DISCUSSION PAPER SERIES

DP17512

Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions

Andrea Carriero, Todd Clark and Massimiliano Marcellino

MONETARY ECONOMICS AND FLUCTUATIONS

CEPR

ISSN 0265-8003

Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions

Andrea Carriero, Todd Clark and Massimiliano Marcellino

Discussion Paper DP17512
Published 28 July 2022
Submitted 26 July 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions

Abstract

A rapidly growing body of research has examined tail risks in macroeconomic outcomes, commonly using 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. 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). We show that Bayesian vector autoregressions (BVARs) with stochastic volatility are able to capture tail risks in macroeconomic forecast distributions and outcomes. Even though the 1-step-ahead conditional predictive distributions from the conventional stochastic volatility specification are symmetric, forecasts of downside risks to output growth are more variable than upside risks, and the reverse applies in the case of inflation and unemployment. 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: C53, E17, E37, F47

Keywords: Forecasting, Downside risk, Asymmetries

Andrea Carriero - a.carriero@qmul.ac.uk
Queen Mary College, London

Todd Clark - todd.clark@researchfed.org
Federal Reserve Bank of Cleveland

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

Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions*

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

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

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

May 2022

Abstract

A rapidly growing body of research has examined tail risks in macroeconomic outcomes, commonly using 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. 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). We show that Bayesian vector autoregressions (BVARs) with stochastic volatility are able to capture tail risks in macroeconomic forecast distributions and outcomes. Even though the 1-step-ahead conditional predictive distributions from the conventional stochastic volatility specification are symmetric, forecasts of downside risks to output growth are more variable than upside risks, and the reverse applies in the case of inflation and unemployment. Overall, the BVAR models perform comparably to quantile regression for estimating and forecasting tail risks, complementing BVARs' established performance for forecasting and structural analysis.

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

*We gratefully acknowledge research assistance from Matthew Gordon, Jared Steinberg, and John Zito and helpful comments from the Editor Ken West and three anonymous referees, Andras Fulop, Domenico Giannone, Sune Karlsson, Elmar Mertens, James Mitchell, Simon Van Norden, Shaun Vahey, Saeed Zaman, Molin Zhong, and seminar participants at the Bank of Canada, the Econometric Society's World Congress, Indiana University, the International Association of Applied Econometrics, King's College, and the University of Strathclyde. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System. Marcellino thanks for financial support MIUR – PRIN Bando 2017 – prot. 2017TA7TYC.

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 (2019), Adrian, et al. (2020), De Nicolo and Lucchetta (2017), and Giglio, Kelly, and Pruitt (2016)). This work has emphasized the link of tail risks to output stemming from poor financial conditions. Other work has considered tail risks to other variables, such as inflation (e.g., Ghysels, Iania, and Striaukas (2018) and Lopez-Salido and Loria (2021)), unemployment (e.g., Galbraith and van Norden (2019) and Kiley (2022)), or used other methods, such as copula modeling (e.g., Smith and Vahey (2016) and Loaiza-Maya and Smith (2019)).[1] 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.

The interest in tail risks reflects an underlying perception or assumption of asymmetries in distributions of outcomes.[2] Some form of asymmetry has long been incorporated in particular economic models: As examples, 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. A body of research has also examined asymmetries in the unemployment rate (e.g., Galbraith and van Norden (2019) and references therein). In the context of the recent literature on tail risks to output growth, Delle Monache, De Polis, and Petrella (2020) develop score-driven parametric models to model time-varying skewness in predictive distributions, and Caldara, et al. (2021) and Lhuissier (2022) apply Markov switching models to capture tail risks. Empirically combining forecasts from quantile regressions, Mitchell, Poon, and Zhu (2022) find more evidence of multi-modalities in predictive distributions than asymmetries. Combining non-parametric and Monte Carlo methods to

[1] Karagedikli, Vahey, and Wakerly (2019) use copula-based combinations of forecasts to quantify tail risks. Focusing on forecasts from the Survey of Professional Forecasters, Ganics, Sekhposyan, and Rossi (2021) develop a density combination approach to obtain fixed-horizon density forecasts from fixed-event forecasts; their combined densities display some skewness and asymmetries. Other examples of studies of quantile forecasts in macroeconomics include Gaglianone and Lima (2012), Korobilis (2017), and Manzan and Zerom (2013, 2015).

[2] 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. In the US, 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."

assess the joint distribution of economic and financial conditions, Adrian, Boyarchenko, and Giannone (2021) also find multi-modalities.

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. (2020) and Adams, et al. (2021) 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.[3] 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 1-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 1-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 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 examine the ability of Bayesian vector autoregressions (BVARs) with stochastic volatility to capture tail risks in macroeconomic forecast distributions and outcomes.[4] BVARs are commonly used for point and density

[3]ABG consider an alternative approach based on an econometric model in which GDP volatility is a deterministic function of financial conditions, which also implies symmetric conditional distributions and asymmetric unconditional distributions. Caldara, Scotti, and Zhong (2021) 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.

[4]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 sets of predictors. In their nowcast setting, Bayesian regressions with stochastic volatility perform better than quantile regressions. This paper differs from Carriero, Clark, and Marcellino (2020) by focusing on multi-step forecasts as opposed to current quarter nowcasts, by focusing on the quarterly variables commonly in BVARs rather than mixed frequency indicators of economic activity and financial conditions that may be informative about the current quarter, by considering a different set of models that includes a BVAR with

forecasting, are known to have a successful track record compared to structural models and survey-based forecasts, and can be easily adapted to include a range of variables and produce a variety of forecast density measures. BVARs 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.

A conventional BVAR with stochastic volatility may be capable of 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. At the 1-step-ahead horizon, BVARs with conventional stochastic volatility will generally yield conditional predictive distributions that are symmetric.[5] At longer horizons, because of parameter uncertainty, the conditional predictive distributions may not be symmetric.[6] But as noted above, asymmetries in conditional distributions are not necessary to obtain more time variation in downside risks than upside risks (or 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 BVARs 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 4-quarter average for smoothing) is -0.26.

To get at tail risks with models that allow for asymmetries in conditional distributions, we also consider a BVAR with stochastic volatility that features a contemporaneous correlation between shocks to the levels and volatilities of variables. We take such a model off the shelf, so to speak: Drawing on the model used in Carriero, Clark, and Marcellino (2018) to measure macroeconomic uncertainty and its effects, we rely on a BVAR with a common factor in volatility that enters the BVAR's conditional mean. 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 BVAR, and the conditional predictive distribution can be skewed. This specification has the advantage that it easily scales to allow moder-

stochastic volatility in mean, and by examining tail risks to not only output growth but also unemployment and inflation.

[5]Caldara, Scotti, and Zhong (2021) work through (in results in the paper's appendix) analytics for a BVAR with stochastic volatility, abstracting from parameter uncertainty.

[6]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. 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.

ate or even large variable sets. In a bivariate BVAR setting, Caldara, Scotti, and Zhong (2021) — henceforth, CSZ — take a different avenue to embedding the ingredient necessary for asymmetries in conditional distributions at the 1-step-ahead horizon: Their model has only lagged (not contemporaneous) volatility in the BVAR's conditional mean but allows a correlation between the BVAR's innovations and the volatility innovations.

Our paper also contributes to the recent macroeconomic tail risk literature by providing more formal evaluations of tail risk forecasts and the performance of alternative models than has much of the recent literature.[7] We formally evaluate tail risk forecasts using the quantile score function and the quantile-weighted continuous ranked probability score developed by Gneiting and Ranjan (2011). Our focus on forecasting with vector autoregressions and formal risk forecast evaluation distinguishes our paper from a contemporary analysis by CSZ. Including both measurement and structural analysis, CSZ focus instead on the mechanisms by which a bivariate BVAR 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.

Reflecting the combination of common practice in the BVAR-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, we measure financial conditions with the Chicago Fed's national financial conditions index. 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 but also provide results for the unemployment rate and inflation.

Our analysis yields the following main results. First, GDP growth, unemployment, and inflation are subject to asymmetries in tail risks, with downside risks more variable than upside risks for GDP growth and upside risks more variable than downside for unemployment and inflation. Second, 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 some important aspects of tail risks to output growth, unemployment, and inflation. In formal forecast scoring, there is little to distinguish our BVAR performance from quantile regression performance (and vice versa). In the BVARs,

[7] Although other work in the tail risk literature, such as ABG, considers forecast accuracy, their measures focus on predictive scores and probability integral transforms, which get at general density accuracy and calibration, whereas we focus on tail forecast accuracy specifically. The recent work of Brownlees and Souza (2021) 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. Brownlees and Souza (2021) conclude that, for output growth, AR models with GARCH outperform quantile regression.

the time variation in downside tail risks as compared to upside risks for output growth (and the reverse for unemployment and inflation) 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. 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 BVAR's conditional mean.

Based on these results, we do not mean to claim that, in truth, there are no asymmetries (possibly time-varying) in conditional predictive distributions for macroeconomic variables. Rather, one aspect of tail risk that has received some emphasis in the literature — downside risks varying more over time than upside risks — can be captured as well with a BVAR with conventional stochastic volatility that yields symmetric conditional distributions as with other models that allow asymmetries in conditional distributions. In addition, in formal metrics of tail forecasts, the conventional BVAR is comparable in accuracy to the specifications that allow asymmetries in conditional distributions. We take this as suggestive evidence that conditional asymmetries are not necessarily a strong, regular feature of predictive distributions for output growth. As noted below in our model presentations, tests for symmetry applied to the conditional predictive distributions from our BVAR model with the volatility factor in the conditional mean are generally consistent with symmetry, even though the same tests detect consistent asymmetries in the predictive distributions for some of the other variables in the BVAR. Our interpretation is consistent with the cautionary findings of Plagborg-Moller, et al. (2020) that no predictors considered yielded "useful advance warnings of tail risks or indeed about any features of the GDP growth distribution other than the mean." But we recognize that other methods or analyses may reach a different conclusion, and we leave to further research whether some of these other methods under development can establish gains over formulations of BVARs with stochastic volatility.

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. 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 BVAR's conditional mean (BVAR-SVF-M), as in Carriero, Clark, and Marcellino (2018); and quantile regression (QR) as in ABG. Throughout, we focus on quarterly data and forecast horizons of 1 and 4 quarters.

Our BVAR specifications include five variables, at a quarterly frequency: GDP growth (annualized, as $400\Delta \ln$ GDP), the unemployment rate, inflation in the GDP price index (annualized, as $400\Delta \ln P$), the federal funds rate, and the Chicago Fed's NFCI.[8] The first four variables are very commonly used in small BVARs in the forecasting literature (see, e.g., Clark and Ravazzolo (2015)). 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). The supplemental appendix includes the results of a robustness check in which we replaced the NFCI with the turbulence measure of financial market volatility considered in Giglio, Kelly, and Pruitt (2016). In these results, our BVAR models forecast tail risks better than QR, suggesting some sensitivity of QR results to the choice of financial indicator.

Although we focus on small models, one feature of BVAR-based models is their easy scalability: We could easily add more financial measures to our BVARs. 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 BVAR specifications (which also include the NFCI) yields long-rise and shortfall estimates and forecast accuracy very similar to our baseline estimates reported herein.

As detailed below, our quantile regression specifications for GDP growth, unemployment, and inflation use subsets of the five variables in the model, with the addition of a long-run survey-based measure of inflation expectations taken from the Federal Reserve Board's FRB/US model (denoted PTR in the model).

[8] Although we focus on the five-variable model as compared to simple quantile regression, an earlier version of this paper (Carriero, Clark, and Marcellino (2020b)) obtained qualitatively similar results with just a bivariate model in GDP growth and the NFCI and with a 15-variable model.

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. This model implies that the reduced-form variance-covariance matrix of innovations to the BVAR is $\text{var}(v_t) \equiv \Sigma_t = A^{-1}\Lambda_t A^{-1\prime}$. Note that, as in Primiceri's (2005) implementation, innovations 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 BVAR's coefficients. In implementation, we include four lags in the BVAR.

Regarding the priors for the BVAR-SV model, we use settings like those common in the forecasting literature. For the BVAR coefficients contained in Π, we use a Minnesota-type multivariate normal prior. With the variables of interest transformed for stationarity, we set the prior mean of all the BVAR 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) in using a fairly uninformative multivariate normal prior, with means of 0 and variances of 10 for all coefficients. For the Gaussian priors on 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 10 degrees of freedom. 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). Volatility is sampled with a Gibbs step based on Kim, Shephard, and Chib (1998). Estimates and forecasts from the BVAR-SV model are based on 25,000 retained draws, obtained by sampling a total of 30,000 draws and discarding the first 5,000. As detailed below in the quantile regression section, at the 4-quarters-ahead horizon, the forecasts of interest are average growth rates or 4-quarter changes. In the BVAR case, we transform the underlying quarterly forecast draws as needed to obtain average growth rates or 4-quarter changes.

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 BVAR with stochastic volatility, links the (unobservable) factor in volatility to the last quarter's levels of the BVAR's variables, and allows the volatility factor to enter the BVAR's conditional mean. 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). Hence, the empirical results will be also informative on whether or not uncertainty plays a major role in generating asymmetries in predictive distributions. This model takes the form:

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

The log volatility of each variable i 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 BVAR'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.

We use priors for the BVAR-SVF-M model aligned with those of the BVAR-SV specification. Regarding the unique components of the BVAR-SVF-M model, for the coefficients Π_m on uncertainty in the BVAR'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 (independent) Gaussian prior on 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 Gaussian 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 use a normal distribution with mean 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 (also Gaussian) 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 an inverse Gamma prior with 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 implementation details. First, the uncertainty shock $u_{m,t}$ is independent of the conditional errors ϵ_t as well as the elements of $\nu_t = (e_{1,t}, \ldots, e_{n,t})'$, which are distributed independently of one another as i.i.d. $N(0, \Phi_\nu)$, with $\Phi_\nu = diag(\phi_1, \ldots, \phi_n)$. 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, we set the model's lag orders at $p = 4$ and $p_m = 2$. Finally, we estimate the model with a Gibbs sampler. 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 (see Carriero, Clark, and Marcellino (2018) for more information on the particle Gibbs step). Estimates and forecasts from the BVAR-SVF-M model are based on 25,000 retained draws, obtained by sampling a total of 30,000 draws and discarding the first 5,000.

2.3 Asymmetries and BVARs with Stochastic Volatility

In what respects can the BVARs just described yield asymmetries in tail risk estimates and forecasts?[9] First consider the possibility of asymmetries in conditional predictive distribu-

[9]Although we have maintained the paper's focus on the efficacy of tail forecasts, we have used marginal likelihood estimates to compare the historical fit of the BVAR-SV and BVAR-SVF-M models (using the computational approach detailed in the supplemental appendix to Carriero, et al. (2022), we obtained the estimates as sums of 1-step-ahead predictive likelihoods, for a sample from 1980:Q1 through 2019:Q4). These fit measures indicate that the BVAR-SV model fits the data better for the full sample. The advantage of this model over the BVAR-SVF-M specification largely accrued in the period since the onset of the Great

tions. In the case of the BVAR-SV model, for the reasons analyzed in CSZ, its conditional 1-step-ahead predictive distributions 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) — a test that is based on the empirical distribution — to the forecast draws from the predictive densities of our estimated BVAR-SV model yields results broadly consistent with symmetry. In the out-of-sample forecasts of GDP growth detailed below, with a 5 percent critical value the Bai-Ng test rejects conditional symmetry in only 8 percent of forecast origins at the 1-step-ahead horizon and 9 percent at the 4-steps-ahead horizon. 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-ahead horizon.

Unlike the BVAR-SV model, the BVAR-SVF-M specification is capable of producing or capture asymmetries in the conditional predictive distribution, including at the 1-step ahead horizon. With a common factor in volatility that enters the BVAR'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. Below we describe some Monte Carlo experiments that verify the ability of this model to capture asymmetries in predictive distributions. In our empirical results, applying at each forecast origin the conditional symmetry test of Bai and Ng (2001) 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 out-of-sample forecasts of GDP growth detailed below, with a 5 percent critical value the Bai-Ng test rejects conditional symmetry in only 7 percent of forecast origins at the 1-step-ahead horizon but 19 percent at the 4-steps-ahead horizon; corresponding figures for the unemployment (inflation) rate are 10 (37) percent and 28 (53) percent, respectively. At both horizons, the rejection rates for the federal funds rate are 62 percent, and rates for the NFCI are 90 percent or more. We read these results as indicating that, although the model does not pick up notable asymmetries in GDP growth's conditional predictive distributions, it is capable of picking them up, as evidenced by the results for other variables in the BVAR.

Notwithstanding these findings, the results presented below will capture a form of asym-

Recession; the models fit the data about equally well up to about 2007.

metry that has received attention in the tail risk literature but does not actually require asymmetries in conditional predictive distributions: We will obtain more time variation in downside risks than upside (for output growth) or the reverse (for unemployment and inflation) because both BVAR models — including the BVAR-SV model — are able to capture simultaneous shifts in the means and variances of the conditional distributions. In the estimates for GDP growth, the simultaneous downward shift of the conditional mean and upward shift of the conditional variance occur over short periods as the model captures patterns of correlated shocks to levels and volatilities.

In the supplemental appendix, we provide the results of Monte Carlo experiments that corroborate this finding. In particular, with data generated by an empirically-parameterized BVAR-SVF-M specification, 1-step-ahead predictive distributions display more time variation in downside risks than upside risks, with comparable estimates from 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.

In these experiments, for each data set, we also compute (5 percent) quantile scores and qwCRPS-left scores for forecasts from the BVAR models and quantile regression (the next subsection details the basic quantile regression). These results show patterns like those seen in our empirical results below, in which quantile regression and the BVAR-SV and BVAR-SVF-M models are broadly comparable in tail risk forecast accuracy: In Monte Carlo data from an empirically-parameterized BVAR-SVF-M specification, forecasts from the BVAR-SV and BVAR-SVF-M models are about equally accurate and as accurate or slightly more accurate than quantile regression.

To verify that the BVAR-SVF-M model would capture asymmetries truly present in the data, in another set of Monte Carlo experiments we generated data from a BVAR-SV model with a strong negative correlation between the BVAR's shock to the level of the output variable and the shock to volatility, which creates a significant asymmetry or skewness in the conditional predictive distribution (for reasons detailed in Caldara, Scotti, and Zhong (2021)). In this case, shortfall and long-rise estimates obtained from the Monte Carlo data display time series asymmetries sharper than those seen in the first DGP. In this case, with the DGP featuring strong asymmetries, while both of the estimated BVAR models yield estimates of shortfall more volatile than estimates of long-rise, the BVAR-SVF-M model does so by a larger magnitude than does the BVAR-SV specification. Using the Monte Carlo-generated data to compare tail risk forecast accuracy for the quantile regression, BVAR-SV, and BVAR-SVF-M models, the BVAR-SVF-M specification is more accurate than quantile regression or the BVAR-SV model. While the BVAR-SV model continues to be at least as successful as QR, neither match the accuracy of the BVAR-SVF-M model. Overall, the results of the second set of Monte Carlo experiments indicate that, for a DGP featuring strong asymmetry, it continues to be the case that the BVAR models we consider — including the BVAR-SV specification — can capture tail risk at least as well as quantile regression. In fact, in this case, the BVAR-SVF-M model is most accurate, offering relatively sizable gains over QR.

2.4 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 (QR)** 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 predictand $y_{t+h}^{(h)}$ is horizon-specific, and the vector of predictors x_t includes a constant, one lag of the variable being predicted, and the NFCI_t; the inflation application includes some additional predictors detailed below.

In all cases, the parameter vector β_τ is obtained with standard 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 a range of quantiles, from $\tau = 0.05$ to $\tau = 0.95$. Following ABG, to obtain the expected shortfall and long-rise measures considered (detailed below), at each point in time we use the estimates of the four quantiles of $\tau = 0.05$, 0.25, 0.75, and 0.95 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

Table 1: Quantile regression model specifications

Application	Dependent variable	Predictors
Output growth		
$h = 1$	$400 \ln(\text{GDP}_{t+1}/\text{GDP}_t)$	$400 \ln(\text{GDP}_t/\text{GDP}_{t-1})$, NFCI_t
$h = 4$	$100 \ln(\text{GDP}_{t+4}/\text{GDP}_t)$	$400 \ln(\text{GDP}_t/\text{GDP}_{t-1})$, NFCI_t
Unemployment		
$h = 4$	$\text{UR}_{t+4} - \text{UR}_t$	UR_t, NFCI_t
Inflation		
$h = 4$	$100 \ln(\text{PGDP}_{t+4}/\text{PGDP}_t)$	$100 \ln(\text{PGDP}_t/\text{PGDP}_{t-4})$, UR_t, PTR_t, NFCI_t

Note: UR refers to the unemployment rate, PGDP refers to inflation in the GDP price index, and PTR refers to a survey-based measure of long-run inflation expectations.

comparisons.[10]

Table 1 lists the specifications of the predictand and predictors we use in our quantile regression applications to growth, unemployment, and inflation. Our choices are informed by recent precedents in the literature. Following ABG, 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 $y_{t+h}^{(h)} \equiv h^{-1} \sum_{i=1}^{h} y_{t+i}$. For unemployment, we follow Kiley (2022) by modeling and forecasting the multi-step change in the unemployment rate, in our case the 4-quarter change in unemployment, using the lagged level of unemployment and the NFCI as predictors. For inflation, we follow Lopez-Salido and Loria (2021) in modeling and forecasting the 4-quarter rate of inflation, using lagged inflation, the unemployment rate, inflation expectations as measured with PTR, and the NFCI as predictors.

3 Data

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).

[10]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.

In full-sample estimates described below, GDP, the GDP deflator, and unemployment rate are measured with the 2022:Q1 series from the RTDSM, the last in our sample of vintages.

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 FRED database of the Federal Reserve Bank of St. Louis. In the case of the NFCI, its construction from a factor model means it will be subject to revision over time, but to maximize the historical sample for our forecast evaluation (to have it start no later than 1985) we abstract from the real-time aspect of the NFCI and use a current vintage series obtained from FRED. After constructing real-time vintages of the NFCI starting in 1988, Amburgey and McCracken (2022) find that using their real-time NFCI yields results comparable to those obtained with a final-vintage measure abstracting from data revisions (although using the real-time NFCI showed some advantages around recessions). We obtained PTR, the measure of long-run inflation expectations, from the public data files for the Federal Reserve Board's FRB/US model.

Our analysis of real-time forecasts uses real-time data vintages from 1985:Q1 through 2022:Q1. 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 2022:Q1 contains data ending in 2021:Q4. The out-of-sample forecast evaluation uses a sample of forecasts produced starting in 1985:Q1 and ending in 2021:Q4; however, in many results, we end the sample in 2019 to avoid distortions from the extreme volatility induced by the COVID-19 pandemic.

To evaluate the accuracy of the real-time forecasts, following studies such as Clark (2011) and Faust and Wright (2009) that have used early estimates, we use the first available (in the quarterly vintages of the RTDSM) estimates of the real-time measured variables as actuals in evaluating forecast accuracy.

Finally, in our main results all models are estimated with data samples that start in 1971:Q1, reflecting the starting point of the NFCI. With four lags in the BVARs, the starting point of the estimation sample is then 1972:Q1. For QR, reflecting the lag structures and forecast horizons, the starting point of the estimation sample is 1971:Q2 for the 1-quarter-ahead GDP growth application and 1972:Q1 for all the 4-quarters-ahead applications.

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. For GDP growth, following conventions in recent literature we will mostly focus on the left tail but provide upper tail results for completeness. For unemployment, some may view upper tail risk as more interesting than lower tail (e.g., Kiley (2022) focuses on the upper tail), but again, we address both for completeness. For inflation, both upside and downside risks can be seen as important.

In comparing the models, we first consider 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) — e.g., 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's fit of a skewed t-distribution to selected quantiles to obtain the complete density function and, in turn, the shortfall and long-rise.[11]

To more formally assess the efficacy of the models in quantifying tail risks, we consider two measures of the accuracy of the lower and upper tail quantile estimates. 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 τ, where in the main results, $\tau = 0.05$ and 0.95. Applied to these quantile estimates, the first accuracy measure is the quantile score, commonly associated with the tick loss function (see, e.g., Giacomini and Komunjer (2005)). The quantile score is computed as

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

[11]Mitchell, Poon, and Mazzi (2022) and Mitchell, Poon, and Zhu (2022) examine a range of approaches to obtaining complete density estimates from quantile regression.

where: y_{t+h} is the actual outcome for GDP growth, the change in unemployment, or inflation; $Q_{\tau,t+h}$ is the forecast quantile at quantile τ; 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 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.

We also formally evaluate the QR, BVAR-SV, and BVAR-SVF-M forecasts with the quantile-weighted continuous ranked probability score (qwCRPS). Gneiting and Ranjan (2011) develop the qwCRPS as a proper scoring function of the entire predictive density with quantile weighting that allows the researcher to emphasize selected portions of the density. The qwCRPS is computed as a weighted sum of quantile scores over 19 quantiles:

$$\text{qwCRPS}_{t+h} = \frac{2}{J-1}\sum_{j=1}^{J-1} v(\tau_j)\text{QS}_{\tau_j,t+h} \tag{6}$$

with $J = 20$ and $\tau_j = j/J = 0.05, 0.10, 0.15, \ldots, 0.90, 0.95$. We consider a left-tail-weighted version (qwCRPS-left) with the weighting function set to $v(\tau_j) = (1-\tau_j)^2$ and a right-tail-weighted version (qwCRPS-right) with the weighting function set to $v(\tau_j) = \tau_j^2$.

