-SV specification, at horizons of one and four quarters ahead. The charts compare the empirical quantiles of the time series of residuals or forecast errors (again, posterior medians) against the quantiles of the normal distribution, with the sloped line tracing out where the empirical quantiles would lie if they matched the normal quantiles. In Figure 1's results for residuals, the upper panel displays some notable departures from normality, most dramatically for the federal funds rate and the NFCI. Normalizing the residuals by their volatilities (lower panel) helps move the empirical quantiles toward the normal case, although still with some departures, most notably for the unemployment rate and the NFCI. Figure 2's results for forecast errors (both horizons) are qualitatively similar in suggesting departures from normality, although a little less so for the funds rate in the out-of-sample results than in the in-sample residuals case.

Overall, from these results we conclude that, consistent with some prior results in the literature, there is some suggestive evidence of asymmetries in macroeconomic data, but the formal statistical evidence is hardly overwhelming. Notably, in these estimates there does not appear to be much evidence of asymmetries in GDP growth (data, VAR residuals, or forecast errors).

5.2 Empirical Challenges with Approaches Based on Quantile Regression

Although quantile regression-based approaches are drawing considerable interest for quantifying tail risks in macroeconomics, they can come with some challenges in data samples of the size typically available. One challenge is that a quantile regression approach can yield quantile estimates that cross one another. In developing results for this paper, crossing occurred occasionally in real-time forecasting. Figure 3 provides two examples, reporting time series of forecast quantiles (5 percent, 25 percent, 75 percent, and 95 percent, dated by the forecast origin) obtained recursively using real-time data vintages from 1985:Q1 through 2018:Q2. The first, reported in the top panel of the figure, occurs with a one-quarter-ahead model of GDP growth relating growth to a constant, lagged growth, and lagged turbulence (as part of the robustness check provided in Section 5.5). In this case, crossing occurs in that the 95 percent quantile drops below the 75 percent quantile in two different periods of the sample. In a case such as this, with the focus on bottom tail risks, the crossing in upper tail quantiles might not be a great concern, but this kind of crossing can pose challenges with the second-step fitting of a skewed distribution to the few quantiles, and in turn contaminate the shortfall and long-rise estimates obtained on the basis of these second-step estimates. The second crossing example, in the bottom panel of the figure, occurs with a model (similar to the Phillips curve-motivated model of Lopez-Salido and Loria (2019)) relating the four-quarters-ahead change in GDP inflation to a constant and to the gap between inflation and a survey-based long-horizon expectation, the unemployment rate, and the NFCI. (Although we did not use this model in the paper's results, we produced these results in exploratory analysis of tail risks to inflation.) In this case, the crossing occurs in the early years of the forecast sample, in the 75 percent and 95 percent quantiles. In unreported estimates, additional crossing or meeting of quantile estimates occurs around 2010, in the 75 percent-90 percent and 10 percent-25 percent quantiles.

Another challenge with quantile regression-based methods in macroeconomic time series is that small samples can create precision challenges with tail quantile estimates. In keeping with some of the overall volatility in our out-of-sample quantile regression estimates presented later in this section, we observed some sharp changes in quantile estimates that could occur with small changes in sample. In one of the more striking examples, when (in unreported results) we added a few years of data to the sample of ABG as a check of our data and code, the estimated coefficient estimate on the NFCI for the 75 percent quantile changed sharply (from about -1 to about 0.1). Of course, this particular change does not affect the lower tail quantile, but it can affect the shortfall estimate due to the second step involved in smoothing the quantile estimates by fitting a skewed-t distribution. The top panel of Figure 4 provides the time series of estimates of the coefficient on the lagged NFCI in the one-step-ahead quantile regression model of GDP growth, for the median and 5 percent, 25 percent, 75 percent, and 95 percent quantiles. The estimates, obtained recursively using real-time data vintages from 1985:Q1 through 2018:Q2, are dated by the forecast origin. Of course, even in conventional linear models, estimates of the conditional mean parameters can shift around over time. Consistent with that, the median coefficient estimate rises some early in the sample before stabilizing around a value of about -1. The estimates of coefficients at the 25 percent and 75 percent quantiles display a comparable degree of time variation. Time variation is much greater for the estimates of coefficients at the 5 percent and 95 percent quantiles. This kind of variability may contribute to weaker forecast performance out-of-sample than in-sample. The bottom panel of Figure 4 compares the corresponding quantile estimates from the QR model to those from the BVAR-SV specification. The variability of the QR model's upper-tail coefficient on the NFCI in the late 1980s and 1990s translates into some volatility in the estimate of the 95 percent quantile relative to the estimate obtained with the BVAR-SV model. The variability of the QR and BVAR-SV quantile estimates are more comparable for the median and lower tail quantiles. As we will detail below in results using turbulence as an alternative financial indicator, this small-sample challenge with tail quantiles can manifest itself as a quantile estimate that can be seen as unrealistically extreme.

In the quantile literature, the challenges of estimating extreme quantiles with small samples of observations are well known. Researchers have developed extremal quantile methods of bias correction and inference that improve on the performance of conventional estimators and conventional Gaussian inference in such cases. According to a rule of thumb summarized in the survey of Chernozhukov, Fernandez-Val, and Kaji (2017), extreme value methods should be used when $\tau T/k \leq 15$ to 20, where τ denotes the quantile, T is the sample size, and k is the number of regressors. In an application like that of ABG and this paper, with about 40 years of quarterly data, a quantile of 0.05, and 3 regressors (constant, lagged GDP growth, and lagged NFCI), $\tau T/k \approx 2.7$. This suggests some challenges with using standard, simple quantile regression for extreme quantile estimation, tail forecasting, and inference in quarterly macroeconomic time series.

In light of these challenges, in unreported results we also considered imposing some shrinkage on the quantile regression estimates through Bayesian quantile regression (a tool used in the nowcasting application of Carriero, Clark, and Marcellino (2020)).[25] 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). In our (limited, in the interest of brevity) applications forecasting GDP growth one step ahead and the unemployment rate four steps ahead, the use of Bayesian estimation reduced the variability of coefficient estimates across the recursive, out-of-sample forecasting windows. The resulting quantile estimates tended to move in fairly close alignment across quantiles (more so than the conventional quantile regression estimates), likely reducing the scope for the crossing of quantile estimates. Nonetheless, on an out-of-sample basis, the time series of the 5 percent quantile estimate obtained from Bayesian quantile regression is very similar to the estimate obtained from simple, conventional quantile regression. So in our main application of interest, Bayesian estimation appears unlikely to improve lower tail forecast accuracy. Of course, there could be other applications in which it could have a better payoff.[26]

5.3 Predictive Distributions

To compare the abilities of the quantile regression, BVAR-SV model, and BVAR-SVF-M model to capture downside risks to GDP growth, Figures 5 and 6 report time series of expected shortfall (at 5 percent) and long-rise (at 95 percent) estimates at the one-step- and four-steps-ahead forecast

[25] Korobilis (2017) uses a Bayesian quantile regression in an inflation forecasting application.

[26] In fact, in the tail risk nowcasting application with a large information set of Carriero, Clark and Marcellino (2020), Bayesian quantile regression performs much better than standard, simple quantile regression.

horizons. The in-sample forecast estimates of Figure 5 display some of the asymmetries highlighted by ABG. For example, with GDP growth estimates from quantile regression and the BVAR-SV model (top panel), the shortfall drops sharply around the Great Recession of 2007-2009, whereas the long-rise changes relatively little. The same occurs in some episodes around recessions in the 1970s and early 1980s. For GDP growth, these asymmetries hold up at the four-steps-ahead horizon. As noted above, the BVAR-SV estimates are showing more variation in downside risk as compared to upside risk even though the underlying conditional predictive distributions are symmetric. Simultaneous shifts in conditional means and variances suffice to produce this relative time variation without asymmetry in conditional distributions being necessary. As noted above, in our estimates, the simple correlation between the estimated volatility of GDP growth and the level of GDP growth (using a four-quarter average for smoothing) is -0.3. This comovement obtains even though it is not directly built into the model. Again, as far as we know, our paper is the first to make this point about simultaneous shifts in means and variances yielding asymmetries in upside and downside risks.

On the other hand, as indicated in the results in the upper panel, the four-steps-ahead estimates from the quantile regression specification are overall a fair amount more variable than the BVAR-SV model's estimates, with much more symmetry than in the one-step-ahead case. The lower panel of Figure 5 directly compares in-sample shortfall and long-rise estimates from the BVAR-SV and BVAR-SVF-M specifications. Qualitatively, these models yield similar estimates, although the BVAR-SVF-M specifications yield larger moves in long-rise and shortfall estimates at the four-step horizon — but with even more symmetry than in the SV specifications. In all cases, the expected shortfall is more variable than the long-rise, in keeping with one of the asymmetries patterns noted in ABG and CSZ. For example, at the four-step horizon, the standard deviation of shortfall divided by the standard deviation of long-rise is 1.5 for the quantile regression estimates and 1.9 for the BVAR-SV estimates (details are provided in the supplemental appendix's Table A2). In addition, results provided in the appendix (Figure A3) show that a finding of CSZ obtains in our BVAR-SV estimates: For the NFCI (financial conditions), the asymmetry is in an upside direction, with long-rise more variable than shortfall.

Moving from in-sample to out-of-sample forecasts as reported in Figure 6 weakens somewhat the picture of asymmetries between the expected shortfall and long-rise. The out-of-sample estimates show some of the same asymmetries as do the corresponding in-sample results, but not as many. For GDP growth, the BVAR-SV model captures some asymmetries around the Great Recession.[27]

[27]Some supplementary analysis indicates that the increase in the BVAR-SV estimate of the long-rise around the

At the one-step horizon, the QR-based estimates actually display some **upside** asymmetry in long-rise in the early or mid-1990s; this is not as evident in the BVAR-SV estimates. In addition, over the out-of-sample period, the differences in volatilities of shortfall versus long-rise are more modest than in the in-sample period. For example, at the one-step (four-step) horizon, the standard deviation of shortfall divided by the standard deviation of long-rise is 0.8 (1.2) for the quantile regression estimates and 1.1 (0.9) for the BVAR-SV estimates. In the out-of-sample case, as in the in-sample case, from the perspective of capturing downside asymmetries there appears to be no broad advantage to the BVAR-SVF-M model over the BVAR-SV specification, although, in the in-sample case, shortfall estimates fall more in the recessions of the early 1980s with BVAR-SVF-M than with BVAR-SV.

In supplemental results included in the appendix, we have also directly compared the in-sample estimates from each model to its corresponding out-of-sample estimates. With the BVAR-SV and BVAR-SVF-M estimates, we obtain the pattern that might be expected: The out-of-sample estimates of shortfall and long-rise tend to show more variability than the in-sample estimates, with the out-of-sample measures looking like noisy estimates of the in-sample measures. In contrast, with the quantile regression approach, the out-of-sample estimates of shortfall and long-rise are generally less variable than the in-sample forecast estimates, although there is something of a notable exception with long-rise in the early 1990s, when the in-sample estimate spiked higher for a time. In the out-of-sample case, as in the in-sample, for the NFCI (financial conditions), the asymmetry is in an upside direction, with long-rise more variable than shortfall (see appendix Figure A3).

Differences in in-sample and out-of-sample forecast estimates could be driven partly by instabilities in the parameters of the VAR. In the in-sample case, the parameter estimates will average across any breaks, and the forecasts will reflect these averages. In the out-of-sample case, instabilities in underlying parameters can get partially accommodated as parameter estimates are recursively updated as the sample expands with forecasting moving forward in time. We have checked the time profiles of recursive parameter estimates from the BVAR-SV and BVAR-SVF-M

Great Recession is not driven by the zero lower bound (ZLB) on the federal funds rate. The baseline estimates in the figure do not impose the ZLB. When we use conditional forecasting to impose a non-zero bound on the federal funds rate for forecasts produced between 2009 and 2014, we obtain a qualitatively very similar path of quantile, shortfall, and long-rise estimates. In these computations, to impose the ZLB on forecasts produced at origins between 2009:Q1 and 2014:Q4, we follow the approach of Aastveit, et al. (2017) and compute BVAR-SV forecasts conditional on the federal funds rate staying at 0.15 percent over the forecast horizon. By early 2009, verbal forward guidance from the Federal Reserve had led financial markets to expect the federal funds rate to remain near zero for at least a year, as evidenced in (early 2009 and subsequent) survey-based projections of short-term interest rates. That remained the case well into 2014. The conditional forecasts are produced using a Kalman filter implementation of the minimum-MSE approach to conditional forecasting that is standard with VARs.

models and found that these do not display clear breaks. A few coefficients gradually drift by small or modest amounts, although these changes are not generally large relative to the imprecision around each estimate. One might also worry that the asymmetries in tail risks over time are driven by changes in parameter estimates around recessions. But the recursive parameter estimates don't show shifts around recessions, either. This is consistent with the asymmetries we find in shortfall relative to long-rise as genuine. Although at a conceptual level one might prefer the BVAR-SVF-M model over the BVAR-SV because it directly allows for asymmetries in conditional predictive distributions, in practice a BVAR-SV model appears to be able to capture similar behavior in asymmetries over time in shortfall as compared to long-rise.

In the supplemental appendix, we provide the results of Monte Carlo experiments that corroborate this finding. In particular, with data generated by a BVAR-SVF-M specification, one-step-ahead predictive distributions yield more time variation in downside risks than upside risks, with comparable performance of BVAR-SV and BVAR-SVF-M models. This pattern stems from simultaneous shifts in conditional means (down) and conditional variances (up). In the BVAR-SVF-M case, these simultaneous shifts can occur with shocks to the common volatility factor that enter both the conditional mean and variance. In the BVAR-SV estimates, although the BVAR-SV specification assumes that "levels" innovations to the data y_t are independent of innovations to log volatility, in the data and estimates for the Monte Carlo data, it appears that over short periods the model captures patterns of correlated shocks to levels and volatilities that push conditional means and variances in opposite directions and lead to more variability in the lower tail quantile than the upper tail quantile.

As this discussion suggests, one aspect of model specification that appears to be important for capturing a pattern of time variation in downside risks to output growth like that of ABG is including in the model (in its conditional mean) a measure of financial conditions. These conditions appear to bear significantly on the timing of the model's estimates of shifts in the means and variances of the conditional predictive distributions. For example, in the in-sample case, if we exclude financial conditions and consider just a four-variable macroeconomic VAR, we still obtain more time variation in downside risks as compared to upside, but the timing of the asymmetric shifts in the bottom tail compared to the top is considerably different than in the estimates presented in this section.

5.4 Forecast Accuracy

This subsection provides a more formal analysis of forecast performance, both in-sample and out-of-sample, for point forecasts, density forecasts, interval and quantile forecasts, and the joint VaR-ES forecasts. Admittedly, the in-sample forecasts are not really forecasts. But with the short sample (especially with respect to tail risks) and other studies in the risk assessment literature relying on in-sample forecasts, the in-sample analysis can help shed some light on the relative performance of the models we consider. In light of a potentially large role for the recent Great Recession (associated with a business cycle peak of 2007:Q4 and trough of 2009:Q2) and subsequently slow recovery, we provide results for a full sample of 1985-2018 and a subsample of 1985-2007.

Table 2 provides results for the in-sample forecasts from the quantile regression and BVAR-based models. To facilitate comparisons, except in the case of the coverage rates, the results for the BVAR-SV and BVAR-SVF-M models are reported as relative to those for quantile regression. In particular, for these models, we report RMSE as a ratio of the quantile regression's RMSE, the quantile score as a ratio of the quantile regression's score, and the qwCRPS as a ratio of the corresponding score for quantile regression — in all three cases, a ratio below 1 means the BVAR is more accurate. We report the log predictive and VaR-ES scores as differences with respect to the quantile regression, so that a differential above 0 means the BVAR is more accurate. To gauge statistical significance (roughly, given that the test in question is intended for out-of-sample forecasts and the first table of results are for in-sample forecasts), we estimate Diebold and Mariano (1995)–West (1996) t-tests for equality of the average loss (with loss defined as squared error, log score, quantile score, VaR-ES score, or qwCRPS). We also compute t-tests for the empirical coverage rate equaling the nominal rate of 5 percent. In the tables, differences in accuracy 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).

In terms of conventional point forecasting and density forecasts as captured by log scores, the BVARs and quantile regression are broadly similar in accuracy. In point forecasting, the BVARs have RMSEs similar to one another and slightly better than quantile regression, although the differences do not appear to be large enough to be statistically significant and some of the gain in RMSE appears to be driven by the post-2007 part of the sample. The similarity of log score results means that, even if quantile regression were capturing downside risks, it is not fitting the overall density better than do the BVAR-SV and BVAR-SVF-M specifications (a higher number means a better forecast). With respect to the 5 percent (left tail) interval forecasts, all of the models yield coverage rates reasonably close to the nominal rates, although the BVAR-SV and BVAR-SVF-M rates get down to about 1 percent (implying bands too wide) at the one-step horizon in the 1985-2007 period.

Turning to the quantile score, VaR-ES score, and qwCRPS results that bear more directly on the efficacy of the models in capturing downside tail risks, by the quantile score, the quantile regression and BVAR-based models perform similarly. There are some small differences between the BVAR models and quantile regression, but none statistically significant by the rough (for in-sample purposes) metric of the Diebold and Mariano–West statistics. For the VaR-ES score, quantile regression and the BVAR-SVF-M model perform pretty similarly, with no statistically significant differences. By this metric, the BVAR-SVF-M specification scores a little better than does the BVAR-SV model, but again, few differences (only one) are statistically significant. According to the qwCRPS, the BVAR-SV and BVAR-SVF-M models are consistently more accurate than quantile regression, but only by a little, with only one difference statistically significant.

