

w o r k i n g
p a p e r

21 08

**Tail Forecasting with Multivariate
Bayesian Additive Regression Trees**

Todd E. Clark, Florian Huber,
Gary Koop, Massimiliano Marcellino, and
Michael Pfarrhofer



FEDERAL RESERVE BANK OF CLEVELAND

ISSN: 2573-7953

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment on research in progress. They may not have been subject to the formal editorial review accorded official Federal Reserve Bank of Cleveland publications. 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 Board of Governors of the Federal Reserve System.

Working papers are available on the Cleveland Fed's website at:

www.clevelandfed.org/research.

Tail Forecasting with Multivariate Bayesian Additive Regression Trees

Todd E. Clark, Florian Huber, Gary Koop,
Massimiliano Marcellino, and Michael Pfarrhofer

We develop novel multivariate time series models using Bayesian additive regression trees that posit nonlinear relationships among macroeconomic variables, their lags, and possibly the lags of the errors. The variance of the errors can be stable, driven by stochastic volatility (SV), or follow a novel nonparametric specification. Estimation is carried out using scalable Markov chain Monte Carlo estimation algorithms for each specification. We evaluate the real-time density and tail forecasting performance of the various models for a set of US macroeconomic and financial indicators. Our results suggest that using nonparametric models generally leads to improved forecast accuracy. In particular, when interest centers on the tails of the posterior predictive, flexible models improve upon standard VAR models with SV. Another key finding is that if we allow for nonlinearities in the conditional mean, allowing for heteroskedasticity becomes less important. A scenario analysis reveals highly nonlinear relations between the predictive distribution and financial conditions.

JEL: C11, C32, C53.

Keywords: nonparametric VAR, regression trees, macroeconomic forecasting.

Suggested citation: Clark, Todd E., Florian Huber, Gary Koop, Massimiliano Marcellino, and Michael Pfarrhofer. 2021. “Tail Forecasting with Multivariate Bayesian Additive Regression Trees.” Federal Reserve Bank of Cleveland, Working Paper No. 21-08. <https://doi.org/10.26509/frbc-wp-202108>.

Todd E. Clark is at the Federal Reserve Bank of Cleveland (todd.clark@clev.frb.org). Florian Huber is at the University of Salzburg. Gary Koop is at the University of Strathclyde. Massimiliano Marcellino is at Bocconi University, IGIER, and CEPR. Michael Pfarrhofer is at the University of Salzburg. Marcellino thanks MIUR – PRIN Bando 2017 – prot. 2017TA7TYC for financial support. Huber and Pfarrhofer gratefully acknowledge financial support from the Austrian Science Fund (FWF, grant no. ZK 35).

1 Introduction

Two recent events, the financial crisis and the COVID-19 pandemic, have increased the interest in tail risks in macroeconomic outcomes. A fast-growing literature has focused on the risks of significant declines in GDP, with quantile regression as the main method to estimate tail risks (see, e.g., [Adrian, Boyarchenko, and Giannone \(2019\)](#); [Adrian, et al. \(2018\)](#); [Cook and Doh \(2019\)](#); [De Nicolò and Lucchetta \(2017\)](#); [Ferrara, Mogliani, and Sahuc \(2019\)](#); [Giglio, Kelly, and Pruitt \(2016\)](#); [González-Rivera, Maldonado, and Ruiz \(2019\)](#); [Delle Monache, De Polis, and Petrella \(2020\)](#); [Plagborg-Møller, et al. \(2020\)](#); [Reichlin, Ricco, and Hasenzagl \(2020\)](#); and [Mitchell, Poon, and Mazzi \(forthcoming\)](#)). For output growth, forecasting tail risks has some precedent in the literature on forecasting recessions or just periods of negative growth (see, e.g., [Aastveit, Ravazzolo, and van Dijk \(2018\)](#)).

In related quantile forecasting work, some studies have considered tail risks to other variables, such as unemployment (e.g., [Galbraith and van Norden \(2019\)](#) and [Kiley \(2018\)](#)) and inflation (e.g., [Ghysels, Iania, and Striaukas \(2018\)](#)). The 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. Other examples of studies of quantile forecasts in macroeconomics include [Gaglianone and Lima \(2012\)](#), [Korobilis \(2017\)](#), and [Manzan and Zerom \(2013, 2015\)](#).

The present paper departs from this existing literature by using Bayesian parametric and nonparametric time series models instead of quantile regression. Our focus is motivated by [Carriero, Clark, and Marcellino \(2020b, CCM\)](#), who provide empirical evidence in favor of Bayesian methods and parametric approaches. In particular, CCM evaluate the ability of alternative econometric methods to produce accurate nowcasts of tail risks to GDP growth, possibly in the presence of a large information set. They find that Bayesian quantile regression performs much better than classical quantile regression and that Bayesian linear regression performs similarly or sometimes better for tail forecasting, once endowed with stochastic volatility (SV). The intuition for this finding, explained more formally in [Carriero, Clark, and Marcellino \(2020a\)](#), is that the explanatory variables drive changes in the conditional mean of growth, which decreases during crisis times, while SV permits an increase in the conditional variance. Thus, the left tail of the conditional distribution of growth can decrease more than the right tail during crisis times, generating the kind of asymmetries emphasized in the quantile regression-based literature. [Caldara, Scotti, and Zhong \(2020\)](#) make a similar point using a model with leverage, where the estimated SV enters the conditional mean with a negative coefficient.

A parallel, and also fast-growing, literature evaluates the use of machine learning techniques for macroeconomic forecasting, with random forests (see [Breiman \(2001\)](#) and, e.g., [Masini, Medeiros, and Mendes \(2021\)](#), for a survey) performing particularly well, also during crisis times, in a variety of studies and for key variables such as GDP growth and inflation; see, e.g., [Goulet Coulombe \(2020\)](#), [Goulet Coulombe, et al. \(2020\)](#); [Goulet Coulombe, Marcellino, and Stevanovic \(2021\)](#), and [Medeiros, et al. \(2021\)](#). While these papers adopt classical methods, Bayesian techniques are also available. In particular, Bayesian additive regression trees (BART; see [Chipman, George, and McCulloch \(2010\)](#)) provide a flexible and popular approach in many fields of statistics. [Huber and Rossini \(2020, HR\)](#) develop Bayesian methods that build BART into a VAR, leading to the Bayesian additive vector autoregressive tree model, and demonstrate that it forecasts well. [Huber, et al. \(2020, HKOPS\)](#) develop Bayesian methods for the mixed-frequency version of this model, showing that it also forecasts well, particularly during the COVID-19 pandemic.

The present paper combines the tail forecasting focus of CCM with the BART methodology of HR and HKOPS. The first contribution of this paper lies in model development. We use four different nonparametric VARs. The first of these is the original model of HR (we use the acronym BART for this model). In addition to this, we introduce three novel alternative BART-based nonparametric VARs that, we argue, have properties that make them potentially useful for macroeconomic forecasting, particularly in unstable times. In the first alternative specification (mixBART), the variables depend on their lags both linearly and nonlinearly, so that the nonlinear component captures what is left unexplained by the linear component. In the second specification (errorBART), the variables depend linearly on their past, but nonlinearly on the past errors, permitting a flexible adjustment of the conditional mean in the presence of large past shocks. In the third specification (fullBART), in addition to a fully nonlinear mean function, we allow for a nonlinear variance-covariance matrix for the errors, which quickly adjusts in the presence of large shocks. Importantly, the nonparametric elements of the models mean they could capture nonlinearities or multi-modalities like those emphasized by [Adrian, Boyarchenko, and Giannone \(2021\)](#), another recent analysis of tail risks to output growth.

The flexible modeling of the conditional mean in BART-based specifications could make the error variance stable and we do consider homoskedastic versions of our nonparametric models. But this is not necessarily the case. Hence, we also consider versions of our models complemented either with SV or with a novel nonparametric specification for the time variation in the conditional variance, related to that in [Pratola, et al. \(2020\)](#), labeled heteroBART.

Our second contribution is the development of Markov chain Monte Carlo (MCMC) es-

timization algorithms for each (homoskedastic and heteroskedastic) BART specification. The algorithms we propose combine state-of-the-art techniques for fast estimation of VAR models (see [Carriero, Clark, and Marcellino \(2019\)](#)) with the Bayesian backfitting step proposed in [Chipman, George, and McCulloch \(2010\)](#). As opposed to [Pratola, et al. \(2020\)](#), we also propose a novel updating step for heteroBART based on using the auxiliary sampler for SV models developed in [Omori, et al. \(2007\)](#). The resulting MCMC algorithm is scalable to large dimension and thus allows for estimating large semi- and nonparametric VAR models. Indeed, our empirical work uses VARs larger than those of [Huber and Rossini \(2020\)](#).

Our final contribution, using (real-time) data for a set of US macroeconomic and financial indicators, is the assessment of the performance of the various BART models for density and tail forecasting, relative to that of a BVAR-SV, a strong benchmark according to, e.g., [Clark \(2011\)](#), [Clark and Ravazzolo \(2015\)](#), [Koop \(2013\)](#) and CCM. The comparison is based on the cumulative ranked probability score (CRPS), with equal weights for all quantiles of the predictive distribution, the quantile-weighted CRPS (qwCRPS) proposed by [Gneiting and Ranjan \(2011\)](#), and the quantile score (QS). The baseline results are for a large VAR with 23 variables. Results for small and moderately sized models are included in an empirical appendix. We also assess the stability of the results over time, with a specific interest in detecting periods where BART forecasts are better than BVAR forecasts.

The empirical results can be summarized as follows. First, most of the BART-based models improve upon the competing linear BVAR with SV. These improvements are especially pronounced for longer horizon forecasts and when the different variants of the CRPS are being considered. When we focus on tail forecasting performance, we also find that controlling for nonlinearities in the conditional mean improves predictive accuracy. Second, when using BART to accommodate nonlinearities, it is less important to allow for heteroskedasticity (for density and tail risk forecasts), with the relevant exception of the unemployment rate. Third, with the caveat in the previous point, when allowing for heteroskedasticity, BART with heteroBART is overall slightly better than BART-SV. Fourth, the more complex BART specifications can be as good as or a little better than the basic BART model of HR, but there are no major systematic gains. Fifth, notwithstanding the rich nonlinear structure, the out-of-sample predictive density charts do not show much downside risk asymmetry, contrary to [Adrian, Boyarchenko, and Giannone \(2019\)](#) and more in line with [Carriero, Clark, and Marcellino \(2020a\)](#). Sixth, the relative ranking of the models is overall stable over time. Seventh, for the various types of BART models, the nonlinear features of the predictive distributions increase with the forecast horizon but are also present in some periods at the one-step-ahead horizon, in particular at the

end of the sample, in association with the COVID-19 period. Finally, conditioning on different values of the national financial conditions index reveals highly nonlinear interactions between one-step-ahead predictive distributions and financial conditions.

The paper proceeds as follows. Section 2 describes the various multivariate Bayesian additive regression tree models and discusses estimation. Section 3 considers the data, forecast design, and evaluation metrics used in the empirical application and discusses empirical findings. Section 4 summarizes and concludes. A supplemental appendix provides additional technical details and empirical results.

2 Nonparametric modeling of VARs using Bayesian additive regression trees

This section explains the BART formulations considered in this paper. In a multivariate time series model such as a VAR, specification choices are made for conditional means and conditional variances. For instance, in the classic BVAR-SV the conditional means are linear and log conditional variances follow random walks. In this paper, we compare this model to various models that are partially or completely nonparametric. In various combinations, the models include nonparametric representations of the conditional mean of a VAR and as well as parametric and nonparametric representations of the conditional variance.

2.1 nonparametric VARs

Let $\{\mathbf{y}_t\}_{t=1}^T$ denote an M -dimensional vector of macroeconomic and financial time series with typical i^{th} element y_{it} . We assume that \mathbf{y}_t depends on its p lags, which we store in a $K(=Mp)$ -dimensional vector $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})'$. The relationship between \mathbf{y}_t and \mathbf{x}_t is assumed to be unknown and potentially highly nonlinear. This is captured through the following general multivariate model:

$$\mathbf{y}_t = F(\mathbf{x}_t) + \boldsymbol{\eta}_t, \tag{1}$$

$$\boldsymbol{\eta}_t = G(\mathbf{z}_t) + \boldsymbol{\varepsilon}_t, \tag{2}$$

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t). \tag{3}$$

Here, we let $F : \mathbb{R}^K \rightarrow \mathbb{R}^M$ and $G : \mathbb{R}^K \rightarrow \mathbb{R}^M$ denote unknown functions with $F(\mathbf{x}_t) = (f_1(\mathbf{x}_t), \dots, f_M(\mathbf{x}_t))'$ and $G(\mathbf{z}_t) = (g_1(\mathbf{z}_t), \dots, g_M(\mathbf{z}_t))'$ while f_j and g_j are equation-specific scalar-valued functions. \mathbf{z}_t is a vector of additional explanatory variables with dimension $N \times 1$

that is defined below in the context of our different models. The shocks in ε_t follow a multivariate Gaussian distribution with a time-varying variance-covariance matrix Σ_t . For the law of motion of Σ_t , we will use both a standard SV model and a more flexible specification closely related to the heteroskedastic BART model proposed in [Pratola, et al. \(2020\)](#). More details are provided in [Section 2.3](#).