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 studies such as ABG and Kiley (2022) 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 as seen in some previous work. 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 use final-vintage data for the in-sample results.[12]

The supplemental appendix provides results (qualitatively similar to the results herein)

[12]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 BVAR forward starting from the data y_{t-1} to obtain draws of forecasts for periods t through $t+H-1$.

for some additional forecast metrics. These include root mean square errors of point forecasts and simple coverage measures for interval forecasts (i.e., the percentages of outcomes falling below the 5 and 95 percent quantiles of the forecast distribution). These additional results also include the joint value at risk-expected shortfall score (targeting the left tail) employed in Carriero, Clark, and Marcellino (2020), which fits in the general class of scoring functions developed in Fissler and Ziegel (2016). Finally, the appendix provides results for dynamic quantile (DQ) tests as developed in Engle and Manganelli (2004) to assess whether quantile forecasts meet basic requirements of unbiasedness and, at the 1-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.

5 Empirical Results

This section begins with estimates — for GDP growth — 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 out-of-sample forecast accuracy for GDP growth. The final subsection examines results for forecasting the unemployment rate and inflation. The supplemental appendix includes results on tests of skewness and kurtosis in the raw data, BVAR-SV residuals, and BVAR-SV out-of-sample forecast errors, as well as results on in-sample forecast accuracy and other results mentioned above.

In presenting results, we distinguish the 2020-2021 period of the COVID-19 pandemic from the rest of the sample. The extreme volatility of the first few quarters of the pandemic produced volatility in some parameter estimates and tail risk forecasts (the latter reflecting both unusual jumping off points as well as some shifts in parameters). For example, in the out-of-sample estimates for the 1-quarter-ahead QR model of GDP growth, at the $\tau = 0.95$ quantile the coefficient on lagged growth swung from 0.45 in the estimate using data through 2020:Q1 to -0.44 in the estimate using data through 2020:Q3.[13] Accordingly, in reporting estimates of in-sample forecasts of shortfall and long-rise, we end the sample in 2019. In reporting out-of-sample results, for chart readability we separately provide results for samples ending in 2019 and samples ending in 2021. For the longer sample, our charts report 5 and 95 percent quantiles rather than long-rise and expected shortfall. The second-step smoothing used by ABG to obtain long-rise and expected shortfall is made difficult

[13] Although the BVAR's parameter estimates are less dramatically impacted by the COVID-19 extremes, they are not entirely unaffected. Carriero, et al. (2022) develop a model extension to accommodate an outlier volatility state. In unreported estimates, in this tail risk analysis, extending our models to include the outlier treatment yields qualitatively similar BVAR results to those reported for 2020-2021.

by the quantile crossings that occur in 2020 with simple QR estimates for GDP growth and unemployment; although there are various steps one might take to prevent or fix the crossing, we take the simpler approach of just directly reporting the tail quantile forecasts.

Note also that, throughout, we date forecasts by the forecast origin rather than the outcome date. So in our charts, for a forecast horizon of h quarters, a forecast date of t refers to a forecast outcome date of $t + h - 1$.

5.1 Predictive Distributions

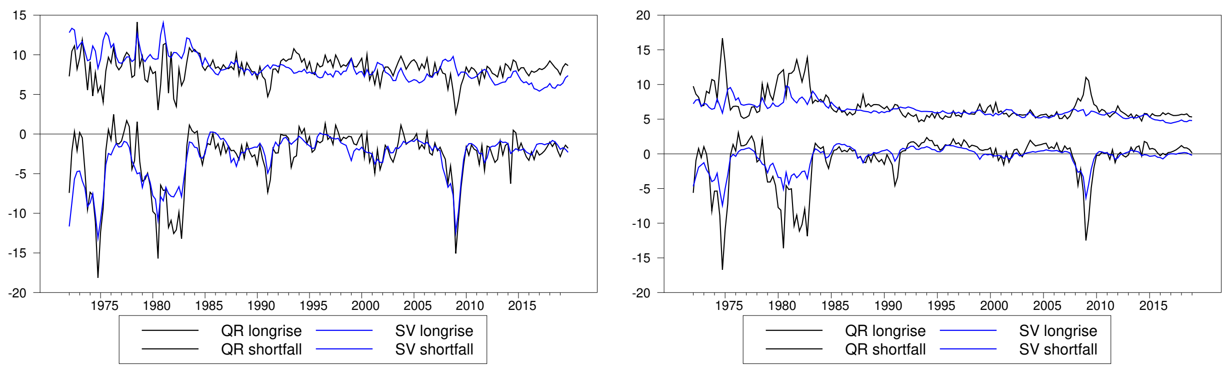
To compare the abilities of the quantile regression, BVAR-SV model, and BVAR-SVF-M model to capture downside risks to GDP growth, Figures 1 and 2 report time series of expected shortfall (at 5 percent) and long-rise (at 95 percent) estimates at the 1-step-ahead and 4-steps-ahead forecast horizons. The in-sample forecast estimates of Figure 1 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 4-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. In our estimates, the simple correlation between the estimated volatility of GDP growth and the level of GDP growth (using a 4-quarter average for smoothing) is -0.26. This comovement obtains even though it is not directly built into the model.

On the other hand, as indicated in the results in the upper panel of Figure 1, the 4-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 1-step-ahead case. The lower panel of Figure 1 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 in the 1970s and early 1980s the BVAR-SVF-M specifications yield larger moves in long-rise and shortfall estimates at the 4-steps-ahead horizon. 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 4-steps-ahead horizon, the standard deviation of shortfall divided by the standard deviation of long-rise is 1.7 for the quantile regression estimates and 1.6 for the BVAR-SV estimates (details are

Figure 1: In-sample forecasts of GDP growth for 1972-2019

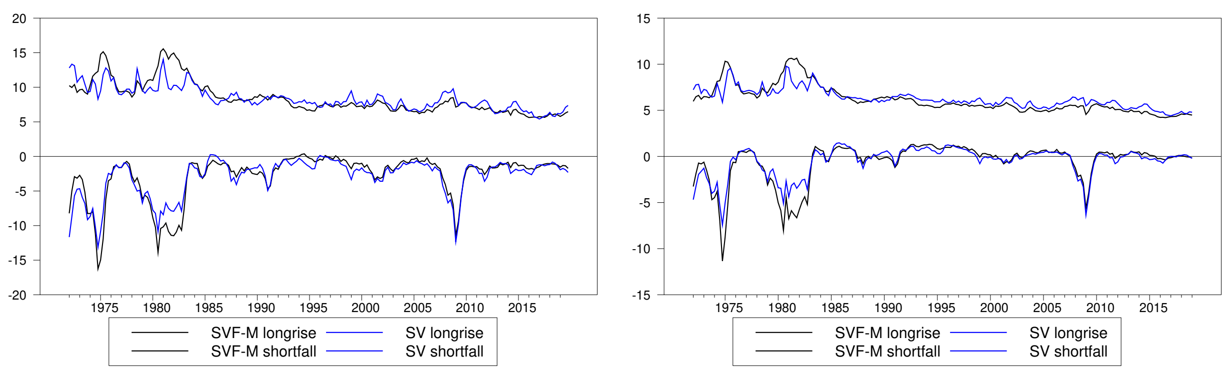
Long-rise and expected shortfall: QR vs. BVAR-SV

(a) $h = 1$ **(b)** $h = 4$



Long-rise and expected shortfall: BVAR-SV vs. BVAR-SVF-M

(c) $h = 1$ **(d)** $h = 4$



provided in the supplemental appendix's Table A4). In addition, results provided in the appendix (Figure A4) 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 2 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. At the 1-step-ahead 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 1-step-ahead (4-steps-ahead) horizon, the standard deviation of shortfall divided by the standard deviation of long-rise is 0.8 (1.4) for the quantile regression estimates and 0.9 (0.8) 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, as noted earlier, in the in-sample case, shortfall estimates fall more in the recessions of the 1970s and 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.

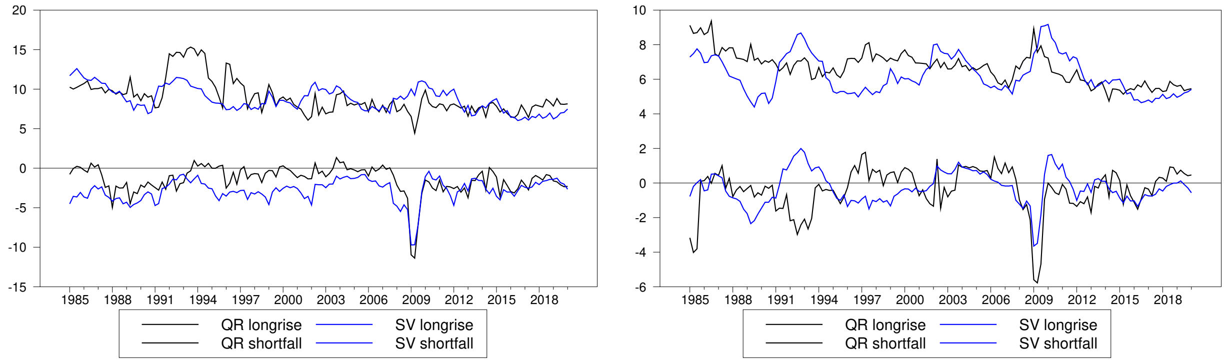
Differences in in-sample and out-of-sample forecast estimates could be driven partly by instabilities in the parameters of the BVAR. In the in-sample case, the parameter estimates will average across any breaks, and the forecasts will reflect these averages. In the out-of-sample case, instabilities in 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 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. As noted above, we have conducted Monte Carlo experiments that corroborate this finding.

The last row of Figure 2 extends the sample, reporting out-of-sample forecasts of 5 and 95 percent tail quantiles (for brevity, from just the QR and BVAR-SV specifications) to go through 2021:Q4, covering the period of the COVID-19 pandemic. In this unprecedented

Figure 2: Out-of-sample forecasts of GDP growth

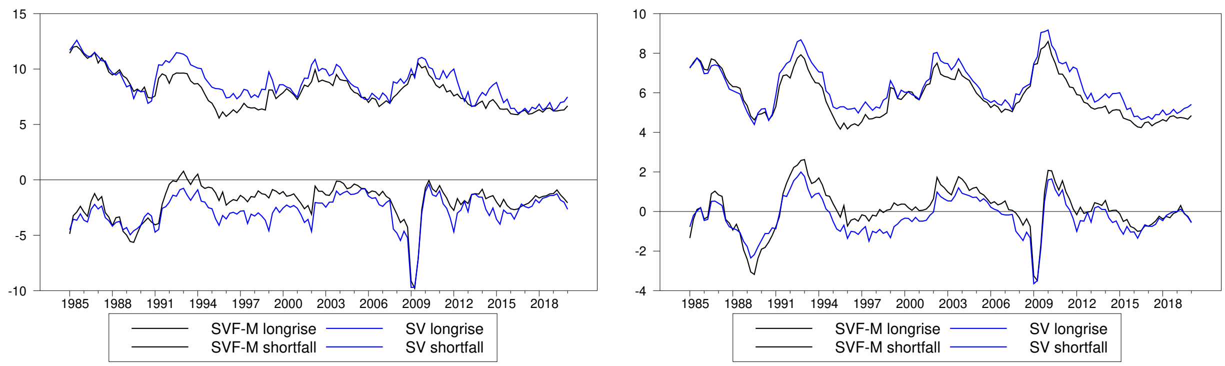
Long-rise and expected shortfall: QR vs. BVAR-SV, 1985-2019

(a) $h = 1$ **(b)** $h = 4$



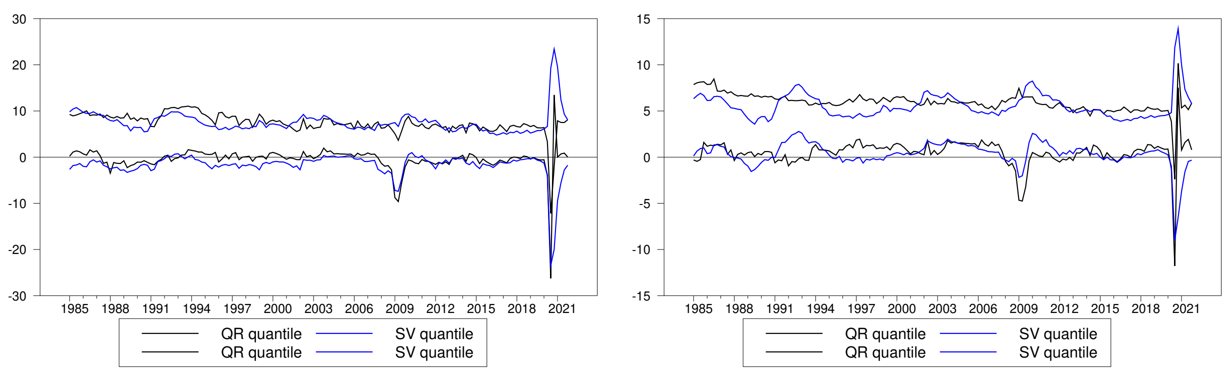
Long-rise and expected shortfall: BVAR-SV vs. BVAR-SVF-M, 1985-201

(c) $h = 1$ **(d)** $h = 4$



Quantiles: QR vs. BVAR-SV, 1985-2021

(e) $h = 1$ **(f)** $h = 4$



episode, financial conditions deteriorated relatively modestly in early 2020 and then recovered quickly. With lagged GDP growth included among each model's predictors, the extreme movements of GDP growth in 2020 drove unprecedented movements in expected shortfall and long-rise.[14] Although not obvious from the chart due to the large axis range, the period displayed the type of asymmetry noted above, with the lower tail quantile falling more than the upper tail quantile changed. For example, in the 1-step-ahead QR forecasts, the 5 percent quantile dropped from -0.57 percent in 2020:Q1 to -3.99 percent in 2020:Q2 and -26.25 percent in 2020:Q3, and then jumped up to 13.48 percent in 2020:Q4. QR forecasts of the 95 percent quantile fell from 6.40 percent in 2020:Q1 to 3.33 percent in 2020:Q2, dropped to -12.19 percent in 2020:Q3, and moved up to -2.09 percent in 2020:Q4 (so the 5 and 95 percent quantiles from QR crossed sharply in this quarter). In the 1-step-ahead BVAR-SV forecasts, the 5 percent quantile forecast declined from -1.28 percent in 2020:Q1 to -3.89 percent in 2020:Q2 and -23.45 percent in 2020:Q3, and then remained steeply negative (-20.00 percent) in 2020:Q4. BVAR-SV forecasts of the 95 percent quantile were roughly 6 percent in both 2020:Q1 and 2020:Q2 and rose to 19.27 percent in 2020:Q3 and 23.41 percent in 2020:Q4.

5.2 Forecast Accuracy: GDP growth

This subsection provides a formal analysis of out-of-sample forecast performance for tail risks to GDP growth, for samples of 1985-2019 and 1985-2021 (see the appendix for in-sample forecast results in the former sample).

Table 3 reports quantile scores and qwCRPS (one version with more weight in the left tail and the other with more weight in the right tail) results for out-of-sample forecasts. Broadly, the BVAR-SV tail-risk forecasts are as accurate or more accurate than the QR forecasts. For the sample ending in 2019, the BVAR-SV's quantile scores for the left tail ($\tau = 0.05$) are essentially the same as the QR's scores for $h = 4$ and moderately worse for $h = 1$, but not statistically different in either case. Using the qwCRPS-left metric that considers the entire density with more weight on the left tail than the right, the BVAR-SV's scores are slightly better than the QR's at both horizons, although not significantly different. For the sample ending in 2021 and thereby covering the period of the pandemic, the BVAR-SV model has a consistent modest advantage over QR in forecasting left-tail risk, but not by a statistically significant margin. Although right-tail risks to output growth have generally not been a focus of the recent literature, the results show that the BVAR-SV model has a more consistent and larger advantage over QR than in the case of left-tail risks, with gains that are statistically significant in a few sample-horizon combinations, and quantitatively large in the pandemic as measured by the 95 percent quantile score. The overall performance of the

[14] Using final vintage estimates for simplicity of illustration, GDP growth registered -5.2, -37.4, and 29.1 percent in 2020:Q1, 2020:Q2, and 2020:Q3, respectively.

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

	1985-2021		**1985-2019**	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	Quantile score: 5 percent quantile			
QR	0.503	0.313	0.168	0.169
BVAR-SV	0.941	0.927	1.135	1.016
BVAR-SVF-M	0.918	0.964	1.006	1.111
	Quantile score: 95 percent quantile			
QR	0.575	0.276	0.267	0.179
BVAR-SV	0.567	0.603	0.913^{**}	0.877^{**}
BVAR-SVF-M	0.638	0.660	0.847^{***}	0.833^{***}
	qwCRPS-left			
QR	0.518	0.371	0.320	0.282
BVAR-SV	0.978	0.950	0.991	0.978
BVAR-SVF-M	0.989	0.965	1.000	1.013
	qwCRPS-right			
QR	0.573	0.374	0.387	0.296
BVAR-SV	0.840	0.820	0.923^{*}	0.903
BVAR-SVF-M	0.871	0.840^{*}	0.933^{*}	0.910

Notes: To facilitate accuracy comparisons the results for the BVAR models are reported as ratios relative to scores for quantile regression (an entry less than 1 means the BVAR is more accurate than QR). Statistical significance of the differences in scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

BVAR-SVF-M model is very similar to that of the BVAR-SV specification, in some cases a little more accurate and in others a little less accurate. So, forecasts from the BVAR-SVF-M model, like those from the BVAR–SV specification, are as accurate or more accurate than the QR forecasts.

What should one take away from these forecast comparisons? As noted above, our paper is one of the few 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 either case, stochastic volatility plays a very important role. 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. (2020), 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 so 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.

5.3 Results for the Unemployment Rate and Inflation

This subsection provides results for forecasts of the change in the unemployment rate and forecasts of inflation, at the 4-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 (2022), and our specification of 4-quarters-ahead inflation is patterned on that of Lopez-Salido and Loria (2021). The forecasts we report for our BVAR-SV and BVAR-SVF-M models are also for the 4-quarters-ahead change in unemployment and for the 4-quarters-ahead inflation rate. In these cases, we use the same model in quarterly rates described above; we transform the quarterly forecasts to obtain the needed 4-quarters-ahead change in unemployment and the 4-quarters-ahead inflation rate.

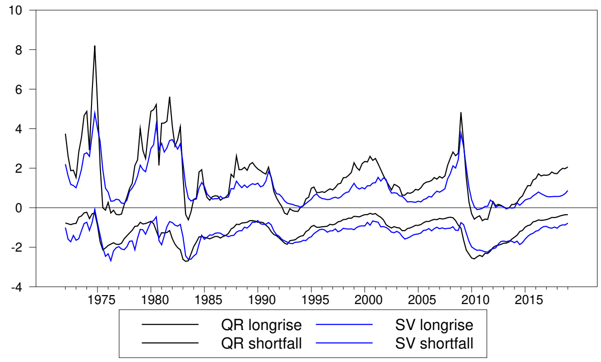
Figure 3 reports time series of expected shortfall (at 5 percent) and long-rise (at 95 percent) estimates at the 4-steps-ahead forecast horizon, for the quantile regression and BVAR-SV specifications (BVAR-SVF-M and BVAR-SV estimates are very similar). The in-sample estimates in the top panel display some notable upside asymmetries. These imply 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, along with upside risks to inflation, particularly in the 1970s and early 1980s. For both unemployment and inflation, the standard deviation of the long-rise estimate is about twice that of the expected shortfall. The contours of the QR and BVAR-SV estimates are quite similar, especially for unemployment rate changes. For inflation, the broad contours of QR and BVAR-SV estimates are similar, although the QR estimate shows more upside asymmetry to inflation in the 1970s and 1980s.

Moving from in-sample (top panel) to out-of-sample forecasts (the middle panel of Figure 3 provides results for 1985-2019) mostly preserves the pattern of asymmetries between

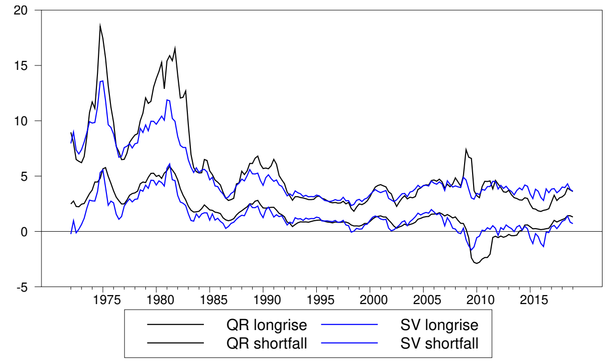
Figure 3: Forecasts of unemployment and inflation, 4-quarters ahead

Long-rise and expected shortfall: In-sample forecasts, 1985-2019

(a) Unemployment change

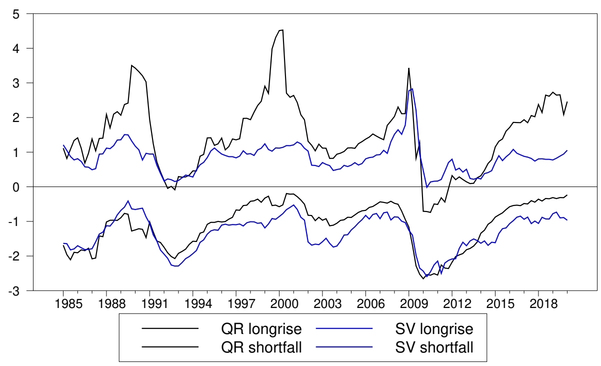


(b) 4-quarter inflation rate

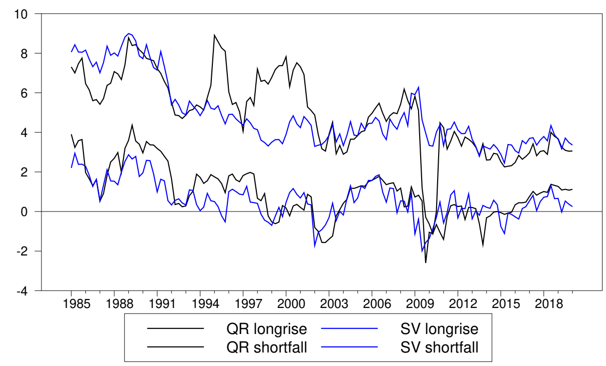


Long-rise and expected shortfall: Out-of-sample forecasts, 1985-2019

(c) Unemployment change

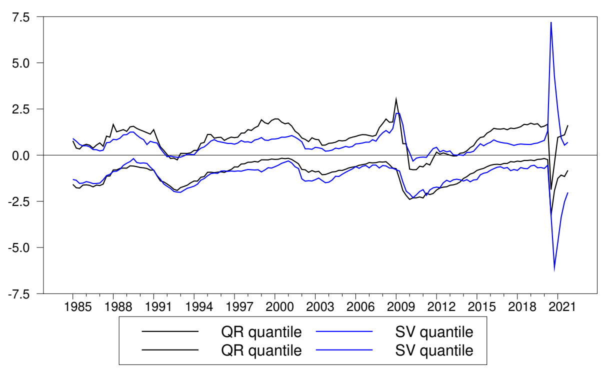


(d) 4-quarter inflation rate

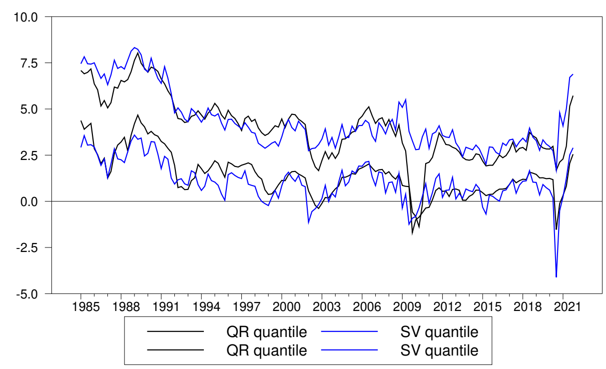


Long-rise and expected shortfall: Out-of-sample forecasts, 1985-2021

(e) Unemployment change



(f) 4-quarter inflation rate



the expected shortfall and long-rise, with the latter more variable than the former for unemployment and inflation. In the QR estimates, long-rise is particularly variable, showing some increases larger than observed in the in-sample estimates. The long-rise estimate for inflation also plummets around the Great Recession. In the BVAR-SV estimates, long-rise is more variable than expected shortfall in the inflation estimates but not the unemployment forecasts. Still, even for unemployment, the BVAR-SV estimates show an asymmetry around the Great Recession, with long-rise moving up sharply relative to expected shortfall. In broad contours, the BVAR-SV and QR estimates move together relatively closely for shortfall and noticeably less so for long-rise, in part reflecting the relatively sharp volatility of long-rise observed for the QR estimates. In the bottom panel's forecasts of 5 and 95 percent quantiles that include the period of the COVID-19 pandemic, in the case of inflation the QR and BVAR-SV estimates display qualitatively similar changes in the pandemic's volatility, with downside risk rising more sharply than upside risk to inflation in the depths of the pandemic. For example, from 2020:Q2 to 2020:Q3 the BVAR-SV estimate of the 5 percent quantile dropped from 0.18 percent to -4.11 percent, while the 95 percent quantile declined from 2.73 percent to 1.72 percent. In the case of unemployment, the QR estimates might be seen as displaying symmetric (albeit different) moves: The forecasts of the tail quantiles fell by similar amounts (both QR forecasts fell by about 3 percentage points in 2020:Q3). The BVAR-SV estimates arguably show a little more asymmetry: From 2020:Q2 to 2020:Q3, the 5 percent quantile estimate fell from -0.55 percent to -3.24 percent (similar to the QR estimates), whereas the 95 percent quantile forecast rose from 1.31 to 7.21 percent (while the QR estimate fell from 1.68 to -1.86 percent).