Table 3 provides corresponding results for out-of-sample forecasts. In both point and density forecasts of GDP growth, the BVARs are somewhat better than quantile regression, at both horizons. In point forecast accuracy, the BVAR-based specifications are similar to one another and both are modestly better than quantile regression, more noticeably than in the in-sample forecast results. In overall density forecast accuracy as captured by log scores, the BVARs continue to have a small to modest advantage over quantile regression, with BVAR-SVF-M achieving statistically significant gains above 10 percent at the one-step-ahead horizon. With both point and density forecasts for the 1985-2018 sample, the Diebold and Mariano–West tests indicate that the gains with the BVAR-SVF-M model over quantile regression are statistically significant at the shorter forecast horizon.

As to the tail risk-focused measures, by the quantile score (again, at the 5 percent quantile), the BVAR-SVF-M model is as good as (h=1) or better (h=4) than quantile regression. At the longer horizon, the performance of quantile regression on an out-of-sample basis is influenced by a period around the 1990-1991 recession when quantile regression does not do well at the four-step horizon. The BVAR-SV specification is not quite as good as the BVAR-SVF-M model, but not really very different, either. With the quantile score measure, none of the differences for the BVARs relative to the quantile regression are large enough to achieve statistical significance with the

Diebold and Mariano–West test. By the joint VaR-ES score, the BVAR-SVF-M model is broadly comparable to quantile regression, with quantile regression better at the one-step horizon and the BVAR-SVF-M model better at the longer horizon. But the differences are not large enough to be statistically significant. By this score, too, for most horizon-sample combinations, the BVAR-SV model is not quite as good as the BVAR-SVF-M specification, but it is not really different, either. By the qwCRPS metric, both the BVAR-SV and BVAR-SVF-M models consistently improve on the accuracy of quantile regression, by very comparable amounts, but none of the differences are statistically significant.

What should one take away from these forecast comparisons? As noted above, our paper is the first in the recent macroeconomic tail risk literature to formally evaluate the accuracy of the tail risk forecasts. In these results for GDP growth, quantile regression doesn't seem to offer any advantages in forecast accuracy over a BVAR-SV or BVAR-SVF-M specification. In some dimensions, one or another of the BVAR specifications is better. Of course, quantile regression itself is quite simple, but if one wants to assess tail risks with expected shortfall or assess shortfall forecasts, as opposed to just compute a tail quantile and take it as a measure of GDP-at-risk as in Adrian, et al. (2018), the second-step smoothing of ABG becomes necessary, and arguably, for those already familiar with BVARs, that step adds at least some complication. In addition, to obtain results for more than a single forecast horizon or for more than one variable, one must specify and estimate different models for each horizon and each variable. Our results show that one can keep the rich and broadly useful features of BVARs with stochastic volatility for forecasting and structural analysis while still using them for the risk assessments now of interest — and obtain tail risk assessments quite comparable to what would be obtained with quantile regression. And one can do with a single BVAR covering all variables and horizons rather than multiple models covering each different variable-horizon combination. As noted above, the BVAR specifications yield this asymmetry in the time variation of downside versus upside tail risks even though the conditional predictive distributions are symmetric, because they are subject to simultaneous shifts in means and variances that the model estimates accommodate through periods of correlation in shocks to levels and volatilities.

5.5 Results Using Financial Turbulence

To assess the robustness of our findings to the choice of the NFCI as the measure of financial conditions, this subsection presents the key results of interest obtained by measuring financial conditions with the turbulence indicator of Giglio, Kelly, and Pruitt (2016). In these results, turbulence replaces the NFCI as the financial variable in the VAR and QR models and as the variable to which volatility is attached in the BVAR-SVF-M specification. Reflecting availability, the turbulence-based results use earlier data (raw data back to 1959 in estimation and a 1972:Q1 starting vintage in real-time forecasting) and a sample that ends in 2011 instead of 2018.

Figures 7 (in-sample) and 8 (out-of-sample) compare estimates of expected shortfall (at 5 percent) and long-rise (at 95 percent) at the one-step- and four-steps-ahead forecast horizons obtained with the baseline models including the NFCI and the alternative models including turbulence. In each figure, the top panel compares estimates from the baseline QR specification including the NFCI to estimates from the QR specification including turbulence, and the bottom panel compares estimates from the baseline five-variable BVAR-SV model including the NFCI and the BVAR-SV specification including turbulence. In the case of the BVAR-SV model, the shortfall and long-rise estimates are quite similar across the specifications including the NFCI and turbulence. For example, regardless of this choice of financial indicator, the in-sample estimates of shortfall are considerably more variable than those of long-rise. So the BVAR-SV specification seems to have some robustness to the choice of financial indicator included in the model. However, with quantile regression, the shortfall and long-rise estimates are more sensitive to the choice of financial indicator. Shortfall estimates are considerably more variable with turbulence as the financial indicator than with the NFCI as the indicator, whereas long-rise estimates are less variable with turbulence included than with the NFCI included. The difference across the NFCI and turbulence specifications is especially sharp with the out-of-sample estimates of shortfall (top panel of Figure 8). In fact, the shortfall estimates (of less than -30 at the one-step horizon and about -40 at the four-step horizon) obtained from the quantile regression including turbulence might strain credulity in the eyes of some readers. This may be due to some regressor-specific peculiarities in fitting the tail quantiles of GDP growth with quantile regression; the median projections from one quantile regression using the NFCI versus another with turbulence are much more similar than are the shortfall estimates.

Tables 4 (in-sample) and 5 (out-of-sample) compare the forecast accuracy of the BVAR-based models with turbulence to the quantile regression with turbulence. In the quantile score and VaR-ES score results for models including turbulence as the indicator of financial conditions, the BVAR-SV and BVAR-SVF-M models are in most cases significantly more accurate than the QR specification. This applies with both the in-sample and out-of-sample forecasts, and across the evaluation samples of 1972-2011, 1985-2011, and 1985-2007. The patterns are broadly similar with the qwCRPS metric, although with fewer instances of statistical significance. In the quantile score, the BVAR models have a score value about 30 percent lower than the QR's score value, a difference large enough to often achieve significance by the Diebold-Mariano test metric. In the VaR-ES score results, the BVAR-SV and BVAR-SVF-M scores are commonly 0.3 (or more) higher than the QR score, and often significant according to the Diebold-Mariano test. The better performance of the BVAR models versus QR when turbulence is included than in the baseline case of the NFCI being included reflects the fact that, with QR, results are sensitive to the choice of financial indicator — better with the NFCI than with turbulence.

One other consideration with the use of VAR-based models is their easy scalability: although we include only the NFCI as a measure of financial conditions in our baseline specification, we can easily add other measures to the model. For example, in the forecasting and tail risk literature (e.g., CSZ), it is common to use a credit spread to measure financial conditions. We have verified that adding the credit spread of Gilchrist and Zakrajsek (2012) to our baseline BVAR-SV specification (which also includes the NFCI) yields long-rise and shortfall estimates and forecast accuracy very similar to our baseline estimates reported herein.[28] This indicates another dimension along which a VAR model can be robust to some specification change.

5.6 Results for the Unemployment Rate

This subsection provides a subset of results for forecasts of the change in the unemployment rate, at the four-quarters-ahead horizon. As explained in Section 3's exposition of the quantile regression specification, our use of a multi-step change in the unemployment rate is patterned on the specification of Kiley (2018). The forecasts we report for our BVAR-SV and BVAR-SVF-M models are also for the four-quarters-ahead change in unemployment. In these cases, we use the same model in quarterly rates described above; we transform the quarterly forecasts to obtain the needed four-quarters-ahead change in unemployment. One other important point to note is that, with these unemployment results, we focus on the upper tail (95 percent and above) outcomes, rather than the lower tail as in the GDP growth results. To facilitate results (particularly since the VaR-ES score measure is defined for the negative tail), for this purpose, before computing the quantile and VaR-ES scores, we flip the 5-95 percent quantiles and multiply the data and quantiles by -1, and then compute the scores for the bottom 5 percent tail (or in some cases we produce the results by multiplying the change in the unemployment rate by -1).

Figure 9 reports time series of expected shortfall (at 5 percent) and long-rise (at 95 percent)

[28]We obtained data on the credit spread from 1973 through 2018 from the website of the Federal Reserve Board of Governors, published as regular updates associated with Favara, Gilchrist, Lewis, and Zakrajsek (2016).

estimates at the four-steps-ahead forecast horizon, for the quantile regression and BVAR-SV specifications (BVAR-SVF-M and BVAR-SV estimates are very similar). The in-sample estimates in the upper panel display some notable upside asymmetries, implying upside risk of increases in the unemployment rate, most notably in the mid-1970s, in the late 70s-early 80s, and around the Great Recession. The standard deviation of the long-rise estimate is about twice that of the expected shortfall. The contours of the quantile regression and BVAR-SV estimates are quite similar. Moving from in-sample (top panel) to out-of-sample forecasts (bottom panel) weakens somewhat the picture of asymmetries between the expected shortfall and long-rise. In the shorter sample of real-time, out-of-sample forecasts, shortfall and long-rise move more symmetrically together, with the exception of the choppy spikes in the quantile regression's estimated long-rise around the late 1990s.

Tables 6 and 7 report interval coverage rates, quantile scores, VaR-ES scores, and qwCRPS results for in-sample and out-of-sample forecasts, respectively. In the in-sample case, the percentages of observations falling in the upper 5 percent tail of the predictive distributions are reasonably close to 5 percent for all three models, such that none of the departures from 5 percent coverage are statistically significant. The score measures indicate that the BVAR-based specifications are about as accurate as quantile regression, with few statistically significant differences in accuracy, except that the qwCRPS results indicate the BVARs to be significantly better than quantile regression. In the out-of-sample case, the estimated 95 percent quantile of the forecast distribution appears to not be quite high enough; modestly more than 5 percent of unemployment rate outcomes fall in the upper 5 percent tail of the forecast distributions. However, none of the departures from 5 percent coverage appear to be statistically significant. In the out-of-sample scores, the performance of the BVAR-SVF-M model is very similar to that of the BVAR-SV specification, and both are comparable in accuracy to quantile regression, with score differences that are statistically insignificant.

These results for one-year changes in the unemployment rate corroborate our main results on (negative) tail risks to GDP growth; with (positive) tail risks to unemployment, too, quantile regression doesn't seem to offer any advantages in forecast accuracy over a BVAR-SV or BVAR-SVF-M specification.

6 Conclusions

A rapidly growing body of research has examined tail risks in macroeconomic outcomes. Most of this work has focused on the risks of significant declines in GDP, and relied on quantile regression methods to estimate tail risks. Although much of the recent work on macroeconomic tail risks hasn't cleanly distinguished symmetry in conditional predictive distributions from unconditional distributions, the evidence of downside risk varying more than upside risk that has become a focus of this work can obtain with conditional predictive distributions that are symmetric but subject to simultaneous shifts in conditional means and variances. Accordingly, in this paper we examine the ability of Bayesian VARs (BVARs) with stochastic volatility (SV) to capture tail risks in macroeconomic forecast distributions and outcomes. In particular, we focus on capturing asymmetries in the time series behavior of measures of upside and downside risks that imply asymmetries in unconditional distributions but do not necessarily require asymmetries in conditional predictive distributions. Another novelty of our paper in this tail risk literature is that we include formal evaluation of tail risk forecasts, using the quantile score and a joint test of value-at-risk and expected shortfall that has a conceptual advantage in that the test is based on an elicitable loss function.

We find, first, that the evidence of skewness in output growth is not all that strong, statistically speaking. Second, although quantile regression-based approaches can be a useful tool for quantifying tail risks in macroeconomics, they can come with some challenges in data samples of the size typically available in macroeconomics. Challenges we document include quantile crossing and considerable variability in the coefficient estimates of quantile regressions. In addition, our quantile regression results show some fragility with respect to the choice of financial indicator. Third, with our BVAR specifications featuring time-varying volatility, we are able to capture more time variation in downside risk as compared to upside risk for output growth (vice versa for the change in the unemployment rate) in line with the results of quantile regression, and in formal evaluation of tail risk forecasts, our BVAR specifications perform comparably to quantile regression (with some gains in standard point and density forecasts). Finally, our findings on the effectiveness of BVARs in capturing tail risks apply with both a conventional BVAR with stochastic volatility and a model using a BVAR-SVF-M specification, with a common volatility factor that enters the conditional mean of the VAR. These models appear to be equally well suited to capturing asymmetries in the unconditional predictive distribution of GDP growth. The BVAR-SV captures simultaneity in mean and variance shifts with sporadic correlation between the empirical estimates of level and volatility shocks, whereas in the BVAR-SVF-M model, a shock to the volatility factor also represents a shock to the levels of the macroeconomic variables.

Our findings imply that, at least for GDP growth (and the change in the unemployment rate), quantile regression doesn't seem to offer any advantages in forecast accuracy over a BVAR-SV or

BVAR-SVF-M specification. Actually, in some dimensions, one or another of the BVAR specifications is better. Moreover, quantile regression itself is quite simple, but if one wants to assess tail risks with expected shortfall or assess shortfall forecasts, as opposed to just compute a tail quantile and take it as a measure of GDP-at-risk as in Adrian, et al. (2018), the second-step smoothing of ABG becomes necessary, and that step adds complications. Our results show that one can keep the rich and broadly useful features of BVARs for forecasting while still using them for the risk assessments of interest and obtain tail risk assessments quite comparable to what would be obtained with quantile regression. Finally, another possible challenge with quantile regression-based methods in macroeconomic time series is the availability of small samples that create precision challenges with tail quantile estimates. Applying recently developed extremal quantile methods of bias correction and inference that improve on the performance of conventional estimators could be an interesting topic for further research in this area.

References

[1] Aastveit, Knut Are, Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino (2017), "Have Standard VARs Remained Stable Since the Crisis?" *Journal of Applied Econometrics* 32, 931-951. https://doi.org/10.1002/jae.2555

[2] Adams, Patrick, Tobias Adrian, Nina Boyarchenko, and Domenico Giannone (2020), "Forecasting Macroeconomic Risks," CEPR Discussion Paper DP 14436.

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

[4] Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019b), "Multimodality in Macro-Financial Dynamics," Federal Reserve Bank of New York Staff Report 903.

[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] Alessandri, Piergiorgio, and Haroon Mumtaz (2017), "Financial Conditions and Density Forecasts for US Output and Inflation," *Review of Economic Dynamics* 24, 66-78. https://doi.org/10.1016/j.red.2017.01.003.

[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] Azzalini, Adelchi, and Antonella Capitanio (2003), "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-distribution," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65, 367-389. https://doi.org/10.1111/1467-9868.00391.

[9] Bai, Jushan, and Serena Ng (2001), "A Consistent Test for Conditional Symmetry in Time Series Models," *Journal of Econometrics* 103, 225-258. https://doi.org/10.1016/S0304-4076(01)00044-6.

[10] Bai, Jushan, and Serena Ng (2005), "Tests for Skewness, Kurtosis, and Normality for Time Series Data," *Journal of Business and Economic Statistics* 23, 49-60. https://doi.org/10.1198/073500104000000271.

[11] Brownlees, Christian, and Andre B.M. Souza (2020), "Backtesting Global Growth-at-Risk," manuscript.

[12] Caldara, Dario, Chiara Scotti, and Molin Zhong (2020), "Macroeconomic and Financial Risks: A Tale of Volatility," manuscript.

[13] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2016), "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business and Economic Statistics* 34, 375-390. https://doi.org/10.1080/07350015.2015.1040116.

[14] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2018), "Measuring Uncertainty and Its Impact on the Economy," *Review of Economics and Statistics* 100, 799-815. https://doi.org/10.1162/REST_a_00693.

[15] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019), "Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors," *Journal of Econometrics* 212, 137-154. https://doi.org/10.1016/j.jeconom.2019.04.024.

[16] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2020), "Nowcasting Tail Risks to Economic Activity with Many Indicators," Federal Reserve Bank of Cleveland, Working Paper No. 20-13R. https://doi.org/10.26509/frbc-wp-202013r.

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

[18] Chernozhukov, Victor, Ivan Fernandez-Val, and Tetsuya Kaji (2017), "Extremal Quantile Regression," in *Handbook of Quantile Regression*, Roger Koenker, Victor Chernozhukov, Xuming He, Limin Peng (eds.), Chapman and Hall/CRC: New York. https://doi.org/10.1201/9781315120256-18.

[19] 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.

[20] 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.

[21] Cogley, Timothy, Sergei Morozov, and Thomas J. Sargent (2005), "Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System," *Journal of Economic Dynamics and Control* 29, 1893-1925. https://doi.org/10.1016/j.jedc.2005.06.005.

[22] 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.

[23] 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.

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

[25] Croushore, Dean, and Tom Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-130. https://doi.org/10.1016/S0304-4076(01)00072-0.

[26] Delle Monache, Davide, Andrea De Polis, and Ivan Petrella (2020), "Modeling and Forecasting Macroeconomic Downside Risk," manuscript.

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

[28] 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.

[29] Engle, Robert F., and Simone Manganelli (2004), "CAViaR," *Journal of Business and Economic Statistics*, 22, 367-381. https://doi.org/10.1198/073500104000000370.

[30] 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.

[31] Favara, Giovanni, Simon Gilchrist, Kurt F. Lewis, and Egon Zakrajsek (2016), "Updating the Recession Risk and the Excess Bond Premium," FEDS Notes, Board of Governors of the Federal Reserve System, October 6. https://doi.org/10.17016/2380-7172.1836

[32] Figueres, Juan Manuel, and Marek Jarocinski (2020), "Vulnerable growth in the Euro Area: Measuring the Financial Conditions," *Economics Letters* 191,109-126. 10.1016/j.econlet.2020.109126.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

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

[38] 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.

[39] 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.