The specific choices made for \mathbf{z}_t , F , and G allow for a wide range of flexible models. We discriminate between models that assume that either F or G (or both) is unknown and potentially nonlinear functions. The key notion is that \mathbf{x}_t and \mathbf{z}_t differ in the way they impact \mathbf{y}_t .

In this paper, we obtain four different model specifications that differ in the choice of F , G , and \mathbf{z}_t . The first model is a multivariate nonparametric VAR model. Assuming that $G(\mathbf{z}_t) = 0$ for all t , the model in [Eq. \(1\)](#) and [Eq. \(2\)](#) reduces to:

$$\mathbf{y}_t = F(\mathbf{x}_t) + \varepsilon_t,$$

which posits a nonlinear relationship between \mathbf{y}_t and \mathbf{x}_t and no effect of \mathbf{z}_t on \mathbf{y}_t . If F is approximated using BART, we end up with the model proposed in [Huber and Rossini \(2020\)](#) and applied to the mixed-frequency case in [Huber, et al. \(2020\)](#). In the remainder of the paper, this model is labeled the BART model.

The second model we propose assumes that $\mathbf{z}_t = \mathbf{x}_t$ and $G(\mathbf{x}_t)$ is unknown and nonlinear, while $F(\mathbf{x}_t)$ is linear and depends on an $M \times K$ coefficient matrix \mathbf{A} . The corresponding model reads:

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + G(\mathbf{x}_t) + \varepsilon_t, \tag{4}$$

which is a multivariate additive regression model that assumes that there exists a linear part ($\mathbf{A}\mathbf{x}_t$) and some unknown nonlinear part ($G(\mathbf{x}_t)$). Intuitively speaking, this model assumes that the shocks $\boldsymbol{\eta}_t$ follow a nonlinear regression specification that serves to control for any nonlinear effects that persist after controlling for linear relations. In the remainder of the paper we will estimate G using BART and call this the mixture BART (mixBART) model.

If we set $\mathbf{z}_t = (\boldsymbol{\eta}'_{t-1}, \dots, \boldsymbol{\eta}'_{t-p})'$ and $F(\mathbf{x}_t) = \mathbf{A}\mathbf{x}_t$, the resulting model implies that the reduced-form shocks $\boldsymbol{\eta}_t$ depend nonlinearly on their recent past. This specification allows for flexible adjustments of the conditional mean by exploiting information contained in past reduced-form shocks. During recessions such as the COVID-19 pandemic, this feature could help to

quickly adjust forecasts in the presence of large historical forecast errors. Again, we use BART to approximate G , leading to the errorBART model. Another feasible option, which would require approximation-based techniques along the lines used in [Huber, et al. \(2020\)](#), would be to specify \mathbf{x}_t equal to the lags of $\boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_t$. This would be a nonparametric variant of a multivariate ARMA model (for a Bayesian treatment of ARMA models, see [Chib and Greenberg \(1994\)](#)).

Finally, the last model we consider assumes that $\boldsymbol{\Sigma}_t$ is a diagonal matrix, implying that the shocks $\boldsymbol{\varepsilon}_t$ are independent. This allows us to write Eqs. (1) to (2) as a system of unrelated regression models. The first equation of the model is:

$$y_{1t} = f_1(\mathbf{x}_t) + \varepsilon_{1t}.$$

The second equation depends nonlinearly on \mathbf{x}_t and ε_{1t} as follows:

$$y_{2t} = f_2(\mathbf{x}_t) + g_2(\varepsilon_{1t}) + \varepsilon_{2t}.$$

In general, the i^{th} equation is given by:

$$y_{it} = f_i(\mathbf{x}_t) + g_i(\mathbf{r}_{it}) + \varepsilon_{it}, \tag{5}$$

with $\mathbf{r}_{it} = (\varepsilon_{1t}, \dots, \varepsilon_{i-1,t})'$ being an $(i - 1)$ -dimensional vector of shocks. This model assumes that the contemporaneous relations across the shocks take a nonlinear form. This specification implicitly assumes a nonlinear variance-covariance matrix that quickly adjusts to large shocks. Moreover, it also implies that covariances might change over time, since different configurations of \mathbf{r}_{it} can yield different fitted values. Across all models considered, this model provides the largest degree of flexibility, since it allows for a nonlinear mean function F as well as a nonlinear covariance function G with its argument differing across equations. In what follows, both F and G are again approximated using BART, leading to what we call the fullBART model.

The models we propose are all extremely flexible and thus well suited to capture outliers such as the ones observed during the pandemic. Up to this point, we noted that BART will be used to infer the unknown functions F and G . Many choices are possible, but due to its excellent empirical properties, we adopt BART. The next section introduces BART more formally and briefly summarizes and illustrates its main empirical features.

2.2 Bayesian additive regression trees

BART approximates the unknown functions F and G using a sum of regression trees. In what follows, our focus will be on estimating the function associated with the i^{th} equation.

Let $\mathbf{y}_{i\bullet}$ denote the i^{th} column of $\mathbf{Y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_T)'$, $\mathbf{X} = (\mathbf{x}'_1, \dots, \mathbf{x}'_T)'$, $\mathbf{Z} = (\mathbf{z}'_1, \dots, \mathbf{z}'_T)'$ (with $\mathbf{x}_{\bullet j}$ denoting the j^{th} column of \mathbf{X}) and $\boldsymbol{\varepsilon}_{\bullet i} = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$. The i^{th} equation of the nonlinear VAR outlined in Eqs. (1) to (2) is:

$$\mathbf{y}_{i\bullet} = f_i(\mathbf{X}) + g_i(\mathbf{Z}) + \boldsymbol{\varepsilon}_{\bullet i}. \quad (6)$$

BART replaces f_i and g_i with a sum-of-trees approximation:

$$\begin{aligned} f_i(\mathbf{X}) &\approx \sum_{s=1}^S l_{is}(\mathbf{X} | \mathcal{T}_{is}^f, \boldsymbol{\mu}_{is}^f), \\ g_i(\mathbf{X}) &\approx \sum_{s=1}^S l_{is}(\mathbf{Z} | \mathcal{T}_{is}^g, \boldsymbol{\mu}_{is}^g). \end{aligned}$$

Here, we let l_{is} denote a (single) regression tree function that depends on a so-called tree structure \mathcal{T}_{is}^j for $j \in \{f, g\}$ and a vector of tree-specific terminal node parameters $\boldsymbol{\mu}_{is}^j$ of dimension b_{is}^j . Notice that the trees are determined exclusively by \mathcal{T}_{is}^j and $\boldsymbol{\mu}_{is}^j$.

To illustrate what BART does, we focus on a special case of Eq. (6) assuming a single tree ($S = 1$). To simplify the notation, we drop the tree and equation-specific subscripts i and s as well as the superscript j . The corresponding regression tree model is then given by:

$$\mathbf{y} = l(\mathbf{X} | \mathcal{T}, \boldsymbol{\mu}) + \boldsymbol{\varepsilon}. \quad (7)$$

The tree structure \mathcal{T} is comprised of so-called *interior nodes* that are coupled with decision rules and a set of b *terminal nodes* with terminal node parameter vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_b)'$. The decision rules serve to split the covariate space into several disjoint sets. This is achieved by constructing partition sets associated with each terminal node \mathcal{S}_n ($n = 1, \dots, b$), which depend on whether $\mathbf{x}_{\bullet j} \leq c$ or $\mathbf{x}_{\bullet j} > c$ with c denoting a threshold in the range of $\mathbf{x}_{\bullet j}$.

The conditional mean of this model is given by:

$$E(\mathbf{y} | \mathbf{x}) = l(\mathbf{X} | \mathcal{T}, \boldsymbol{\mu}) = \sum_{n=1}^b \mu_n \mathbb{I}(\mathbf{X} \in \mathcal{S}_n),$$

with \mathbb{I} denoting an indicator function that is equal to one if its argument is true or zero otherwise.

This equation suggests that the conditional mean under a single tree is a piece-wise constant function that assigns μ_n if a specific configuration of \mathbf{X} is in the set \mathcal{S}_n . Notice that this is a simple analysis of variance (ANOVA) model that can be stated in terms of a multivariate regression model conditional on the indicators.

In the case where the tree is simple (i.e., if b is small), the corresponding conditional mean will feature a relatively low number of breaks. Hence, such a model only explains a small fraction of the variation in \mathbf{y} and thus acts as a weak learner. Instead of fitting more complex trees, BART builds on the notion that summing over many simple trees (which are pruned using Bayesian shrinkage) improves upon using a single complex tree. This is illustrated in Eq. (8):

$$E(\mathbf{y}|\mathbf{x}) = \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + \dots + \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \diagup \quad \diagdown \quad \diagup \quad \diagdown \\ \bullet \quad \bullet \quad \bullet \quad \bullet \end{array} \quad (8)$$

The conditional mean of the BART model is composed of the sum of simpler trees, which, when viewed together, allow for capturing rich dynamics in \mathbf{y} . The corresponding additive model has strong explanatory power but regularization helps to avoid issues related to overfitting. Chipman, George, and McCulloch (2010) show that adopting this strategy yields favorable forecasts for 42 different data sets.

2.3 Adding heteroskedasticity to the model

Up to this point we did not discuss the specific law of motion for Σ_t . Several recent papers have shown that allowing for conditional heteroskedasticity sharply improves density forecasts of macroeconomic aggregates (see, among others, Clark (2011); Clark and Ravazzolo (2015); and Carriero, Clark, and Marcellino (2016)). In one set of models, we pair the BART formulations described above with conventional stochastic volatility of the innovations to the model. Standard SV models assume that the latent volatility process evolves according to a simple stochastic process that is persistent (such as a random walk or an AR(1) model with a persistence parameter close to one). During a pandemic, this high persistence in the volatility process could be detrimental for predictive accuracy, since the predictive variance only slowly adjusts to new information. As a solution, Carriero, et al. (2021) discuss several alternative volatility models that allow for combining transitory and persistent changes in the volatility. These models allow for richer volatility dynamics but also assume a parametric and known law of motion.

Accordingly, in another set of results for BART models, we propose an alternative volatility specification based on heteroskedastic BART (Pratola, et al. (2020)). We decompose Σ_t as follows:

$$\Sigma_t = \mathbf{Q}\mathbf{H}_t\mathbf{Q}', \quad (9)$$

with \mathbf{Q} denoting a lower triangular matrix with ones along the diagonal, and $\mathbf{H}_t = \text{diag}(e^{v_1(\mathbf{w}_t)}, \dots, e^{v_M(\mathbf{w}_t)})$ is a diagonal matrix with $v_j : \mathbb{R}^q \rightarrow \mathbb{R}$ being an unknown function that describes how the error variance is related to a set of q covariates in \mathbf{w}_t .¹ The function v_i is again approximated using BART:

$$v_i(\mathbf{w}_t) \approx \sum_{s=1}^S l_{is}(\mathbf{w} | \mathcal{T}_{is}^v, \boldsymbol{\mu}_{is}^v). \quad (10)$$

Selecting appropriate predictors \mathbf{w}_t is crucial. In our empirical work, we consider $\mathbf{w}_t = (t, \mathbf{x}_t')'$. This choice has the advantage that our model allows for a (potentially) nonlinear trend and it assumes that the lagged values of \mathbf{y}_t impact not only the conditional mean but also the error variances. Since the different decision rules might only depend on selected elements in \mathbf{w}_t , we do not risk overfitting if M or p is large. Moreover, and this turns out to be a key advantage, our choice of \mathbf{w}_t allows for multi-steps-ahead predictions of the error variances. More precisely, this is achieved by using Eqs. (1) and (2) to obtain a draw from the one-step-ahead predictive distribution, labeled $\hat{\mathbf{y}}_{T+1}$, which is then used to compute \mathbf{H}_{T+2} based on $\mathbf{w}_{T+2} = (T+2, \hat{\mathbf{y}}_{T+1}', \mathbf{y}_T', \dots, \mathbf{y}_{T-p+1}')'$. \mathbf{H}_{T+2} , in turn, allows us to generate a draw from the two-steps-ahead predictive distribution, $\tilde{\mathbf{y}}_{T+2}$, which is based on \mathbf{H}_{T+2} . In general, the h -steps-ahead forecast distribution can be obtained analogously.

We also consider versions of our models that have conventional SV where the log-volatilities in each equation are assumed to follow AR(1) processes.

These BART-based models are compared to various linear BVARs. All of our BVARs are specified (including prior choices where relevant) as special cases of the mixBART model with G removed and separate horseshoe priors used on the linear VAR coefficients and the free elements in \mathbf{Q} (for details, see Appendix A).

¹The assumption of a time-invariant \mathbf{Q} is common in the forecasting literature and builds on the empirical evidence in Primiceri (2005) that time variation in the covariances is typically limited.