Table 3: Accuracy of out-of-sample forecasts of unemployment and inflation, 4-quarters ahead

	Unemployment change		Inflation	
	1985-2021	**1985-2019**	**1985-2021**	**1985-2019**
	Quantile score: 5 percent quantile			
QR	0.099	0.049	0.151	0.149
BVAR-SV	0.823	0.995	0.545**	0.518**
BVAR-SVF-M	0.743	0.892	0.757*	0.747*
	Quantile score: 95 percent quantile			
QR	0.195	0.107	0.211	0.165
BVAR-SV	1.253	1.247	0.647	0.697
BVAR-SVF-M	1.207	1.219	0.684	0.707
	qwCRPS-left			
QR	0.152	0.105	0.208	0.199
BVAR-SV	0.936	0.931*	0.886	0.872
BVAR-SVF-M	0.934	0.953	0.972	0.963
	qwCRPS-right			
QR	0.200	0.139	0.246	0.222
BVAR-SV	1.038	0.983	0.835	0.844
BVAR-SVF-M	1.030	1.003	0.878	0.879

Notes: To facilitate accuracy comparisons the results for the BVAR models are reported as ratios relative to scores for quantile regression (an entry less than 1 means the BVAR is more accurate than QR). Statistical significance of the differences in scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table 3 reports quantile scores and qwCRPS results for out-of-sample forecasts (see the appendix for in-sample forecasts). In the case of inflation, the BVAR-SV scores are consistently better than the QR scores, by relatively large quantitative margins, in both the left and right tails. The advantages of the BVAR-SV model are statistically significant by the 5 percent quantile score but not statistically significant in other cases. For inflation, results are very similar for the longer sample including the pandemic as for the 1985-2019 sample. In the case of the change in the unemployment rate, the QR and BVAR-SV forecasts are broadly comparable in accuracy. In the left tail, the BVAR-SV model has a small advantage over QR. But in the right tail, the BVAR-SV model is less accurate than QR by the 95 percent quantile score and about the same in accuracy by the qwCRPS-right score. For unemployment, as for inflation, results are very similar for the longer sample including the pandemic as for the 1985-2019 sample. Overall, for unemployment and for inflation, the BVAR-SV model can be seen to be at least as accurate as QR for forecasting tail risks. Once again, within the BVAR class, the BVAR-SVF-M specification performs similarly to the BVAR-SV model, sometimes offering modest improvements and other times being not quite as good.

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 some 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.

In this paper we examine the ability of BVARs with stochastic volatility to capture tail risks in macroeconomic forecast distributions and outcomes, for GDP growth, unemployment, and inflation. A conventional BVAR-SV formulation is capable of 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. We also consider a model extended to feature a common volatility factor that enters the conditional mean of the BVAR, so as to allow a contemporaneous correlation between shocks to the levels and volatilities of variables and thereby be capable of producing conditional predictive distributions with asymmetries. 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 the quantile-weighted continuous ranked probability score.

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 and inflation) in line with the results of quantile regression, and in formal evaluation of tail risk forecasts, our BVAR specifications perform comparably to quantile regression. Our findings on the effectiveness of BVARs in capturing tail risks apply with both the conventional BVAR with stochastic volatility and the model extended to feature a common volatility factor that enters the conditional mean of the BVAR. These models appear to be equally well suited to capturing asymmetries in the unconditional predictive distribution of GDP growth, unemployment, and inflation. The BVAR-SV model 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 immediately affects the levels of the macroeconomic variables.

Our findings imply that, at least for GDP growth, unemployment, and inflation, quantile regression doesn't seem to offer any advantages in forecast accuracy over a BVAR-SV or BVAR-SVF-M specification. 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. Based on these results, we do not mean to claim that, in truth, there are no asymmetries (possibly time-varying) in conditional predictive distributions. Rather, one aspect of tail risk that has received some emphasis in the literature — downside risks varying more over time than upside risks for output growth — can be captured as well with a BVAR with conventional stochastic volatility that yields symmetric conditional distributions as with other models that allow asymmetries in conditional distributions. In addition, in formal metrics of tail forecasts, the conventional BVAR is comparable in accuracy to the specifications that allow asymmetries in conditional distributions. We take this as suggestive evidence that conditional asymmetries are not necessarily a strong, regular feature of predictive distributions for output growth, unemployment, and inflation, in keep with the cautionary findings of Plagborg-Moller, et al. (2020) for output growth. But as we noted earlier, other methods or analyses may reach a different conclusion, and we leave to further research whether some of these other methods under development can establish gains over formulations of BVARs with stochastic volatility.

References

Adams, Patrick A., Tobias Adrian, Nina Boyarchenko, and Domenico Giannone (2021), "Forecasting Macroeconomic Risks," *International Journal of Forecasting* 37, 1173-1191.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019), "Vulnerable Growth," *American Economic Review* 109, 1263-1289.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2021), "Multimodality in Macrofinancial Dynamics," *International Economic Review* 62, 861-886.

Adrian, Tobias, Federico Grinberg, Nellie Liang, and Sheheryar Malik (2020), "The Term Structure of Growth-at-Risk," *American Economic Journal: Macroeconomics*, forthcoming.

Alessandri, Piergiorgio, and Haroon Mumtaz (2017), "Financial Conditions and Density Forecasts for US Output and Inflation," *Review of Economic Dynamics* 24, 66-78.

Amburgey, Aaron, and Michael W. McCracken (2022), "On the Real-Time Predictive Content of Financial Conditions Indices for Growth," Federal Reserve Bank of St. Louis Working Paper 2022-003.

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.

Bai, Jushan, and Serena Ng (2001), "A Consistent Test for Conditional Symmetry in Time Series Models," *Journal of Econometrics* 103, 225-258.

Brownlees, Christian, and Andre B.M. Souza (2021), "Backtesting Global Growth-at-Risk," *Journal of Monetary Economics* 118, 312-330.

Caldara, Dario, Danilo Cascaldi-Garcia, Pablo Cuba-Borda, and Francesca Loria (2021), "Understanding Growth-at-Risk: A Markov Switching Approach," manuscript.

Caldara, Dario, Chiara Scotti, and Molin Zhong (2021), "Macroeconomic and Financial Risks: A Tale of Mean and Volatility," International Finance Discussion Papers 1326, Board of Governors of the Federal Reserve System.

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.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2020), "Nowcasting Tail Risks to Economic Activity with Many Indicators," *Journal of Applied Econometrics*, forthcoming.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2020b), "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.

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2022), "Addressing COVID-19 Outliers in BVARs with Stochastic Volatility," *Review of Economics and Statistics*, forthcoming.

Clark, Todd E. (2011), "Real-Time Density Forecasts from BVARs with Stochastic Volatility," *Journal of Business and Economic Statistics* 29, 327-341.

Clark, Todd E., and Francesco Ravazzolo (2015), "Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics* 30, 551-575.

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.

Croushore, Dean, and Tom Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-130.

Delle Monache, Davide, Andrea De Polis, and Ivan Petrella (2020), "Modeling and Forecasting Macroeconomic Downside Risk," EMF Research Papers 34, Economic Modelling and Forecasting Group.

De Nicolo, Gianni, and Marcella Lucchetta (2017), "Forecasting Tail Risks," *Journal of Applied Econometrics* 32, 159-170.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

Engle, Robert F., and Simone Manganelli (2004), "CAViaR," *Journal of Business and Economic Statistics*, 22, 367-381.

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.

Fissler, Tobias, and Johanna F. Ziegel (2016), "Higher Order Elicitability and Osband's Principle," *Annals of Statistics* 44, 1680-1707.

Gaglianone, Wagner P., and Luiz Renato Lima (2012), "Constructing Density Forecasts from Quantile Regressions," *Journal of Money, Credit and Banking* 44, 1589-1607.

Galbraith, John W., and Simon van Norden (2019), "Asymmetry in Unemployment Rate Forecast Errors," *International Journal of Forecasting* 35, 1613-1626.

Ganics, Gergely, Tatevik Sekhposyan, and Barbara Rossi (2021), ""From Fixed-event to Fixed-horizon Density Forecasts: Obtaining Measures of Multi-horizon Uncertainty from Survey Density Forecasts," *Journal of Money, Credit, and Banking*, forthcoming.

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

Giacomini, Raffaella, and Ivana Komunjer (2005), "Evaluation and Combination of Conditional Quantile Forecasts," *Journal of Business and Economic Statistics* 23, 416-431.

Giglio, Stefano, Bryan Kelly, and Seth Pruitt (2016), "Systemic Risk and the Macroeconomy: An Empirical Evaluation," *Journal of Financial Economics* 119, 457-471.

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.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng (2015),"Measuring Uncertainty," *American Economic Review* 105, 1177-1216.

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.

Kiley, Michael T. (2022), "Unemployment Risk," *Journal of Money, Credit and Banking*, forthcoming.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies* 65, 361-393.

Korobilis, Dimitris (2017), "Quantile Regression Forecasts of Inflation under Model Uncertainty," *International Journal of Forecasting* 33, 11-20.

Lhuissier, Stephane (2022), "Financial Conditions and Macroeconomic Downside Risks in the Euro Area," *European Economic Review* 143,

Loaiza-Maya, Ruben, and Michael Stanley Smith (2019), "Real-Time Macroeconomic Forecasting With a Heteroskedastic Inversion Copula," *Journal of Business and Economic Statistics*.

Lopez-Salido, J. David, and Francesca Loria (2020), "Inflation at Risk," Working Paper 2020-013, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System.

Manzan, Sebastiano (2015), "Forecasting the Distribution of Economic Variables in a Data-Rich Environment," *Journal of Business and Economic Statistics* 33, 144-164.

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.

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.

Mitchell, James, Aubrey Poon, and Gian Luigi Mazzi (2022), "Nowcasting Euro Area GDP Growth Using Bayesian Quantile Regression," in *Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling*, *Advances in Econometrics*, vol. 43A (Alexander Chudik, Cheng Hsiao, and Allan Timmermann, eds.), 51-72. Emerald Publishing Limited, Bingley.

Mitchell, James, Aubrey Poon, and Dan Zhu (2022), "Multimodality in Macroeconomic

Dynamics: Constructing Density Forecasts from Quantile Regressions," manuscript.

Morley, James and Jeremy Piger (2012), "The Asymmetric Business Cycle, *Review of Economics and Statistics* 94, 208-221.

Plagborg-Moller, Mikkel, Lucrezia Reichlin, Giovanni Ricco, and Thomas Hasenzagl (2020), "When Is Growth at Risk?" *Brookings Papers on Economic Activity*, Spring, 167-229.

Primiceri, Giorgio E. (2005), "Time-Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies* 72, 821-852.

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.

West, Kenneth D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica* 64, 1067-1084.

Table 4: Small GDP application

Metric	QR	$QR-ridge$	BQR $Minn$	BQR $est.Minn$	BQR HS	BQR $ENET$
			h=1			
0.05 QS	0.252	1.057	1.005	0.983	0.998	0.984
0.10 QS	0.385	1.058	1.032	0.993	1.003	0.995
0.20 QS	0.584	0.994	0.981	0.990	0.980^{**}	0.981^{**}
0.30 QS	0.723	0.964	0.959	0.967^{**}	0.964^{***}	0.951^{***}
0.40 QS	0.806	0.939^{**}	0.930^{***}	0.954^{***}	0.963^{***}	0.938^{***}
0.50 QS	0.830	0.919^{***}	0.926^{***}	0.965^{***}	0.975^{***}	0.948^{***}
0.60 QS	0.823	0.911^{***}	0.926^{***}	0.969^{**}	0.978^{**}	0.952^{***}
0.70 QS	0.779	0.935^{**}	0.938^{***}	0.978^{*}	0.978^{*}	0.957^{***}
0.80 QS	0.662	0.964^{**}	0.976^{*}	0.996	0.982^{**}	0.965^{***}
0.90 QS	0.429	0.975^{**}	0.986	1.004	0.985^{*}	0.969^{***}
0.95 QS	0.268	0.962	0.957	0.972	0.959^{*}	0.949^{***}
CRPS center	0.264	0.945^{**}	0.946^{***}	0.973^{***}	0.975^{***}	0.955^{***}
CRPS left	0.393	0.973	0.965^{**}	0.975^{***}	0.977^{***}	0.963^{***}
CRPS right	0.417	0.946^{***}	0.953^{***}	0.982^{**}	0.978^{***}	0.958^{***}
			h=4			
0.05 QS	0.199	1.111	1.100	1.023	0.993	0.994
0.10 QS	0.306	1.053	1.066	1.026	0.996	0.992
0.20 QS	0.473	0.927	0.941	0.936^{*}	0.950^{*}	0.931^{*}
0.30 QS	0.569	0.897	0.904	0.952^{**}	0.974^{***}	0.953^{**}
0.40 QS	0.631	0.903^{**}	0.907^{*}	0.956^{**}	0.969^{**}	0.950^{***}
0.50 QS	0.657	0.916^{***}	0.909^{***}	0.953^{**}	0.963^{***}	0.949^{***}
0.60 QS	0.635	0.941^{***}	0.922^{***}	0.965^{***}	0.971^{***}	0.957^{***}
0.70 QS	0.574	0.948^{*}	0.931^{**}	0.964^{**}	0.967^{**}	0.956^{***}
0.80 QS	0.452	0.966^{**}	0.969^{**}	0.988	0.984^{*}	0.975^{***}
0.90 QS	0.295	1.043	1.019	1.016	1.008	0.983
0.95 QS	0.183	1.082	1.068	1.039	1.020	0.993
CRPS center	0.202	0.933^{*}	0.928^{*}	0.962^{**}	0.970^{***}	0.955^{***}
CRPS left	0.310	0.941	0.945	0.964^{**}	0.970^{**}	0.955^{**}
CRPS right	0.306	0.958^{*}	0.948^{*}	0.975^{*}	0.978^{**}	0.963^{***}
			h=12			
0.05 QS	0.269	1.067	1.060	1.099	1.070	1.045
0.10 QS	0.416	1.028	1.009	1.035	1.019	1.001
0.20 QS	0.543	1.000	0.992	1.003	1.037	1.028
0.30 QS	0.640	0.962^{*}	0.954^{**}	0.964^{**}	0.994	1.012
0.40 QS	0.678	0.979	0.954^{***}	0.968^{**}	1.001	1.021
0.50 QS	0.674	0.989	0.988	0.996	1.018	1.029
0.60 QS	0.680	0.967	0.942	0.963	0.984	1.001
0.70 QS	0.573	0.962	0.957	0.961	0.992	1.027
0.80 QS	0.429	0.993	0.992	0.999	1.000	0.993
0.90 QS	0.259	0.979	0.984	0.986	0.990	0.982
0.95 QS	0.140	1.123	1.090	1.032	1.001	0.983
CRPS center	0.214	0.979	0.967	0.979	1.003	1.015
CRPS left	0.353	0.991	0.977	0.992	1.012	1.016
CRPS right	0.304	0.977	0.970	0.979	0.997	1.008

Note: Comparison of tail forecast accuracy of QR-ridge and BQR specifications relative to QR (baseline, in denominator of relative comparisons). Score levels given in QR columns. For specifications other than QR, values below 1 indicate improvement over baseline. Evaluation window from 1985:Q1 through 2019:Q4.

Supplemental Appendix to "Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions'

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

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

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

May 2022

Abstract

This appendix first provides some supplemental results, including: skewness and kurtosis estimates; additional comparisons of tail risk forecasts, ranging from the relative volatilities of expected shortfall and long-rise to accuracy assessments of in-sample forecasts; assessments of tail risk using 10 and 90 percent quantiles; results for some additional forecast metrics; evaluations of forecast accuracy comparing Bayesian quantile regression to the QR and BVAR methods in the paper; and a robustness check of tail risks to GDP growth using an alternative measure of financial conditions. It then details the Monte Carlo experiments summarized in the paper.

Contents

List of Tables

List of Figures

1 Skewness and Kurtosis Results

Some recent research has considered simple evidence and concluded that GDP growth outcomes feature skewness (e.g., Jensen, et al. (2020), 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.[1] 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. In this skewness assessment, we focus on samples ending in 2019:Q4 to avoid the extreme volatility in some variables that resulted from the COVID-19 pandemic, but we include a table of results using data through 2021:Q4 (which show that the pandemic caused some skewness and kurtosis estimates to soar). The residuals used in the estimates presented 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.[2] 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 means, at forecast horizons of one and four quarters.

In Table A3'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 -0.369, 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 modestly more evidence of kurtosis than skewness, with rejections of no-kurtosis for GDP growth, unemployment, and the NFCI. When the BVAR residuals are normalized by their time-varying volatilities, the evidence of kurtosis declines, and the null of no-skewness is rejected for the unemployment rate and NFCI. As indicated in the bottom two panels of the table, the evidence of 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 described in the paper). But the Bai-Ng test for conditional symmetry provides some evidence for asymmetries in forecast errors for a few variables.

When we shorten the sample to 1985-2019 (but without changing the 1972-2019 sample of model estimates used to obtain the residuals), it remains the case (see estimates in A3) that the normalized residuals show significant skewness for the unemployment rate and NFCI, although the

[1] 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 BVAR residuals or forecast errors, we use the residuals or errors in question without further transformation.

[2] 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.

evidence of significant skewness or kurtosis in the data and raw residuals is a little weaker than in the 1972-2019 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.

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, reported in Figures A1 and A2, compare the empirical quantiles of the residuals or forecast errors with quantiles of the normal distribution. These results, too, end the sample in 2019:Q4 to avoid COVID-19 distortions. The results for residuals (results for out-of-sample forecast errors are similar) display some notable departures from normality, most dramatically for the federal funds rate and the NFCI. Normalizing the residuals by their volatilities helps move the empirical quantiles toward the normal case, although still with some departures.

Overall, these results indicate 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, BVAR residuals, or forecast errors). However, it is entirely possible that this subsection's standard tests of skewness over full samples of data will not capture important asymmetries that emerge just at certain points in time and are not evident on average. The paper's results on tail risks may pick up features of interest.

Table A1: Skewness and kurtosis statistics, data and BVAR-SV residuals, through 2019:Q4

	skewness	kurtosis	Bai-Ng skewness	Bai-Ng kurtosis	Bai-Ng normality	conditional symmetry
Data, 1972-2019						
GDP growth	−0.369	5.706	−0.845	2.141**	5.296*	0.792
Unemployment	0.634	2.750	0.578	−0.066	0.339	1.910*
GDP inflation	1.428	4.726	1.971**	1.077	5.045*	0.918
Fed funds rate	0.759	3.455	0.846	0.235	0.770	1.783
NFCI	2.002	6.774	2.045**	1.990**	8.141**	4.980***
BVAR-SV residuals, 1972-2019						
GDP growth	0.139	6.629	0.244	1.651*	2.787	1.940*
Unemployment	0.769	7.014	1.243	1.821*	4.859*	1.574
GDP inflation	0.470	5.011	1.175	1.644	4.085	1.721
Fed funds rate	1.380	23.715	0.625	1.338	2.181	2.235**
NFCI	1.582	10.414	1.516	1.652*	5.028*	6.383***
BVAR-SV residuals normalized by SV, 1972-2019						
GDP growth	0.110	2.878	0.640	−0.280	0.488	1.378
Unemployment	0.275	2.860	2.233**	−0.381	5.133*	1.186
GDP inflation	0.055	2.622	0.449	−1.282	1.846	1.746
Fed funds rate	−0.165	2.809	−1.120	−0.538	1.545	2.070*
NFCI	0.236	2.401	1.879*	−1.832*	6.887**	1.934*
BVAR-SV forecast errors, horizon = 1 quarter, 1985-2019						
GDP growth	−0.213	3.382	−0.631	0.495	0.643	0.840
Unemployment	0.951	4.669	1.594	1.432	4.592	2.665**
GDP inflation	−0.339	3.059	−1.782*	0.118	3.189	1.057
Fed funds rate	−0.047	4.619	−0.098	1.266	1.613	2.380**
NFCI	2.654	22.984	0.988	1.170	2.345	2.195*
BVAR-SV forecast errors, horizon = 4 quarters, 1985-2019						
GDP growth	−0.652	4.794	−0.807	0.825	1.332	1.456
Unemployment	2.162	9.247	0.734	0.720	1.057	1.164
GDP inflation	−0.626	3.052	−2.170**	0.084	4.717*	1.897
Fed funds rate	−0.115	2.970	−0.439	−0.033	0.194	1.812
NFCI	1.390	9.648	0.704	0.803	1.141	2.455**

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

Table A2: Skewness and kurtosis statistics, data and BVAR-SV residuals, through 2021:Q4

	skewness	kurtosis	Bai-Ng skewness	Bai-Ng kurtosis	Bai-Ng normality	conditional symmetry
Data, 1972-2021						
GDP growth	−2.490	36.130	−1.071	1.081	2.315	0.799
Unemployment	0.826	3.609	1.440	0.308	2.167	5.103***
GDP inflation	1.331	4.576	1.987**	1.008	4.962*	1.485
Fed funds rate	0.780	3.424	0.770	0.118	0.607	1.687
NFCI	2.060	7.057	1.989**	2.120**	8.453**	5.020***
BVAR-SV residuals, 1972-2021						
GDP growth	−1.709	45.285	−0.845	1.079	1.877	1.355
Unemployment	6.170	101.355	1.035	1.093	2.266	1.606
GDP inflation	0.459	5.417	1.433	1.918*	5.732*	1.880
Fed funds rate	0.974	22.560	0.484	1.385	2.152	3.626***
NFCI	1.436	10.346	1.421	1.636	4.696*	5.869***
BVAR-SV residuals normalized by SV, 1972-2021						
GDP growth	0.064	2.451	0.554	−1.767*	3.430	1.841
Unemployment	0.255	2.855	1.686*	−0.312	2.941	0.808
GDP inflation	0.095	2.590	0.807	−1.430	2.695	1.084
Fed funds rate	−0.195	2.506	−1.452	−1.667*	4.888*	2.126*
NFCI	0.305	2.426	2.405**	−1.762*	8.888**	2.294**
BVAR-SV forecast errors, horizon = 1 quarter, 1985-2021						
GDP growth	−2.430	56.425	−0.947	1.100	2.107	1.113
Unemployment	5.185	75.832	1.073	1.136	2.441	2.737**
GDP inflation	0.411	5.449	0.813	1.180	2.054	1.084
Fed funds rate	0.623	6.780	0.923	1.529	3.191	4.587***
NFCI	2.552	22.066	1.003	1.157	2.343	2.391**
BVAR-SV forecast errors, horizon = 4 quarters, 1985-2021						
GDP growth	−0.772	14.645	−0.481	1.596	2.780	2.024*
Unemployment	0.543	27.850	0.300	1.713*	3.023	4.565***
GDP inflation	1.126	9.213	0.863	0.919	1.590	1.349
Fed funds rate	0.363	4.460	0.793	0.835	1.325	2.803***
NFCI	1.345	9.603	0.725	0.856	1.258	1.835

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

Table A3: Skewness and kurtosis statistics, data and BVAR-SV residuals, 1985-2019

	skewness	kurtosis	Bai-Ng skewness	Bai-Ng kurtosis	Bai-Ng normality	Bai-Ng condit. symmetry
Data, 1985-2019						
GDP growth	-1.294	7.839	-1.126	1.148	2.584	1.420
Unemployment	0.904	3.282	0.622	0.078	0.392	1.241
GDP inflation	0.046	3.655	0.162	1.295	1.703	1.087
Fed funds rate	0.284	1.921	0.138	-0.548	0.319	1.495
NFCI	2.778	14.740	0.740	0.805	1.195	3.928 ***
BVAR-SV residuals, 1985-2019						
	skewness	kurtosis	Bai-Ng skewness	Bai-Ng kurtosis	Bai-Ng normality	Bai-Ng condit. symmetry
GDP growth	-0.920	6.572	-1.270	1.300	3.303	1.704
Unemployment	1.266	5.925	1.291	1.107	2.893	2.607 **
GDP inflation	0.065	3.561	0.272	1.097	1.276	1.015
Fed funds rate	-0.705	4.795	-2.540 **	2.247 **	11.502 ***	1.944 *
NFCI	5.006	40.652	1.214	1.192	2.894	5.494 ***
BVAR-SV residuals normalized by SV, 1985-2019						
GDP growth	0.054	2.797	0.359	-0.572	0.456	1.579
Unemployment	0.271	2.590	2.105 **	-1.094	5.628 *	1.018
GDP inflation	0.103	2.764	0.606	-0.663	0.806	1.690
Fed funds rate	-0.191	2.705	-1.400	-0.750	2.523	1.498
NFCI	0.388	2.716	2.523 **	-0.666	6.807 **	2.222 **

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.