[40] Gneiting, Tilmann, and Roopesh Ranjan (2011), "Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules," *Journal of Business and Economic Statistics* 29, 411-422. https://doi.org/10.1198/jbes.2010.08110

[41] Jensen, Henrik, Ivan Petrella, Soren Hove Ravn, and Emiliano Santoro (2020), "Leverage and Deepening Business-Cycle Skewness," *American Economic Journal: Macroeconomics* 12, 245-281. https://doi.org/10.1257/mac.20170319.

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

[43] Karagedikli, Ozer, Shaun P. Vahey, and Liz Wakerly (2019), "Improved Methods for Combining Point Forecasts for an Asymmetrically Distributed Variable," CAMA Working Paper No. 15/2019. `https://doi.org/10.2139/ssrn.3332882`.

[44] 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`

[45] 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`.

[46] Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies* 65, 361-393. `https://doi.org/10.1111/1467-937X.00050`.

[47] 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`.

[48] Kozeniauskas, Nicholas, Anna Orlik and Laura Veldkamp (2018), "What Are Uncertainty Shocks?" *Journal of Monetary Economics* 100, 1-15. `https://doi.org/10.1016/j.jmoneco.2018.06.004`.

[49] Kritzman, Mark, and Yuanzhen Li (2010), "Skulls, Financial Turbulence, and Risk Management," *Financial Analysts Journal* 66, 30-41. `https://doi.org/10.2469/faj.v66.n5.3`.

[50] Loaiza-Maya, Ruben, and Michael Stanley Smith (2019), "Real-Time Macroeconomic Forecasting With a Heteroskedastic Inversion Copula," *Journal of Business and Economic Statistics*. `https://doi.org/10.1080/07350015.2018.1514309`.

[51] Lopez-Salido, J. David, and Francesca Loria (2019), "Inflation at Risk," CEPR Discussion Paper DP14074.

[52] Loria, Francesca, Christian Matthes, and Donghai Zhang (2019), "Assessing Macroeconomic Tail Risk," Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2019-026. https://doi.org/10.17016/FEDS.2019.026.

[53] 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.

[54] 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.

[55] 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.

[56] Mitchell, James, and Martin Weale (2019), "Forecasting with Unknown Unknowns: Censoring and Fat Tails on the Bank of England's Monetary Policy Committee," manuscript, University of Warwick.

[57] Morley, James and Jeremy Piger (2012), "The Asymmetric Business Cycle, *Review of Economics and Statistics* 94, 208-221. https://doi.org/10.1162/REST_a_00169.

[58] Nieto, Maria Rosa, and Esther Ruiz (2016), "Frontiers in VaR Forecasting and Backtesting," *International Journal of Forecasting* 32, 475-501. https://doi.org/10.1016/j.ijforecast.2015.08.003.

[59] Orlik, Anna and Laura Veldkamp (2015), "Understanding Uncertainty Shocks and the Role of the Black Swan," National Bureau of Economic Research Working Paper no. 20445. https://doi.org/10.3386/w20445.

[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] 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.

[63] Salgado, Sergio, Fatih Guvenen, and Nicholas Bloom (2020), "Skewed Business Cycles," manuscript.

[64] 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.

[65] Smith, Michael S., and Shaun P. Vahey (2016), "Asymmetric Forecast Densities for U.S. Macroeconomic Variables from a Gaussian Copula Model of Cross-Sectional and Serial Dependence," *Journal of Business and Economic Statistics* 34, 416-434. https://doi.org/10.1080/07350015.2015.1044533.

[66] 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.

[67] 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

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

Table 1: **Skewness and kurtosis statistics, data and BVAR-SV residuals**

	skewness	kurtosis	Bai-Ng skewness	Bai-Ng kurtosis	Bai-Ng normality	conditional symmetry
Data, 1972-2018						
GDP growth	−0.364	5.499	−0.865	2.095**	5.139*	0.742
Unemployment	0.683	2.765	0.644	−0.063	0.419	1.419
GDP inflation	1.402	4.614	1.995**	0.998	4.977*	1.019
Fed funds rate	0.709	3.392	0.787	0.249	0.681	1.998*
NFCI	1.979	6.555	2.016**	1.796*	7.290**	4.957***
BVAR-SV residuals, 1972-2018						
GDP growth	0.237	6.505	0.406	1.514	2.458	0.920
Unemployment	0.542	6.231	1.073	1.869*	4.646*	1.667
GDP inflation	0.196	4.798	0.589	1.862*	3.814	1.658
Fed funds rate	1.422	23.294	0.644	1.336	2.201	1.483
NFCI	−0.186	9.861	−0.258	2.148**	4.682*	0.769
BVAR-SV residuals normalized by SV, 1972-2018						
GDP growth	0.129	2.976	0.664	−0.047	0.443	1.294
Unemployment	0.326	2.815	2.783***	−0.498	7.992**	2.306**
GDP inflation	0.119	2.627	0.956	−1.268	2.520	1.286
Fed funds rate	−0.119	2.921	−0.754	−0.208	0.611	1.171
NFCI	0.341	2.483	2.557**	−1.586	9.053**	2.324**
BVAR-SV forecast errors, horizon = 1 quarter, 1985-2018						
GDP growth	0.042	3.127	0.162	0.226	0.077	1.154
Unemployment	0.850	4.295	1.560	1.313	4.157	2.452**
GDP inflation	−0.367	3.030	−1.839*	0.058	3.386	1.838
Fed funds rate	−0.115	4.645	−0.260	1.329	1.833	2.737**
NFCI	1.632	17.525	0.826	1.292	2.351	1.353
BVAR-SV forecast errors, horizon = 4 quarters, 1985-2018						
GDP growth	−0.473	4.470	−0.706	0.852	1.224	1.196
Unemployment	2.132	9.275	0.807	0.755	1.221	1.258
GDP inflation	−0.577	2.850	−2.183**	−0.215	4.812*	2.127*
Fed funds rate	−0.083	2.976	−0.317	−0.029	0.101	1.285
NFCI	0.862	7.487	0.595	0.841	1.061	0.787

Notes: Statistical significance of the Bai-Ng test statistics is indicated by *** (1%), ** (5%), or * (10%).

Table 2: **Accuracy of in-sample forecasts of GDP growth**

Point forecasts: RMSEs				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	2.117	1.505	1.888	1.218
BVAR-SV	0.943	0.962	0.994	1.023
BVAR-SVF-M	0.940^{*}	0.959	0.985	1.015
Density forecasts: average log scores				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	−2.096	−1.718	−2.054	−1.601
BVAR-SV	0.044^{*}	0.031	0.004	−0.030
BVAR-SVF-M	0.041^{**}	0.061	0.014	0.003
Interval coverage: 5 percent tail				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.067	0.069	0.054	0.045
BVAR-SV	0.030	0.069	0.011^{***}	0.056
BVAR-SVF-M	0.030	0.053	0.011^{***}	0.034
Quantile score (5% quantile)				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.225	0.166	0.197	0.142
BVAR-SV	0.982	1.081	1.031	1.100
BVAR-SVF-M	0.953	1.068	0.974	1.003
VaR-ES score (5% quantile)				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.370	−0.041	0.169	−0.232
BVAR-SV	−0.073	−0.143	-0.151^{*}	−0.227
BVAR-SVF-M	−0.031	−0.036	−0.076	0.000
qwCRPS				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.361	0.262	0.333	0.229
BVAR-SV	0.959	0.978	0.988	1.012
BVAR-SVF-M	0.956^{*}	0.968	0.980	0.997

Notes: The point forecasts used for the quantile regressions are the medians. The log scores of the quantile regression predictions are computed as in ABG, with a skewed-t density estimated from the quantiles in a second step. Except in the case of the 5 percent coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for RMSE, quantile score, and qwCRPS (an entry less than 1 means the BVAR is more accurate than QR), and score differences for the average log score and VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 5 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table 3: **Accuracy of out-of-sample point and density forecasts of GDP growth**

	1985-2018		1985-2007	
Point forecasts: RMSEs				
	$h=1Q$	$h=4Q$	$h=1Q$	$h=4Q$
Quantile regression	2.096	1.789	2.116	1.682
BVAR-SV	0.875*	0.873	0.867	0.819
BVAR-SVF-M	0.883*	0.885	0.868	0.816
Density forecasts: average log scores				
	$h=1Q$	$h=4Q$	$h=1Q$	$h=4Q$
Quantile regression	−2.149	−1.976	−2.184	−1.938
BVAR-SV	0.062*	0.146	0.065	0.162
BVAR-SVF-M	0.094***	0.144	0.097**	0.180
Interval coverage: 5 percent tail				
	$h=1Q$	$h=4Q$	$h=1Q$	$h=4Q$
Quantile regression	0.045	0.084	0.043	0.090
BVAR-SV	0.015***	0.069	0.011***	0.034
BVAR-SVF-M	0.015***	0.084	0.000***	0.045
Quantile score (5% quantile)				
	$h=1Q$	$h=4Q$	$h=1Q$	$h=4Q$
Quantile regression	0.173	0.189	0.168	0.166
BVAR-SV	1.099	0.877	1.158	0.767
BVAR-SVF-M	0.978	0.899	0.981	0.683
VaR-ES score (5% quantile)				
	$h=1Q$	$h=4Q$	$h=1Q$	$h=4Q$
Quantile regression	0.269	0.484	0.148	0.333
BVAR-SV	−0.251*	0.420	−0.355*	0.591
BVAR-SVF-M	−0.074	0.303	−0.116	0.734
qwCRPS				
	$h=1Q$	$h=4Q$	$h=1Q$	$h=4Q$
Quantile regression	0.332	0.299	0.341	0.294
BVAR-SV	0.941	0.909	0.927	0.839
BVAR-SVF-M	0.935	0.923	0.913	0.831

Notes: The point forecasts used for the quantile regressions are the medians. The log scores of the quantile regression predictions are computed as in ABG, with a skewed-t density estimated from the quantiles in a second step. Except in the case of the 5 percent coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for RMSE, quantile score, and qwCRPS (an entry less than 1 means the BVAR is more accurate than QR), and score differences for the average log score and VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 5 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table 4: **Accuracy of in-sample forecasts of GDP growth, models using turbulence**

	1972-2011		1985-2011		1985-2007	
Point forecasts: RMSEs						
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	3.006	2.159	2.088	1.619	1.957	1.272
BVAR-SV	0.958	0.905*	0.987	1.009	0.935**	0.902**
BVAR-SVF-M	0.963	0.913	0.992	1.019	0.928**	0.905**
Density forecasts: average log scores						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	−2.442	−2.150	−2.229	−1.866	−2.164	−1.757
BVAR-SV	0.118***	0.190***	0.121***	0.085	0.128***	0.173***
BVAR-SVF-M	0.082**	0.137*	0.089***	0.028	0.088***	0.118***
Interval coverage: 5 percent tail						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.063	0.057	0.019**	0.019	0.011***	0.022
BVAR-SV	0.031	0.045	0.028	0.067	0.022*	0.022
BVAR-SVF-M	0.019***	0.038	0.028	0.057	0.000***	0.011***
Quantile score (5% quantile)						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.328	0.226	0.265	0.178	0.258	0.175
BVAR-SV	0.754**	0.912	0.711***	1.098	0.709***	0.772**
BVAR-SVF-M	0.796*	0.976	0.765***	1.114	0.745***	0.761***
VaR-ES score (5% quantile)						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.794	0.431	0.710	0.209	0.677	0.147
BVAR-SV	0.320***	0.278	0.394***	0.183	0.456***	0.493***
BVAR-SVF-M	0.240***	0.192	0.280***	0.094	0.336***	0.455***
qwCRPS						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.514	0.381	0.368	0.282	0.352	0.243
BVAR-SV	0.897**	0.887	0.942	0.999	0.890***	0.854***
BVAR-SVF-M	0.910*	0.899	0.960	1.013	0.905***	0.861***

Notes: The point forecasts used for the quantile regressions are the medians. The log scores of the quantile regression predictions are computed as in ABG, with a skewed-t density estimated from the quantiles in a second step. Except in the case of the 5 percent coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for RMSE, quantile score, and qwCRPS (an entry less than 1 means the BVAR is more accurate than QR), and score differences for the average log score and VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 5 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table 5: **Accuracy of out-of-sample point and density forecasts of GDP growth, models using turbulence**

	1972-2011		1985-2011		1985-2007	
Point forecasts: RMSEs						
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	2.815	2.509	1.727	1.670	1.721	1.416
BVAR-SV	0.980	0.888	1.272	1.222	1.103	1.008
BVAR-SVF-M	0.981	0.871	1.298^{*}	1.245	1.128^{*}	1.047
Density forecasts: average log scores						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	−2.637	−2.410	−2.186	−1.967	−2.159	−1.906
BVAR-SV	0.259	0.310	−0.038	−0.044	−0.007	0.111
BVAR-SVF-M	0.252	0.234	−0.052	−0.196	−0.024	0.072
Interval coverage: 5 percent tail						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.031	0.064	0.000	0.000	0.000^{***}	0.000
BVAR-SV	0.025^{*}	0.127	0.000	0.143	0.000^{***}	0.079
BVAR-SVF-M	0.019^{**}	0.108	0.019	0.143	0.000^{***}	0.067
Quantile score (5% quantile)						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.396	0.360	0.286	0.206	0.271	0.198
BVAR-SV	0.662^{***}	0.679	0.654^{***}	1.025	0.703^{***}	0.601^{***}
BVAR-SVF-M	0.640^{***}	0.729	0.663^{***}	1.261	0.692^{***}	0.594^{***}
VaR-ES score (5% quantile)						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	1.191	1.654	0.739	0.463	0.708	0.441
BVAR-SV	0.587^{**}	1.100	0.279^{***}	−0.154	0.271^{***}	0.742^{***}
BVAR-SVF-M	0.666^{**}	0.313	0.377^{***}	−1.320	0.332^{***}	0.677^{**}
qwCRPS						
	1972-2011		1985-2011		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.498	0.442	0.339	0.286	0.330	0.257
BVAR-SV	0.930	0.870	1.064	1.210	1.001	0.956
BVAR-SVF-M	0.928	0.875	1.086	1.287	1.003	0.988

Notes: The point forecasts used for the quantile regressions are the medians. The log scores of the quantile regression predictions are computed as in ABG, with a skewed-t density estimated from the quantiles in a second step. Except in the case of the 5 percent coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for RMSE, quantile score, and qwCRPS (an entry less than 1 means the BVAR is more accurate than QR), and score differences for the average log score and VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 5 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table 6: **In-sample forecast results for unemployment rate changes**

Interval coverage: 95 percent tail		
	1985-2018	1985-2007
Quantile regression	0.046	0.034
BVAR-SV	0.076	0.056
BVAR-SVF-M	0.061	0.045
Quantile score (95% quantile)		
	1985-2018	1985-2007
Quantile regression	0.086	0.064
BVAR-SV	1.004	0.828
BVAR-SVF-M	0.934	0.758**
VaR-ES score (95% quantile)		
	1985-2018	1985-2007
Quantile regression	0.486	0.456
BVAR-SV	0.058	0.089
BVAR-SVF-M	0.087***	0.112***
qwCRPS		
	1985-2018	1985-2007
Quantile regression	0.117	0.097
BVAR-SV	0.861***	0.799***
BVAR-SVF-M	0.835**	0.776**

Notes: Results are for four-steps-ahead forecasts of the four-quarter change in the unemployment rate. Except in the case of the coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for quantile score and qwCRPS (an entry less than 1 means the BVAR is more accurate than QR), and score differences for the VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 95 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table 7: **Out-of-sample forecast results for unemployment rate changes**

Interval coverage: 95 percent tail		
	1985-2018	1985-2007
Quantile regression	0.130	0.079
BVAR-SV	0.130	0.124
BVAR-SVF-M	0.145	0.146
Quantile score (95% quantile)		
	1985-2018	1985-2007
Quantile regression	0.113	0.072
BVAR-SV	1.092	1.125
BVAR-SVF-M	1.096	1.072
VaR-ES score (95% quantile)		
	1985-2018	1985-2007
Quantile regression	0.690	0.525
BVAR-SV	−0.014	−0.069
BVAR-SVF-M	−0.037	−0.058
qwCRPS		
	1985-2018	1985-2007
Quantile regression	0.139	0.117
BVAR-SV	0.947	0.906
BVAR-SVF-M	0.984	0.904

Notes: Results are for four-steps-ahead forecasts of the four-quarter change in the unemployment rate. Except in the case of the coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for quantile score and qwCRPS (an entry less than 1 means the BVAR is more accurate than QR), and score differences for VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 95 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