2.4 Bayesian inference

We estimate our model using Bayesian techniques. The prior setup closely follows [Huber, et al. \(2020\)](#). Here, we focus on the prior associated with the tree structures \mathcal{T}_{is}^j and the terminal node parameters μ_{is}^j . [Chipman, George, and McCulloch \(2010\)](#) build on [Chipman, George, and McCulloch \(1998\)](#) and propose a benchmark prior that induces shrinkage on the trees as well as the terminal node parameters. We adopt this prior since it has been shown to work well for a wide variety of different data sets and for both in- and out-of-sample applications. Since the priors on the remaining coefficients are relatively standard, we provide additional information in [Appendix A](#).

2.4.1 Priors on the trees and terminal node parameters

We do not specify a prior directly on the trees but instead design a tree-generating stochastic process that serves as a prior. This process features three aspects. The first is related to the probability that a given node is nonterminal. Let $d = 0, 1, \dots$ denote the depth of some tree and $\alpha \in (0, 1)$ and $\beta \in \mathbb{R}^+$ be hyperparameters. The probability that a node at depth d is nonterminal is given by:

$$\frac{\alpha}{(1 + d)^\beta}.$$

In our empirical work, we set $\alpha = 0.95$ and $\beta = 2$ for the trees \mathcal{T}_{is}^j for all i, s, j . [Chipman, George, and McCulloch \(2010\)](#) recommend these values for α and β as a standard choice that works well across a wide variety of different data sets. This prior implies that the probability that trees grow large decreases in d and thus favors smaller trees.

The second aspect of the prior is concerned with the selection of the variables that are used in a splitting rule. Here, we use a discrete uniform prior, which implies that we do not introduce prior information on which variables show up in a splitting rule. Finally, the third component is concerned with the specific value of the thresholds in the splitting rule. For these, we use a uniform prior over the range of the splitting variable as well.

On the terminal node parameters, we use independent Gaussian priors that are specified as follows:

$$\mu_{is,k}^j \sim \mathcal{N}(0, \phi_{is,k}^j), \text{ for } k = 1, \dots, b_{is}^j. \quad (11)$$

Following [Chipman, George, and McCulloch \(2010\)](#) we set the prior variance $\phi_{is,k}^j$ in a data-

based way. The key idea is to specify the prior such that a certain amount of prior mass is placed on the range of the data but at the same time set the prior in a way that it introduces more shrinkage if S is large. A specification for $\phi_{is,k}^j$ that achieves this is:

$$\sqrt{\phi_{is,k}^j} = \frac{\max(\mathbf{z}_i^j) - \min(\mathbf{z}_i^j)}{2\gamma\sqrt{S}}, \quad (12)$$

where \mathbf{z}_i^j is a T -dimensional vector that is equal to $\mathbf{z}_i^f = \mathbf{y}_{i\bullet} - g_i(\mathbf{Z})$ if $j = f$, $\mathbf{z}_i^g = \mathbf{y}_{i\bullet} - f_i(\mathbf{Z})$ if $j = g$, and $\mathbf{z}_i^v = \log((\mathbf{y}_{i\bullet} - f(\mathbf{Z}) - g_i(\mathbf{Z}))^2)$. The parameter γ controls the tightness of the prior, with smaller values leading to a prior that puts more prior mass on the range of \mathbf{z}_i^j . As noted by [Huber, et al. \(2020\)](#) this prior has the advantage of becoming increasingly loose (for fixed values of S and γ) if \mathbf{z}_i^j includes outliers. This leads to a wider predictive distribution and thus a higher likelihood of observing outlying values. [Chipman, George, and McCulloch \(2010\)](#) propose $\gamma = 2$ in combination with transforming the data such that \mathbf{z}_i^j ranges from -0.5 to 0.5 (implying that the numerator in [Eq. \(12\)](#) is equal to 1). These choices translate into a 95 percent probability that $\mu_{is,k}^j$ is in the range of \mathbf{z}_i^j .

In our empirical work, we will use the same prior hyperparameters γ , α , and β for all equations in the VAR and for all $j \in \{f, g, v\}$. This choice reflects findings in [Pratola, et al. \(2020\)](#) that the choice of the hyperparameters also works well for heteroskedastic BART.

Posterior and predictive inference is done using MCMC methods. The full conditional posterior distributions of the model parameters are mostly available in closed form or can be obtained using a Metropolis-Hastings (MH) step. The conditional posteriors of the covariance parameters in \mathbf{Q} and the VAR coefficients \mathbf{A} take a well-known conditionally Gaussian form and are thus discussed in the technical appendix. Here, we focus on how to sample the tree-specific structure used to approximate G and F .

Our MCMC algorithm is based on the algorithm proposed in [Carriero, Clark, and Marcellino \(2019\)](#). This implies that the model in [Eqs. \(1\) and \(2\)](#) can be, conditional on Σ_t , written as a system of M independent regression models. The i^{th} equation closely resembles [Eq. \(5\)](#):

$$y_{it} = f_i(\mathbf{x}_t) + g_i(\mathbf{z}_t) + \mathbf{q}_i' \mathbf{r}_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}\left(0, e^{v_j(\mathbf{w}_t)}\right), \quad (13)$$

where $\mathbf{q}_i = (q_{i1}, \dots, q_{ii-1})'$ denotes the free elements of the i^{th} row in \mathbf{Q} and \mathbf{r}_{it} is the $(i-1)$ -vector of shocks defined below [Eq. \(13\)](#). For later convenience we let $\mathbf{r}_i = (\mathbf{r}_{i1}, \dots, \mathbf{r}_{iT})'$ denote a $T \times (i-1)$ matrix of stacked shocks.

Equation [\(13\)](#) reduces to [Eq. \(5\)](#) if $\mathbf{q}_i = \mathbf{0}_{i-1}$ and $\mathbf{z}_t = \mathbf{r}_{it}$. The other model specifications

follow by suitably choosing g_i and \mathbf{z}_t . Equation (13) is a very general regression model with a scalar response. In what follows, we will discuss how to simulate the trees and terminal node parameters using Eq. (13) (i.e., on an equation-by-equation basis).

We sample the trees using the Bayesian backfitting strategy discussed in Chipman, George, and McCulloch (2010). This step samples each tree conditional on the remaining $S - 1$ trees. Let $\tilde{\mathbf{z}}_{in}^f = \mathbf{z}_i^f - \sum_{n \neq s} l_{is}(\mathbf{X} | \mathcal{T}_{is}^f, \boldsymbol{\mu}_{is}^f)$, $\tilde{\mathbf{z}}_{in}^g = \mathbf{z}_i^g - \sum_{n \neq s} l_{is}(\mathbf{Z} | \mathcal{T}_{is}^g, \boldsymbol{\mu}_{is}^g) - \mathbf{r}_i \mathbf{q}_i$, and $\tilde{\mathbf{z}}_{in}^v = \mathbf{z}_i^v - \sum_{n \neq s} l_{is}(\mathbf{w} | \mathcal{T}_{is}^v, \boldsymbol{\mu}_{is}^v) - \mathbf{r}_i \mathbf{q}_i$ denote partial residual vectors with the n^{th} tree l_{in} excluded.

Conditional on the partial residual vector $\tilde{\mathbf{z}}_{in}^j$ (for $j \in \{f, g\}$) and the full history of the latent error variances $\mathbf{h}_i = (v_i(\mathbf{w}_1), \dots, v_i(\mathbf{w}_T))'$ (which are modeled using heteroBART or SV), we simulate the tree structures \mathcal{T}_{in}^j and terminal node parameters $\boldsymbol{\mu}_{is}^j$. Chipman, George, and McCulloch (2010) draw \mathcal{T}_{in}^j marginally of $\boldsymbol{\mu}_{is}^j$:

$$p(\mathcal{T}_{in}^j | \tilde{\mathbf{z}}_{in}^j, \mathbf{h}_i) \propto p(\mathcal{T}_{in}^j) \int p(\tilde{\mathbf{z}}_{in}^j | \mathcal{T}_{in}^j, \boldsymbol{\mu}_{in}^j, \mathbf{h}_i) p(\boldsymbol{\mu}_{in}^j | \mathcal{T}_{in}^j, \mathbf{h}_i) d\boldsymbol{\mu}_{in}^j.$$

This integral can be solved analytically (up to a normalizing constant). To draw from $p(\mathcal{T}_{in}^j | \tilde{\mathbf{z}}_{in}^j, \mathbf{h}_i)$ we use the MH algorithm originally proposed in Chipman, George, and McCulloch (1998). Since the tree structure features a discrete state space, the MH algorithm specifies a transition kernel $q(\mathcal{T}_{in}^{j(a)}, \mathcal{T}_{in}^{j*})$ that is used to grow new trees \mathcal{T}_{in}^{j*} , conditional on the previously accepted tree structure $(\mathcal{T}_{in}^{j(a)})$, using one of four distinct moves with prespecified probabilities:

- Grow** The first possible move is to grow a terminal node. This move randomly selects a terminal node of $\mathcal{T}_{in}^{j(a)}$ and then proposes to split this terminal node into two new terminal nodes based on a random splitting rule. This move is selected with probability 0.25.
- Prune** The second move prunes a terminal node. It selects two terminal nodes and merges by collapsing the nodes below. This move is selected with probability 0.25.
- Change** This step random selects an interior node and changes the previously used splitting rule by assigning a splitting rule. This splitting rule is obtained by randomly drawing a splitting variable from the prior (which follows a discrete uniform distribution) and a corresponding threshold. We select this move with probability 0.40.
- Swap** The final step swaps a splitting rule between parent and child nodes (a child node is one that arises from some other node). This move is used with probability 0.10.

These four moves yield a tree \mathcal{T}_{in}^{j*} that is then accepted with probability:

$$\min \left(\frac{p(\mathcal{T}_{in}^{j*} | \tilde{\mathbf{z}}_{in}^j, \mathbf{h}_i)}{p(\mathcal{T}_{in}^{j(a)} | \tilde{\mathbf{z}}_{in}^j, \mathbf{h}_i)} \frac{q(\mathcal{T}_{in}^{j*}, \mathcal{T}_{in}^{j(a)})}{q(\mathcal{T}_{in}^{j(a)}, \mathcal{T}_{in}^{j*})}, 1 \right). \quad (14)$$

This MH update has the advantage of being independent of the terminal node parameters and thus avoids issues with computationally involved reversible jump MCMC algorithms.

In the case of \mathcal{T}_{is}^v we first render the model conditionally Gaussian using the approximation proposed in [Omori, et al. \(2007\)](#). Squaring and taking logs of ε_{it} yields:

$$\tilde{y}_{it} = \log(\varepsilon_{it}^2) = \sum_{is}^S l_i(\mathbf{w}_t | \mathcal{T}_{is}^v, \boldsymbol{\mu}_{is}^v) + \varpi_{it}, \quad \varpi_{it} \sim \log \chi_1^2.$$

ϖ_{it} is then simply approximated using a scale-location mixture of Gaussians with 10 components. The resulting model is a standard BART model with heteroskedasticity and a time-varying intercept. More precisely,

$$\tilde{y}_{it} | \xi_t = j \sim \mathcal{N} \left(\sum_{is}^S l_i(\mathbf{w}_t | \mathcal{T}_{is}^v, \boldsymbol{\mu}_{is}^v) + \mathbf{m}_j, \mathbf{s}_j^2 \right), \quad (15)$$

with \mathbf{m}_j and \mathbf{s}_j^2 being the mean and variance of the j^{th} Gaussian component, respectively. ξ_t denotes a component indicator that takes values between 1 and 10 with $\text{Prob}(\xi_t = j) = \mathbf{q}_j$. The values of \mathbf{m}_j , \mathbf{s}_j^2 , and \mathbf{q}_j are known and can be read off Table 1 in [Omori, et al. \(2007\)](#). Equation (15) is a standard BART model with time-varying intercept and variance, and the trees \mathcal{T}_{is}^v can be sampled with the same MH step outlined above. The main difference with respect to the model outlined in [Pratola, et al. \(2020\)](#) is that they restrict the trees to be nonnegative and then, instead of assuming a sum, approximate the unknown positive function using a product of trees. We found in limited experiments that if one of the trees is sufficiently close to zero, mixing issues arise. Our approach circumvents these issues by estimating a conditional BART model for the log volatilities.

Conditional on the tree structures, the terminal node parameters for all different types of BART models we consider are easily simulated from independent Gaussian distributions. These take a standard form and resemble the one of a simple intercept model. The tree structure serves to allocate observations to different terminal nodes, and these observations are then consequently used to compute the posterior moments. If \mathbf{y}_t contains severe outliers (such as the ones observed during the pandemic), BART will most likely group them together and the corresponding terminal node parameter will have a posterior variance that is equal to the inverse

of the number of outliers plus the prior precision (which will be low; see Eq. (12)). Hence, the corresponding posterior variance will be large, which leads to wider predictive intervals and thus a higher probability of observing outliers under the posterior predictive distribution.

In case of the standard BART-based VAR, these methods are the same as the ones discussed in Huber and Rossini (2020) and a special case of the one developed in Huber, et al. (2020). The steps necessary to simulate each tree and the corresponding terminal node parameters individually are then combined with the steps outlined in Appendix A. This yields an MCMC algorithm that operates on an equation-by-equation basis by iteratively sampling from the relevant full conditionals. The equation-by-equation nature appreciably speeds up estimation, especially if we use a linear VAR component in the model. The steps necessary for sampling the trees and terminal node parameters depend on K only indirectly through the decision rules, leading to a larger state space that the MH algorithm needs to explore. This might cause mixing issues, but we have noticed that in cases where K is moderate to large (i.e., K up to 100), no mixing issues arise. Specifically, in our empirical work we repeat our algorithm 30,000 times and discard the first 15,000 draws as burn-in. MCMC convergence diagnostics based on the full sample corroborate findings in Chipman, George, and McCulloch (2010) and illustrate that our algorithm quickly converges toward the desired stationary distribution.