Figure A1: Q-Q plots of BVAR-SV residuals and SV-normalized residuals, 1972-2019 sample

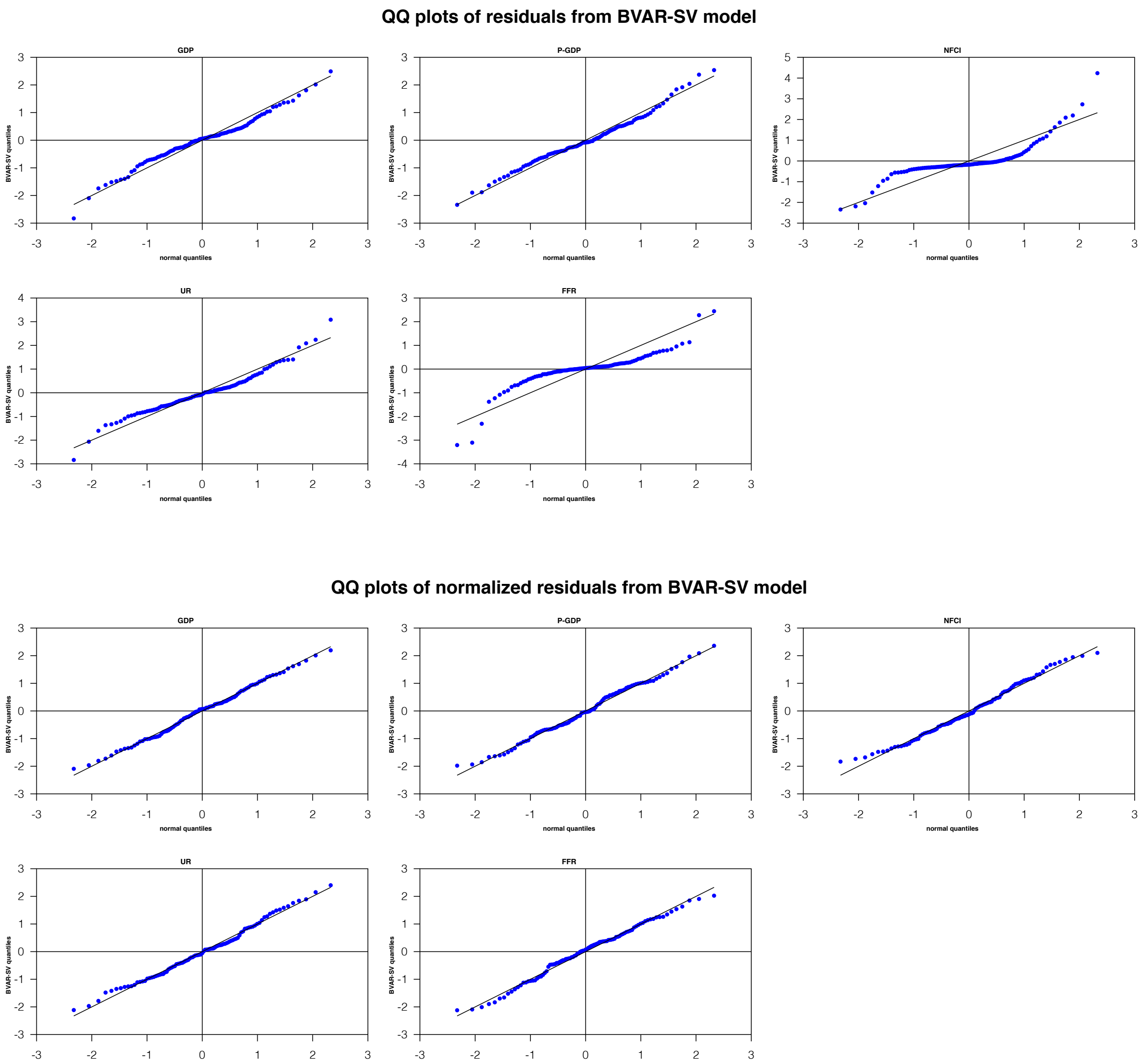
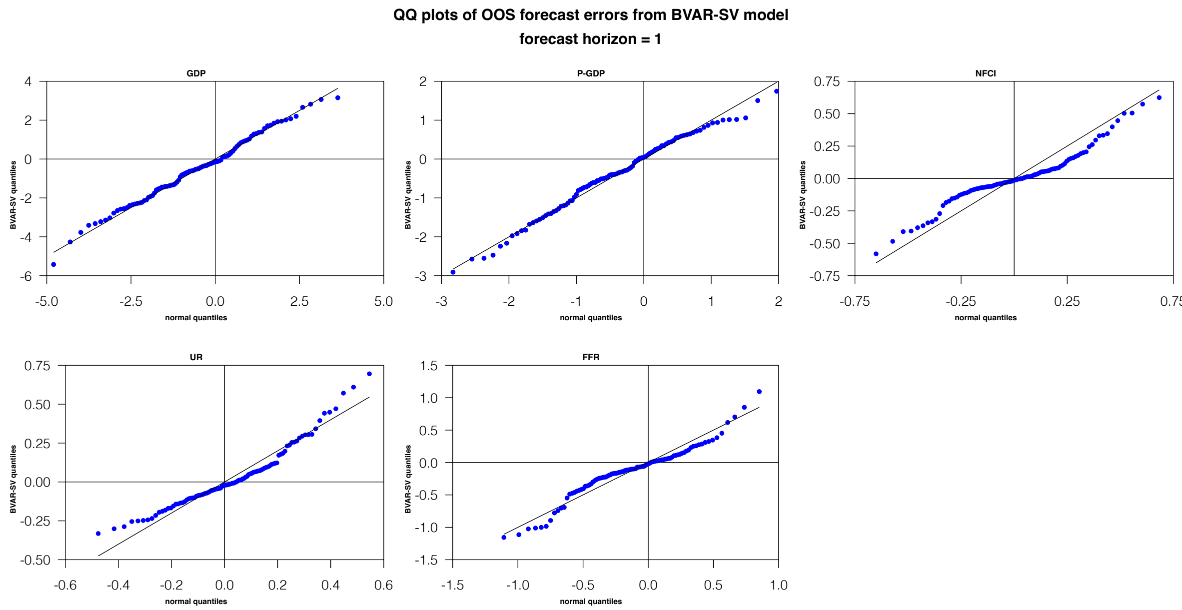
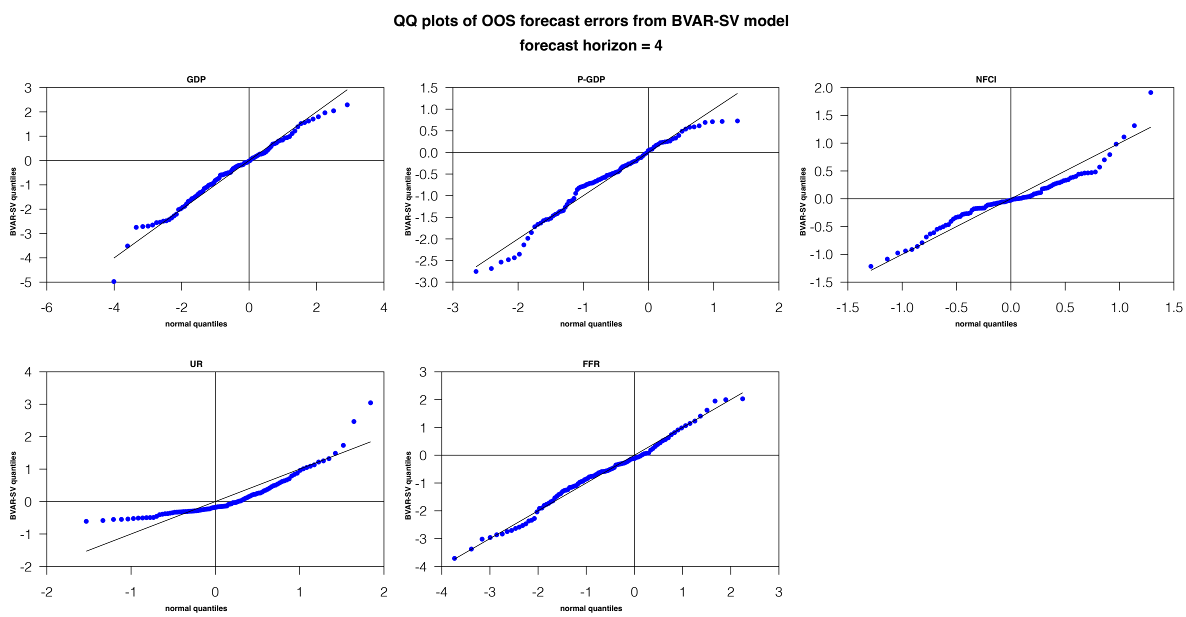


Figure A2: Q-Q plots of BVAR-SV OOS forecast errors, 1985-2019 sample





2 Additional Comparisons

This section provides the following additional comparisons: relative volatilities; in-sample versus out-of-sample forecasts of expected shortfall and long-rise for GDP growth from the QR, BVAR-SV, and BVAR-SVF-M specifications; in-sample and out-of-sample forecasts of expected shortfall and long-rise for the NFCI obtained with the BVAR-SV model; and the baseline accuracy comparisons for in-sample forecasts.

Table A4: Relative volatilities (ratio ES/LR) of expected shortfall and long-rise

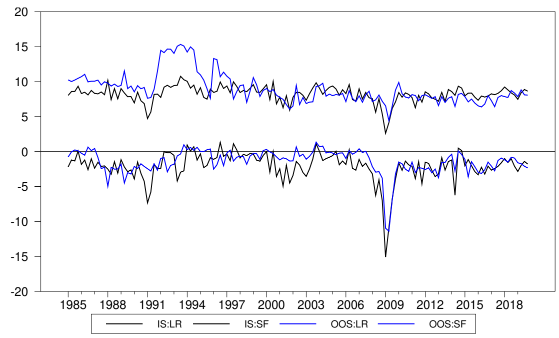
In-sample forecasts, 1972:Q1-2019:Q4				
	GDP $h = 1Q$	GDP $h = 4Q$	Unemployment $h = 4Q$	Inflation $h = 4Q$
Quantile regression	2.216	1.663	0.412	0.466
BVAR-SV	1.626	1.558	0.489	0.556
BVAR-SVF-M	1.417	1.465	0.405	0.461
Out-of-sample forecasts, 1985:Q1-2020:Q1				
	GDP $h = 1Q$	GDP $h = 4Q$	Unemployment $h = 4Q$	Inflation $h = 4Q$
Quantile regression	0.817	1.390	0.623	0.649
BVAR-SV	0.926	0.842	1.151	0.576
BVAR-SVF-M	1.035	1.003	0.929	0.503

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.

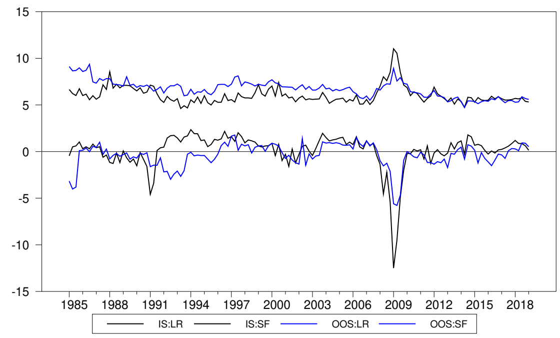
Figure A3: Long-rise and expected shortfall, in-sample (black lines) vs. out-of-sample (blue lines) forecasts of GDP growth for 1985-2019

QR

(a) $h = 1$

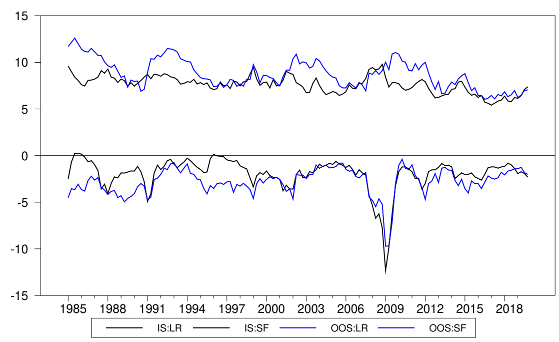


(b) $h = 4$

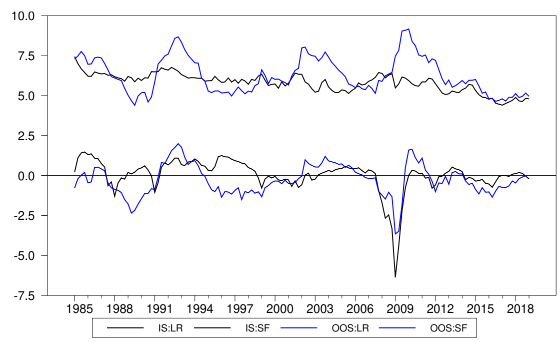


BVAR-SV

(c) $h = 1$

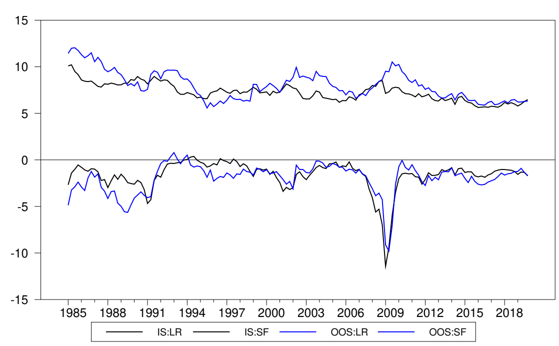


(d) $h = 4$



BVAR-SVF-M

(e) $h = 1$



(f) $h = 4$

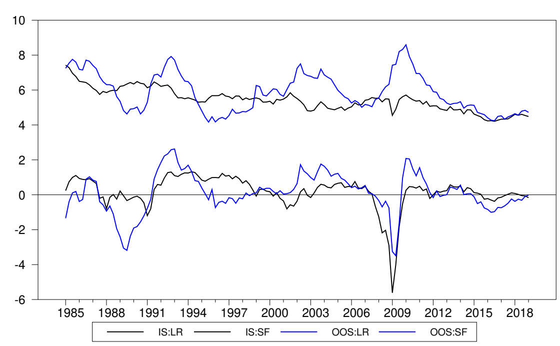
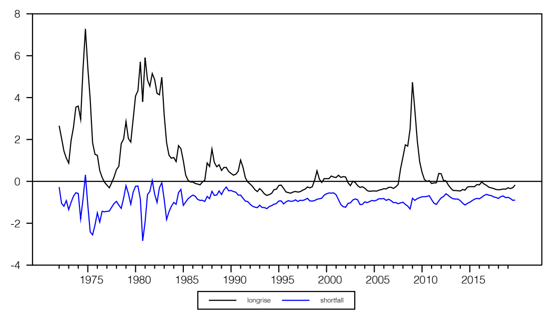


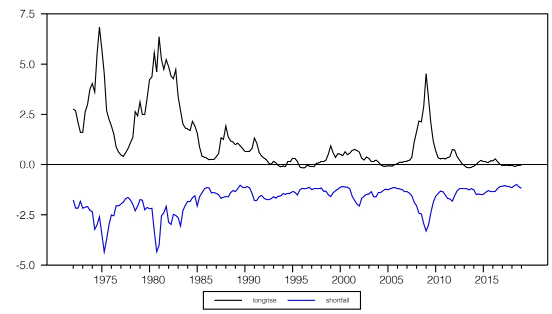
Figure A4: Long-rise and expected shortfall, in-sample and out-of-sample forecasts of NFCI from the BVAR-SV specification

In-sample forecasts

(a) $h = 1$

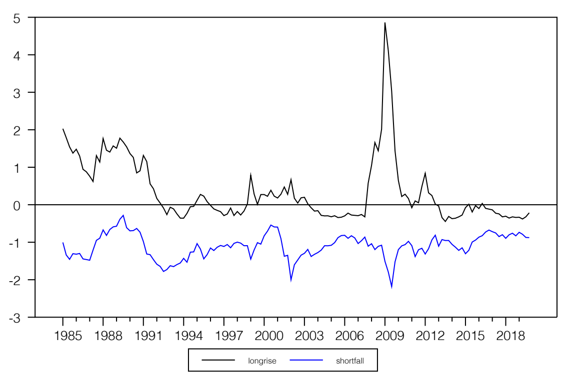


(b) $h = 4$



Out-of-sample forecasts

(c) $h = 1$



(d) $h = 4$

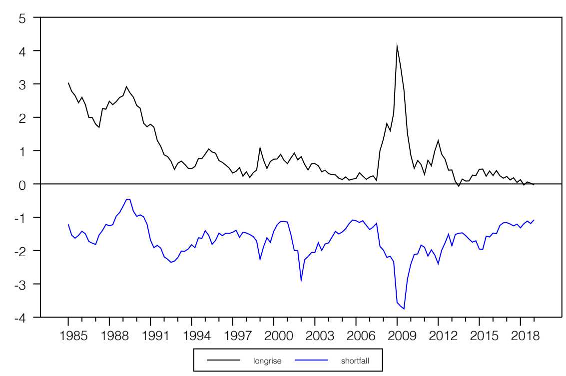


Table A5: Accuracy of in-sample forecasts of GDP growth

	1985-2021		1985-2019	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	Quantile score: 5 percent quantile			
QR	0.492	0.282	0.218	0.153
BVAR-SV	0.890	0.969	0.996	1.231
BVAR-SVF-M	0.780	0.724	0.973	1.148
	Quantile score: 95 percent quantile			
QR	0.382	0.202	0.214	0.141
BVAR-SV	0.786**	0.782*	0.946	0.887*
BVAR-SVF-M	0.677	0.728*	0.941	0.833*
	qwCRPS-left			
QR	0.524	0.328	0.347	0.248
BVAR-SV	0.963	0.979	0.971	1.021
BVAR-SVF-M	0.931	0.905	0.966	0.995
	qwCRPS-right			
QR	0.499	0.296	0.345	0.234
BVAR-SV	0.901*	0.917**	0.952**	0.951
BVAR-SVF-M	0.890	0.898**	0.962*	0.942

Notes: Results for 1985-2021 use models estimated with a data sample of 1972:Q1-2021:Q4. Results for 1985-2019 use models estimated with a data sample of 1972:Q1-2019:Q4. To facilitate accuracy comparisons the results for the BVAR models are reported as ratios relative to scores for quantile regression (an entry less than 1 means the BVAR is more accurate than QR). Statistical significance of the differences in scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table A6: Accuracy of in-sample forecasts of unemployment and inflation, 4-quarters ahead

	Unemployment change		**Inflation**	
	1985-2021	**1985-2019**	**1985-2021**	**1985-2019**
	Quantile score: 5 percent quantile			
QR	0.091	0.044	0.064	0.058
BVAR-SV	0.930	1.065	1.037	1.028
BVAR-SVF-M	0.648	0.947	1.004	0.986
	Quantile score: 95 percent quantile			
QR	0.160	0.082	0.097	0.065
BVAR-SV	1.191	1.185	0.826	0.952
BVAR-SVF-M	0.904	0.979	0.741	0.980
	qwCRPS-left			
QR	0.141	0.093	0.122	0.113
BVAR-SV	0.926	0.895**	0.934	0.870
BVAR-SVF-M	0.856	0.854	0.937	0.870
	qwCRPS-right			
QR	0.176	0.116	0.143	0.122
BVAR-SV	0.987	0.912	0.868	0.835
BVAR-SVF-M	0.899*	0.846*	0.842	0.841

Notes: Results for 1985-2021 use models estimated with a data sample of 1972:Q1-2021:Q4. Results for 1985-2019 use models estimated with a data sample of 1972:Q1-2019:Q4. To facilitate accuracy comparisons the results for the BVAR models are reported as ratios relative to scores for quantile regression (an entry less than 1 means the BVAR is more accurate than QR). Statistical significance of the differences in scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

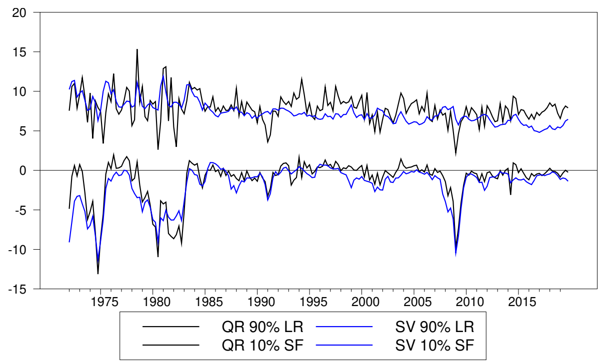
3 Results for 10 and 90 Percent Quantiles

Using 10 and 90 percent quantiles (rather than the paper's 5 and 95 percent quantiles), this section reports forecasts of expected shortfall and long-rise for GDP growth and forecast accuracy comparisons for GDP growth, unemployment, and inflation. Accuracy results include both in-sample and out-of-sample forecasts. Note that, in estimating the expected shortfall and long-rise for quantile regression at the 10 and 90 percent quantiles, in the second step smoothing step of ABG applied, we use the 10 and 90 percent quantiles as moments, rather than the 5 and 95 percent quantiles.