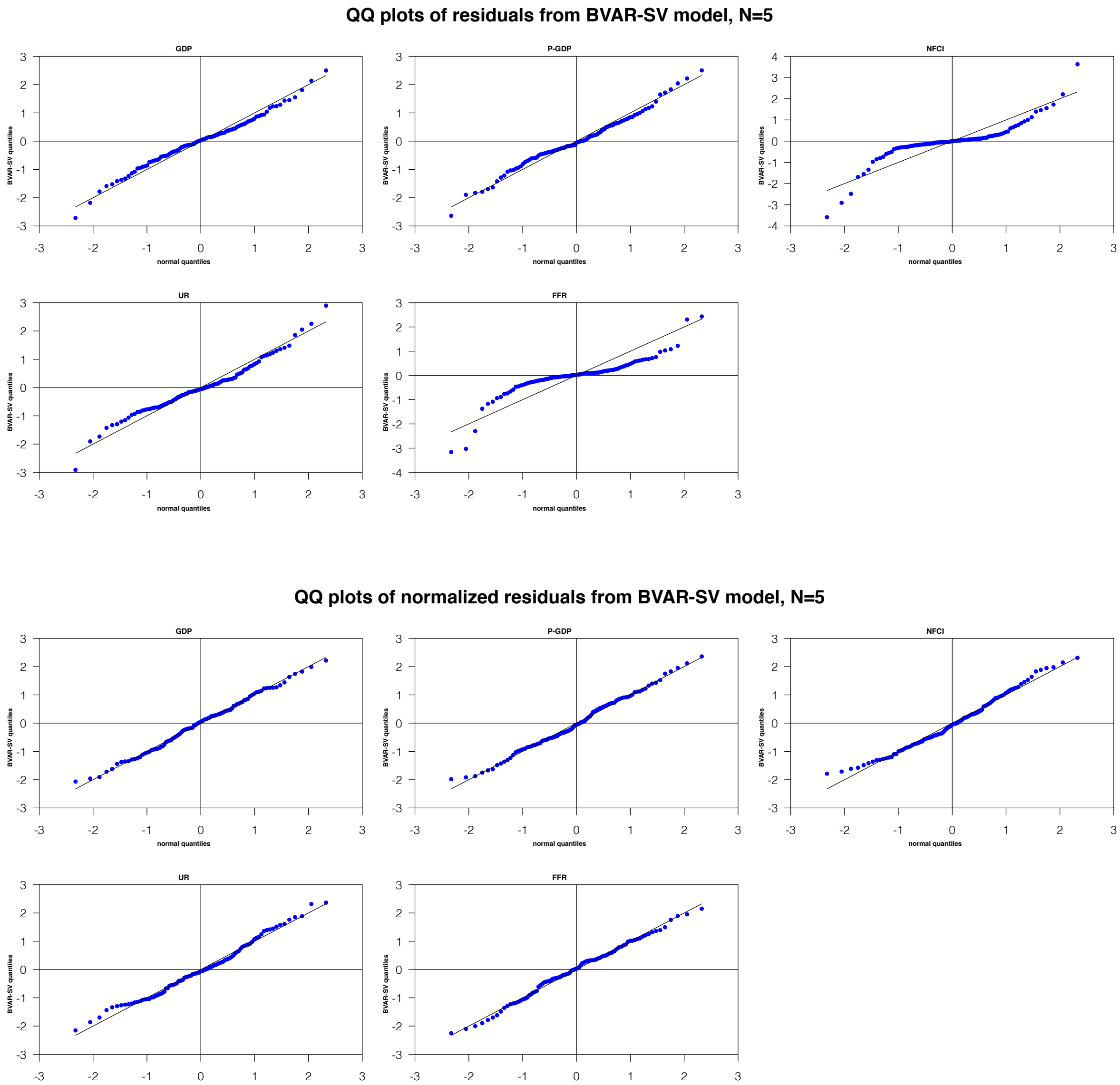


Figure 1: Q-Q plots of residuals and SV-normalized residuals, BVAR-SV model with five variables, 1972-2018 sample

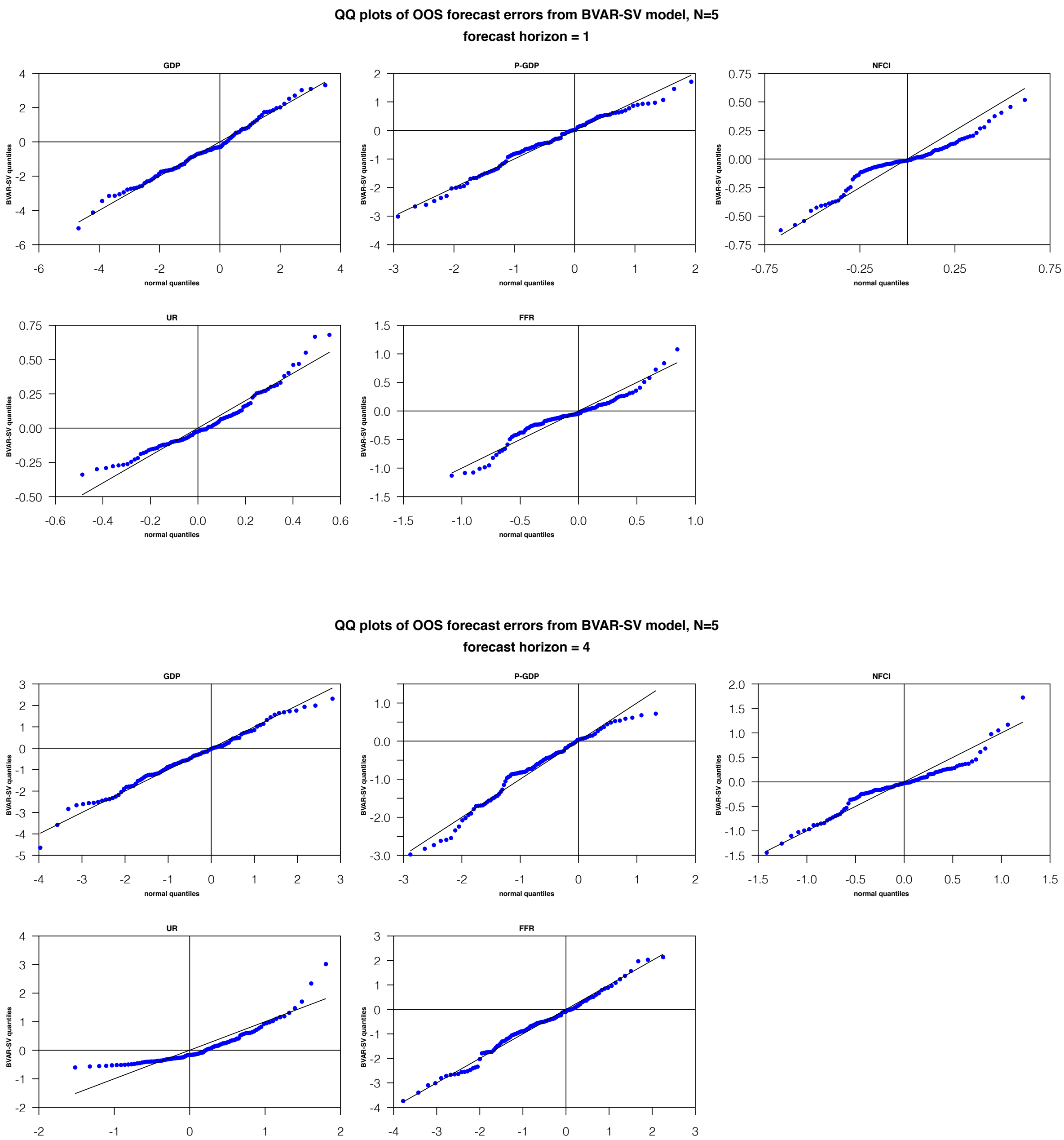
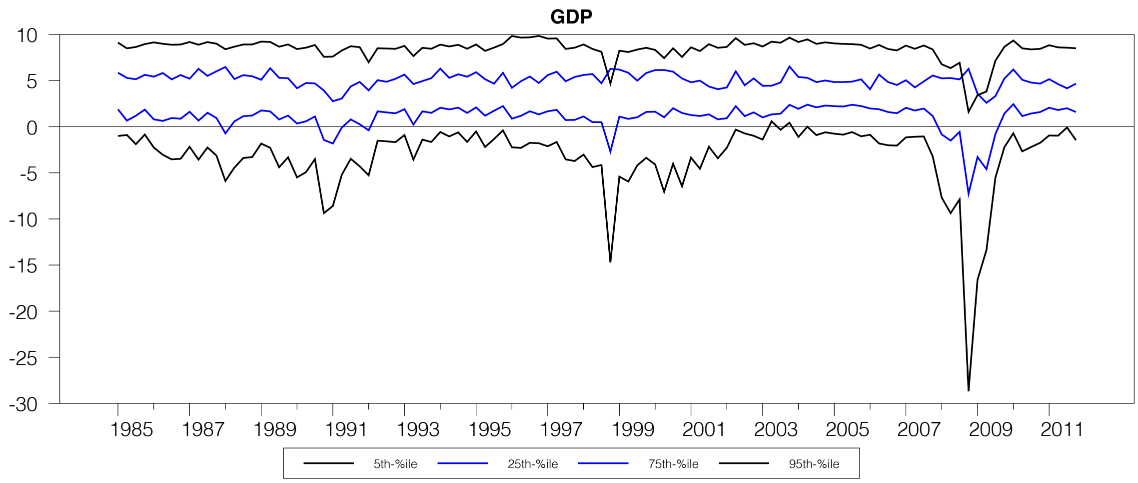


Figure 2: Q-Q plots of OOS forecast errors, BVAR-SV model with five variables, 1985-2018 sample



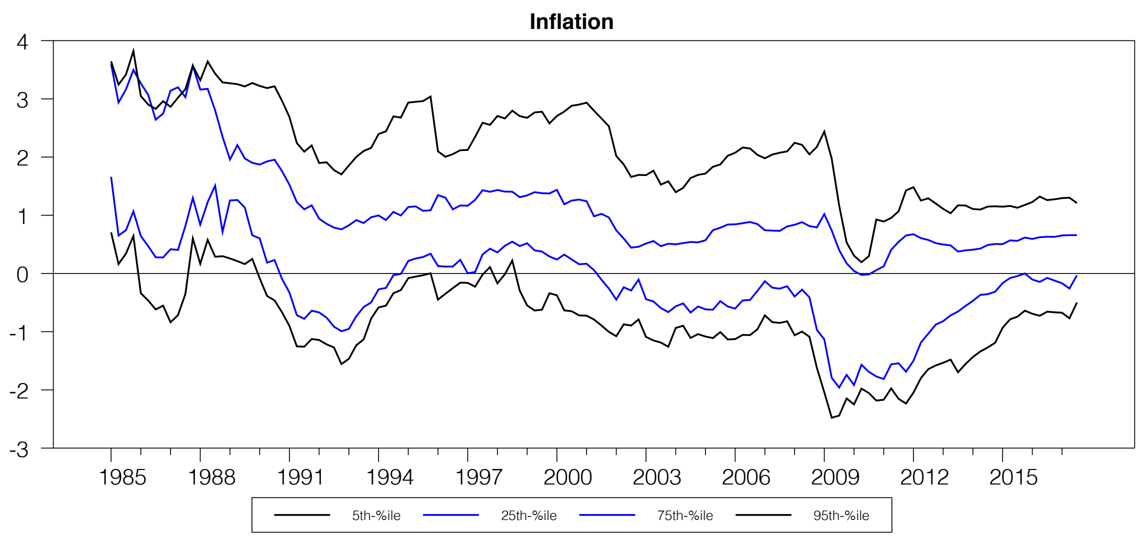
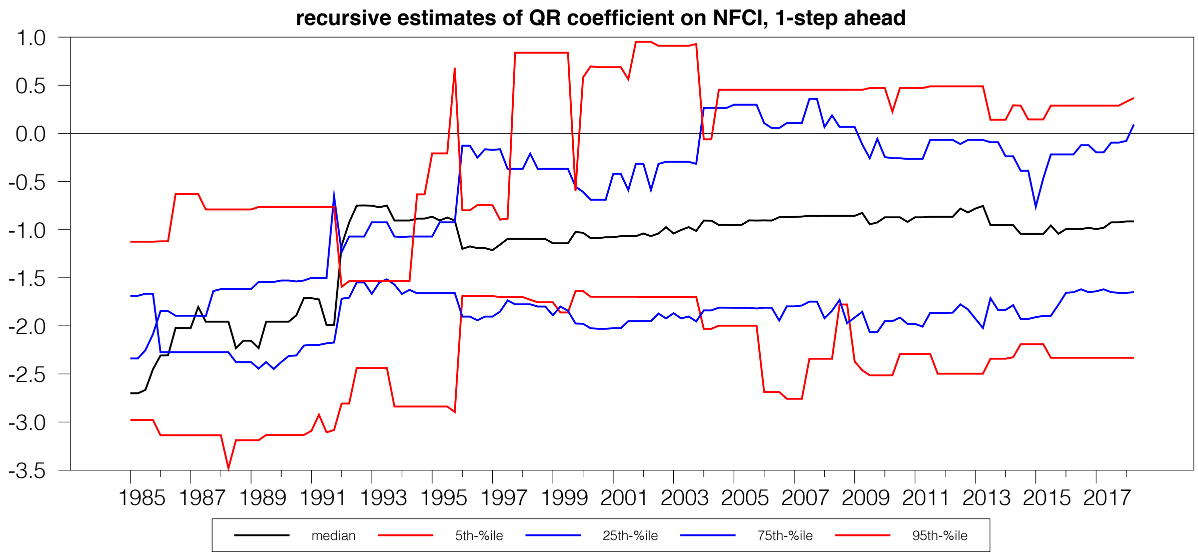


Figure 3: Examples of crossing of quantiles estimated for out-of-sample forecasting. The top panel provides the time series of quantiles (5%, 25%, 75%, and 95%) from a one-quarter-ahead model of GDP growth relating growth to a constant, lagged growth, and lagged turbulence. The estimates, obtained recursively using real-time data vintages from 1985:Q1 through 2018:Q2, are dated by the forecast origin. The bottom panel provides the time series of quantiles (5%, 25%, 75%, and 95%) for the four-quarter change in GDP inflation estimated recursively using real-time data vintages from 1985:Q1 through 2018:Q2. This quantile regression model relates the change in inflation to a constant and the lagged change in inflation.



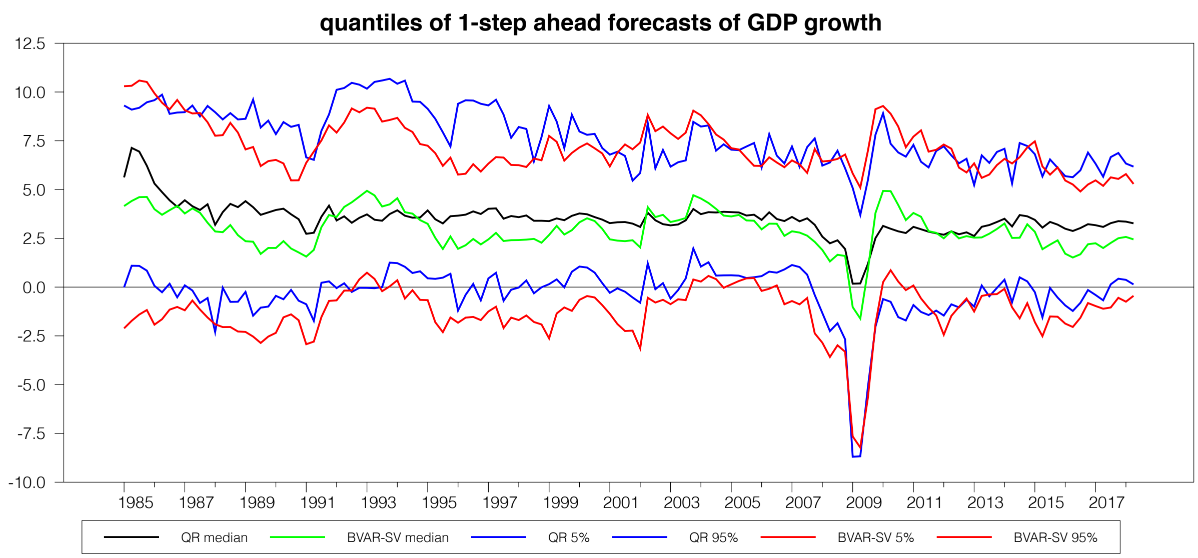


Figure 4: The top panel provides time series of estimates of the coefficient on the lagged NFCI in the one-step-ahead quantile regression model of GDP growth, for the median and 5%, 25%, 75%, and 95% quantiles. The estimates, obtained recursively using real-time data vintages from 1985:Q1 through 2018:Q2, are dated by the forecast origin. The bottom panel compares out-of-sample one-step-ahead quantile estimates (median, 5%, and 95%) from the baseline quantile regression and BVAR-SV models.