3 Empirical application

To assess the efficacy of BART-based models for macroeconomic forecasting, we evaluate the accuracy of real-time density and tail risk forecasts from VARs and BART models estimated with a set of 23 quarterly variables for the US. A number of studies have found that larger VARs of this dimension forecast as well as or better than smaller VARs (e.g., Banbura, Giannone, and Reichlin (2010); Koop (2013); and Carriero, Clark, and Marcellino (2019)). Our variable set is patterned after that of Giannone, Lenza, and Primiceri (2015). Note that we add to their variable set the broad index of financial conditions published by the Federal Reserve Bank of Chicago (NFCI), which, starting with the work of Adrian, Boyarchenko, and Giannone (2019), is frequently used in the literature on assessing macroeconomic tail risks. With an eye to brevity, we focus our results on a few broad key indicators: GDP growth, inflation in the GDP price index, and the unemployment rate.

Table 1: Data, description and information set.

FRED-Code	Series	Trans.	Small	Medium	Large
GDPC1	Real gross domestic product (GDP)	$400\Delta \ln$	x	x	x
GDPCTPI	GDP price index	$400\Delta \ln$	x	x	x
FEDFUNDS	Federal funds rate	level	x	x	x
UNRATE	Unemployment rate	$(-1) \times \Delta$	x	x	x
CPIAUCSL	Consumer price index (CPI)	$400\Delta \ln$			x
PPIACO	Producer price index (PPI) for all commodities	$400\Delta \ln$			x
INDPRO	Industrial production	$400\Delta \ln$			x
PAYEMS	Payroll employment	$400\Delta \ln$			x
CES0800000001	Payroll employment, services	$400\Delta \ln$			x
PCECC96	Real personal consumption expenditures	$400\Delta \ln$		x	x
A008RA3Q086SBEA	Gross private domestic fixed investment: Nonres.	$400\Delta \ln$			x
A011RA3Q086SBEA	Gross private domestic fixed investment: Res.	$400\Delta \ln$			x
PCEPTI	PCE chain price index	$400\Delta \ln$			x
GPDICTPI	Gross private domestic investment	$400\Delta \ln$		x	x
CUMFNS	Capacity utilization, manufacturing	level			x
HOANBS	Nonfarm business sector: Hours of all persons	$400\Delta \ln$		x	x
COMPRNFB	Nonfarm bus. sector: Real compensation per hour	$400\Delta \ln$		x	x
GS1	1-Year Treasury bond yield	level			x
GS5	5-Year Treasury bond yield	level			x
EXUSUK	US / UK exchange rate	$400\Delta \ln$			x
M2REAL	Real M2 money stock	$400\Delta \ln$			x
SP500	S&P 500	$400\Delta \ln$			x
NFCI	Chicago Fed index of financial conditions	level	x	x	x

Notes: “FRED-Code” refers to the code of the respective series at fred.stlouisfed.org. Transformations (“Trans.”): Δ indicates first differences and \ln is the natural logarithm. For model sizes, x marks inclusion in the respective information set (Small, Medium, and Large).

3.1 Data description

Table 1 lists the variables we use (alongside codes and transformations).² Data are obtained from fred.stlouisfed.org. To facilitate the evaluation of tail forecasts (downside risks in particular), we include the unemployment rate with a sign switch.³ All the models we consider set $p = 5$.

With real-time data vintages available beginning with 1996:Q4, our real-time forecast sample begins with 1997:Q1 and ends with 2020:Q4.⁴ However, for some variables, real-time data vintages begin later in the sample. In these cases, we use the first vintage to fill in artificial vintages for earlier years, truncating it according to the release calendar. In all cases, the data sample for model estimation starts with 1973:Q2. In evaluating forecasts, we measure the actual values of the variables as those of the final available vintage, which is 2021:Q1.⁵

²We work with a large data set involving all the listed variables. However, we have also experimented with small and medium sized data sets. Results for these are available in the online appendix. Table 1 lists which variables are in the small and medium sized data sets.

³Because the S&P 500 index of stock prices is unavailable prior to 2011 in the online FRED database, we obtained data for this series prior to 2011 from the compiled “FRED-QD” data set, also available from the St. Louis Fed’s website.

⁴If release frequency is higher than quarterly, we use the final vintage per respective quarter for producing forecasts.

⁵Our data set, thus, includes observations during the pandemic. The usefulness of BART for pandemic forecasting was established in previous work by several of the authors; see HKOPS. To convince the reader that our findings in the present paper are not unduly influenced by the pandemic observations, we have also repeated the forecast exercise using data through the end of 2019. Results are similar to those presented here and are available in the online appendix to this paper.

3.2 Forecast evaluation metrics

In evaluating real-time out-of-sample forecasts, we consider a range of metrics, with a focus on tail risk.

As a baseline assessment of overall density accuracy, we use the cumulative ranked probability score (CRPS), developed in [Gneiting and Raftery \(2007\)](#) (with equal weights for all quantiles of the predictive distribution). The CRPS, defined such that a lower number is a better score, is given by

$$CRPS_t(y_{it}) = \int_{-\infty}^{\infty} (\mathfrak{F}(z) - \mathbb{I}\{y_{it} \leq z\})^2 dz = E_{\mathfrak{f}}|\hat{y}_{it} - y_{it}| - 0.5E_{\mathfrak{f}}|\hat{y}_{it} - \hat{y}'_{it}|,$$

where \mathfrak{F} denotes the cumulative distribution function associated with the predictive density \mathfrak{f} , y_{it} ($1, \dots, M$) is the realization of the forecasted variable, $\mathbb{I}\{y_{it} \leq z\}$ is an indicator function taking value 1 if $y_{it} \leq z$ and 0 otherwise, and \hat{y}_{it} and \hat{y}'_{it} are independent random draws from the posterior predictive density.

As a basic measure of accuracy of the lower tail risk forecast, we use the quantile score, commonly associated with the tick loss function (see, e.g., [Giacomini and Komunjer \(2005\)](#)). The quantile score is computed as

$$QS_{\tau t} = (y_{it} - \mathcal{Q}_{\tau t}) (\tau - \mathbb{I}\{y_{it} \leq \mathcal{Q}_{\tau t}\}),$$

where $\mathcal{Q}_{\tau t}$ is the forecast quantile at quantile τ , and the indicator function $\mathbb{I}\{y_t \leq \mathcal{Q}_{\tau t}\}$ has a value of 1 if the outcome is at or below the forecast quantile and 0 otherwise. We evaluate the QS using $\tau = 0.05, 0.10$, and 0.25 .

We also evaluate tail forecast accuracy using two implementations of the quantile-weighted CRPS (qwCRPS) developed by [Gneiting and Ranjan \(2011\)](#) as a proper scoring function of the entire predictive density. The qwCRPS is computed as a weighted sum of quantile scores at a range of J quantiles:

$$\text{qwCRPS}_t = \frac{2}{J-1} \sum_{j=1}^{J-1} \omega(\tau_j) QS_{\tau_j t}, \quad (16)$$

with $\tau_j = j/J$. We rely on a grid of 19 quantiles $\tau \in \{0.05, 0.10, \dots, 0.90, 0.95\}$ with $J = 20$ to compute these weighted scores. In one implementation (denoted qwCRPS-tails), we set the weights as $\omega(\tau_j) = (2\tau_j - 1)^2$ in order to target both tails of the predictive distribution, and in the other (denoted qwCRPS-left), we set the weights to $\omega(\tau_j) = (1 - \tau_j)^2$ in order to target the

left tail (downside risk).

The tables provided below report averages of these score measures over the 1997-2020 sample. We also produce results using a sample that ends in 2019. These are available in the online appendix and are similar to those using the full sample. To facilitate comparisons, we report scores relative to those of a benchmark BVAR with stochastic volatility. By all metrics, a ratio of less than 1 means that a given model is improving on the accuracy of the BVAR-SV baseline. To roughly gauge statistical significance, we rely on [Diebold and Mariano \(1995\)](#) and [West \(1996\)](#) t -tests of significance differences of scores, for each model compared to the benchmark.

To give some sense of performance over time, for each forecast metric we also report figures of mean scores computed recursively, relative to the recursive mean for the BVAR-SV benchmark. These recursive scores can be expressed as $FM_{it} = 1/\tilde{t} \times \sum_{t=1}^{\tilde{t}} x_{it}$ with the values shown calculated as $100(FM_{it}/FM_{\text{bench},t} - 1)$. For example, a value of -20 indicates that the respective model has a 20 percent lower measure. The first two years of the holdout are not shown; we use the first eight observations to initialize the recursive means.

3.3 Full density forecast performance

[Table 2](#) and [Figure 1](#) provide results on accuracy measures that refer to the entire predictive density, including the CRPS, qwCRPS-tails, and qwCRPS-left. Throughout, we report results for horizons of $h \in \{1, 4, 8, 12\}$ quarters. We use an iterative approach to produce higher-order forecasts. The acronyms in the tables can be understood by noting they combine the acronyms with the various specifications for the conditional mean (BVAR, BART, mixBART, errorBART, and fullBART) with the acronyms for different treatments of the conditional variance (SV and heteroBART). We also present results for homoskedastic versions of the models that are labeled “cons” in the tables.

Model		CRPS				qwCRPS-tails				qwCRPS-left			
		h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
GDPC1													
	BVAR cons	1.087***	1.078**	1.107***	1.126***	1.140***	1.135**	1.164***	1.183***	1.056*	1.038	1.067	1.082
	BART cons	1.021	0.971	0.987	0.998	1.040	0.975	0.982	0.983***	0.998	0.967	1.002	1.003
	mixBART cons	1.016	1.001	1.051*	1.087***	1.015	1.003	1.041	1.078**	1.012	0.998	1.038	1.052
	errorBART cons	0.997	1.015	1.031	1.026	1.007	1.013	1.017	1.005	0.970	0.999	1.012	0.997
	fullBART cons	1.022	0.969	0.988	1.004	1.041	0.972	0.982	0.989***	1.001	0.966	1.004	1.008
	BVAR SV	1.981	2.051	2.074	2.098	0.213	0.225	0.230	0.236	0.309	0.318	0.319	0.323
	BART SV	0.990	0.971	0.982	0.994	1.012	0.975	0.981	0.982**	0.988	0.975	1.003	1.009
	mixBART SV	0.980	1.001	1.019	1.042***	0.996	1.005	1.018	1.037*	0.986	0.999	1.015	1.027
	errorBART SV	0.970	1.007	1.009	1.004	0.981	1.005	1.006	0.996	0.976	1.010	1.011	1.004
	fullBART SV	0.993	0.986	1.012	1.022***	1.025	1.004	1.021	1.021	0.975	0.974	1.014	1.014
	BVAR heteroBART	1.004	0.985**	0.986*	0.984**	1.010	0.982*	0.980	0.971***	1.004	0.995	0.992	0.986**
	BART heteroBART	0.980	0.971	0.971**	0.981**	1.001	0.973	0.967***	0.968***	0.981	0.975	0.990	0.995
	mixBART heteroBART	0.977	0.996	1.008	1.025*	0.989	0.995	1.002	1.014	0.986	0.997	1.007	1.012
	errorBART heteroBART	1.004	1.003	1.017	1.017	1.013	0.999	1.004	0.996	0.979	0.997	1.009	1.000
	fullBART heteroBART	0.984	0.965	0.972**	0.988*	1.003	0.967	0.968***	0.972**	0.986	0.970	0.992	1.000
GDPC1PI													
	BVAR cons	1.022	1.091**	1.082	1.085	1.051	1.143**	1.153**	1.152	1.003	1.005	0.969	0.958
	BART cons	1.047	0.866**	0.754***	0.709***	1.041	0.857***	0.744***	0.703***	1.040	0.879**	0.766***	0.738***
	mixBART cons	1.007	0.954	0.823**	0.788***	0.997	0.925	0.828**	0.848*	1.000	0.982	0.842*	0.824*
	errorBART cons	0.990	1.029	0.927	0.839**	0.966	0.991	0.910	0.834***	0.976	1.019	0.941	0.896**
	fullBART cons	1.022	0.867***	0.757***	0.714***	1.019	0.859***	0.748***	0.710***	1.015	0.884**	0.772***	0.748***
	BVAR SV	0.567	0.690	0.892	1.035	0.058	0.070	0.089	0.105	0.087	0.106	0.128	0.144
	BART SV	1.010	0.858***	0.776***	0.739***	0.995	0.853***	0.771***	0.729***	1.000	0.861**	0.776***	0.758***
	mixBART SV	1.015	0.924	0.819**	0.781***	1.008	0.905**	0.824***	0.826**	1.008	0.947	0.835*	0.814*
	errorBART SV	0.990	0.957	0.849*	0.769**	0.973	0.928	0.835**	0.778***	0.986	0.967	0.880	0.830*
	fullBART SV	1.144**	0.872**	0.744***	0.710***	1.166**	0.870**	0.739***	0.705***	1.153*	0.883*	0.758***	0.739***
	BVAR heteroBART	0.980	0.978	0.973*	0.958**	0.972	0.979	0.975	0.975	0.970	1.003	1.016	1.014
	BART heteroBART	1.121*	0.875**	0.772***	0.730***	1.138**	0.871***	0.763***	0.720***	1.107	0.900*	0.792***	0.773***
	mixBART heteroBART	0.985	0.914*	0.808***	0.778***	0.981	0.890**	0.805***	0.813***	0.973	0.944	0.832**	0.818**
	errorBART heteroBART	0.990	1.026	0.927	0.848**	0.968	0.988	0.908	0.841***	0.979	1.021	0.951	0.920
	fullBART heteroBART	1.089	0.869***	0.753***	0.713***	1.080	0.865***	0.750***	0.707***	1.084	0.884**	0.776***	0.754***
UNRATE													
	BVAR cons	1.177*	1.063**	1.046	1.041	1.232*	1.085**	1.067	1.066	1.100**	1.043	1.021	1.026
	BART cons	1.051	0.992	1.024	1.029*	1.069	0.981	1.009	1.011	0.965	0.969	1.009	1.021
	mixBART cons	1.196	1.010	1.040	1.049*	1.218	0.998	1.025	1.038	1.067	0.992	1.022	1.038
	errorBART cons	1.123	1.008	1.000	0.989	1.154	1.001	0.994	0.986	1.018	0.997	0.986	0.980
	fullBART cons	1.049	0.989	1.027	1.031*	1.061	0.978	1.012	1.014	0.967	0.966	1.010	1.023
	BVAR SV	0.268	0.300	0.320	0.334	0.030	0.034	0.037	0.038	0.045	0.048	0.052	0.054
	BART SV	0.899*	1.001	1.028	1.036*	0.875	0.990	1.013	1.016	0.932**	0.982	1.014	1.029
	mixBART SV	0.938*	0.999	1.020	1.023	0.919	0.996	1.014	1.019	0.972	0.988	1.009	1.020
	errorBART SV	0.967	0.998	0.993	0.985	0.947	0.999	0.998	0.991	1.004	0.999	0.996	0.987
	fullBART SV	0.970	0.996	1.027	1.036*	0.981	0.983	1.010	1.019	0.925	0.976	1.012	1.027
	BVAR heteroBART	1.088	0.980***	0.985**	0.992	1.097	0.977***	0.985	0.990	1.011	0.968***	0.976	0.985
	BART heteroBART	0.958	0.988	1.011	1.019***	0.964	0.977	1.002	1.008	0.923*	0.969	0.997	1.012***
	mixBART heteroBART	0.967	0.988	1.010	1.015	0.970	0.980	1.005	1.012	0.939**	0.969	0.995	1.008
	errorBART heteroBART	1.113	1.003	0.999	0.992	1.146	0.998	0.994	0.990	1.015	0.993	0.986	0.983
	fullBART heteroBART	0.952	0.995	1.014	1.024**	0.960*	0.978	1.002	1.012	0.917*	0.971	0.997	1.017*