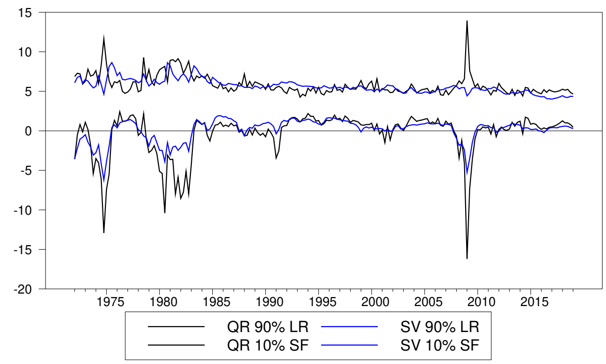
Figure A5: Long-rise and expected shortfall using 10 and 90 percent quantiles, respectively, in-sample forecasts of GDP growth for 1972-2019

QR vs. BVAR-SV

(a) $h = 1$

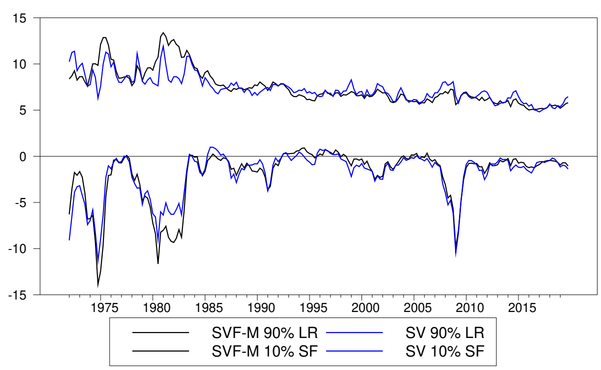


(b) $h = 4$



BVAR-SV vs. BVAR-SVF-M

(c) $h = 1$



(d) $h = 4$

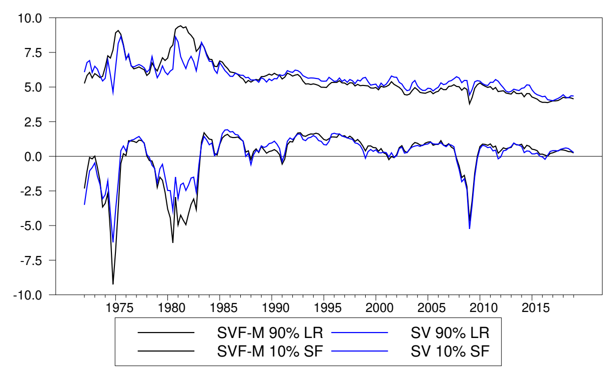
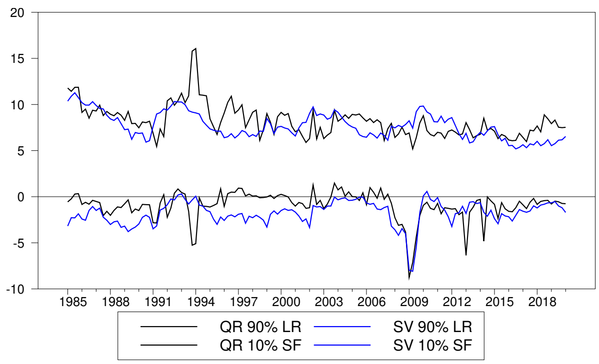


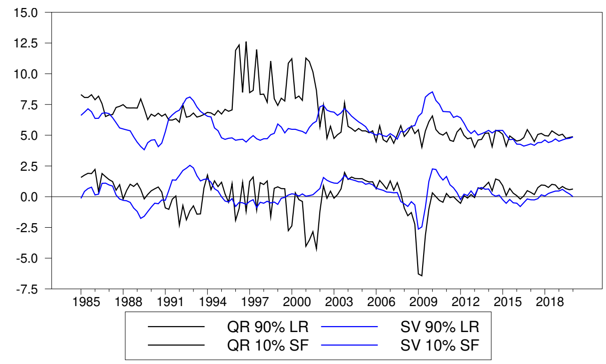
Figure A6: Long-rise and expected shortfall using 10 and 90 percent quantiles, respectively, out-of-sample forecasts of GDP growth for 1985-2019

QR vs. BVAR-SV

(a) $h = 1$

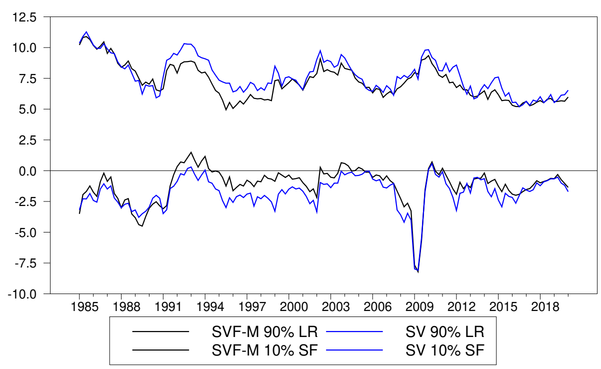


(b) $h = 4$



BVAR-SV vs. BVAR-SVF-M

(c) $h = 1$



(d) $h = 4$

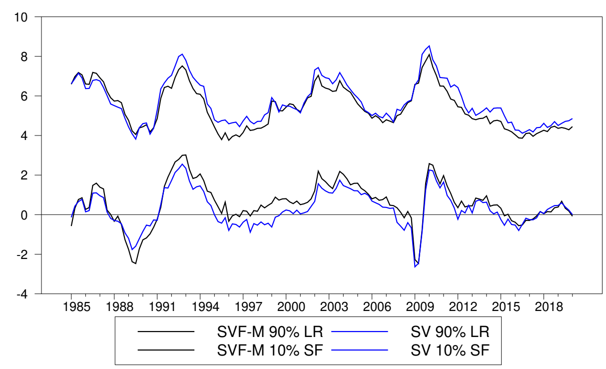


Table A7: Accuracy of in-sample forecasts of GDP growth, 10-90 tails

	1985-2021		**1985-2019**	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	Interval coverage: 10 percent tail			
QR	0.108	0.110	0.107	0.117
BVAR-SV	0.074	0.124	0.071	0.131
BVAR-SVF-M	0.074	0.110	0.086	0.117
	Interval coverage: 90 percent tail			
QR	0.959^{***}	0.966^{**}	0.943^{**}	0.956^{***}
BVAR-SV	0.939^{*}	0.986^{***}	0.936^{*}	0.964^{***}
BVAR-SVF-M	0.939^{**}	0.959^{***}	0.936^{*}	0.934
	Quantile score (10 percent quantile)			
QR	0.640	0.388	0.367	0.262
BVAR-SV	0.946	0.992	0.935	1.086
BVAR-SVF-M	0.875	0.850	0.929	1.044
	Quantile score (90 percent quantile)			
QR	0.575	0.287	0.359	0.213
BVAR-SV	0.828^{**}	0.901	0.947	0.958
BVAR-SVF-M	0.782	0.846	0.973	0.952
	VaR-ES score (10 percent quantile)			
QR	2.973	1.930	2.991	1.840
BVAR-SV	−0.433	−1.319	0.269	−0.355
BVAR-SVF-M	−0.123	−0.340	0.293	−0.170

Notes: Results for 1985-2021 use models estimated with a data sample of 1972:Q1-2021:Q4. Results for 1985-2019 use models estimated with a data sample of 1972:Q1-2019:Q4. Except in the case of the coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as ratios relative to scores for quantile regression (an entry less than 1 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal coverage rate is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table A8: Accuracy of in-sample forecasts of unemployment and inflation, 4-quarters ahead, 10-90 tails

	Unemployment change		**Inflation**	
	1985-2021	**1985-2019**	**1985-2021**	**1985-2019**
	Interval coverage: 10 percent tail			
QR	0.083	0.066	0.110	0.109
BVAR-SV	0.014***	0.022***	0.062	0.066
BVAR-SVF-M	0.034***	0.029***	0.055**	0.073
	Interval coverage: 90 percent tail			
QR	0.869	0.869	0.917	0.905
BVAR-SV	0.855	0.854	0.966***	0.956***
BVAR-SVF-M	0.862	0.869	0.979***	0.971***
	Quantile score (10 percent quantile)			
QR	0.139	0.077	0.112	0.105
BVAR-SV	0.923	0.981	0.965	0.901
BVAR-SVF-M	0.735	0.886	0.955	0.877
	Quantile score (90 percent quantile)			
QR	0.225	0.135	0.154	0.112
BVAR-SV	1.063	0.996	0.835	0.902*
BVAR-SVF-M	0.888	0.873	0.747	0.916

Notes: Results for 1985-2021 use models estimated with a data sample of 1972:Q1-2021:Q4. Results for 1985-2019 use models estimated with a data sample of 1972:Q1-2019:Q4. Except in the case of the coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as ratios relative to scores for quantile regression (an entry less than 1 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal coverage rate is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table A9: Accuracy of out-of-sample forecasts of GDP growth, 10-90 tails

	1985-2021		**1985-2019**	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	Interval coverage: 10 percent tail			
QR	0.108	0.200	0.093	0.175
BVAR-SV	0.074	0.186	0.064	0.168
BVAR-SVF-M	0.101	0.207	0.093	0.190
	Interval coverage: 90 percent tail			
QR	0.973^{***}	0.972^{***}	0.986^{***}	0.985^{***}
BVAR-SV	0.980^{***}	0.952^{**}	0.993^{***}	0.956^{**}
BVAR-SVF-M	0.959^{***}	0.924	0.964^{***}	0.927
	Quantile score (10 percent quantile)			
QR	0.571	0.415	0.265	0.271
BVAR-SV	1.067	1.004	1.158^{**}	1.070
BVAR-SVF-M	1.060	1.026	1.113	1.126
	Quantile score (90 percent quantile)			
QR	0.701	0.400	0.427	0.292
BVAR-SV	0.732	0.726	0.892^{***}	0.894
BVAR-SVF-M	0.787	0.769	0.880^{***}	0.892
	VaR-ES score (10 percent quantile)			
QR	3.950	4.822	2.473	2.449
BVAR-SV	−0.306	0.538	−0.358	−0.266
BVAR-SVF-M	−0.411	0.199	−0.199	−0.694

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

Table A10: Accuracy of out-of-sample forecasts of unemployment and inflation, 4-quarters ahead, 10-90 tails

	Unemployment change		**Inflation**	
	1985-2021	**1985-2019**	**1985-2021**	**1985-2019**
	Interval coverage: 10 percent tail			
QR	0.166	0.153	0.276**	0.270**
BVAR-SV	0.055	0.051	0.200**	0.204**
BVAR-SVF-M	0.097	0.095	0.221**	0.226**
	Interval coverage: 90 percent tail			
QR	0.828	0.847	0.903	0.920
BVAR-SV	0.821	0.839	0.979	1.000
BVAR-SVF-M	0.779	0.796	0.979	1.000
	Quantile score (10 percent quantile)			
QR	0.147	0.089	0.213	0.209
BVAR-SV	0.865	0.916	0.774	0.751
BVAR-SVF-M	0.849	0.886**	0.904	0.890
	Quantile score (90 percent quantile)			
QR	0.247	0.155	0.300	0.259
BVAR-SV	1.191	1.156	0.726	0.733
BVAR-SVF-M	1.156	1.142	0.762	0.750

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

4 Results for Other Forecast Metrics

This section provides results for additional forecast metrics, applied to both in-sample and out-of-sample forecasts. The metrics include root mean square error (RMSE), a joint VaR-ES score for the 5 percent tail, and dynamic quantile tests. In evaluating point forecast accuracy with RMSE, for QR the point forecast is measured as the forecast for $\tau = 0.5$, and for the BVARs, the point forecasts are measured as the mean of the posterior predictive distribution.

We evaluate the shortfall forecasts using the joint value at risk-expected shortfall (VaR-ES) score employed in Carriero, Clark, and Marcellino (2020). 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. Fissler and Ziegel (2016) derive a general class of such scoring functions, and studies such as Patton, Ziegel, and Chen (2019) and Taylor (2019) develop specific functions within this general class. However, some of these functions are designed with asset returns in mind and embed a restriction that ES is strictly negative. Tail forecasts of GDP growth often (in periods of economic expansion) violate such a restriction, with a positive ES. Accordingly, we consider the following VaR-ES scoring function that allows ES to be positive or negative:[3]

$$\begin{aligned} \mathrm{S}_{\tau,t+h} &= b + (Q_{\tau,t+h} - y_{t+h}) \left(\mathbf{1}_{\{y_{t+h} \leq Q_{\tau,t+h}\}} - \tau\right) \\ &+ \frac{1}{\tau} e^{a^{-1}\mathrm{ES}_{\tau,t+h}} \left(\mathbf{1}_{\{y_{t+h} \leq Q_{\tau,t+h}\}} (Q_{\tau,t+h} - y_{t+h}) + \tau \left(\mathrm{ES}_{\tau,t+h} - Q_{\tau,t+h} - a\right)\right), \end{aligned}$$

where $\tau = 0.05$ and $\mathrm{ES}_{\tau,t+h}$ denotes the expected shortfall forecast at quantile τ. In implementation, the scoring function coefficient a is set to 4, and the constant b is set to 6 to ensure a positive score for GDP growth (for simplicity in reporting some results; this setting is irrelevant for the score differences across models).

Finally, this section provides results for dynamic quantile (DQ) tests as developed in Engle and Manganelli (2004) to assess whether quantile forecasts meet basic requirements of unbiasedness and, at the 1-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 (2021): 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, these results use the 10 percent quantile. At the 1-step-ahead horizon, the tests never reject the null of 0 coefficients; although some rejections (for GDP growth, unemployment, and inflation) occur at the 4-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

[3] In the general notation of Fissler and Ziegel (2016), this scoring function uses $G_1(x) = x$ and $G_2(x) = e^{a^{-1}x}$. Because the $G_2(x)$ function is not homogenous, the scoring function values are specific to the units of the variable of interest

Table A11: Accuracy of in-sample forecasts of GDP growth, other measures

	1985-2021		1985-2019	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	RMSE			
QR	4.630	2.153	2.047	1.462
BVAR-SV	0.986	0.944	0.968	0.998
BVAR-SVF-M	0.965	0.928	0.959*	0.980
	VaR-ES score: 5 percent quantile			
QR	3.363	2.460	3.229	2.086
BVAR-SV	−0.479	−1.780	0.004	−0.759
BVAR-SVF-M	−0.076	−0.123	0.094	−0.418
	Interval coverage: 5 percent tail			
QR	0.047	0.062	0.043	0.058
BVAR-SV	0.047	0.083	0.036	0.073
BVAR-SVF-M	0.041	0.055	0.029	0.066
	Interval coverage: 95 percent tail			
QR	0.986**	0.986**	0.979**	1.000
BVAR-SV	0.966	0.986*	0.971	0.993*
BVAR-SVF-M	0.959	0.993*	0.964	0.993*

Notes: Results for 1985-2021 use models estimated with a data sample of 1972:Q1-2021:Q4. Results for 1985-2019 use models estimated with a data sample of 1972:Q1-2019:Q4. Except in the case of the coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to the accuracy of the quantile regression, as a ratio for the RMSE (an entry less than 1 means the BVAR is more accurate than QR) and score difference for the VaR-ES score (a positive entry means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal coverage rate is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

(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.

Table A12: Accuracy of in-sample forecasts of unemployment and inflation, 4-quarters ahead, other measures

	Unemployment change		**Inflation**	
	1985-2021	**1985-2019**	**1985-2021**	**1985-2019**
	RMSE			
QR	1.304	0.726	0.794	0.638
BVAR-SV	1.039	0.894^{*}	0.979	0.906
BVAR-SVF-M	1.035	0.815	0.961	0.925
	VaR-ES score: 5 percent quantile			
QR	2.995	2.896	0.906	0.902
BVAR-SV	-0.281^{***}	−0.058	−0.141	−0.020
BVAR-SVF-M	−0.077	0.035	−0.064	0.043
	Interval coverage: 5 percent tail			
QR	0.041	0.029	0.041	0.044
BVAR-SV	0.007^{*}	0.000	0.034	0.022^{**}
BVAR-SVF-M	0.000	0.000	0.021^{**}	0.022^{**}
	Interval coverage: 95 percent tail			
QR	0.924	0.927	0.959	0.956
BVAR-SV	0.910	0.898	0.979^{**}	1.000
BVAR-SVF-M	0.924	0.927	1.000	1.000

Notes: Results for 1985-2021 use models estimated with a data sample of 1972:Q1-2021:Q4. Results for 1985-2019 use models estimated with a data sample of 1972:Q1-2019:Q4. Except in the case of the coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as relative to the accuracy of the quantile regression, as a ratio for the RMSE (an entry less than 1 means the BVAR is more accurate than QR) and score difference for the VaR-ES score (a positive entry means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal coverage rate is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table A13: Accuracy of out-of-sample forecasts of GDP growth, other measures

	1985-2021		**1985-2019**	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	RMSE			
QR	4.906	2.411	2.064	1.699
BVAR-SV	0.955	0.889	0.910	0.930
BVAR-SVF-M	0.950	0.889	0.925	0.943
	VaR-ES score: 5 percent quantile			
QR	8.834	7.625	2.784	2.771
BVAR-SV	3.690	2.425	-0.598^{**}	-0.017
BVAR-SVF-M	3.597	1.854	-0.133	-0.669
	Interval coverage: 5 percent tail			
QR	0.074	0.083	0.057	0.051
BVAR-SV	0.027	0.110	0.014^{***}	0.088
BVAR-SVF-M	0.027	0.159^{*}	0.014^{***}	0.139
	Interval coverage: 95 percent tail			
QR	0.980^{*}	0.986^{***}	0.993^{***}	1.000
BVAR-SV	0.993^{***}	0.993^{***}	1.000	0.993^{***}
BVAR-SVF-M	0.993^{***}	0.966	1.000	0.971

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

Table A14: Accuracy of out-of-sample forecasts of unemployment and inflation, 4-quarters ahead, other measures

	Unemployment change		**Inflation**	
	1985-2021	**1985-2019**	**1985-2021**	**1985-2019**
	RMSE			
QR	1.351	0.781	1.293	1.192
BVAR-SV	1.057	0.946	0.919	0.899
BVAR-SVF-M	1.026	0.943	0.988	0.968
	VaR-ES score: 5 percent quantile			
QR	3.431	2.991	4.617	4.647
BVAR-SV	0.249	0.003	2.998**	3.112**
BVAR-SVF-M	0.413	0.090	1.917**	1.960**
	Interval coverage: 5 percent tail			
QR	0.103	0.088	0.214**	0.212**
BVAR-SV	0.007***	0.000	0.083	0.080
BVAR-SVF-M	0.028	0.022	0.131*	0.131*
	Interval coverage: 95 percent tail			
QR	0.876	0.898	0.938	0.956
BVAR-SV	0.834	0.854	0.986***	1.000
BVAR-SVF-M	0.834	0.854	0.979	1.000

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

Table A15: p-values of dynamic quantile tests, 10 percent quantile, GDP growth

In-sample forecasts of GDP growth, models with NFCI, 1985-2019				
	Lagged hit rates		Lagged NFCI	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
QR	0.705	0.982	0.968	0.762
BVAR-SV	0.913	0.473	0.986	0.550
BVAR-SVF-M	0.840	0.023	0.964	0.523
Out-of-sample forecasts of GDP growth, models with NFCI, 1985-2019				
	Lagged hit rates		Lagged NFCI	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
QR	0.996	0.228	0.997	0.374
BVAR-SV	0.886	0.373	0.649	0.000
BVAR-SVF-M	0.865	0.347	0.720	0.000
In-sample forecasts of GDP growth, models with turbulence, 1972-20191				
	Lagged hit rates		Lagged turbulence	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
QR	0.329	0.484	0.998	0.822
BVAR-SV	1.000	0.201	0.641	0.804
BVAR-SVF-M	0.968	0.776	0.611	0.849
Out-of-sample forecasts of GDP growth, models with turbulence, 1972-20191				
	Lagged hit rates		Lagged turbulence	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
QR	0.969	0.385	0.992	0.257
BVAR-SV	0.701	0.086	0.472	0.163
BVAR-SVF-M	0.572	0.010	0.568	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 4-steps-ahead horizon incorporate a heteroskedasticity and autocorrelation-robust variance estimator (Newey-West, with six lags).

Table A16: p-values of dynamic quantile tests, unemployment and inflation, 10 and 90 percent quantiles, 4-quarters-ahead forecasts

In-sample forecasts of unemployment changes, models with NFCI, 1985-2019				
	Lagged hit rates		Lagged NFCI	
	$\tau = 0.10$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.90$
QR	0.335	0.262	0.231	0.000
BVAR-SV	0.144	0.169	0.992	0.386
BVAR-SVF-M	0.026	0.335	0.417	0.307
Out-of-sample forecasts of unemployment changes, models with NFCI, 1985-2019				
	Lagged hit rates		Lagged NFCI	
	$\tau = 0.10$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.90$
QR	0.048	0.006	0.117	0.000
BVAR-SV	0.139	0.010	0.387	0.345
BVAR-SVF-M	0.467	0.001	0.125	0.000
In-sample forecasts of inflation, models with NFCI, 1985-2019				
	Lagged hit rates		Lagged NFCI	
	$\tau = 0.10$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.90$
QR	0.014	0.085	0.218	0.016
BVAR-SV	0.017	0.051	0.215	0.415
BVAR-SVF-M	0.010	0.028	0.302	0.415
Out-of-sample forecasts of inflation, models with NFCI, 1985-2019				
	Lagged hit rates		Lagged NFCI	
	$\tau = 0.10$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.90$
QR	0.648	0.131	0.967	0.131
BVAR-SV	0.679	0.679	0.785	0.679
BVAR-SVF-M	0.384	0.384	0.984	0.384

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 4-steps-ahead horizon incorporate a heteroskedasticity and autocorrelation-robust variance estimator (Newey-West, with six lags).

5 Forecast Evaluation Results with Bayesian Quantile Regression

Yu and Moyeed (2001) established that quantile regression has a convenient mixture representation that enables Bayesian estimation. For quantile τ, our BQR formulation takes the form

$$y_{t+h}^{(h)} = x_t' \beta_\tau + \epsilon_{\tau,t+h}, \tag{1}$$

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

$$y_{t+h}^{(h)} = x_t' \beta_\tau + \theta z_{\tau,t+h} + \kappa \sqrt{\sigma_{\tau,h} z_{\tau,t+h}} u_{\tau,t+h}, \tag{2}$$

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

We estimate Bayesian quantile regressions with the Gibbs sampler of Khare and Hobert (2012). We use an independent Normal-Gamma prior, with a normal distribution for the regression coefficients β_τ and a Gamma distribution for the scale parameter $\sigma_{\tau,h}$. One step samples the mixture state time series z from an inverse Gaussian distribution. The next step draws the scale parameter $\sigma_{\tau,w}$ from its inverse Gamma conditional posterior. In the subsequent step, the regression parameter vector β_τ is drawn from its Normal conditional posterior. For each quantile, we take a total of 6000 draws, discard the first 1000, and compute the posterior mean coefficient vector $\hat{\beta}_\tau$ from the remaining 5000 draws. The quantile forecast draw is formed as $x_t'\hat{\beta}_\tau$.

For the Gaussian prior on β_τ, we use a mean of 0 and a variance that is Minnesota-style in the sense that that we take account of the relative scales of variables and shrink coefficients on other variables more than those on the lag of the dependent variable. The shrinkage is controlled by two hyperparameters (smaller numbers mean more shrinkage): λ_1, which controls the overall rate of shrinkage; and λ_2, which controls the rate of shrinkage on variables other than lags of the dependent variable. At each forecast origin, the prior variance for the coefficient on the lagged dependent variable is simply λ_1. The prior variance associated with the coefficient on the i-th variable $x_{i,t}$ of x_t is specified as $\lambda_1 \lambda_2 \frac{\sigma_y^2}{\sigma_i^2}$. Finally, for the intercept, the prior is uninformative, with variance $1000\sigma_y^2$. In setting these components of the prior, for σ_y^2 and σ_i^2 we use simple variances estimated with the regression sample available as of the forecast origin.[4] We fix the hyperparameters at values that may be considered very common in Minnesota-type priors and forecasting: $\lambda_1 = 0.04$ and $\lambda_2 = 0.25$.

To briefly summarize these out-of-sample forecast results, in our GDP growth, unemployment, and inflation applications, Bayesian quantile regression (BQR) yields results similar to those for frequentist quantile regression (QR). Depending on the application, forecast horizon, and sample,

[4] In the vector autoregression literature, it is typical to use the variances of the residual of low-order AR models. With multi-step forecast horizons in a direct-multi-step setup, the obvious analogue is not so clear. We simplify the choice by using simple variances to capture scale differences in the volatilities of variables.

Table A17: Accuracy of out-of-sample forecasts of GDP growth, Bayesian QR as baseline

	1985-2021		1985-2019	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	Quantile score: 5 percent quantile			
BQR	0.502	0.317	0.193	0.201
QR	1.003	0.987	0.868	0.840
BVAR-SV	0.945	0.915	0.986	0.854
BVAR-SVF-M	0.921	0.952	0.873**	0.934
	Quantile score: 95 percent quantile			
BQR	0.535	0.277	0.267	0.194
QR	1.075	0.996	1.001	0.926***
BVAR-SV	0.610	0.601	0.914	0.812***
BVAR-SVF-M	0.686	0.658*	0.848*	0.771***
	qwCRPS-left			
BQR	0.495	0.340	0.299	0.255
QR	1.047**	1.091	1.069**	1.102
BVAR-SV	1.024	1.036	1.060	1.078
BVAR-SVF-M	1.036	1.052	1.069*	1.116
	qwCRPS-right			
BQR	0.538	0.353	0.358	0.277
QR	1.064***	1.062**	1.079**	1.071
BVAR-SV	0.894	0.871	0.996	0.967
BVAR-SVF-M	0.926	0.892	1.007	0.975

Notes: To facilitate accuracy comparisons the results for the QR and BVAR models are reported as ratios relative to scores for Bayesian quantile regression, denoted BQR (an entry less than 1 means the QR or BVAR is more accurate than BQR). Statistical significance of the differences in scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

BQR is sometimes modestly better than QR and sometimes modestly worse. In turn, in tail risk forecasting performance, our BVARs are overall at least as effective as BQR.

Table A18: Accuracy of out-of-sample forecasts of unemployment and inflation, 4-quarters ahead, Bayesian QR as baseline

	Unemployment change		Inflation	
	1985-2021	**1985-2019**	**1985-2021**	**1985-2019**
	Quantile score: 5 percent quantile			
BQR	0.097	0.046	0.130	0.128
QR	1.025^{*}	1.070^{***}	1.167	1.164
BVAR-SV	0.843	1.064	0.636	0.602
BVAR-SVF-M	0.762	0.954	0.884	0.869
	Quantile score: 95 percent quantile			
BQR	0.199	0.117	0.216	0.170
QR	0.984	0.915	0.977	0.969
BVAR-SV	1.233	1.142	0.632	0.675
BVAR-SVF-M	1.187	1.116	0.668	0.685
	qwCRPS-left			
BQR	0.150	0.103	0.182	0.172
QR	1.011	1.019	1.141^{*}	1.154^{*}
BVAR-SV	0.946	0.948	1.011	1.007
BVAR-SVF-M	0.944	0.971	1.110	1.112
	qwCRPS-right			
BQR	0.203	0.143	0.230	0.205
QR	0.983	0.970	1.071	1.081
BVAR-SV	1.020	0.953	0.894	0.912
BVAR-SVF-M	1.012	0.972	0.940	0.950

Notes: To facilitate accuracy comparisons the results for the QR and BVAR models are reported as ratios relative to scores for Bayesian quantile regression, denoted BQR (an entry less than 1 means the QR or BVAR is more accurate than BQR). Statistical significance of the differences in scores is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

6 Results Using Turbulence Indicator of Financial Conditions

This section provides results for GDP growth obtained with models that replace the NFCI as a measure of financial indicators with the turbulence 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. Because the turbulence series, obtained from the data files of Giglio, Kelly, and Pruitt (2016), starts and ends earlier than the NFCI, the results include out-of-sample forecasts in the 1970s.

Figures A7 (in-sample) and A8 (out-of-sample) compare estimates of expected shortfall (at 5 percent) and long-rise (at 95 percent) at the 1-step-ahead and 4-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, the middle panel compares estimates from the baseline BVAR-SV model including the NFCI and the BVAR-SV specification including turbulence, and the bottom panel makes the corresponding comparison for the BVAR-SVF-M estimates. In the cases of the BVAR models, 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 A10). In fact, the shortfall estimates obtained from the quantile regression including turbulence might strain credulity in the eyes of some readers.