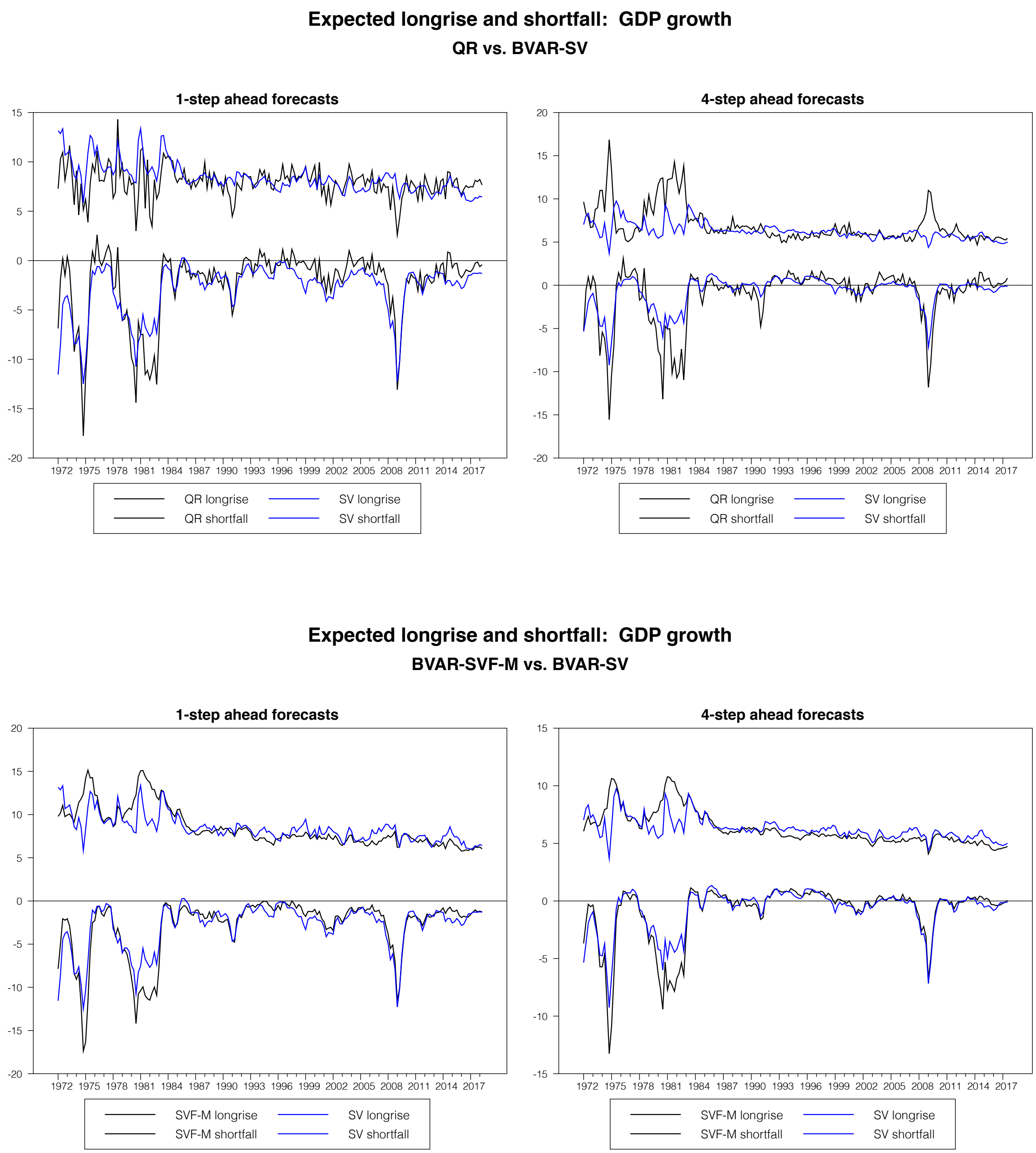
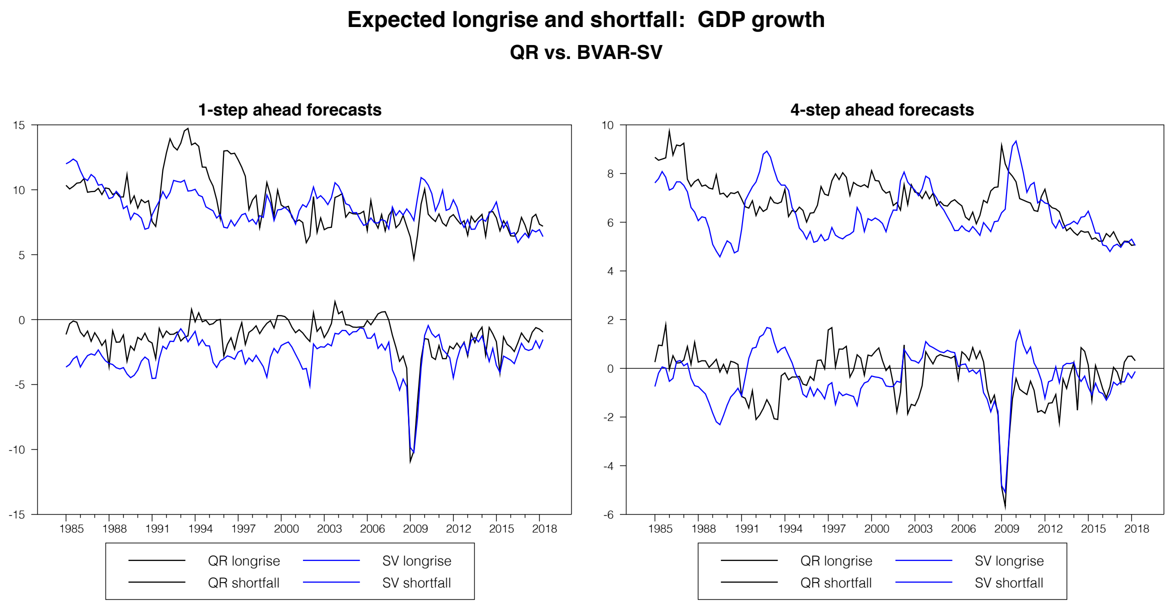


Figure 5: Long-rise and expected shortfall, in-sample forecasts of GDP growth for 1972-2018. The top panel compares estimates from the QR and five-variable BVAR-SV models. The bottom panel compares estimates from the BVAR-SV and BVAR-SVF-M models.



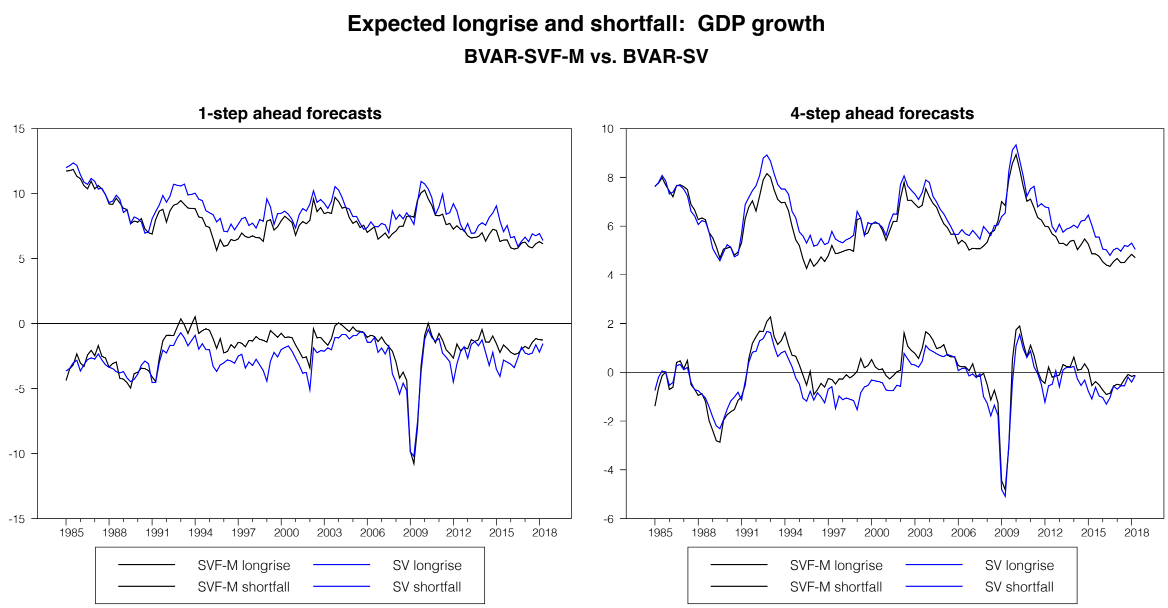


Figure 6: Long-rise and expected shortfall, out-of-sample forecasts of GDP growth for 1985-2018. The top panel compares estimates from the QR and five-variable BVAR-SV models. The bottom panel compares estimates from the BVAR-SV and BVAR-SVF-M models.

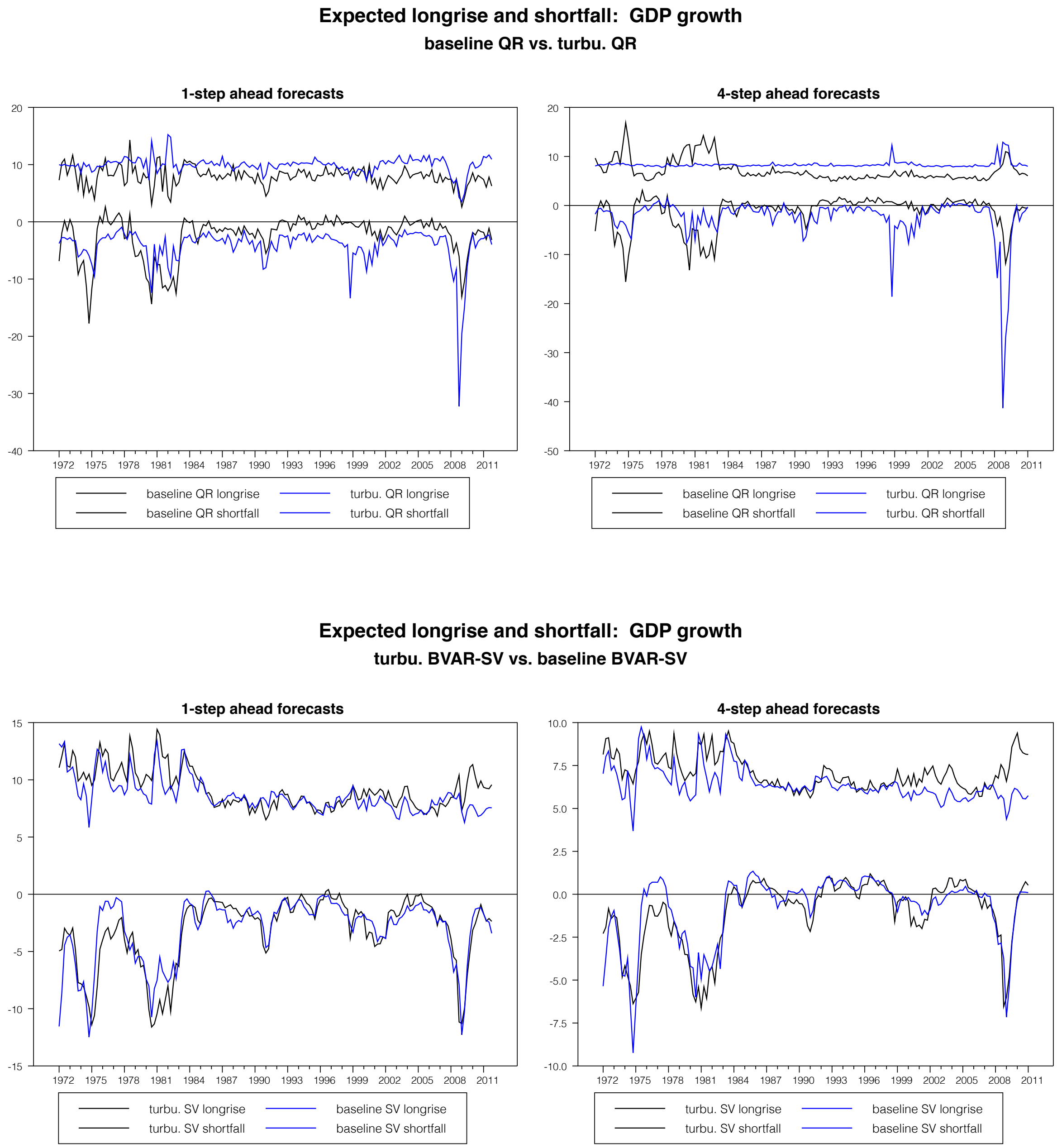


Figure 7: Long-rise and expected shortfall, in-sample forecasts of GDP growth for 1972-2011. The top panel compares estimates from the baseline QR specification including the NFCI to estimates from the QR specification including turbulence. The bottom panel compares estimates from the baseline five-variable BVAR-SV model including the NFCI and the BVAR-SV specification including turbulence.

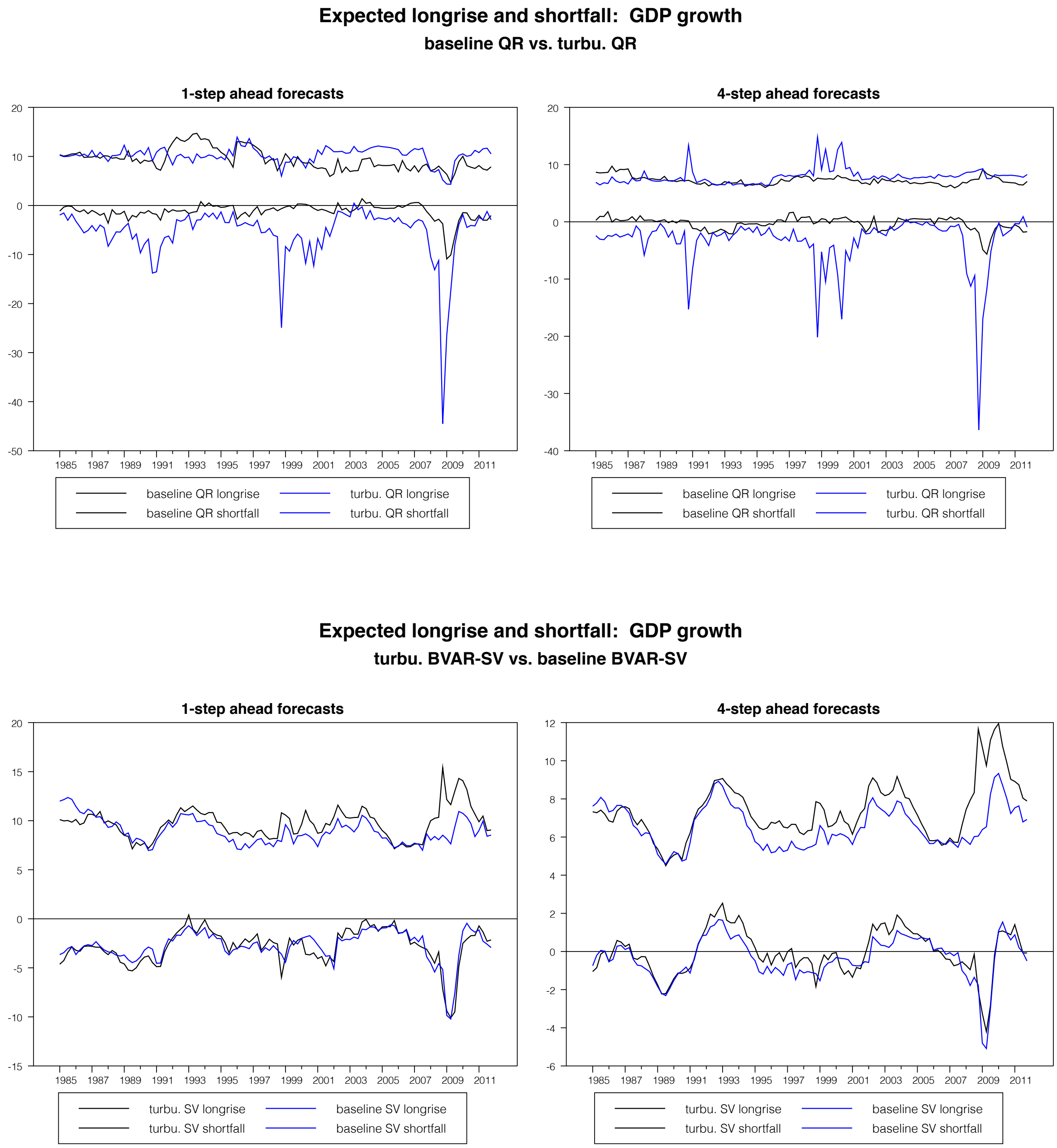
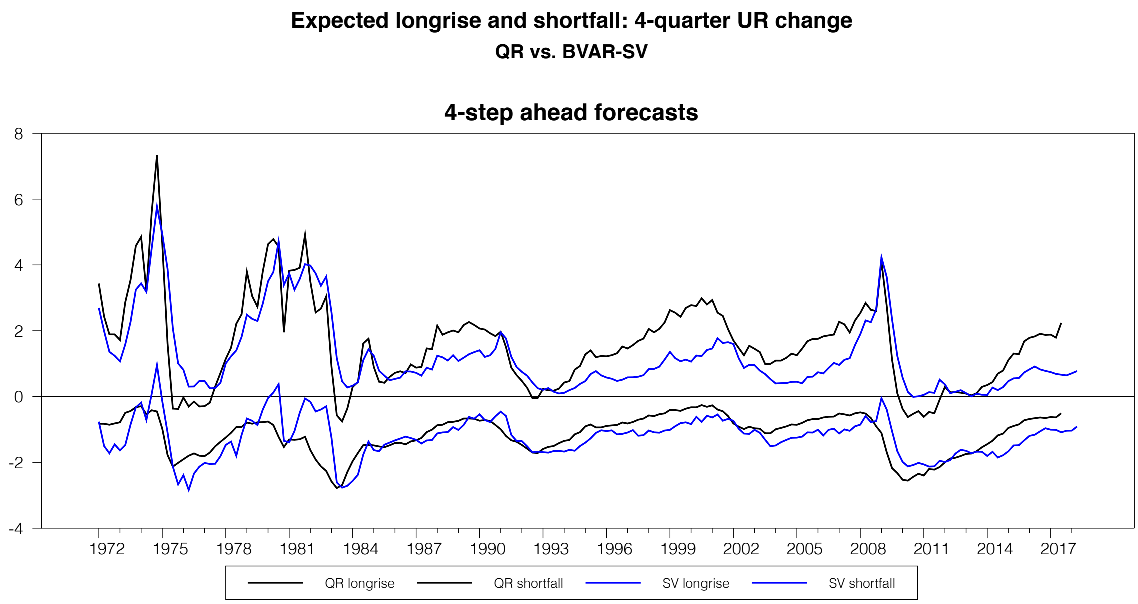


Figure 8: Long-rise and expected shortfall, out-of-sample forecasts of GDP growth for 1985-2011. The top panel compares estimates from the baseline QR specification including the NFCI to estimates from the QR specification including turbulence. The bottom panel compares estimates from the baseline five-variable BVAR-SV model including the NFCI and the BVAR-SV specification including turbulence.