Table 2: Cumulative ranked probability score (CRPS) and quantile weighted CRPSs.

Notes: CRPSs are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Quantile weights 'tail' indicate a weighting scheme capturing both tails; 'left' captures performance for downside risks. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

Starting from the baseline of a homoskedastic linear VAR, our results reflect a common pattern in the literature: Including stochastic volatility in the BVAR improves on the accuracy of a homoskedastic, linear VAR. As indicated in the top row of each panel of Table 2, by the CRPS measure of overall density accuracy, the homoskedastic BVAR is typically about 10 percent less accurate than the BVAR-SV model in the cases of GDP growth and inflation and about 5 percent less accurate in the case of the unemployment rate. The results are broadly similar for the CRPS measures that put more weight on the tails of the predictive distribution, with the BVAR-SV’s advantages a little greater in some cases (e.g., with the qwCRPS-tails metric) and a little smaller in other cases (e.g., with the qwCRPS-left metric and inflation and the unemployment rate). These results indicate the BVAR-SV is a good benchmark to match or beat.

Strikingly, the results in Table 2 indicate that a BART specification with homoskedastic innovations can match or beat the accuracy of the BVAR-SV model. The nonparametric specification of the BART model evidently captures nonlinearities in the conditional mean process in such a way as to match or beat the accuracy gains that come from including the time-varying volatility of innovations in a linear VAR. In the case of GDP growth, across the various CRPS measures and forecast horizons, the score ratio for the homoskedastic BART model relative to the BVAR-SV baseline ranges from 0.967 to 1.040, with more entries below 1 than above. The score patterns are broadly similar for the unemployment rate, with a slightly wider range. But in the case of inflation, the homoskedastic BART specification improves materially on the BVAR-SV’s accuracy at horizons of 4 through 12 quarters. For example, 8 or 12 quarters ahead, the CRPS, qwCRPS-tails, qwCRPS-left ratios show accuracy gains of about 25 to 30 percent. These are quite large (and all highly statistically significant) by the standards of the macro forecasting literature.

Adding SV to the BART specification is neither particularly helpful nor harmful. Comparing the BART-SV specification to the homoskedastic BART model, their respective CRPS ratios are typically quite similar. In a few cases, adding SV helps a little or hurts a little, but in most, differences in ratios are negligible. The most noticeable exception occurs with one-quarter-ahead unemployment forecasts, in which case the BART specification with SV yields scores for CRPS and qwCRPS-tails that are roughly 15 percent lower than those for the homoskedastic BART model (the same is not true for qwCRPS-left).

With the conditional mean modeled as a linear VAR, using the BART-based specification of heteroskedasticity in the innovations — the BVAR with heteroBART — yields overall density accuracy very comparable to that of the baseline BVAR-SV. In the case of GDP growth forecasts,

at horizons of 4 through 12 quarters, the CRPS ratios (for all three measures) for this specification consistently show small gains of about 2 to 3 percent, whereas at the one-quarter-ahead horizon, the CRPS ratios are close to 1. Results for inflation show broadly similar performance of the BVAR with heteroBART model. For the unemployment rate, this model also yields gains of 2-3 percent at horizons of 4 or more quarters, but in this case, the BVAR with heteroBART is less accurate than the BVAR-SV baseline at the one-quarter-ahead horizon.

In light of these results, it is natural to examine whether pairing BART for the conditional mean with heteroBART for the conditional variance does better yet, beating the benchmark BVAR-SV and the models that separately introduce either BART or heteroBART. But it turns out that although the BART with heteroBART performs comparably to each of the BVAR-SV, homoskedastic BART, and BVAR with heteroBART specifications, it does not systematically improve on them. For example, with GDP growth forecasts, CRPS ratios for the BART with heteroBART are very similar to those achieved by the BVAR with heteroBART. The same applies to unemployment rate forecasts, except that, at the one-quarter-ahead horizon, BART with heteroBART avoids the deterioration in accuracy relative to the baseline that occurs with the BVAR with heteroBART model. In the case of inflation forecasts at horizons of 4 or more quarters, BART with heteroBART matches homoskedastic BART and BART with SV in materially improving on the accuracy of BVAR-SV forecasts.

In summary, if we examine the forecast metrics based on the entire predictive density, then BART is as good as and often better than BVAR-SV. However, among the different versions of BART, nothing consistently beats the simplest version of BART that has homoskedastic errors. It seems we can get most of the gains from allowing for nonlinearities simply by using BART for the conditional mean.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

The preceding discussion was related to average performance over the forecast evaluation period. To examine whether differences in performance among any of the methods are related to any particular episodes, Figures 1 through 3 plot recursive averages of our three CRPS-based measures. Most of the lines in these figures are roughly constant over time, indicating a lack of variation in the performance of a model relative to the BVAR-SV benchmark. Note, for instance,

the consistently good performance of the various BART approaches for forecasting inflation for horizons $h = 4, 8, 12$.

There is one main model that exhibits substantial variation over time: the homoskedastic BVAR. Overall, this tends to forecast poorly. But this overall performance combines periods of extremely bad forecasting with times when the model forecasts quite well. See, for instance, the $h = 12$ forecasts from this model. Overall, they tend to be poor. But for all of them, there is a period in the financial crisis when the homoskedastic BVAR is forecasting better than any of the alternatives. This holds true regardless of which version of CRPS is used, and it is due to the larger estimated variance of the homoskedastic BVAR, which helps to accommodate some outlying observations occurring during the financial crisis (see [Carriero, Clark, and Marcellino \(2015\)](#) for a similar finding when nowcasting US GDP growth with univariate mixed-frequency models).

3.4 Tail forecasting performance

We now turn to an examination of the tail forecast performance of our various models. [Table 3](#) has a format similar to that of [Table 2](#), but presents quantile scores instead of CRPSs. Broadly speaking, the quantile score results are similar to the CRPS results, telling a story in which homoskedasticity is a bad assumption to make and BART is as good as and typically better than the BVAR-SV benchmark. However, in some cases the results are even more strongly in favor of the nonparametric approaches.

The previous statement is particularly true for the unemployment rate. Consider, for instance, the one-quarter-ahead forecasts. When judged using CRPS-based measures, one of the BART approaches (typically BART SV) forecasts better than the BVAR SV benchmark, but many of the BART approaches forecast roughly the same as the BVAR-SV benchmark. However, with the quantile scores — which isolate the left tail of the predictive distribution — there is a much more consistently good performance for the BART approaches, particularly in the 5 percent and 10 percent tails.

		QS5				QS10				QS25			
Model		h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
GDPC1													
	BVAR cons	1.025	0.971	0.976	0.977	1.081	1.053	1.072	1.080	1.049	1.055	1.096	1.120
	BART cons	0.951	0.926	0.950	0.932*	0.997	0.962	0.986	0.964	1.001	0.976	1.030	1.022
	mixBART cons	0.974	0.920	0.915	0.942	1.002	0.968	0.987	0.994	1.017	1.021	1.048	1.051
	errorBART cons	0.942*	0.953	0.953**	0.937***	0.956*	0.976	0.973*	0.943***	0.971	0.998	1.006	0.992***
	fullBART cons	0.958	0.922	0.945	0.937*	0.998	0.963	0.987	0.971	1.005	0.973	1.033	1.025
	BVAR SV	0.722	0.773	0.796	0.829	0.880	0.933	0.956	0.993	1.166	1.189	1.177	1.193
	BART SV	0.986	0.955	0.960	0.948	1.010	0.977	0.999	0.979	0.994	0.983	1.029	1.035
	mixBART SV	0.987	0.958	0.949	0.946	1.004	0.993	1.000	0.991	0.995	1.013	1.024	1.037
	errorBART SV	0.986	1.017	1.032	1.004	0.986	1.017	1.011	0.998	0.986	1.008	1.012	1.004
	fullBART SV	0.960	0.916	0.943	0.933	0.985	0.974	0.997	0.978	0.981	0.979	1.041	1.038
	BVAR heteroBART	0.998	0.994	0.995	0.962*	1.006	0.997	0.987	0.963***	1.004	0.997	0.993	0.990
	BART heteroBART	0.995	0.964	0.971	0.947**	1.006	0.975	0.984	0.970***	0.985	0.984	1.011	1.010
	mixBART heteroBART	0.987	0.963	0.960	0.942	1.002	0.985	0.989	0.975	0.998	1.005	1.015	1.017
	errorBART heteroBART	0.958*	0.974	0.968	0.954***	0.961**	0.985	0.982	0.954***	0.981	1.001	1.008	0.999
	fullBART heteroBART	1.005	0.959	0.975	0.950**	1.008	0.968	0.984	0.975**	0.991	0.980	1.016	1.012
GDPC1PI													
	BVAR cons	1.030	1.065	1.257**	1.058	0.952	0.997	1.095	0.969	1.032	0.957	0.836	0.838*
	BART cons	1.146	0.978	0.873	0.867	1.033	0.899*	0.796**	0.770**	1.033	0.876*	0.760***	0.746***
	mixBART cons	0.954	0.948	0.996	0.895	0.940	0.972	0.948	0.925	1.017	0.991	0.822	0.875
	errorBART cons	0.914**	0.934	1.012	0.976	0.906*	0.964	0.975	0.977	0.980	1.009	0.931	0.910**
	fullBART cons	1.133	0.967	0.894	0.874	1.018	0.908*	0.809**	0.794**	1.000	0.885*	0.761***	0.755***
	BVAR SV	0.128	0.144	0.149	0.199	0.214	0.240	0.254	0.318	0.338	0.418	0.495	0.529
	BART SV	1.079	0.944	0.872	0.852	0.974	0.867**	0.807**	0.763**	0.989	0.856**	0.765***	0.758***
	mixBART SV	0.954	0.950	0.978	0.865	0.951	0.958	0.949	0.904	1.024	0.950	0.808*	0.851
	errorBART SV	0.955	0.975	1.001	0.931***	0.933	0.961	0.954	0.933**	0.989	0.958	0.877	0.846*
	fullBART SV	1.476**	0.998	0.875	0.844	1.243*	0.920	0.807**	0.759**	1.135	0.870*	0.749***	0.745***
	BVAR heteroBART	0.999	1.088	1.119	1.156	0.945*	1.040	1.109	1.088*	0.961*	1.004	1.025	1.056
	BART heteroBART	1.313*	1.048	0.913	0.907	1.134	0.954	0.820*	0.805**	1.089	0.887*	0.798***	0.788***
	mixBART heteroBART	0.955	0.948	0.978	0.891	0.925*	0.952	0.928	0.888	0.982	0.949	0.813**	0.857
	errorBART heteroBART	0.925**	0.932	1.048*	1.045	0.910**	0.973	0.987	1.010	0.986	1.012	0.947	0.945
	fullBART heteroBART	1.217	1.025	0.933	0.874	1.067	0.932	0.847	0.792**	1.076	0.876**	0.767***	0.773***
UNRATE													
	BVAR cons	1.004	0.955	0.939	0.951	1.065	1.022	0.999	1.012	1.116***	1.060	1.039	1.045
	BART cons	0.921***	0.929*	0.970*	0.984	0.922***	0.933**	0.983	0.993	0.943**	0.961	1.003	1.019
	mixBART cons	0.963	0.945	0.955	0.972	0.981	0.957	0.988	1.006	1.034	0.989	1.023	1.039
	errorBART cons	0.966**	0.969*	0.974***	0.982	0.972**	0.972*	0.977***	0.978	0.995	0.996	0.979*	0.972
	fullBART cons	0.915***	0.921*	0.971*	0.986	0.917***	0.928*	0.984	0.995	0.947*	0.960	1.004	1.019
	BVAR SV	0.131	0.145	0.154	0.159	0.147	0.161	0.172	0.177	0.168	0.180	0.196	0.204
	BART SV	0.935**	0.941*	0.971	0.991	0.938**	0.949	0.986	0.998	0.936**	0.977	1.012	1.025
	mixBART SV	0.976	0.956	0.967	0.977	0.982	0.969	0.988	1.001	0.980	0.985	1.011	1.025
	errorBART SV	1.015	1.006	1.016	1.006	1.020	1.000	1.010	1.005	1.013	1.003	0.994	0.983
	fullBART SV	0.906***	0.926**	0.973	0.985	0.903***	0.938	0.984	0.998	0.921*	0.970	1.004	1.025
	BVAR heteroBART	0.960*	0.972***	0.995	0.993	0.964	0.963***	0.985	0.988	0.990	0.959***	0.966*	0.976
	BART heteroBART	0.920***	0.938**	0.980**	0.992	0.912**	0.940**	0.981**	0.994	0.910*	0.961	0.989	1.009
	mixBART heteroBART	0.932***	0.943	0.970	0.983	0.932***	0.946	0.980	0.994	0.924**	0.959*	0.992	1.010
	errorBART heteroBART	0.967**	0.973*	0.982**	0.987	0.976**	0.972*	0.979**	0.981	0.996	0.993	0.978*	0.974
	fullBART heteroBART	0.907***	0.923*	0.976***	0.990	0.902***	0.927*	0.978*	1.001	0.909*	0.964	0.988	1.013