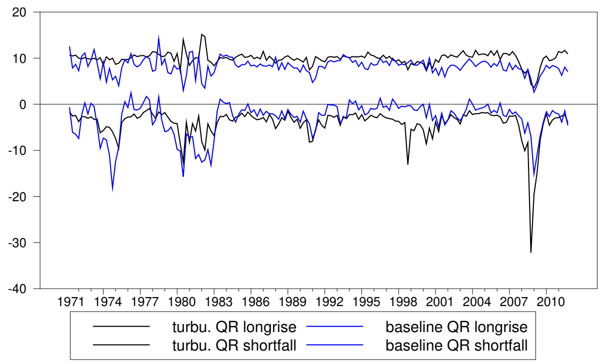
Figures A9 (in-sample) and A10 (out-of-sample) compare shortfall and long-rise estimates across our QR, BVAR-SV, and BVAR-SVF-M specifications including turbulence as the financial indicator. As in the paper's baseline estimates based on the NFCI, estimates are similar across the BVAR-SV and BVAR-SVF-M models. The QR-based estimates are less similar to the BVAR-SV estimates using turbulence than when using the NFCI.

Tables A19 (in-sample) and A20 (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 results for models including turbulence as the indicator of financial conditions, the BVAR-SV and BVAR-SVF-M models are in most cases more accurate than the QR specification, with statistical significance in a number of instances. This applies with both the in-sample and out-of-sample forecasts, and across the evaluation samples of 1972-2011 and 1985-2011. However, the patterns are somewhat weaker with the qwCRPS metrics, with smaller gains and fewer instances of significance (indeed, in the out-of-sample results, QR scores better than the BVARs in the 1985-2011 sample). 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.

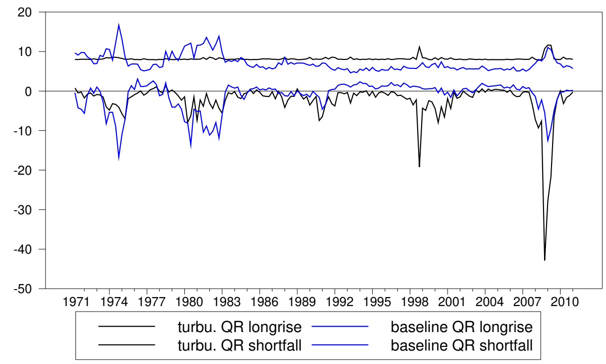
Figure A7: Long-rise and expected shortfall, in-sample forecasts of GDP growth: baseline with NFCI vs. forecasts using turbulence measure of financial conditions

QR

(a) $h = 1$

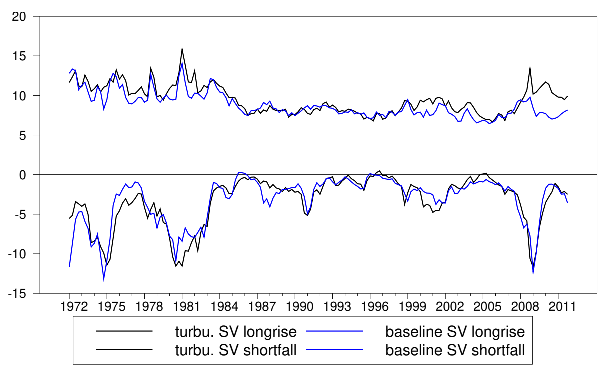


(b) $h = 4$

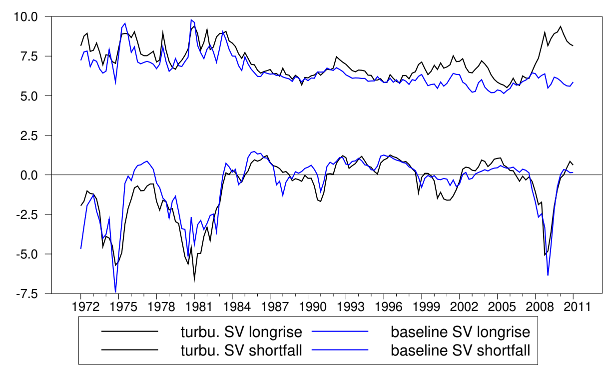


BVAR-SV

(c) $h = 1$

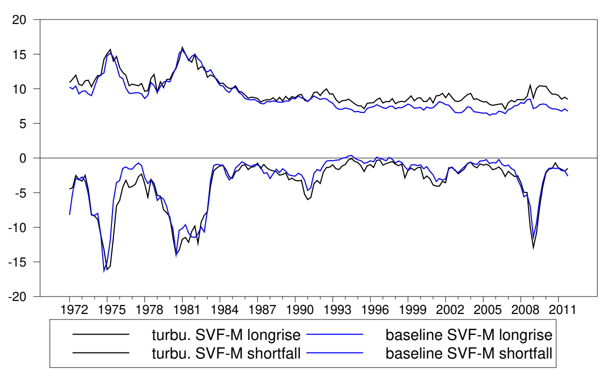


(d) $h = 4$



BVAR-SVF-M

(e) $h = 1$



(f) $h = 4$

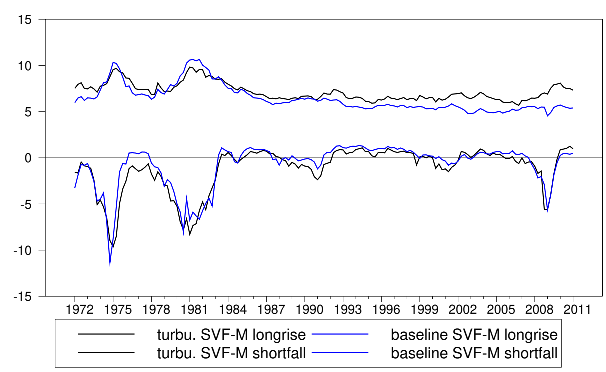
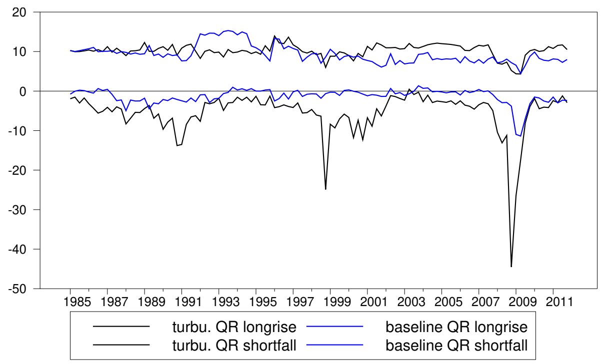


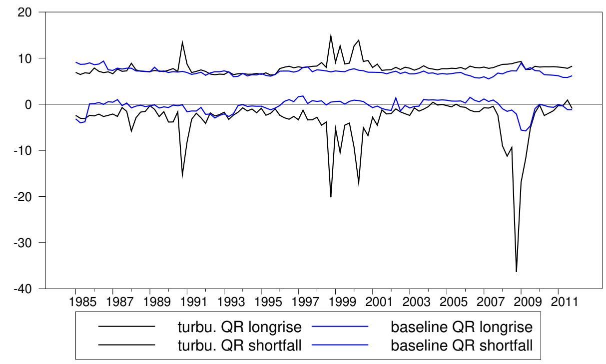
Figure A8: Long-rise and expected shortfall, out-of-sample forecasts of GDP growth: baseline with NFCI vs. forecasts using turbulence measure of financial conditions

QR

(a) $h = 1$

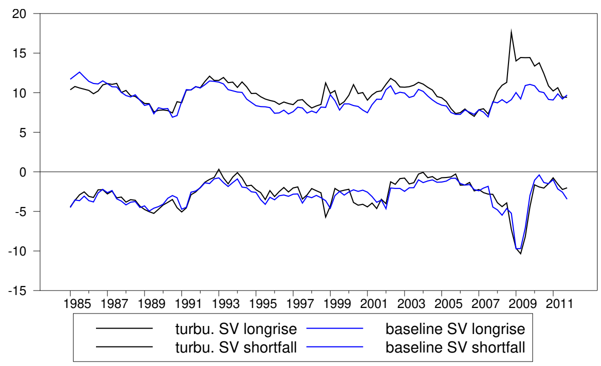


(b) $h = 4$

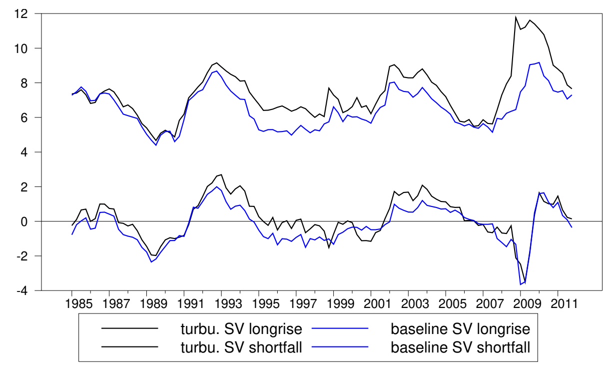


BVAR-SV

(c) $h = 1$

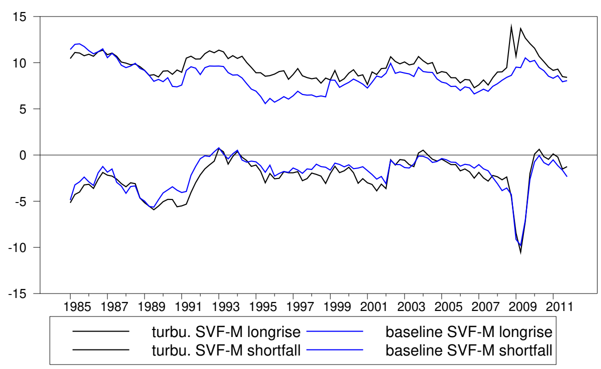


(d) $h = 4$



BVAR-SVF-M

(e) $h = 1$



(f) $h = 4$

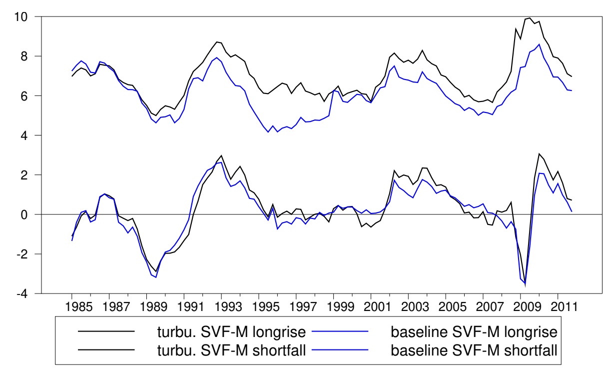
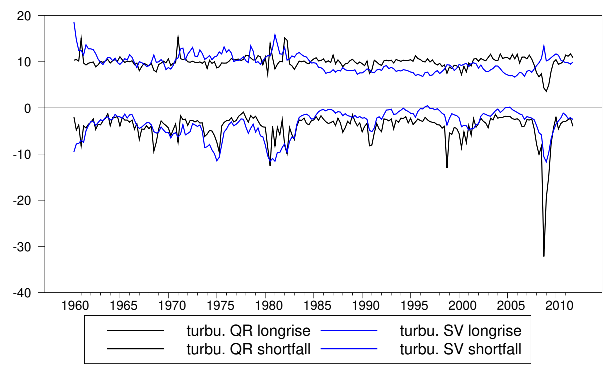


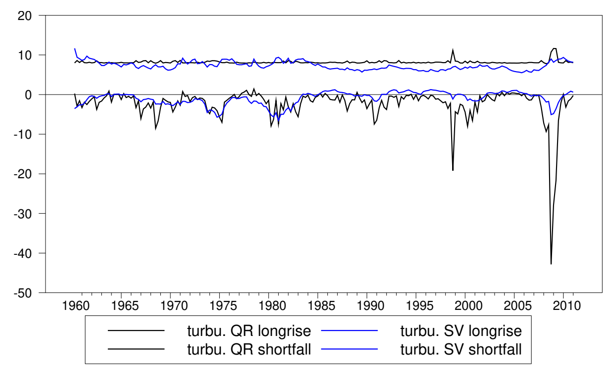
Figure A9: Long-rise and expected shortfall, in-sample forecasts of GDP growth for 1959-2011 using the turbulence measure of financial conditions

QR vs. BVAR-SV

(a) $h = 1$

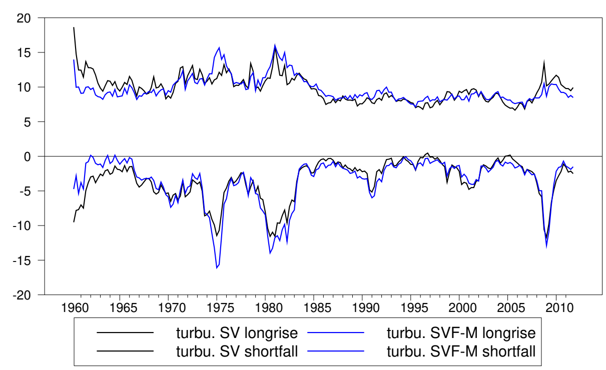


(b) $h = 4$



BVAR-SV vs. BVAR-SVF-M

(c) $h = 1$



(d) $h = 4$

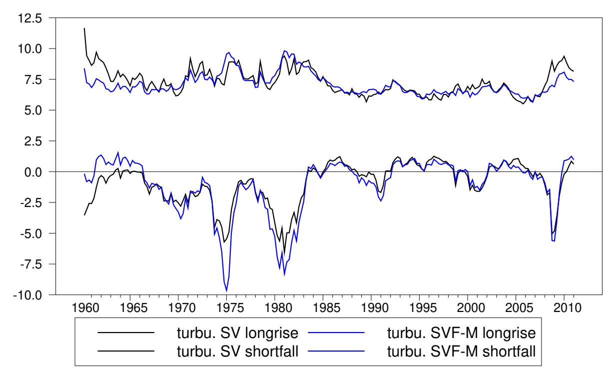
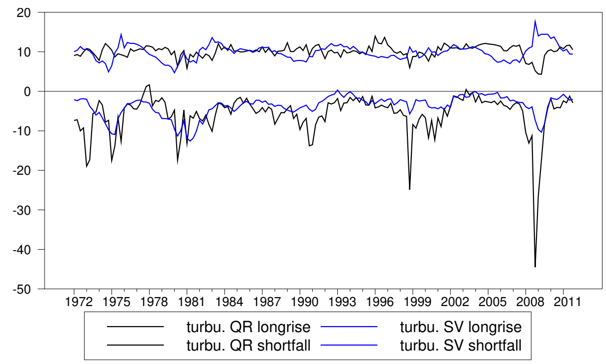


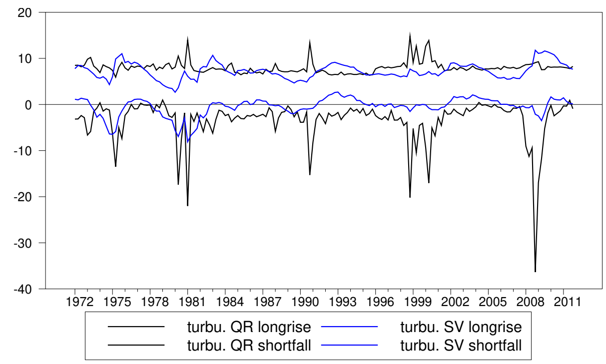
Figure A10: Long-rise and expected shortfall, out-of-sample forecasts of GDP growth for 1972-2011 using the turbulence measure of financial conditions

QR vs. BVAR-SV

(a) $h = 1$

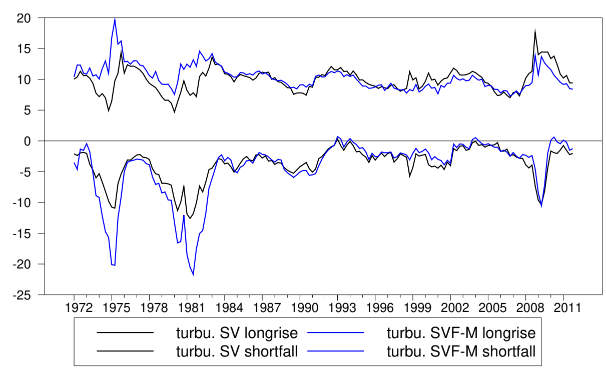


(b) $h = 4$



BVAR-SV vs. BVAR-SVF-M

(c) $h = 1$



(d) $h = 4$

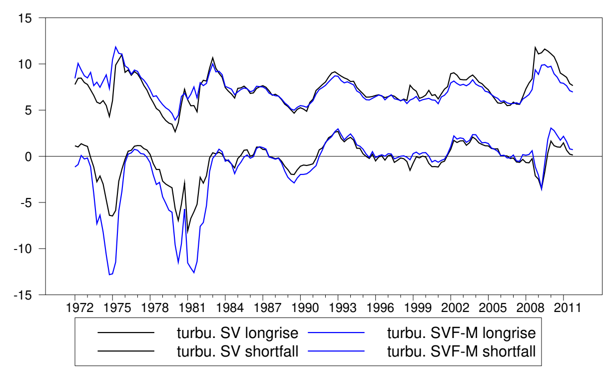


Table A19: Accuracy of in-sample forecasts of GDP growth using the turbulence measure of financial conditions

	1972-2011		**1985-2011**	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	Interval coverage: 5 percent tail			
QR	0.044	0.051	0.009***	0.019
BVAR-SV	0.038	0.051	0.037	0.067
BVAR-SVF-M	0.013***	0.057	0.019**	0.086
	Interval coverage: 95 percent tail			
QR	0.956	0.962	0.981**	1.000
BVAR-SV	0.975**	0.981**	0.981**	1.000
BVAR-SVF-M	0.981***	0.962	0.991***	1.000
	Quantile score: 5 percent quantile			
QR	0.328	0.226	0.264	0.179
BVAR-SV	0.772**	0.924	0.727***	1.153
BVAR-SVF-M	0.801	0.977	0.766***	1.166
	Quantile score: 95 percent quantile			
QR	0.323	0.218	0.278	0.204
BVAR-SV	0.949	0.803**	0.843***	0.786*
BVAR-SVF-M	0.992	0.864*	0.859***	0.804***
	qwCRPS-left			
QR	0.512	0.380	0.364	0.280
BVAR-SV	0.903**	0.912	0.951	1.040
BVAR-SVF-M	0.912*	0.915	0.963	1.045
	qwCRPS-right			
QR	0.505	0.356	0.383	0.268
BVAR-SV	0.969	0.913	0.984	0.995
BVAR-SVF-M	0.981	0.940	0.991	1.008
	VaR-ES score: 5 percent quantile			
QR	4.416	3.374	3.996	2.721
BVAR-SV	0.909***	0.331	1.082***	−0.267
BVAR-SVF-M	0.754**	0.121	0.876***	−0.417

Notes: Results for 1972-2011 use models estimated with data from 1959 through 2011. Results for 1985-2019 use models estimated with a data sample of 1972:Q1-2011:Q4. Except in the case of the coverage rates, to facilitate accuracy comparisons the results for the BVAR models are reported as ratios relative to scores for quantile regression (an entry less than 1 means the BVAR is more accurate than QR). Statistical significance of the differences in scores and of departures of empirical coverage from the nominal coverage rate is indicated by *** (1%), ** (5%), or * (10%), obtained with the Diebold and Mariano–West t-test.

Table A20: Accuracy of out-of-sample forecasts of GDP growth using the turbulence measure of financial conditions

	1972-2011		1985-2011	
	$h = 1Q$	$h = 4Q$	$h = 1Q$	$h = 4Q$
	Interval coverage: 5 percent tail			
QR	0.031	0.064	0.000	0.000
BVAR-SV	0.025^{*}	0.153	0.000	0.162
BVAR-SVF-M	0.019^{**}	0.115	0.028	0.152
	Interval coverage: 95 percent tail			
QR	0.975^{**}	0.981^{**}	0.991^{***}	1.000
BVAR-SV	0.987^{***}	0.981^{***}	1.000	0.990^{***}
BVAR-SVF-M	1.000	0.994^{***}	1.000	0.990^{***}
	Quantile score: 5 percent quantile			
QR	0.396	0.360	0.286	0.206
BVAR-SV	0.664^{***}	0.787	0.640^{***}	1.088
BVAR-SVF-M	0.660^{***}	0.755	0.665^{***}	1.319
	Quantile score: 95 percent quantile			
QR	0.329	0.234	0.307	0.210
BVAR-SV	0.936	0.844^{**}	0.982	0.957
BVAR-SVF-M	0.956	0.853^{***}	0.949	0.918
	qwCRPS-left			
QR	0.498	0.442	0.339	0.286
BVAR-SV	0.961	0.908	1.105	1.280
BVAR-SVF-M	0.950	0.893	1.116	1.313
	qwCRPS-right			
QR	0.491	0.411	0.383	0.307
BVAR-SV	1.022	0.879	1.159^{**}	1.138
BVAR-SVF-M	1.028	0.872	1.162^{**}	1.132
	VaR-ES score: 5 percent quantile			
QR	5.589	6.134	4.247	3.332
BVAR-SV	1.792^{***}	1.694	1.050^{***}	-0.724
BVAR-SVF-M	1.884^{***}	0.736	1.141^{***}	-3.003

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

7 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 section summarizes the results of a Monte Carlo analysis of the performance of quantile regression and the BVAR models. For brevity, this analysis focuses on left tail risks to output growth.

In the Monte Carlo experiments, to make the computational burden manageable, we used a bivariate VAR 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. We conduct two sets of experiments, simulating data from (two different) DGPs featuring stochastic volatility and comparing the tail risk forecasting performance of estimated QR, BVAR-SV, and BVAR-SVF-M models.[5]

7.1 Baseline BVAR-SVF-M DGP

Our first experiment treats the BVAR-SVF-M model as the data-generating process (DGP), with parameters largely taken from empirical estimates based on historical GDP and NFCI data. This 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):

$$\begin{pmatrix} \mathrm{GDP}_t \\ \mathrm{NFCI}_t \end{pmatrix} = \begin{pmatrix} 0.247 & -0.282 & -0.400 & -0.200 \\ 0.001 & 0.781 & 0.300 & -0.050 \end{pmatrix} \begin{pmatrix} \mathrm{GDP}_{t-1} \\ \mathrm{NFCI}_{t-1} \\ \ln m_t \\ \ln m_{t-1} \end{pmatrix} + \begin{pmatrix} 1.000 & 0.000 \\ -0.011 & 1.000 \end{pmatrix}^{-1} \begin{pmatrix} \lambda_{gdp,t}^{0.5}\ \epsilon_{gdp,t} \\ \lambda_{nfci,t}^{0.5}\ \epsilon_{nfci,t} \end{pmatrix},$$

where

$$\begin{pmatrix} \ln \lambda_{gdp,t} \\ \ln \lambda_{nfci,t} \end{pmatrix} = \begin{pmatrix} 1.292 \\ 3.030 \end{pmatrix} \ln m_t + \begin{pmatrix} \ln h_{gdp,t} \\ \ln h_{nfci,t} \end{pmatrix}$$

$$\begin{aligned} \ln m_t &= 0.738 \ln m_{t-1} + 0.014\ \mathrm{GDP}_{t-1} + 0.100\ \mathrm{NFCI}_{t-1} + u_{m,t},\ \ var(u_{m,t}) = 0.03 \\ \ln h_{gdp,t} &= 1.873 - 0.090 \ln h_{gdp,t-1} + e_{gdp,t},\ \ var(e_{gdp,t}) = 0.035 \\ \ln h_{nfci,t} &= -2.646 + 0.215 \ln h_{nfci,t-1} + e_{nfci,t},\ \ var(e_{nfci,t}) = 0.031. \end{aligned}$$

With this DGP, we simulate 100 artificial data sets of a total of 195 observations (a sample length

[5] While not reported, with the first DGP, we also considered a parameterization with a larger (in absolute value) coefficient on the volatility factor in the VAR's conditional mean — which strengthens the contemporaneous link between volatility and the level of GDP and creates more skewness in the conditional predictive distribution — and obtained results essentially the same as those reported below.)

corresponding to our actual empirical sample). For each data set, we estimated quantile regression, BVAR-SV, and BVAR-SVF-M models and formed 1-step-ahead in-sample forecast distributions — deliberately using the 1-step-ahead horizon and in-sample forecasting to make computation tractable. With the in-sample forecasts, we computed quantiles, 5 percent quantile scores, and qwCRPS-left scores, as well as expected shortfall and long-rise at 5 and 95 percent, 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 A11 through A15 present the time series of estimates obtained with the BVAR-SV and BVAR-SVF-M models for the 100 data sets. Qualitatively, these generate periodic (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 (with a median of 1.16) in the BVAR-SV forecasts and 1.45 (with a median of 1.26) in the BVAR-SVF-M forecasts.

In comparing tail risk forecast accuracy, results for QR, BVAR-SV, and BVAR-SVF-M show the basic pattern obtained in the paper. 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 5 percent 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 73 percent of the data sets. With the qwCRPS-left score, the mean ratio for the BVAR-SVF-M model as compared to QR is also 0.98, with the BVAR-SVF-M model having a lower score in 90 percent of the data sets. The corresponding results for the BVAR-SV model compared to quantile regression are broadly similar: Compared to QR, the BVAR-SV model is 2 percent more accurate in quantile score, and equally accurate in the qwCRPS-left score. Overall, the accuracy of BVAR-SV and BVAR-SVF-M tail risk predictions are similar to one another and comparable to QR. 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 1-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.[6] In particular, in visually inspecting the shortfall estimates compared to the levels and volatilities shocks of the BVAR-SV model estimated for data 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 (using our five variable specification in the paper) 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 Figure A21), whereas the rolling window correlation of shocks to the level and volatility of GDP growth turns negative around recessions.[7] As a result, the BVAR-SV and BVAR-SVF-M models 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.