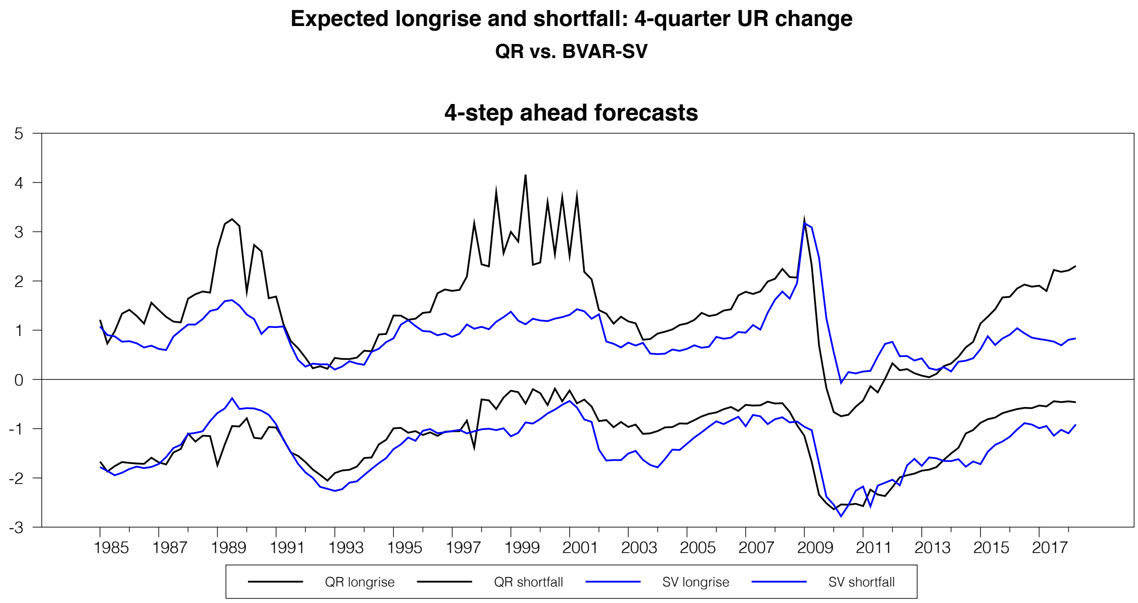


Figure 9: Long-rise and expected shortfall, forecasts of four-quarter changes in the unemployment rate. The top panel provides estimates for in-sample forecasts, and the bottom panel reports estimates for out-of-sample forecasts.

A Appendix

This appendix first provides some supplemental results, including: (1) a table of skewness and kurtosis statistics for the data and BVAR-SV residuals over a shorter sample of 1985-2018; (2) a table of the relative volatilities of expected shortfall and long-rise estimates (for the 5 and 95 percent quantiles, respectively); (3) a table with results of dynamic quantile test results; (4) charts directly comparing in-sample and out-of-sample estimates of expected shortfall and long-rise over a common sample of 1985-2018; and (5) results using 10 and 90 percent quantiles rather than 5 and 95 percent quantiles. It then details the Monte Carlo experiments summarized in the paper.

Table A1: **Skewness and kurtosis statistics, data and BVAR-SV residuals, 1985-2018**

	skewness	kurtosis	Bai-Ng skewness	Bai-Ng kurtosis	Bai-Ng normality	Bai-Ng condit. symmetry
Data, 1985-2018						
GDP growth	-1.244	7.397	-1.158	1.162	2.693	0.965
Unemployment	0.887	3.192	0.634	0.048	0.405	1.094
GDP inflation	0.072	3.655	0.281	1.240	1.617	1.136
Fed funds rate	0.182	1.787	0.084	-0.531	0.289	1.797
NFCI	2.870	15.447	0.666	0.751	1.007	3.455 ***
BVAR-SV residuals, 1985-2018						
	skewness	kurtosis	Bai-Ng skewness	Bai-Ng kurtosis	Bai-Ng normality	Bai-Ng condit. symmetry
GDP growth	-0.767	5.727	-1.271	1.276	3.243	0.746
Unemployment	1.079	5.005	1.398	1.089	3.142	2.551 **
GDP inflation	0.028	3.637	0.115	1.276	1.642	1.414
Fed funds rate	-0.438	4.656	-1.665 *	2.045 **	6.955 **	1.795
NFCI	4.303	35.978	1.135	1.192	2.707	2.851 ***
BVAR-SV residuals normalized by SV, 1985-2018						
GDP growth	0.061	2.787	0.404	-0.593	0.515	1.047
Unemployment	0.311	2.581	2.501 **	-1.028	7.313 **	2.147 *
GDP inflation	0.187	2.798	1.075	-0.541	1.448	1.889
Fed funds rate	-0.162	2.810	-1.053	-0.466	1.325	0.907
NFCI	0.529	2.729	3.163 ***	-0.673	10.459 ***	1.986 *

Notes: Statistical significance of the Bai-Ng test statistics is indicated by *** (1%), ** (5%), or * (10%). The results for BVAR-SV residuals are based on residuals obtained from model estimates using data starting in 1972, but skewness and kurtosis statistics are computed for a sample starting in 1985.

Table A2: **Relative volatilities (ratio ES/LR) of expected shortfall and long-rise**

In-sample forecasts, 1972-2018		
	$h = 1Q$	$h = 4Q$
Quantile regression	2.207	1.521
BVAR-SV	1.728	1.942
BVAR-SVF-M	1.526	1.747
Out-of-sample forecasts, 1985-2018		
	$h = 1Q$	$h = 4Q$
Quantile regression	0.774	1.161
BVAR-SV	1.074	0.949
BVAR-SVF-M	1.155	1.012

Notes: The table reports the standard deviation of expected shortfall divided by the standard deviation of the long-rise, for the forecasts and samples indicated.

Table A3: p-**values of dynamic quantile tests**

	Lagged hit rates		Lagged NFCI	
In-sample forecasts of GDP growth, models with NFCI, 1985-2018				
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.810	0.477	0.962	0.345
BVAR-SV	0.994	0.455	0.958	0.639
BVAR-SVF-M	0.971	0.982	0.965	0.540
Out-of-sample forecasts of GDP growth, models with NFCI, 1985-2018				
	Lagged hit rates		Lagged NFCI	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.969	0.412	0.746	0.341
BVAR-SV	0.854	0.423	0.625	0.000
BVAR-SVF-M	0.801	0.717	0.638	0.000
In-sample forecasts of GDP growth, models with turbulence, 1972-2011				
	Lagged hit rates		Lagged turbulence	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.379	0.668	1.000	0.678
BVAR-SV	0.968	0.008	0.611	0.527
BVAR-SVF-M	0.968	0.059	0.611	0.157
Out-of-sample forecasts of GDP growth, models with turbulence, 1972-2011				
	Lagged hit rates		Lagged turbulence	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.969	0.385	0.992	0.257
BVAR-SV	0.759	0.366	0.433	0.455
BVAR-SVF-M	0.741	0.120	0.486	0.004
In-sample forecasts of unemployment changes, models with NFCI, 1985-2018				
	Lagged hit rates		Lagged NFCI	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	NA	0.632	NA	0.009
BVAR-SV	NA	0.172	NA	0.370
BVAR-SVF-M	NA	0.432	NA	0.529
Out-of-sample forecasts of unemployment changes, models with NFCI, 1985-2018				
	Lagged hit rates		Lagged NFCI	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	NA	0.029	NA	0.000
BVAR-SV	NA	0.010	NA	0.296
BVAR-SVF-M	NA	0.033	NA	0.000

Notes: The table reports the p-values of dynamic quantile tests (Wald statistics) applied to the hit rate series of each indicated forecast of the 10 percent quantile (90 percent for the change in the unemployment rate). The two left-side columns provide results for tests of the significance of two lags of the hit rate; the two right-side columns provide results for tests of the significance of two lags of the NFCI or turbulence variable. Tests at the four-steps-ahead horizon incorporate a heteroskedasticity and autocorrelation-robust variance estimator (Newey-West, with six lags).

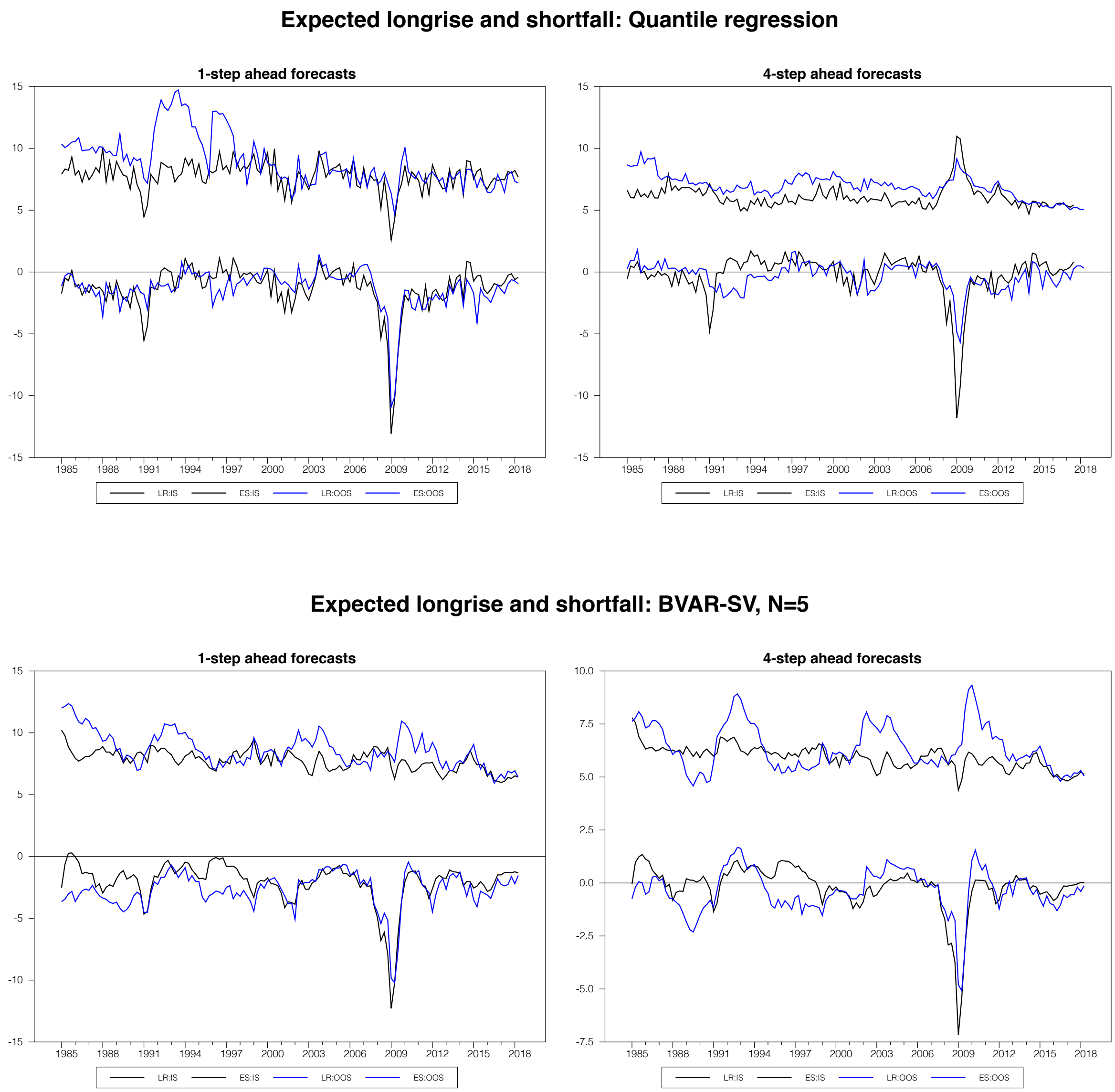


Figure A1: Long-rise and expected shortfall, in-sample and out-of-sample forecasts of GDP growth for 1985-2018. The top panel compares estimates from the QR model. The bottom panel compares estimates from the five-variable BVAR-SV model.

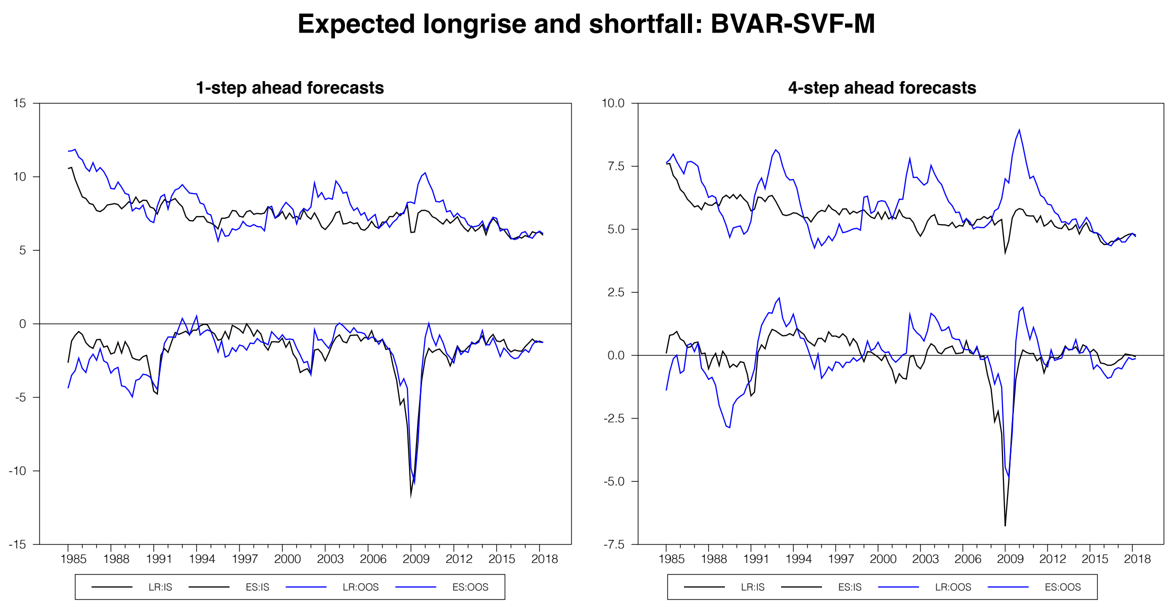


Figure A2: Long-rise and expected shortfall, in-sample and out-of-sample forecasts of GDP growth for 1985-2018. The panel compares estimates from the 5-variable BVAR-SVF-M model.

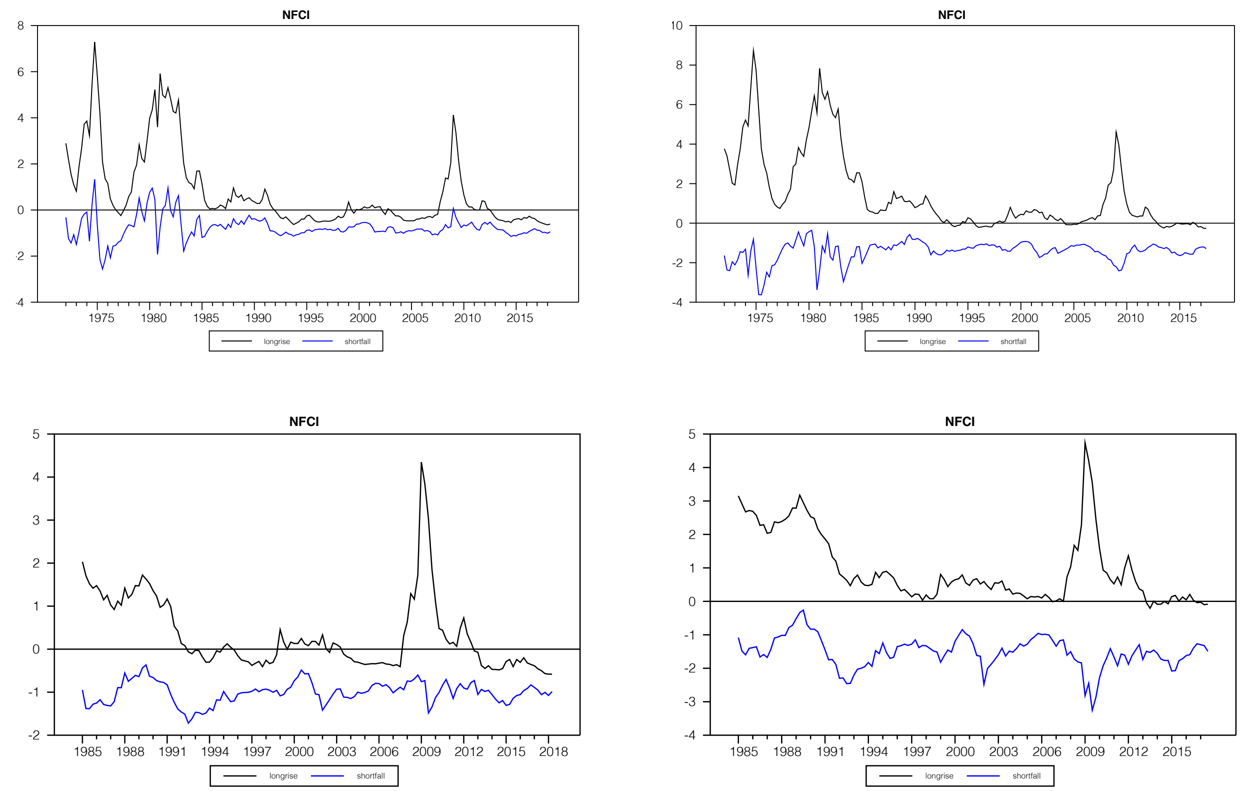


Figure A3: Long-rise and expected shortfall, in-sample and out-of-sample forecasts of NFCI from the BVAR-SV specification. The top left panel provides in-sample estimates at the one-step-ahead horizon, and the top right panel reports estimates for the four-steps-ahead horizon. The lower panels provide the corresponding out-of-sample estimates, with the one-step horizon on the left and the four-step horizon on the right.