Table 3: Quantile scores (QS).

Notes: Qs for the 5th, 10th, and 25th quantiles are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

Turning to the question of whether this strong performance of the different BART approaches holds throughout the hold-out period or is specific to certain time periods, Figures 4 to 6 display recursive averages of the quantile scores for $\tau = \{0.05, 0.10, 0.25\}$. These figures suggest slightly more time variation in the models. For output and the unemployment rate, we find that BART-based models often improve upon the benchmark BVAR-SV during turbulent episodes such as the global financial crisis. This pattern is more pronounced for multi-steps-ahead forecasts.

When we consider one-step-ahead inflation tail forecasts, a great deal of heterogeneity among the different BART models can be observed. Particularly during the recession in 2008-2009, we find that BART and fullBART perform worse than the benchmark. These losses in accuracy are, however, quickly reversed in the expansionary period that started in mid-2009. By contrast, mixBART and errorBART mirror the pattern observed for output and unemployment but to a slightly lesser extent.

In all cases, and similarly to the full density results, we find that the homoskedastic BVAR performs poorly during tranquil periods, while it works well during recessions. Again, this strong performance during recessions can be traced back to the fact that the predictive variance is quite elevated relative to the other models, which harms during normal times but increases the likelihood of observing outliers during recessions.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

3.5 Nonlinear features in predictive densities

To assess if and when predictive distributions exhibit nonlinear features, we rely on a linear approximation of the BART model to obtain a linear VAR representation. Similar approximations have been proposed in Crawford, et al. (2018) and adopted in Huber, et al. (2020).

All results shown up to this point are based on the exact predictive distribution that is available through simulation. In this section, we compare these exact predictive distributions to the ones obtained from a linear approximation. The approximation linearly projects $\Xi = (\mathbf{X}, \mathbf{Z})$ on a $T \times M$ matrix of nonparametric functions, \mathbf{F} , with typical t^{th} row $F(\mathbf{x}_t) + G(\mathbf{z}_t)$:

$$\tilde{\mathbf{A}} = \text{Proj}(\Xi, \mathbf{F}) = \Xi^\dagger \mathbf{F},$$

with Ξ^\dagger being the Moore-Penrose inverse of the matrix of explanatory variables, Ξ . The reader is referred to the discussion in Huber, et al. (2020) to justify this approximation. In essence, it can be shown that, at the T observations, $\mathbf{F} \approx \mathbf{X}\tilde{\mathbf{A}}$.

After having obtained $\tilde{\mathbf{A}}$ we can iteratively compute multi-steps-ahead forecasts using standard formulas for forecasting in VAR models. This yields an approximate predictive distribution $\hat{p}(\mathbf{y}_{t+h}|\mathbf{y}_t)$. In principle, if the DGP is linear we would expect that $\hat{p}(\mathbf{y}_{t+h}|\mathbf{y}_t) \approx p(\mathbf{y}_{t+h}|\mathbf{y}_t)$. Hence, the distance between the approximate and the exact predictive distribution yields insights into the extent of nonlinearities. To formalize the idea of distance between distributions, we will use the Kullback-Leibler divergence (KLD). The KLD will serve as a measure of the importance of nonlinearities. If we observe substantial divergence between the two predictive densities, this indicates that exact distributions feature substantial nonlinearities (since approximation errors become comparatively large). Figures 7 to 9 show the measure for all BART-based models over time and at different forecast horizons as a heatmap. The time axis here refers to the date when the forecast was made.

A common pattern found for all three of the variables is that the errorBART model has less evidence of nonlinearities. This statement is true regardless of how the error variances are modeled. In our forecasting results, errorBART was often found to forecast slightly worse than the other BART-based approaches. Thus it seems that allowing for nonlinearities in the effects of past shocks is relatively unimportant in this data set.

By contrast, the KLD of the fullBART model is large, indicating that the flexible combination between a nonparametric conditional mean and variance function is difficult to linearly approximate. Particularly for the unemployment rate, the KLD is large throughout the evaluation period. Since fullBART is among the set of the best-performing models for unemployment forecasts according to various measures, this suggests that being flexible on the conditional mean and the full variance-covariance matrix seems to pay off.

Another interesting pattern is that there tends to be more evidence of nonlinearity at the longer forecast horizons. This is consistent with our previous evidence that BART-based models forecast particularly well at longer horizons. From a technical perspective, the larger differences for multi-steps-ahead forecasts are driven by the fact that we iteratively forecast in both cases. Since higher-order forecasts are a highly nonlinear function of $\tilde{\mathbf{A}}$, the approximation error increases with the forecast horizon.

Looking at the different time periods, we find that KLD scores increase during the pandemic for GDP growth and the unemployment rate. This finding is also not surprising given that during this period, an unprecedented decline in GDP growth and an increase in unemploy-

ment that was far outside the range of the historical data has been observed. In such situations, nonparametric approaches quickly adapt to these extreme observations.

In the online appendix, we present graphs similar to Figures 7–9, but for VARs of small and medium dimension. One pattern worth noting is that the small and medium models produced less evidence of nonlinearities. Thus, we are finding nonparametric methods to be particularly useful in larger VARs.

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

3.5.1 Predictive distributions

Figures 10-12 provide time series of the 5/95 percent and 10/90 percent quantiles of the predictive distributions from the different models. So that the extreme volatility of 2020 induced by the COVID-19 pandemic does not obscure scales, the charts have separate panels for the 1997-2019 and 2020 periods. GDP growth, inflation, and the unemployment rate are covered in separate figures. In the interest of chart readability, the figures provide results for just the 1- and 4-quarters-ahead horizons; results for the 8- and 12-quarters-ahead horizons are very similar to those for 4 quarters ahead.

Perhaps the most striking feature of the predictive distributions is the strong similarity of the forecast intervals for most models. The estimates that stand out for being somewhat different are in the top row of each figure, in which the forecast intervals from the BVAR with homoskedasticity are much wider than the intervals from the BVARs with stochastic volatility or the heteroBART specification of conditional heteroskedasticity. In this case, in keeping with our findings above on average forecast accuracy, a BVAR with heteroBART performs comparably to a BVAR with stochastic volatility. In the remaining rows of each figure that compare forecast intervals obtained from alternative BART specifications, the intervals are similar across specifications.

Another notable result is that the predictive distributions do not seem to show much of the downside risk asymmetry that has received considerable attention in the tail risks literature, starting with [Adrian, Boyarchenko, and Giannone \(2019\)](#). In particular, in the case of Figure 10's results for GDP growth, around the time of the Great Recession, the one-step ahead forecast intervals show more downward movement in the lower tail than the upper, in keeping with the

emphasis of much work in the tail risks literature. However, at longer horizons, the forecast intervals show few such asymmetries (in fact, at the four-quarters-ahead horizon, the BART-based specifications show an upside asymmetry for a time following the end of the Great Recession). The same pattern prevails in Figure 12’s results for the unemployment rate (recall that the rate is multiplied by -1 so that low unemployment rates are bad). It is worth emphasizing that our nonparametric modeling through BART should be capable of capturing asymmetries if they exist.

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

3.5.2 Volatility forecasts

To shed light on the different ways we have of modeling the volatility process, Figures 13 to 15 present one- and four-steps-ahead forecasts of the volatility itself using our different models. For each model, we plot lines produced for the three volatility treatments (heteroBART, SV, and homoskedastic) so as to offer an easy comparison.

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

The first point worth noting is that there are differences in the volatility estimates produced by the three treatments of the error variances. As expected, homoskedastic modeling of error variances tends to produce volatility forecasts that are relatively smoother and at a higher level. These very gradual changes are produced by our recursive forecasting design that implies almost no discounting of past information.

HeteroBART tends to produce volatility forecasts that are similarly smooth, but much lower than the homoskedastic ones. But volatility forecasts by SV models tend to be more volatile than the other approaches. One exception to this pattern is revealing: the linear BVAR. For this model, SV and heteroBART produce volatility estimates that are quite similar. Informally speaking, in models where both conditional mean and variance are modeled using BART

approaches, the model can “choose” to put nonlinearities in the conditional mean or the conditional variance, and the choice made is typically to put them in the conditional mean. In the linear VAR such a choice is not possible and, thus, the nonlinearities are put in the conditional variances in a similar manner to SV. The fact that the homoskedastic version of BART tends to forecast well also supports the idea that most of the benefits of the nonparametric approach are obtained in its modeling of the conditional mean as opposed to the conditional variance.

There are also interesting differences in the volatility forecasts during the pandemic. Particularly for $h = 1$ and for unemployment and GDP growth, SV models are forecasting much larger increases in volatility than the other approaches. This is consistent with findings in HKOPS where the extreme pandemic observations had a great impact on the conditional mean that was successfully picked up by BART approaches, leaving less variation in the conditional variance.

3.6 The role of financial conditions for tail forecasting

With much of the recent literature emphasizing the role of financial conditions in driving negative tail risks to economic activity (see, e.g., [Adrian, Boyarchenko, and Giannone \(2019\)](#) and [Delle Monache, De Polis, and Petrella \(2020\)](#)), we examine the role of financial conditions in the tail risk forecasts of BART-based specifications. In the interest of brevity, we focus on two models: the BVAR-heteroBART and BART-heteroBART specifications. This comparison helps to shed light on the role of nonparametric treatments of the conditional variance (heteroBART) and conditional mean (BART vs. BVAR). For this assessment, we consider NFCI paths over the forecast horizon that are fixed at selected values. These values are the different quantiles of the NFCI, ranging from 0 (the minimum) to 1 (the maximum) with a step-size of 0.05. This provides 21 paths of the NFCI for which we produce conditional forecasts from the models.

3.6.1 Conditional forecasts using BART

Figure 16 reports time series of the 5 percent and 95 percent quantiles of predictive distributions of GDP growth, unemployment, and inflation obtained for each path of the NFCI, over our entire out-of-sample evaluation period (with 2020 separated from the rest of the sample for chart readability). In these charts, blue lines refer to densities conditioning on low values of the NFCI and red lines refer to densities conditioning on high values of the NFCI.