7.2 BVAR-SV-CSZ DGP

Although we believe that the first Monte Carlo experiment is a reasonably accurate reflection of the properties of the actual data, some might question whether the BVAR-SVF-M model's estimates and associated parameterization of the DGP used in the first experiment either understate or fail to capture asymmetries that actually exist in the data and should be captured in conditional predictive distributions. Accordingly, our second experiment uses a BVAR-SV model with a strong negative correlation between the VAR's shock to the level of GDP and the shock to volatility, which creates a significant asymmetry or skewness in the conditional predictive distribution (for reasons detailed in Caldara, Scotti, and Zhong (2021), with Monte Carlo evidence in their Figure 2).

This DGP modifies the specification of our first experiment to drop the volatility factor out of the VAR's conditional mean and add a negative correlation between the shock to the VAR and the shock to the volatility factor, along the lines of the Monte Carlo specification of Caldara, Scotti, and Zhong (2021). This second DGP takes the following form:

$$\begin{pmatrix} \text{GDP}_t \\ \text{NFCI}_t \end{pmatrix} = \begin{pmatrix} 0.247 & -0.282 \\ 0.001 & 0.781 \end{pmatrix} \begin{pmatrix} \text{GDP}_{t-1} \\ \text{NFCI}_{t-1} \end{pmatrix} + \begin{pmatrix} 1.000 & 0.000 \\ -0.011 & 1.000 \end{pmatrix}^{-1} \begin{pmatrix} \lambda_{gdp,t}^{0.5}\ \epsilon_{gdp,t} \\ \lambda_{nfci,t}^{0.5}\ \epsilon_{nfci,t} \end{pmatrix},$$

[6] 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.

[7] 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.

where

$$\begin{pmatrix} \ln \lambda_{gdp,t} \\ \ln \lambda_{nfci,t} \end{pmatrix} = \begin{pmatrix} 1.292 \\ 3.030 \end{pmatrix} \ln m_t + \begin{pmatrix} \ln h_{gdp,t} \\ \ln h_{nfci,t} \end{pmatrix}$$

$$\begin{aligned} \ln m_t &= 0.738 \ln m_{t-1} + 0.014 \text{ GDP}_{t-1} + 0.100 \text{ NFCI}_{t-1} + 0.03^{0.5} u_{m,t} \\ \ln h_{gdp,t} &= 1.873 - 0.090 \ln h_{gdp,t-1} + e_{gdp,t},\ var(e_{gdp,t}) = 0.015 \\ \ln h_{nfci,t} &= -2.646 + 0.215 \ln h_{nfci,t-1} + e_{nfci,t},\ var(e_{nfci,t}) = 0.015. \end{aligned}$$

and

$$\begin{pmatrix} \epsilon_{gdp,t} \\ \epsilon_{nfci,t} \\ u_{m,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s_{13} & 0 & \sqrt{1 - s_{13}^2} \end{pmatrix} \begin{pmatrix} e_{gdp,t} \\ e_{nfci,t} \\ e_{m,t} \end{pmatrix}, \begin{pmatrix} e_{gdp,t} \\ e_{nfci,t} \\ e_{m,t} \end{pmatrix} \sim i.i.d.\ N(0, I_3).$$

The parameters of this DGP match those of the first experiments, except that we modestly lower the innovation variances of the idiosyncratic volatility factors in order to increase the relative importance of the volatility factor. The covariance s_{13} is set to -0.8, which makes the (contemporaneous) correlation between innovations to the volatility factor and the level of the GDP variable equal to -0.8. This is the same setting used in a simpler DGP in the Monte Carlo experiment of Caldara, Scotti, and Zhong (2021). For simplicity, the DGP does not include a correlation between innovations to the volatility factor and the level of the NFCI.

With this DGP, we again simulate 100 artificial data sets of a total of 195 observations. For each data set, we estimated the quantile regression, BVAR-SV, and BVAR-SVF-M models, formed 1-step-ahead in-sample forecast distributions, and computed quantiles, 5 percent quantile scores, the qwCRPS-left score, and expected shortfall and long-rise at 5 and 95 percent, respectively. Note that, although the DGP does not take the form of the BVAR-SVF-M model, our forecasts use estimates of the BVAR-SVF-M model, not a model of the form of the DGP (the former has the same qualitative aspect, but represented in a different way than in the DGP).

Shortfall and long-rise estimates obtained from the Monte Carlo data display asymmetries sharper than those seen in the first DGP. Figures A16 through A20 present the time series of estimates obtained with the BVAR-SV and BVAR-SVF-M models for the 100 data sets. Qualitatively, these generate periodic (downward) asymmetries in shortfall. Across the 100 data sets, the ratio of the standard deviation of shortfall to the standard deviation of long-rise averages 1.26 (with a median of 1.23) in the BVAR-SV forecasts and 2.02 (with a median of 1.85) in the BVAR-SVF-M forecasts. In this case, with the DGP featuring strong asymmetries, while both of the estimated BVAR models yield estimates of shortfall more volatile than estimates of long-rise, the BVAR-SVF-M model does so more than does the BVAR-SV specification.

In comparing tail risk forecast accuracy for the QR, BVAR-SV, and BVAR-SVF-M models, in this case the BVAR-SVF-M specification is most accurate. While the BVAR-SV model continues to be at least as successful as QR, neither matches the accuracy of the BVAR-SVF-M model. 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 5 percent quantile score, the mean ratio is 0.90 (for an average BVAR-SVF-M gain of 10 percent), with the BVAR-SVF-M model having a lower score in 99 percent of the data sets. With the qwCRPS-left score, the mean ratio for the BVAR-SVF-M model as compared to QR is 0.92, with the BVAR-SVF-M model having a lower score in 100 percent of the data sets. When the BVAR-SV model is compared to QR, its advantages are smaller, although some remain: Compared to QR, the BVAR-SV model is 2 percent more accurate in quantile score (having a lower score in 76 percent of data sets), and equally accurate in the qwCRPS-left score. Overall, these results indicate that, for a DGP featuring strong asymmetry, it continues to be the case that the BVAR models we consider can capture tail risk at least as well as quantile regression. In fact, in this case, the BVAR-SVF-M model is most accurate, offering relatively sizable gains over QR. More importantly, the experiment confirms that, in data that feature strong asymmetries, the BVAR-SVF-M model is capable of capturing them. Even so, it continues to be the case that the BVAR-SV model projects more variation in downside risk to output than upside, consistent with the pattern that has drawn emphasis in recent work on macroeconomic tail risks.

Figure A11: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SVF-M as the DGP: artificial data sets 1 through 20

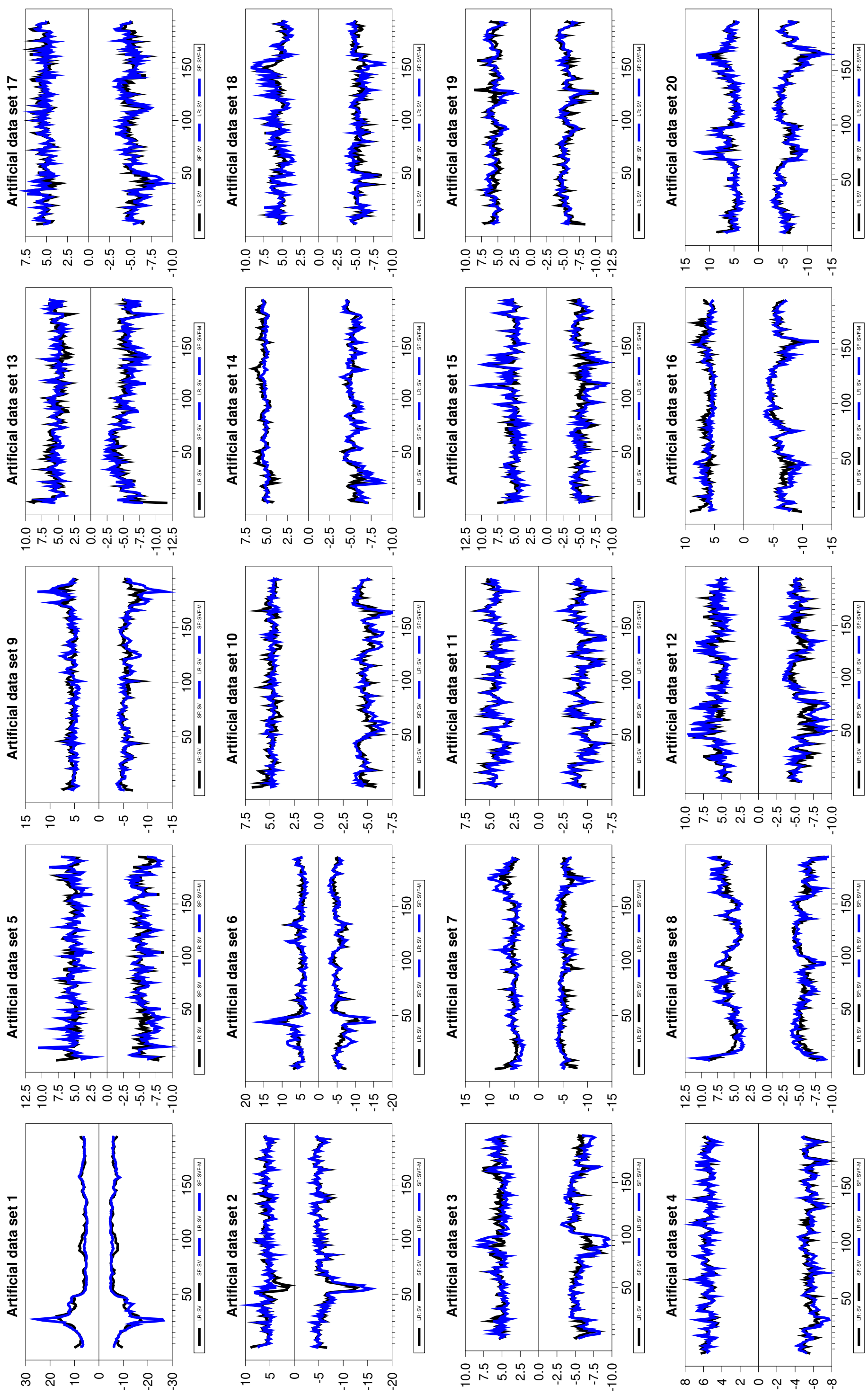


Figure A12: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SVF-M as the DGP: artificial data sets 21 through 40

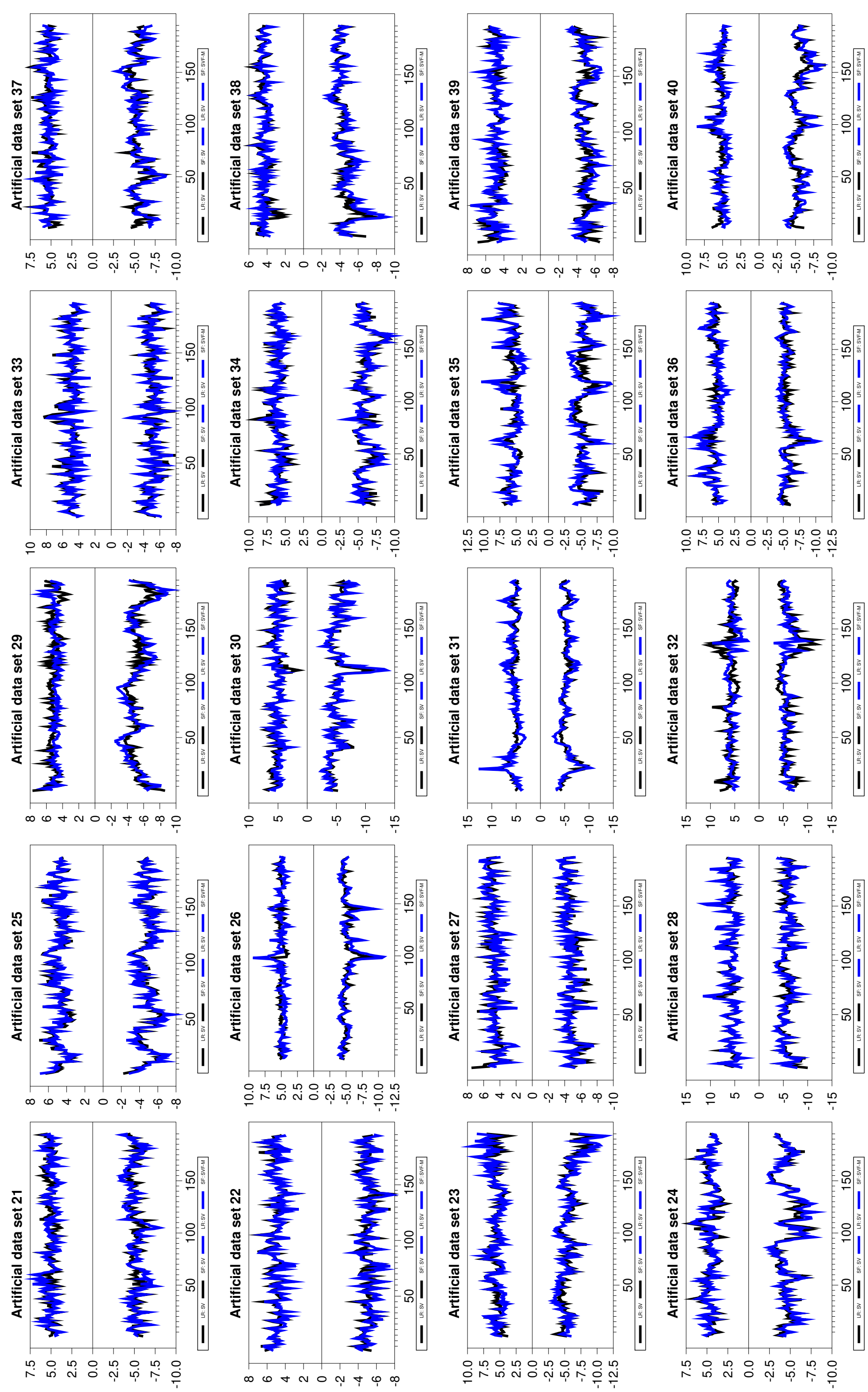


Figure A13: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SVF-M as the DGP: artificial data sets 41 through 60

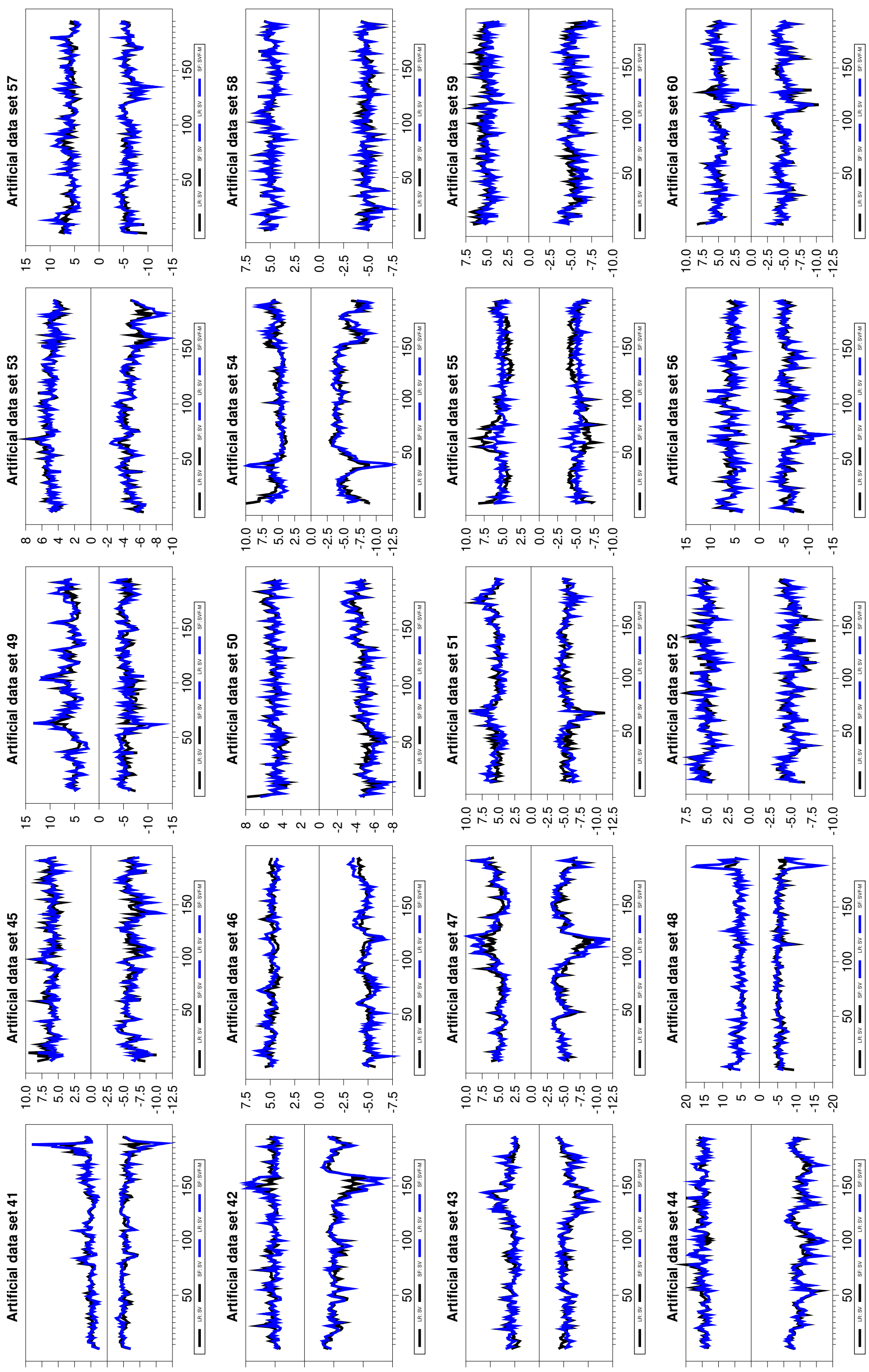


Figure A14: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SVF-M as the DGP: artificial data sets 61 through 80

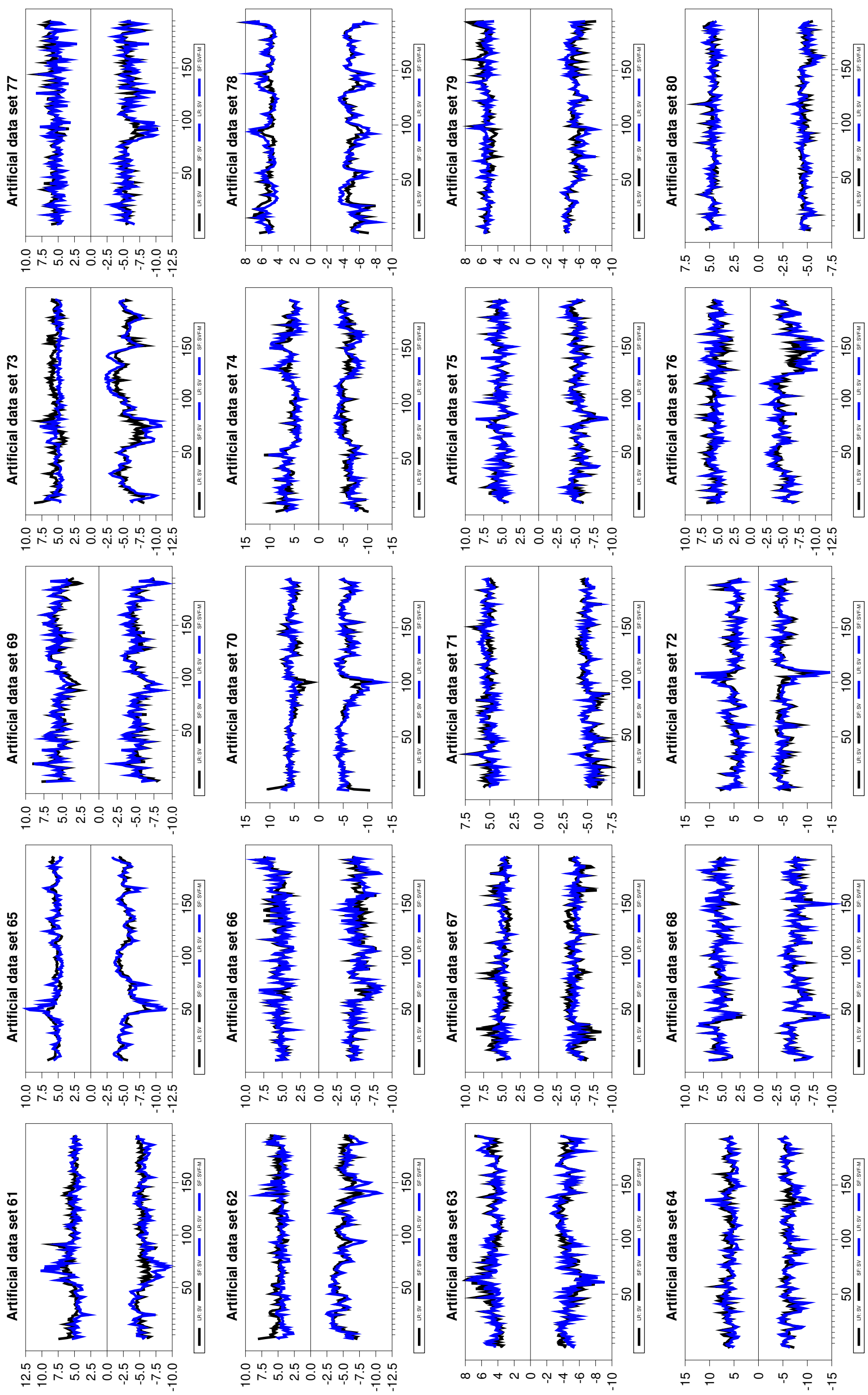


Figure A15: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SVF-M as the DGP: artificial data sets 81 through 100

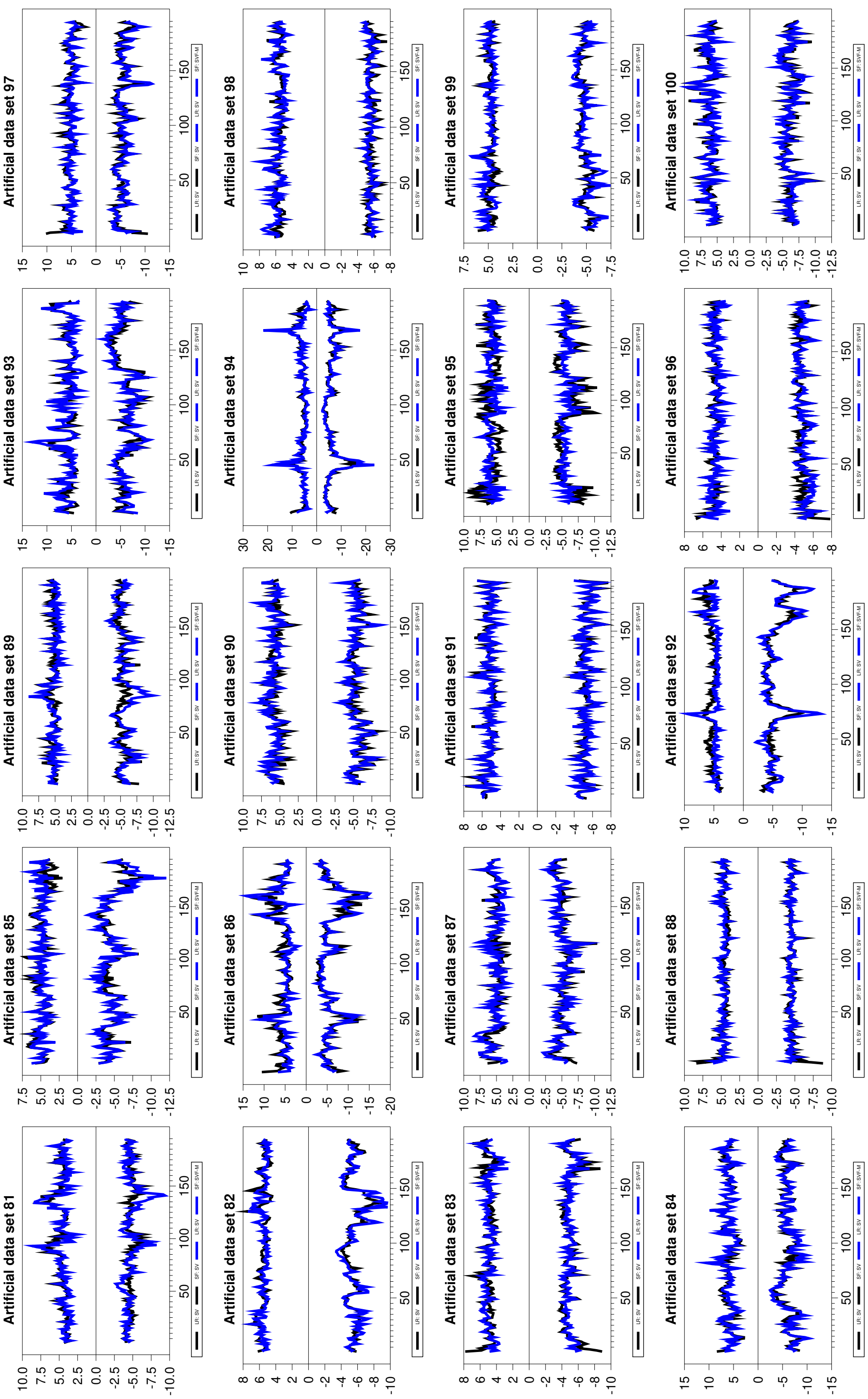


Figure A16: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SV-CSZ DGP: artificial data sets 1 through 20