A.1 Results for 10 and 90 percent quantiles

Table A4: **In-sample forecast results for GDP growth using the 10 percent quantile, 1972-2018**

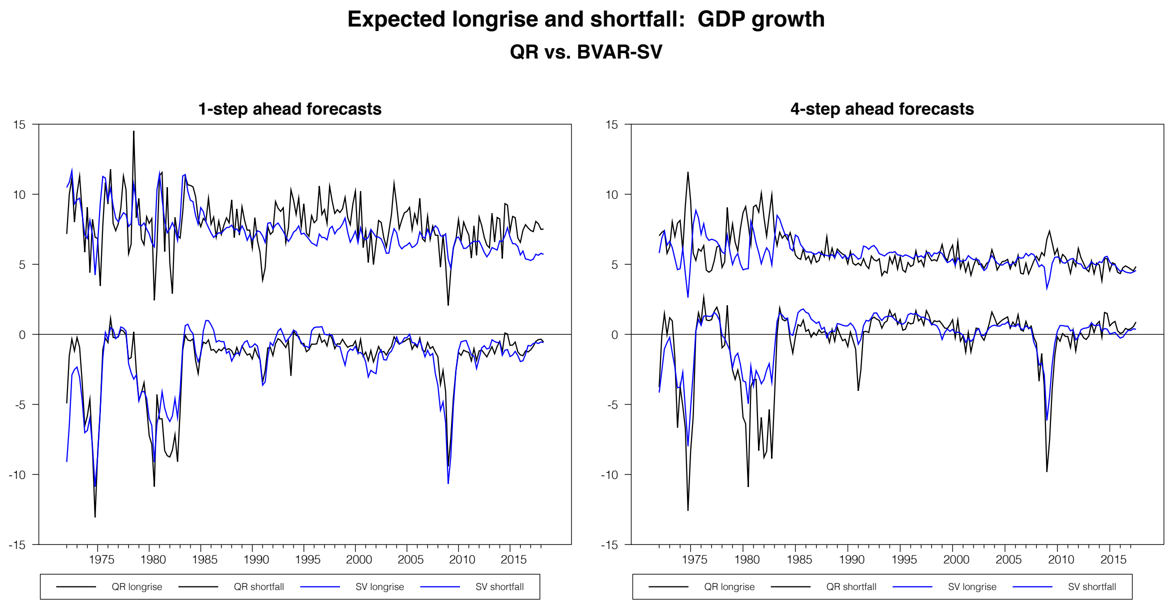
Interval coverage: 10 percent tail				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.112	0.122	0.076	0.067
BVAR-SV	0.075	0.122	0.054 *	0.079
BVAR-SVF-M	0.067	0.115	0.054 **	0.090
Quantile score (10% quantile)				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.377	0.282	0.337	0.242
BVAR-SV	0.936	0.994	0.962	1.023
BVAR-SVF-M	0.932 *	0.986	0.949	0.993
VaR-ES score (10% quantile)				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.320	-0.206	0.111	-0.516
BVAR-SV	0.127 *	0.017	0.139 *	-0.055
BVAR-SVF-M	0.154 ***	0.059	0.185 ***	0.046

Notes: Except in the case of the 10 percent coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for quantile score (an entry less than 1 means the BVAR is more accurate than QR), and score differences for the VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 5 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table A5: **Out-of-sample forecast results for GDP growth using the 10 percent quantile, 1985-2018**

Interval coverage: 10 percent tail				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.119	0.191	0.109	0.202
BVAR-SV	0.067	0.160	0.054 *	0.135
BVAR-SVF-M	0.082	0.191	0.076	0.169
Quantile score (10% quantile)				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.275	0.314	0.273	0.307
BVAR-SV	1.100	0.887	1.114	0.754
BVAR-SVF-M	1.025	0.918	1.022	0.741
VaR-ES score (10% quantile)				
	1985-2018		1985-2007	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
Quantile regression	0.103	0.415	-0.033	0.440
BVAR-SV	-0.168	0.402	-0.233 *	0.848
BVAR-SVF-M	-0.017	0.251	-0.043	0.858

Notes: Except in the case of the 10 percent coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to those for quantile regression, using ratios for quantile score (an entry less than 1 means the BVAR is more accurate than QR), and score differences for VaR-ES score (an entry greater than 0 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal 5 percent is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.



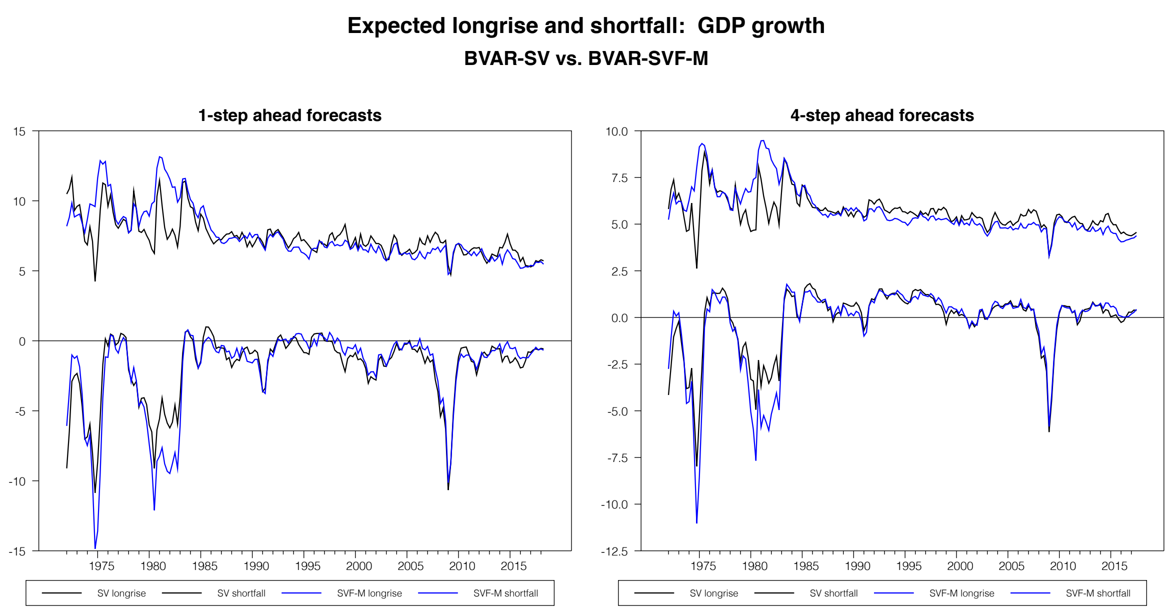
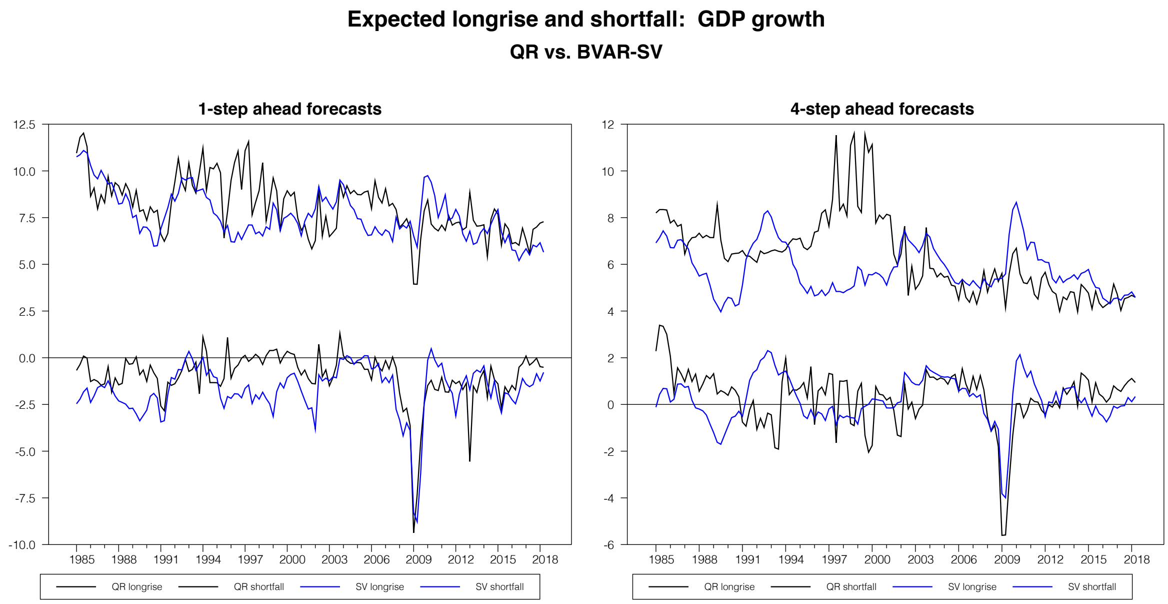


Figure A4: Long-rise and expected shortfall using 10 and 90 percent quantiles, respectively, in-sample forecasts of GDP growth for 1972-2018. The top panel compares estimates from the QR and 5-variable BVAR-SV models. The bottom panel compares estimates from the BVAR-SV and BVAR-SVF-M models.



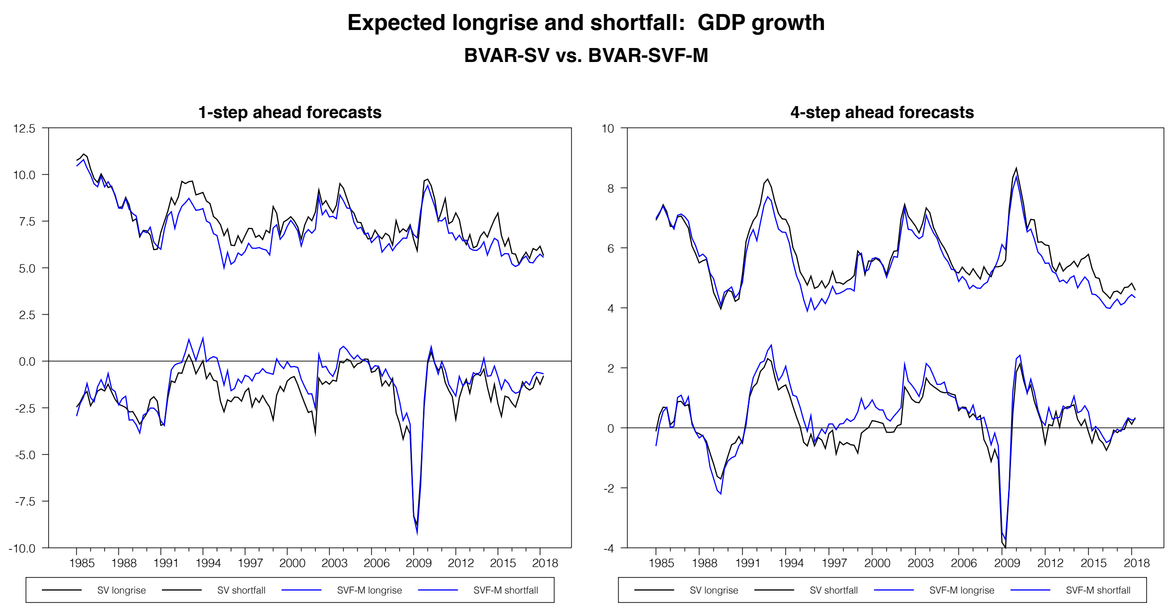


Figure A5: Long-rise and expected shortfall using 10 and 90 percent quantiles, out-of-sample forecasts of GDP growth for 1985-2018. The top panel compares estimates from the QR and 5-variable BVAR-SV models. The bottom panel compares estimates from the BVAR-SV and BVAR-SVF-M models.

A.2 Monte Carlo Assessment of Asymmetries Captured with Stochastic Volatility

At face value, it may seem surprising that the BVAR-SV model yields tail risk estimates comparable to those obtained with quantile regression and the BVAR-SVF-M specification that allows a direct link of macroeconomic volatility to financial conditions. To better understand this outcome, this subsection summarizes the results of a Monte Carlo analysis of the performance of quantile regression and the BVAR models.

In the Monte Carlo experiments, to make the computational burden manageable, we used a bivariate VAR-SVF-M specification with one lag, parameterized so as to reflect some empirical aspects of the features of GDP growth and the NFCI. For simplicity, we will refer to the model's variables as GDP and NFCI. In experiments treating the VAR-SVF-M model as the data-generating process (DGP), the parameterized model takes the following form, in which deterioration in a "financial indicators" variable leads to higher volatility in a "GDP growth" variable and a rise in the uncertainty factor reduces growth and harms financial conditions (raising the NFCI):

$$\begin{pmatrix} \text{GDP}_t \\ \text{NFCI}_t \end{pmatrix} = \begin{pmatrix} 0.249 & -0.269 & -0.400 & -0.200 \\ -0.001 & 0.769 & 0.300 & -0.050 \end{pmatrix} \begin{pmatrix} \text{GDP}_{t-1} \\ \text{NFCI}_{t-1} \\ \ln m_t \\ \ln m_{t-1} \end{pmatrix} + \begin{pmatrix} 1.000 & 0.000 \\ -0.008 & 1.000 \end{pmatrix}^{-1} \begin{pmatrix} \lambda_{gdp,t}^{0.5}\ \epsilon_{gdp,t} \\ \lambda_{nfci,t}^{0.5}\ \epsilon_{nfci,t} \end{pmatrix},$$

where

$$\begin{pmatrix} \ln \lambda_{gdp,t} \\ \ln \lambda_{nfci,t} \end{pmatrix} = \begin{pmatrix} 1.221 \\ 2.956 \end{pmatrix} \ln m_t + \begin{pmatrix} \ln h_{gdp,t} \\ \ln h_{nfci,t} \end{pmatrix}$$

$$\begin{aligned} \ln m_t &= 0.753 \ln m_{t-1} + 0.015\ \text{GDP}_{t-1} + 0.100\ \text{NFCI}_{t-1} + u_{m,t},\ var(u_{m,t}) = 0.03 \\ \ln h_{gdp,t} &= 2.144 - 0.221 \ln h_{gdp,t-1} + e_{gdp,t},\ var(e_{gdp,t}) = 0.034 \\ \ln h_{nfci,t} &= -2.658 + 0.230 \ln h_{nfci,t-1} + e_{nfci,t},\ var(e_{nfci,t}) = 0.032. \end{aligned}$$

With this DGP, we simulate 100 artificial data sets of a total of 190 observations (a sample length corresponding to our actual empirical sample). For each data set, we estimated the quantile regression, BVAR-SV, and BVAR-SVF-M models and formed one-step-ahead in-sample forecast distributions — deliberately using the one-step horizon and in-sample forecasting to make computation tractable. With the in-sample forecasts, we computed quantiles, expected shortfall and long-rise, and the quantile and joint VaR-ES scores (at the 5 percent and 95 percent quantiles, respectively).

Shortfall and long-rise estimates obtained from the Monte Carlo data sets do display asymmetries like those seen in the actual estimates reported in the paper, with periods in which the expected shortfall declines more than the long-rise changes and shortfall is generally more variable than long-rise. To illustrate the asymmetries in shortfall compared to long-rise, Figures A6 and A7 present the time series of estimates obtained with the BVAR-SVF-M model for the first 40 data sets, and Figures A8 and A9 present the estimates obtained with the BVAR-SV model. Qualitatively, these generate period (downward) asymmetries in shortfall as observed in the actual empirical estimates reported in the paper's figures. Tabulations of relative volatilities of shortfall and long-rise confirm the visual impression. Across the 100 data sets, the ratio of the standard deviation of shortfall to the standard deviation of long-rise averages 1.22 in the BVAR-SV forecasts and 1.36 in the BVAR-SVF-M forecasts.

For each model's estimates of the 5 percent quantile and associated shortfall in each data set, we also compute the quantile score and joint VaR-ES score. When we compare the BVAR-SVF-M model to quantile regression with the ratio of the former's score compared to the latter, in the case of the quantile score, the mean ratio is 0.98 (for an average BVAR-SVF-M gain of 2 percent), with the BVAR-SVF-M model having a lower score in 58 percent of the data sets. The corresponding results for the BVAR-SV compared to quantile regression are very similar. As this implies, the accuracy of BVAR-SV and BVAR-SVF-M predictions are very similar, as measured by both the quantile score and the joint VaR-ES score. These patterns align with the paper's empirical findings, in which quantile regression and the BVAR-SV and BVAR-SVF-M models are broadly comparable in tail risk forecast accuracy.

Drawing in part on further investigation of the Monte Carlo estimates, we believe the following two considerations explain these findings in one-step-ahead predictive distributions. First, the BVAR-SV specification appears to be flexible enough that the volatility estimates obtained from it are very similar to those obtained from an estimated BVAR-SVF-M specification corresponding to the DGP. The BVAR-SV model can be seen as a less restrictive form of the BVAR-SVF-M model, in that it does not impose a factor structure on the volatility processes. Of course, the SV setup also does not directly include the link of volatility to financial conditions.

Second, related to this observation, although the BVAR-SV specification assumes that "levels" innovations to the data y_t are independent of innovations to log volatility, in the data and estimates for the Monte Carlo data, it appears that over short periods the model captures patterns of correlated shocks that yield asymmetries.[29] In particular, in visually inspecting the shortfall estimates compared to the levels and volatilities shocks of the BVAR-SV model estimated for data

[29]The correlations in question pertain to short periods and not the overall sample. In the estimates, the shock correlations in question are essentially zero over the full sample of each data set.

generated from the BVAR-SVF-M specification, the downward asymmetries in expected shortfall for output occur when the volatility of output spikes up at about the same time that there are negative shocks to the level of output or adverse shocks to financial conditions. Such a pattern is broadly consistent with the empirical finding of Chavleishvili and Manganelli (2019) that severe financial shocks affect economic activity only when activity is simultaneously hit by a negative shock. Our in-sample estimates of the BVAR-SV model with U.S. data display some correlations between levels and volatilities shocks, particularly for the NFCI and to a lesser extent for GDP growth. For example, over the full sample, the correlation of the shocks to the level and volatility of the NFCI is about 0.2, and over rolling windows of 10 observations, the correlation commonly spikes up around recessions (see the lower panel of the appendix's Figure A10), whereas the rolling window correlation of shocks to the level and volatility of GDP growth turns negative around recessions.[30] As a result, the BVAR-SV and BVAR-SVF-M are well suited to capturing asymmetries in the unconditional predictive distribution of GDP growth. The BVAR-SV captures simultaneity in mean and variance shifts with sporadic correlation between the empirical estimates of level and volatility shocks, whereas in the BVAR-SVF-M model, a shock to the volatility factor also represents a shock to the levels of the macroeconomic variables.