In the period up to the Great Recession, conditioning on higher values of the NFCI (worse financial conditions) tends to widen predictive distributions for GDP growth, more noticeably in these charts with the BVAR-heteroBART specification than the BART-heteroBART model,

with the most visible effect being on the 95th percentiles for the former. For the period since the Great Recession, conditioning on higher values of the NFCI lowers the 5 percent quantile forecast. The patterns are broadly similar for the unemployment rate distributions in the bottom row of charts. In the case of inflation, the middle row of the charts indicates that, with both models, conditioning on higher values of the NFCI significantly boosts the 95 percent quantile of the predictive distribution, implying that more adverse conditions are associated with more upside risk to inflation. The NFCI conditioning has relatively little effect on the lower tail forecast for inflation. These patterns for inflation are relatively consistent over the sample, including in 2020.

[Figure 16 about here.]

3.6.2 Conditional forecasts during the global financial crisis and the pandemic

Focusing on the (negative) unemployment rate, Figure 17 reports posterior predictive densities over a few selected quarters from the depths of the Great Recession (2008:Q3 to 2009:Q1) and the 2020 pandemic (2020:Q1 to 2020:Q3). The densities condition on the quantiles of the NFCI that range from 0 (the minimum, blue lines) to 1 (the maximum, red lines).

In broad terms, the charts in the top two rows show that, with the conditional mean taking the linear form of the BVAR, the nonparametric specification of the innovation process through heteroBART is sufficient to yield predictive distributions that are non-Gaussian. For example, in 2008:Q3, some of the distributions have fat tails, whereas in 2020:Q3, the distributions are sharply peaked rather than bell-shaped. More specifically, in the case of BVAR forecasts during the Great Recession, conditioning on different values of the NFCI impacts the predictive distributions mostly by increasing one of the tails or widening the distributions, with little effect on the mode of the distribution. In 2020, conditioning on different NFCI values has little effect on forecasts for 2020:Q2 and 2020:Q3 but sharply affects the predictive distributions for 2020:Q1, with higher values of the NFCI associated with predictive distributions shifted to the left and widened.

Conditioning on different financial settings has much larger effects on predictive distributions from the BART-heteroBART specification. In this case, pairing a nonparametric specification of the conditional mean with a nonparametric specification of the conditional variance can yield sharply non-Gaussian distributions, with fat tails, asymmetries, or even multi-modality.

Multi-modalities in the predictive distribution for GDP growth is most evident in 2008:Q3 (the quarter in which Lehman Brothers failed). In this case, even under favorable (near-zero)

values of the NFCI, the predictive distributions from the BART specification show two clear peaks. As the conditioning NFCI values are increased, the predictive mean shifts some but the variance rises considerably, with much wider distributions. In 2008:Q4 and 2009:Q1, multimodalities are less prominent, but the predictive distributions retain considerable asymmetries, especially at high values of the NFCI. Higher values of the NFCI are also associated with lower predictive means and modes of GDP growth.

In the period of the pandemic, the BART-based patterns for 2020:Q2 and 2020:Q3 are similar to the BVAR-based patterns. In these periods, conditioning on different NFCI values has relatively little effect on the predictive distributions. But in 2020:Q1 forecasts, the NFCI has a much greater impact on the predictive distributions, resembling that seen in the Great Recession quarters of 2008:Q4 and 2009:Q1, with higher values of the NFCI sharply lowering the predictive mean and raising the variance.

[Figure 17 about here.]

4 Concluding remarks

In this paper we have made three main contributions. First, we have used Bayesian additive regression trees (BART) to introduce novel multivariate models that posit nonlinear relationships among macroeconomic variables, their lags, and possibly the lags of the errors. The errors can be either homoskedastic or heteroskedastic, and in the latter case, we consider both a standard stochastic volatility specification and a novel nonparametric specification. The flexible specifications for the conditional mean and variance could be particularly helpful in the presence of parameter time variation and/or for density and tail forecasting.

Second, we have developed MCMC estimation algorithms for each (homoskedastic and heteroskedastic) BART specification. The algorithms are scalable to large dimension and thus allow for estimating large semi- and nonparametric VAR models.

Finally, we have evaluated the real-time forecasting performance for a set of US macroeconomic and financial indicators of the various BART models, using a variety of loss functions and a BVAR-SV model as a (strong) benchmark. The main findings are that when using BART to accommodate nonlinearities, it is less important to allow for heteroskedasticity; the out-of-sample predictive density charts do not show much downside risk asymmetry; and BART specifications can deliver more accurate tail forecasts than BVAR-SV, in particular for unemployment.

Overall, the models we develop represent an important addition to the toolbox of empirical macroeconomists and forecasters, due to their flexibility, range of applicability, ease of

implementation, and good empirical performance.

In this paper, we have focused on approximating the unknown functions F and G using BART due to its excellent empirical properties. However, there exist several alternative techniques such as Gaussian process and kernel regressions (Quinonero-Candela and Rasmussen (2005) and Adrian, Boyarchenko, and Giannone (2021)), spline-based models (Shin, Bhattacharya, and Johnson (2020)), or infinite mixtures (Kalli and Griffin (2018)) to flexibly model the conditional mean in a multivariate time series model. Assessing whether these techniques can be used to improve forecasts would be a fruitful avenue of further research.

References

- Aastveit, Knut Are, Francesco Ravazzolo, and Herman K. van Dijk (2018), “Combined density nowcasting in an uncertain economic environment,” *Journal of Business and Economic Statistics*, 36, 131–145, <https://doi.org/10.1080/07350015.2015.1137760>.
- Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019), “Vulnerable growth,” *American Economic Review*, 109, 1263–89, <https://doi.org/10.1257/aer.20161923>.
- (2021), “Multimodality in macro-financial dynamics,” *International Economic Review*, in-press, <https://doi.org/10.1111/iere.12501>.
- Adrian, Tobias, Federico Grinberg, Nellie Liang, and Sheheryar Malik (2018), “The term structure of growth-at-risk,” *IMF Working Paper*, 18/180, <https://doi.org/10.5089/9781484372364.001>.
- Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010), “Large Bayesian vector autoregressions,” *Journal of Applied Econometrics*, 25, 71–92, <https://ideas.repec.org/a/jae/japmet/v25y2010i1p71-92.html>, <https://doi.org/10.1002/jae.1137>.
- Breiman, Leo (2001), “Random forests,” *Machine Learning*, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Caldara, Dario, Chiara Scotti, and Molin Zhong (2020), “Macroeconomic and financial risks: A tale of volatility,” Technical report.
- Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2015), “Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility,” *Journal of the Royal Statistical Society: Series A*, 178, p. 837, <https://doi.org/10.1111/rssa.12092>.
- (2016), “Common drifting volatility in large Bayesian VARs,” *Journal of Business & Economic Statistics*, 34, 375–390, <https://doi.org/10.1080/07350015.2015.1040116>.
- (2019), “Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors,” *Journal of Econometrics*, 212, 137–154, <https://ideas.repec.org/a/eee/econom/v212y2019i1p137-154.html>, <https://doi.org/10.1016/j.jeconom.2019.04.024>.
- (2020a), “Capturing macroeconomic tail risks with Bayesian vector autoregressions,” *Federal Reserve Bank of Cleveland Working Paper*, 20-02, <https://doi.org/10.26509/frbc-wp-202002>.
- (2020b), “Nowcasting tail risks to economic activity with many indicators,” *Federal Reserve Bank of Cleveland Working Paper*, 20-13, <https://doi.org/10.26509/frbc-wp-202013r2>.
- Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2021), “Addressing COVID-19 outliers in BVARs with stochastic volatility,” *Federal Reserve Bank of Cleveland Working Papers*, 21-02, <https://doi.org/10.26509/frbc-wp-202102>.
- Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (2010), “The horseshoe estimator for sparse signals,” *Biometrika*, 97, 465–480, <https://doi.org/10.1093/biomet/asq017>.
- Chib, Siddhartha, and Edward Greenberg (1994), “Bayes inference in regression models with ARMA (p, q) errors,” *Journal of Econometrics*, 64, 183–206, [https://doi.org/10.1016/0304-4076\(94\)90063-9](https://doi.org/10.1016/0304-4076(94)90063-9).
- Chipman, Hugh A., Edward I. George, and Robert E. McCulloch (1998), “Bayesian CART model search,” *Journal of the American Statistical Association*, 93, 935–948, <https://doi.org/10.2307/2669832>.
- (2010), “BART: Bayesian additive regression trees,” *The Annals of Applied Statistics*, 4, 266–298, <https://doi.org/10.1214/09-A0AS285>.
- Clark, Todd E. (2011), “Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility,” *Journal of Business & Economic Statistics*, 29, 327–341, <https://doi.org/10.1198/jbes.2010.09248>.
- Clark, Todd E., and Francesco Ravazzolo (2015), “Macroeconomic forecasting performance under alternative specifications of time-varying volatility,” *Journal of Applied Econometrics*, 30, 551–575, <https://doi.org/10.1002/jae.2379>.
- Cook, Thomas, and Taeyoung Doh (2019), “Assessing macroeconomic tail risks in a data-rich environment,” *Federal Reserve Bank of Kansas City Research Working Paper*, 19-12, <https://doi.org/10.18651/RWP2019-12>.
- Crawford, Lorin, Kris C. Wood, Xiang Zhou, and Sayan Mukherjee (2018), “Bayesian approximate kernel regression with variable selection,” *Journal of the American Statistical Association*, 113, 1710–1721, <https://doi.org/10.1080/01621459.2017.1361830>.
- De Nicolò, Gianni, and Marcella Lucchetta (2017), “Forecasting tail risks,” *Journal of Applied Econometrics*, 32, 159–170, <https://doi.org/10.1002/jae.2509>.
- Delle Monache, Davide, Andrea De Polis, and Ivan Petrella (2020), “Modeling and forecasting macroeconomic downside risk,” *CEPR Discussion Paper Series*.
- Diebold, Francis X., and Robert S. Mariano (1995), “Comparing predictive accuracy,” *Journal of Business & Economic Statistics*, 13, 253–263, <https://doi.org/10.2307/1392185>.
- Ferrara, Laurent, M. Mogliani, and J.G. Sahuc (2019), “Real-time high frequency monitoring of growth-at-risk,” Technical report.
- Gaglianone, Wagner Piazza, and Luiz Renato Lima (2012), “Constructing density forecasts from quantile regressions,” *Journal of Money, Credit and Banking*, 44, 1589–1607, <https://doi.org/10.1111/j.1538-4616.2012.00545.x>.

- 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>.
- Ghysels, Eric, Leonardo Iania, and Jonas Striaukas (2018), “Quantile-based inflation risk models,” *National Bank of Belgium Research Working Paper*, 349, <http://hdl.handle.net/10419/207729>.
- Giacomini, Raffaella, and Ivana Komunjer (2005), “Evaluation and combination of conditional quantile forecasts,” *Journal of Business & Economic Statistics*, 23, 416–431, <https://doi.org/10.1198/073500105000000018>.
- Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2015), “Prior selection for vector autoregressions,” *The Review of Economics and Statistics*, 97, 436–451, <https://ideas.repec.org/a/tpr/restat/v97y2015i2p436-451.html>, https://doi.org/10.1162/rest_a_00483.
- 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>.
- Gneiting, Tilmann, and Adrian E. Raftery (2007), “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, 102, 359–378, <https://doi.org/10.1198/016214506000001437>.
- Gneiting, Tilmann, and Roopesh Ranjan (2011), “Comparing density forecasts using threshold- and quantile-weighted scoring rules,” *Journal of Business & Economic Statistics*, 29, 411–422, <https://doi.org/10.1198/jbes.2010.08110>.
- González-Rivera, Gloria, Javier Maldonado, and Esther Ruiz (2019), “Growth in stress,” *International Journal of Forecasting*, 35, 948–966, <https://doi.org/10.1016/j.ijforecast.2019.04.006>.
- Goulet Coulombe, Philippe (2020), “The macroeconomy as a random forest,” *arXiv*, 2006.12724, <https://arxiv.org/abs/2006.12724>.
- Goulet Coulombe, Philippe, Maxime Leroux, Dalibor Stevanovic, and Stephane Surprenant (2020), “How is machine learning useful for macroeconomic forecasting?” *arXiv*, 2008.12477, <https://arxiv.org/abs/2008.12477>.
- Goulet Coulombe, Philippe, Massimiliano Marcellino, and Dalibor Stevanovic (2021), “Can machine learning catch the COVID-19 recession?” *arXiv*, 2103.01201, <https://arxiv.org/abs/2103.01201>.
- Huber, Florian, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner (2020), “Nowcasting in a pandemic using non-parametric mixed frequency VARs,” *Journal of Econometrics*, in-press, <https://doi.org/10.1016/j.jeconom.2020.11.006>.
- Huber, Florian, and Luca Rossini (2020), “Inference in Bayesian additive vector autoregressive tree models,” *arXiv*, 2006.16333, <https://arxiv.org/abs/2006.16333>.
- Kalli, Maria, and Jim E. Griffin (2018), “Bayesian nonparametric vector autoregressive models,” *Journal of Econometrics*, 203, 267–282, <https://doi.org/10.1016/j.jeconom.2017.11.009>.
- Kastner, Gregor, and Sylvia Frühwirth-Schnatter (2014), “Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models,” *Computational Statistics & Data Analysis*, 76, 408–423, <https://doi.org/10.1016/j.csda.2013.01.002>.
- Kiley, Michael T. (2018), “Unemployment risk,” *Board of Governors of the Federal Reserve System Finance and Economics Discussion Series*, 2018-067, <https://doi.org/10.17016/FEDS.2018.067>.
- Koop, Gary (2013), “Forecasting with medium and large Bayesian VARs,” *Journal of Applied Econometrics*, 28, 177–203, <https://ideas.repec.org/a/wly/japmet/v28y2013i2p177-203.html>, <https://doi.org/10.1002/jae.1270>.
- Korobilis, Dimitris (2017), “Quantile regression forecasts of inflation under model uncertainty,” *International Journal of Forecasting*, 33, 11–20, <https://doi.org/10.1016/j.ijforecast.2016.07.005>.
- Makalic, Enes, and Daniel F. Schmidt (2015), “A simple sampler for the horseshoe estimator,” *IEEE Signal Processing Letters*, 23, 179–182, <https://doi.org/10.1109/LSP.2015.2503725>.
- Manzan, Sebastiano (2015), “Forecasting the distribution of economic variables in a data-rich environment,” *Journal of Business and Economic Statistics*, 33, 144–164, <https://doi.org/10.1080/07350015.2014.937436>.
- Manzan, Sebastiano, and Dawit Zerom (2013), “Are macroeconomic variables useful for forecasting the distribution of US inflation?” *International Journal of Forecasting*, 29, 469–478, <https://doi.org/10.1016/j.ijforecast.2013.01.005>.
- (2015), “Asymmetric quantile persistence and predictability: the case of US inflation,” *Oxford Bulletin of Economics and Statistics*, 77, 297–318, <https://doi.org/10.1111/obes.12065>.
- Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes (2021), “Machine learning advances for time series forecasting,” *arXiv*, 2012.12802, <https://arxiv.org/abs/2012.12802>.
- Medeiros, Marcelo C., Gabriel F.R. Vasconcelos, Álvaro Veiga, and Eduardo Zilberman (2021), “Forecasting inflation in a data-rich environment: the benefits of machine learning methods,” *Journal of Business and Economic Statistics*, 39, 98–119, <https://doi.org/10.1080/07350015.2019.1637745>.
- Mitchell, James, Aubrey Poon, and Gian Luigi Mazzi (forthcoming), “Nowcasting euro area GDP growth using quantile regression,” Technical report.
- Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima (2007), “Stochastic volatility with leverage: Fast and efficient likelihood inference,” *Journal of Econometrics*, 140, 425–449, <https://doi.org/10.1016/j.jeconom.2006.07.008>.