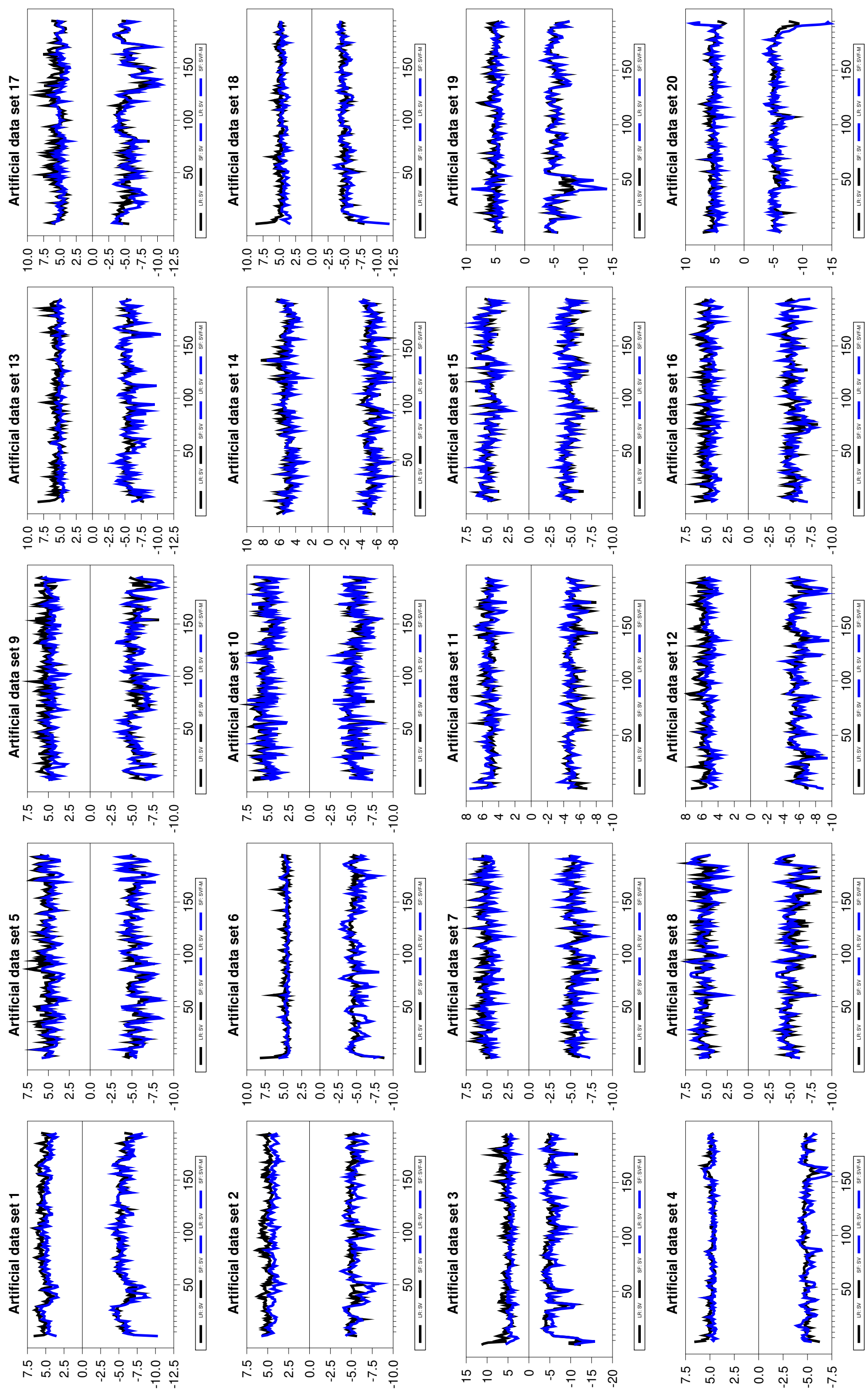


Figure A17: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SV-CSZ DGP: artificial data sets 21 through 40

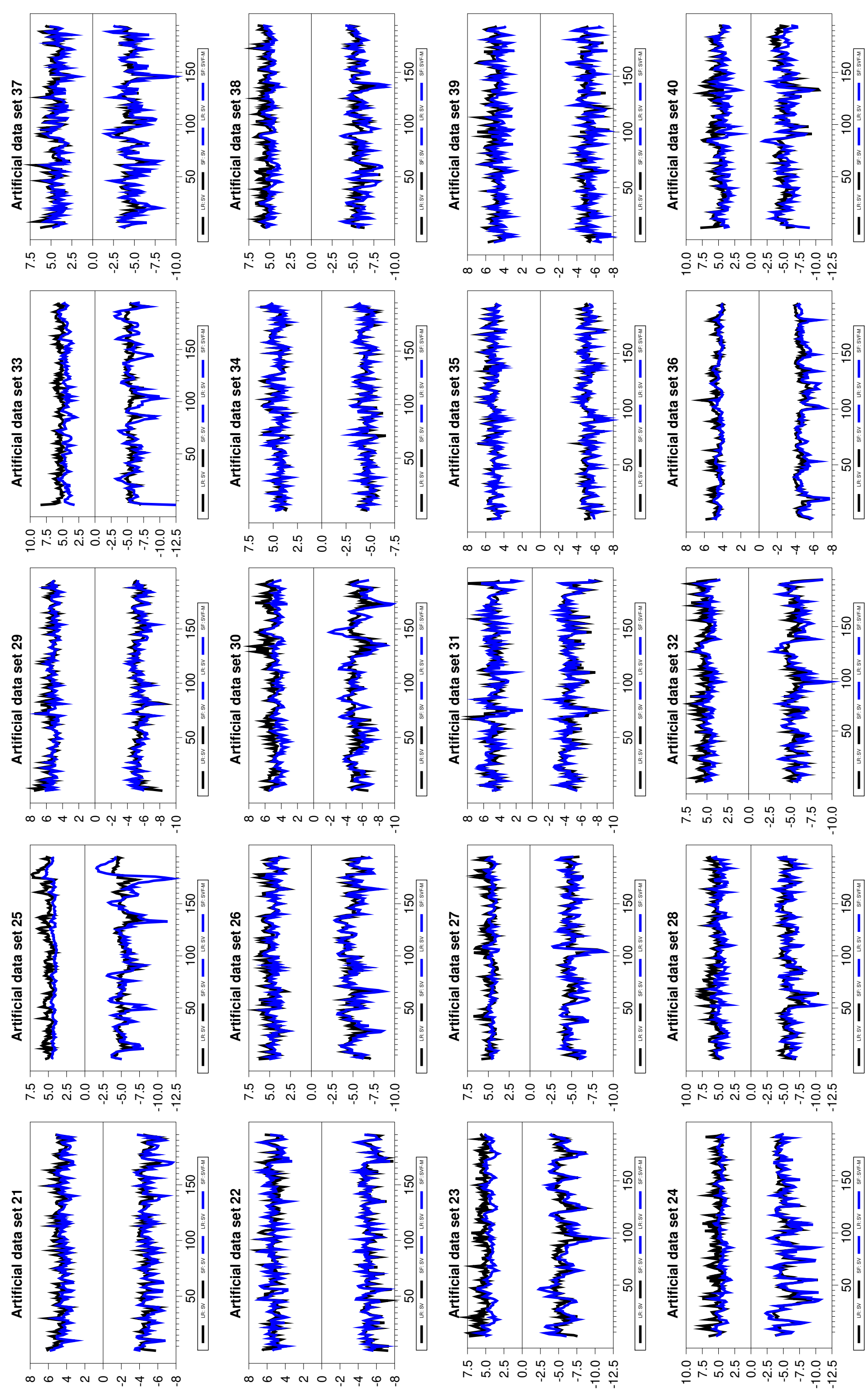


Figure A18: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SV-CSZ DGP: artificial data sets 41 through 60

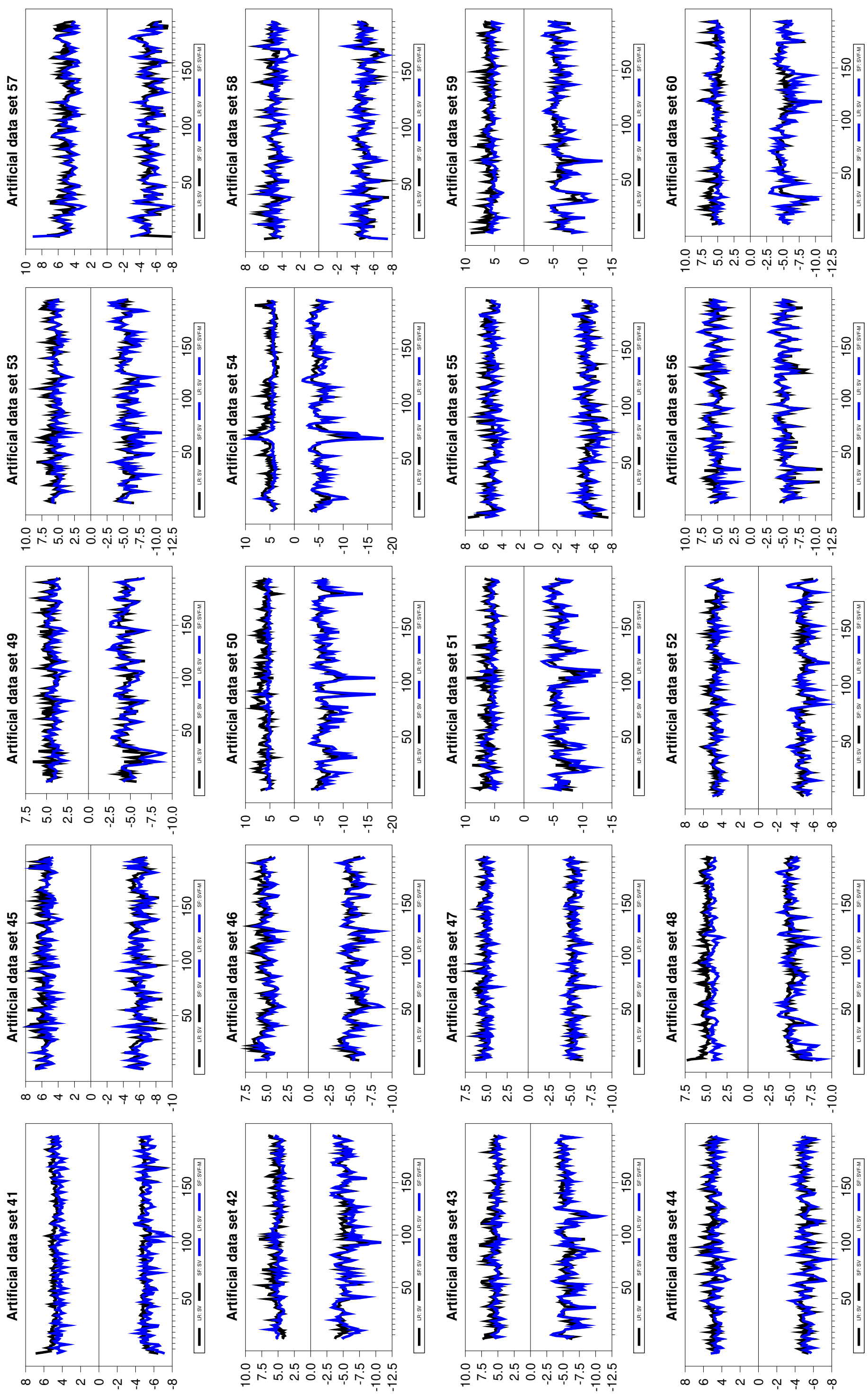


Figure A19: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SV-CSZ DGP: artificial data sets 61 through 80

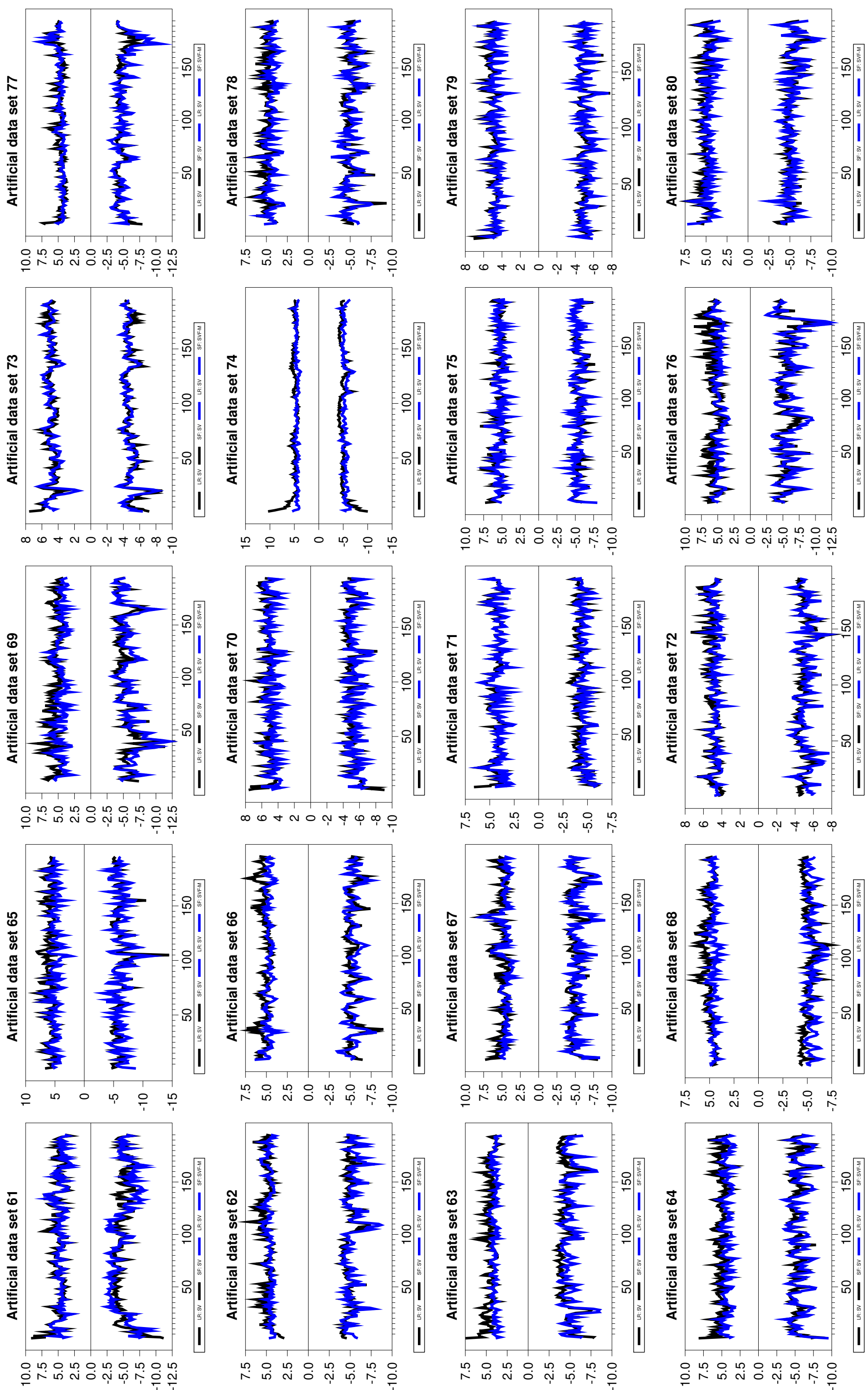


Figure A20: Expected shortfall and long-rise estimates from 1-step-ahead in-sample forecasts obtained with the BVAR-SV and BVAR-SVF-M models, with the BVAR-SV-CSZ DGP: artificial data sets 81 through 100

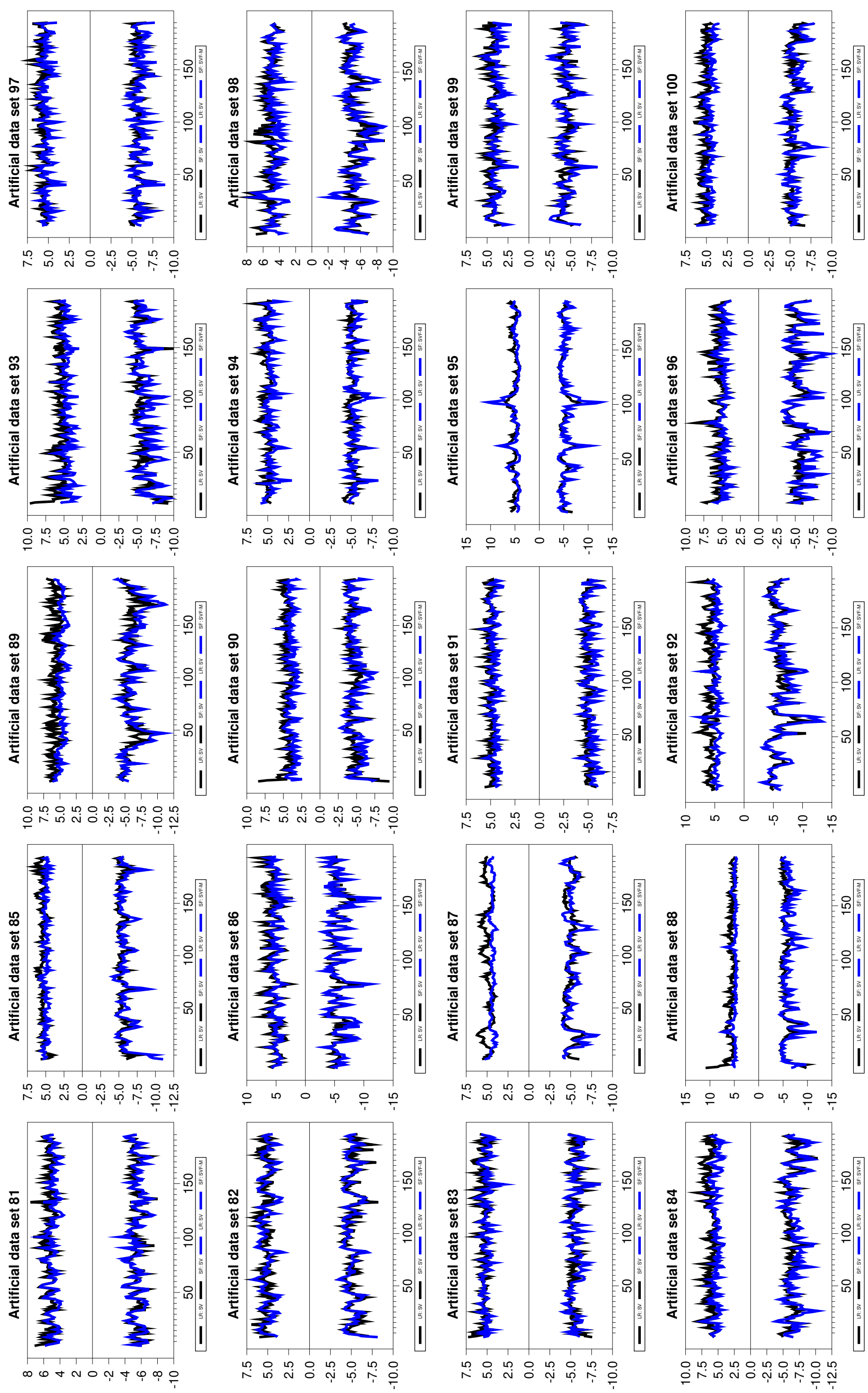
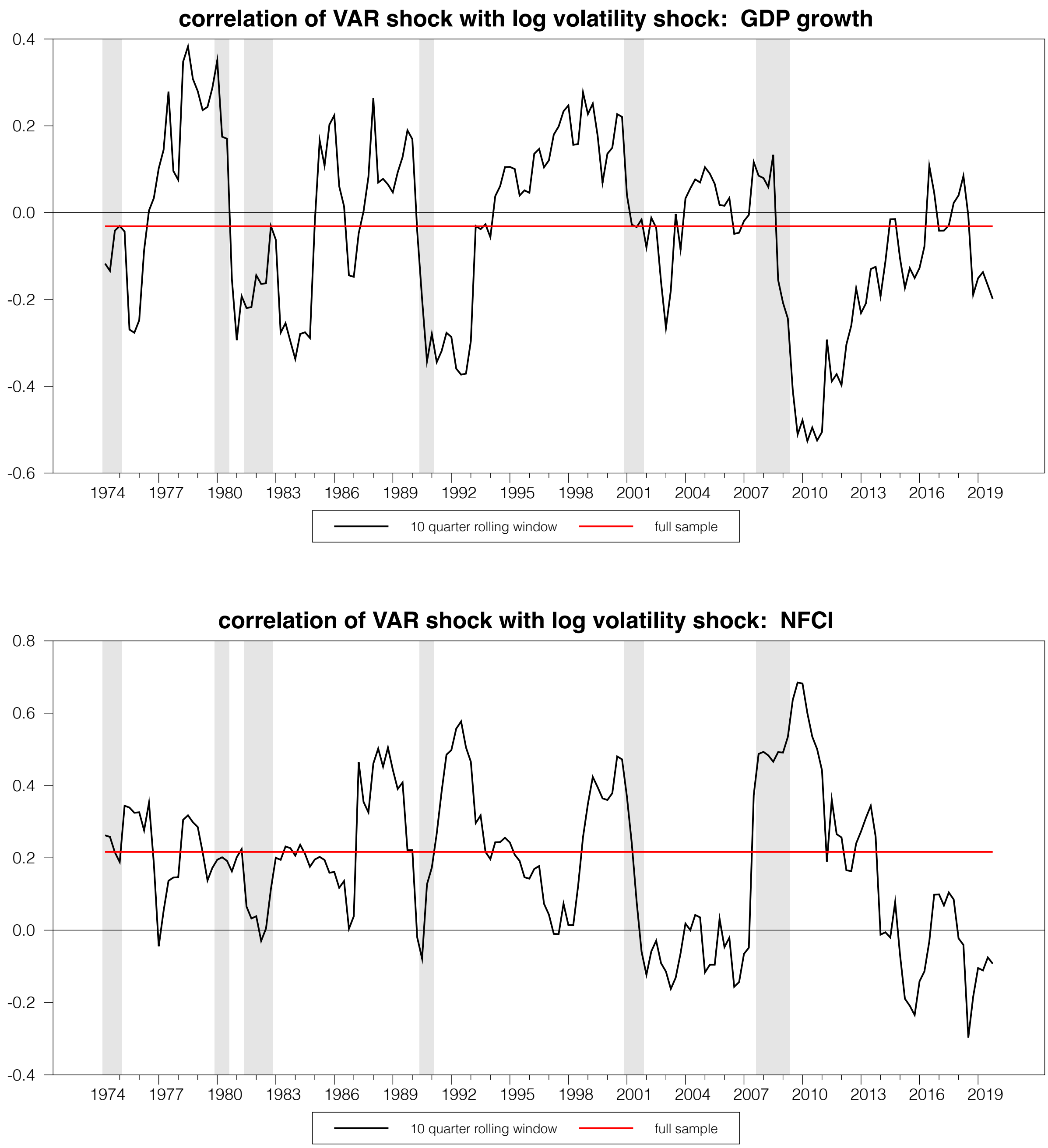


Figure A21: Correlations between levels and volatilities shocks of BVAR-SV model estimated for 1972-2019



Notes: Correlations between levels and volatilities shocks of BVAR-SV model estimated for 1972-2019. 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.

References

[1] 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.

[2] 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.

[3] 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.

[4] Brownlees, Christian, and Andre B.M. Souza (2021), "Backtesting Global Growth-at-Risk," *Journal of Monetary Economics* 118, 312-330. https://doi.org/10.1016/j.jmoneco.2020.11.003.

[5] Caldara, Dario, Chiara Scotti, and Molin Zhong (2021), "Macroeconomic and Financial Risks: A Tale of Mean and Volatility," International Finance Discussion Papers 1326, Board of Governors of the Federal Reserve System. https://doi.org/10.17016/IFDP.2021.1326.

[6] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2020), "Nowcasting Tail Risks to Economic Activity with Many Indicators," *Journal of Applied Econometrics*, forthcoming.

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

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Patton, Andrew J., Johanna F. Ziegel, and Rui Chen (2019), "Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk)," *Journal of Econometrics* 211, 388-413. https://doi.org/10.1016/j.jeconom.2018.10.008.

[19] 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.

[20] 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.