[30]We compute these correlations for draws of the normalized VAR shock $v_{i,t}/\sigma_{i,t}$ and $\nu_{i,t}$ and tabulate the posterior medians of the draws of correlations.

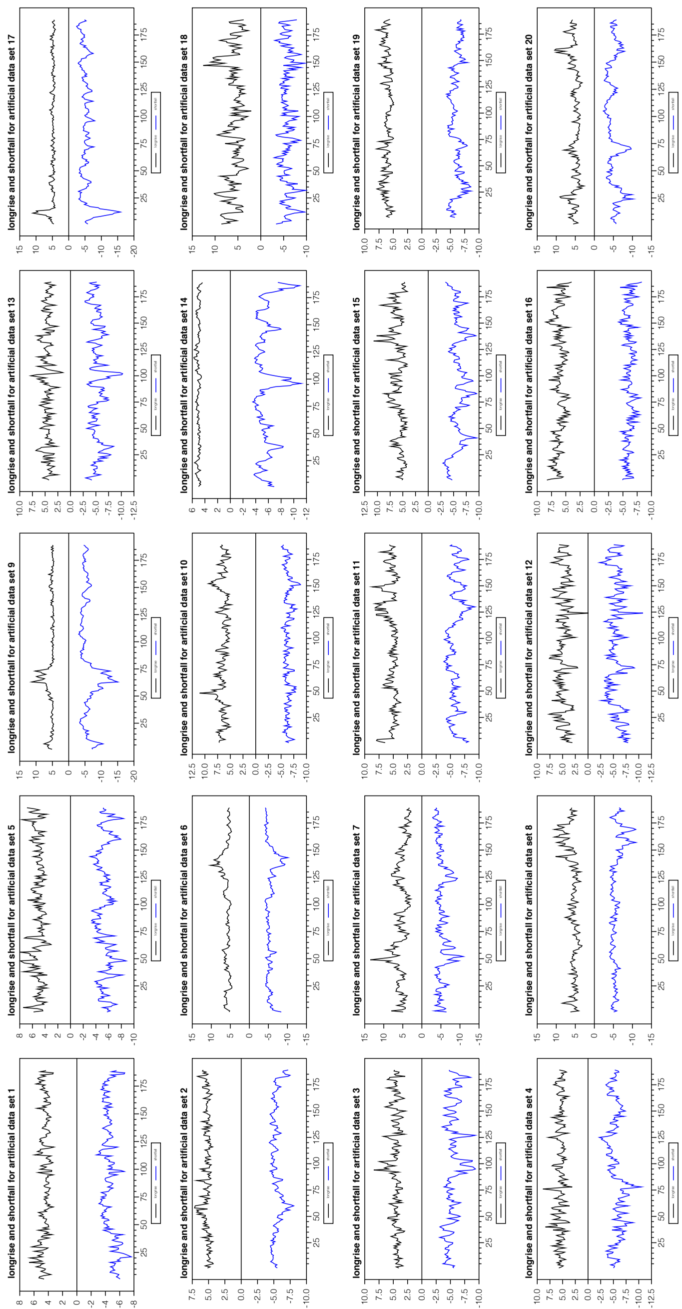


Figure A6: Expected shortfall and long-rise estimates from one-step-ahead in-sample forecasts obtained with the BVAR-SVF-M model, with the BVAR-SVF-M as the DGP: artificial data sets 1 through 20

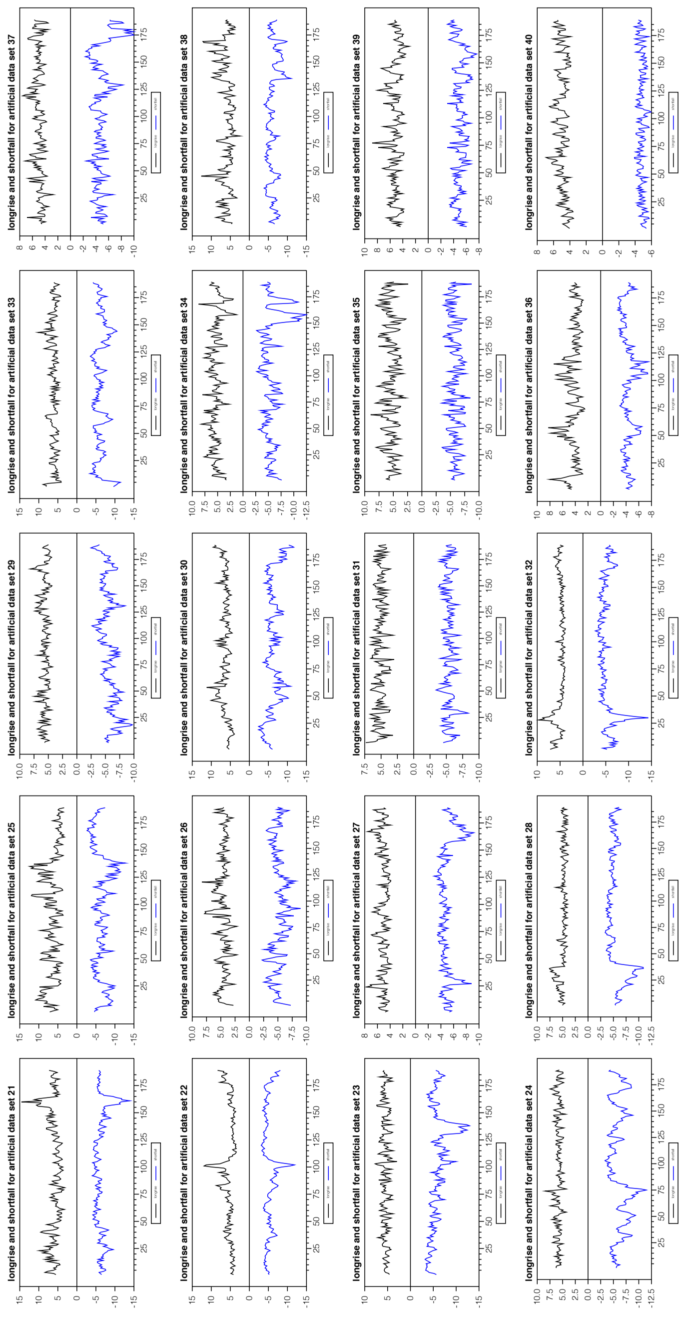


Figure A7: Expected shortfall and long-rise estimates from one-step-ahead in-sample forecasts obtained with the BVAR-SVF-M model, with the BVAR-SVF-M as the DGP: artificial data sets 21 through 40

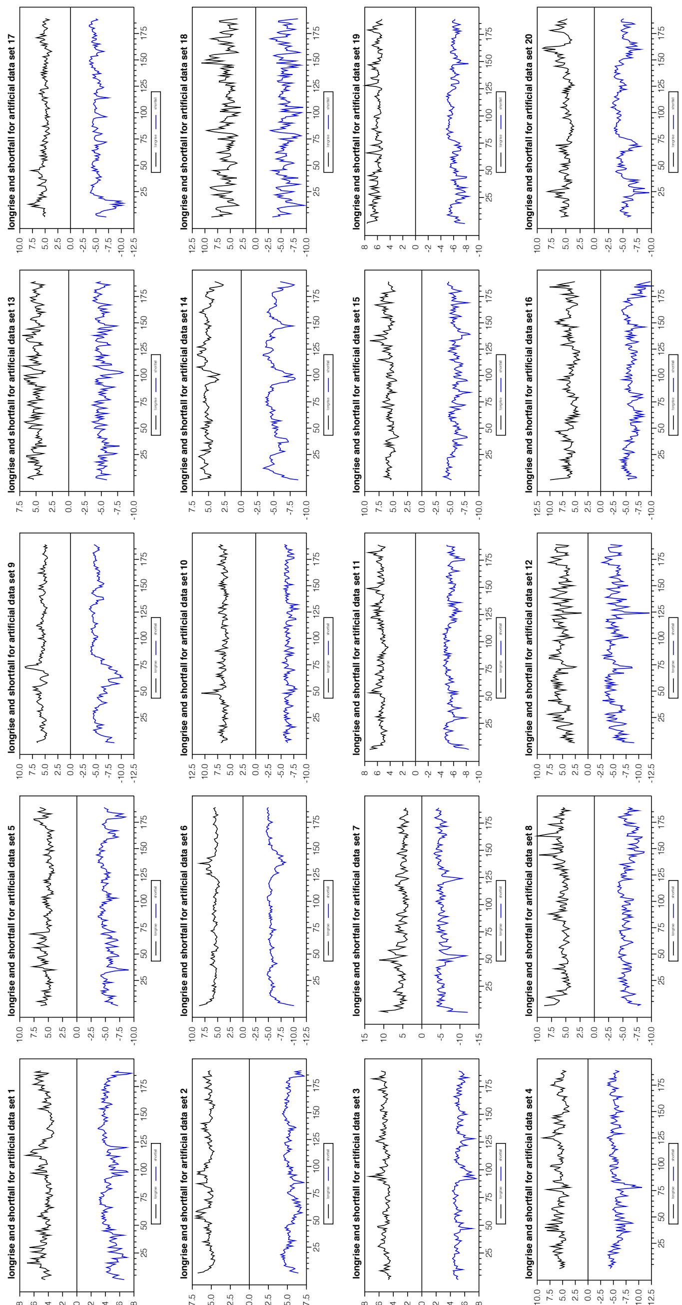


Figure A8: Expected shortfall and long-rise estimates from one-step-ahead in-sample forecasts obtained with the BVAR-SV model, with the BVAR-SVF-M as the DGP: artificial data sets 1 through 20

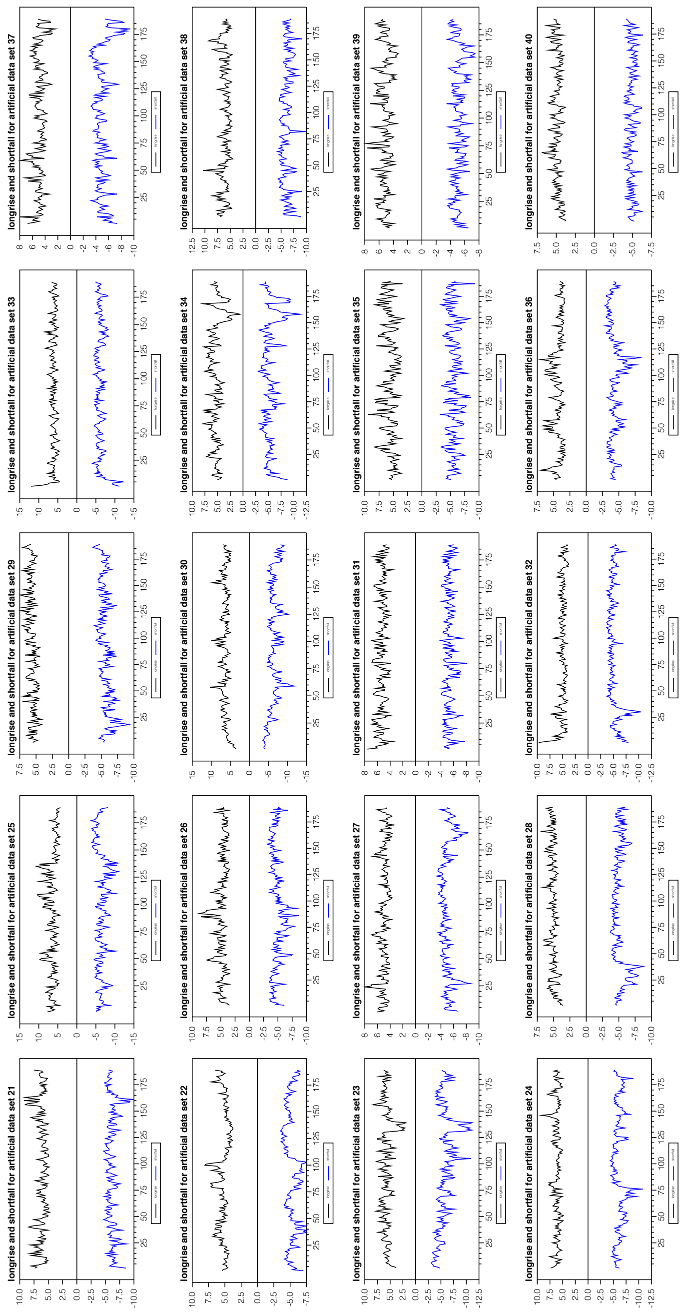
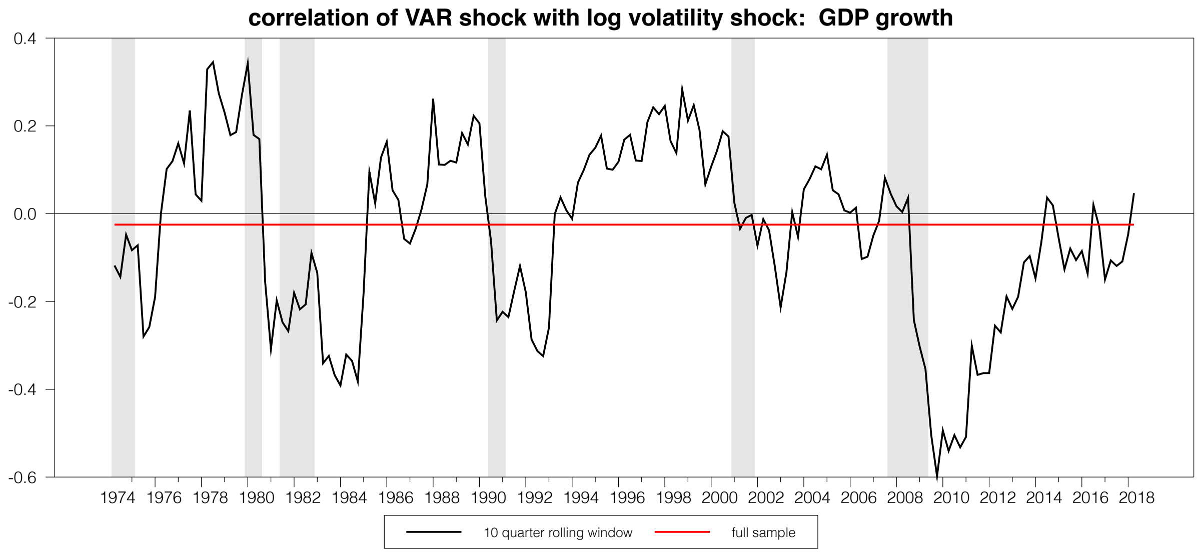


Figure A9: Expected shortfall and long-rise estimates from one-step-ahead in-sample forecasts obtained with the BVAR-SV model, with the BVAR-SVF-M as the DGP: artificial data sets 21 through 40



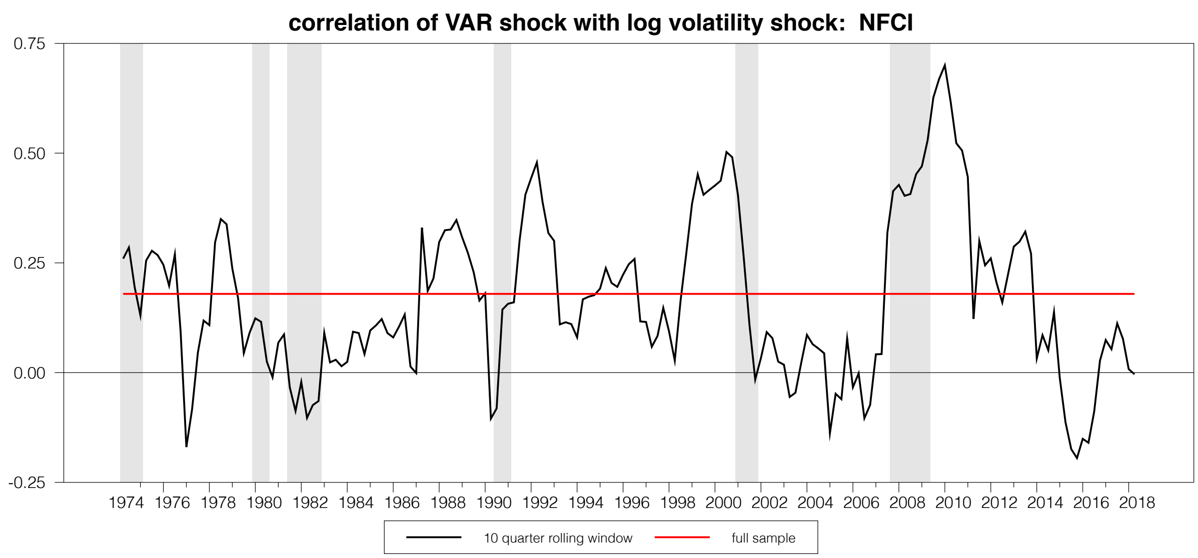


Figure A10: Correlations between levels and volatilities shocks of BVAR-SV model estimated for 1972-2018. The black lines provide correlations computed over rolling windows of 10 observations, and the red lines provide correlations estimated over the full sample of data. The top and bottom panels report the estimates for, respectively, GDP growth and the NFCI. Correlations are computed for each MCMC draw and then tabulated as the medians reported in the charts. Periods shaded in gray denote NBER recessions.