- Plagborg-Møller, Mikkel, Lucrezia Reichlin, Giovanni Ricco, and Thomas Hasenzagl (2020), “When is growth at risk?” *Brookings Papers on Economic Activity*, in-press.
- Pratola, Matthew T., Hugh A. Chipman, Edward I. George, and Robert E. McCulloch (2020), “Heteroscedastic BART via multiplicative regression trees,” *Journal of Computational and Graphical Statistics*, 29, 405–417, <https://doi.org/10.1080/10618600.2019.1677243>.
- Primiceri, Giorgio E (2005), “Time varying structural vector autoregressions and monetary policy,” *The Review of Economic Studies*, 72, 821–852, <https://doi.org/10.1111/j.1467-937x.2005.00353.x>.
- Quinonero-Candela, Joaquin, and Carl Edward Rasmussen (2005), “A unifying view of sparse approximate Gaussian process regression,” *The Journal of Machine Learning Research*, 6, 1939–1959, <https://dl.acm.org/doi/abs/10.5555/1046920.1194909>.
- Reichlin, Lucrezia, Giovanni Ricco, and Thomas Hasenzagl (2020), “Financial variables as predictors of real growth vulnerability,” *Deutsche Bundesbank Discussion Paper*, 05/2020, <http://hdl.handle.net/10419/214829>.
- Shin, Minsuk, Anirban Bhattacharya, and Valen E. Johnson (2020), “Functional horseshoe priors for subspace shrinkage,” *Journal of the American Statistical Association*, 115, 1784–1797, <https://doi.org/10.1080/01621459.2019.1654875>.
- West, Kenneth D. (1996), “Asymptotic inference about predictive ability,” *Econometrica*, 64, 1067–1084, <https://doi.org/10.2307/2171956>.

A Technical Appendix

A.1 Priors on the remaining model parameters

On the VAR coefficients \mathbf{A} we use a horseshoe prior (Carvalho, Polson, and Scott (2010)). Let $\mathbf{a}_i = (a_{i1}, \dots, a_{iK})'$ denote the i^{th} column of \mathbf{A} and a_{ij} the i^{th} element of \mathbf{a}_i . The horseshoe prior is a hierarchical Gaussian prior on a_{ij} :

$$a_{ij} | \lambda_i, \psi_{ij} \sim \mathcal{N}(0, \psi_{ij}^2 \lambda_i^2), \quad \psi_{ij} \sim \mathcal{C}^+(0, 1), \quad \lambda_i \sim \mathcal{C}^+(0, 1) \quad (\text{A.1})$$

with λ_i being a shrinkage hyperparameter that applies to all coefficients in equation i and ψ_{ij} denotes a coefficient-specific scaling parameter. Both λ_i and ψ_{ij} feature a half-Cauchy prior \mathcal{C}^+ . This prior belongs to the general class of global local shrinkage priors that shrink globally (through λ_i) but allow for local deviations (through ψ_{ij}) if λ_i is close to zero.

For the $v = M(M - 1)/2$ free elements in \mathbf{Q} we use a horseshoe prior similar to the one in (A.1). The only exception is that the global shrinkage parameter applies to all elements in \mathbf{Q} (as opposed to the covariance parameters in a specific equation only).

In the case where we use a model that features stochastic volatility, the prior on the unconditional mean is Gaussian with mean zero and variance 10; the prior on the persistence parameter, denoted by ρ_i , is Beta distributed $\frac{\rho_i + 1}{2} \sim \mathcal{B}(25, 5)$; and the prior on the variance of the log-volatility process is Gamma distributed $\mathcal{G}(1/2, 1/2)$.

If we use a model with homoskedastic shocks we use an inverse Gamma prior on the main diagonal elements of $\mathbf{\Sigma}$, σ_i^2 , which we set to be rather uninformative, i.e., $\sigma_i^2 \sim \mathcal{G}^{-1}(c_0, c_1)$. The hyperparameters c_0, c_1 are set equal to 0.01.

A.1.1 Sampling the remaining unknowns of the model

Conditional on the trees and the error volatilities, one can sample the VAR coefficients and the covariance parameters in a single block using standard textbook results for the linear regression model. The joint posterior of \mathbf{a}_i and \mathbf{q}_i is multivariate Gaussian:

$$\begin{pmatrix} \mathbf{a}_i \\ \mathbf{q}_i \end{pmatrix} | \bullet \sim \mathcal{N}(\bar{\boldsymbol{\beta}}_i, \bar{\mathbf{V}}_i),$$

with

$$\begin{aligned}\bar{\mathbf{V}}_i &= (\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i + \underline{\mathbf{V}})^{-1}, \\ \bar{\boldsymbol{\beta}}_i &= \bar{\mathbf{V}}_i \tilde{\mathbf{X}}_i' \tilde{\mathbf{y}}_i,\end{aligned}$$

where $\tilde{\mathbf{X}}_i'$ denotes a $T \times (K + i - 1)$ matrix with typical t^{th} row $\tilde{x}_{it}' = (\mathbf{x}_{it}', \mathbf{r}_i')' / e^{h_{it}/2}$, $\tilde{\mathbf{y}}_i$ has typical t^{th} element $(y_{it} - g_i(\mathbf{z}_t)) / e^{h_{it}/2}$, and $\underline{\mathbf{V}}$ denotes a diagonal prior variance-covariance matrix constructed using the variances described in (A.1). The \bullet notation indicates that we condition on the remaining parameters of the model.

Notice that if we use a model that assumes f_i and g_i to be nonlinear, we simply exclude \mathbf{x}_{it} from $\tilde{\mathbf{X}}_i$. In the case where we estimate the errorBART model, \mathbf{x}_{it} will be replaced with $(\boldsymbol{\eta}'_{t-1}, \dots, \boldsymbol{\eta}'_{t-p})'$.

In models that include stochastic volatility, we use the efficient sampler outlined in [Kastner and Frühwirth-Schnatter \(2014\)](#). This sampler also exploits the 10-component mixture approximation to the $\log \chi_1^2$ distribution but restates the conditionally Gaussian and linear state space model in terms of a big regression model with the regression coefficients being the log-volatilities. This gives rise to an algorithm that samples the volatilities all without a loop from a $T - 1$ -dimensional multivariate Gaussian distribution.

If we use a homoskedastic model, the error variances can easily be sampled from an inverse Gamma posterior with

$$\sigma_i^2 | \bullet \sim \mathcal{G}^{-1} \left(c_0 + \frac{T}{2}, c_1 + \frac{\sum_{t=1}^T \varepsilon_{it}^2}{2} \right).$$

Finally, the hyperparameters of the horseshoe prior are simulated using the auxiliary sampler proposed in [Makalic and Schmidt \(2015\)](#). We will outline the relevant full conditionals for the prior on \mathbf{a}_i only. The hyperparameters for the prior on the free elements in \mathbf{Q} take precisely the same form.

[Makalic and Schmidt \(2015\)](#) introduce auxiliary random variables ζ_i and κ_{ij} . Conditional on these, the posterior of λ_i^2 is inverse Gamma distributed:

$$\psi_{ij}^2 | \bullet \sim \mathcal{G}^{-1} \left(1, \frac{1}{\kappa_{ij}} + \frac{a_{ij}^2}{2\lambda_i^2} \right).$$

The posterior for the equation-specific shrinkage parameter λ_i^2 is also inverse Gamma distributed:

$$\lambda_i^2 | \bullet \sim \mathcal{G}^{-1} \left(\frac{K+1}{2}, \frac{1}{\zeta_i} + \frac{1}{2} \sum_{j=1}^K \frac{a_{ij}^2}{\psi_{ij}^2} \right).$$

The two auxiliary parameters are also inverse Gamma distributed and take particularly simple forms:

$$\begin{aligned} \kappa_{ij} | \bullet &\sim \mathcal{G}^{-1} \left(1, 1 + \frac{1}{\zeta_i} \right), \\ \zeta_i | \bullet &\sim \mathcal{G}^{-1} \left(1, 1 + \frac{1}{\lambda_i^2} \right). \end{aligned}$$

B Additional empirical results

B.1 Holdout ending 2019:Q4: Point and density forecasts

Model		CRPS				qwCRPS-tails				qwCRPS-left			
		h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
SMALL													
	BVAR cons	1.186***	1.073**	1.080**	1.095***	1.320***	1.124**	1.126***	1.124***	1.107**	1.022	1.032	1.038
	BART cons	1.069	0.976	0.999	1.009	1.108**	0.986	1.000	1.002	1.026	0.964	1.013	1.021*
	mixBART cons	1.070	0.996	1.015	1.030	1.099**	1.002	1.018	1.023	1.016	0.976	1.013	1.014
	errorBART cons	1.076*	1.030	1.025	1.034**	1.096***	1.032	1.018	1.016**	1.019	1.005	0.998	1.001
	fullBART cons	1.069	0.977	0.998	1.009	1.105**	0.986	1.000	1.002	1.026	0.966	1.014	1.022
	BVAR SV	1.016	1.001	1.016	1.020**	1.028	1.004	1.019	1.016	1.011	1.004	1.020	1.019
	BART SV	1.063	0.982	1.007	1.007	1.095*	0.990	1.007	1.000	1.032	0.973	1.021	1.014
	mixBART SV	1.050	0.992	1.018	1.026**	1.075	1.000	1.023	1.022	1.014	0.975	1.018	1.011
	errorBART SV	1.031	1.006	1.019	1.026	1.051	1.010	1.015	1.012	1.028	1.008	1.022	1.024
	fullBART SV	1.078*	0.996	1.019	1.024*	1.138***	1.015	1.031	1.026*	1.041	0.979	1.017	1.014
	BVAR heteroBART	1.022	0.989	1.009	1.007	1.040	0.989	1.002	0.989	1.031	0.999	1.025	1.017
	BART heteroBART	1.067	0.981	1.002	0.998	1.096*	0.986	0.994	0.987	1.042	0.978	1.020	1.013
	mixBART heteroBART	1.063	0.993	1.012	1.015	1.081	0.994	1.007	1.003	1.034	0.983	1.023	1.013
	errorBART heteroBART	1.045	1.008	1.007	1.019	1.058	1.006	0.997	0.998	1.014	1.000	0.999	1.005
	fullBART heteroBART	1.074	0.981	1.000	1.000	1.104*	0.986	0.992	0.987	1.051	0.977	1.018	1.012
MEDIUM													
	BVAR cons	1.183***	1.068**	1.081**	1.094***	1.316***	1.123**	1.129***	1.127***	1.134**	1.024	1.039	1.043
	BART cons	1.074	0.984	0.999	1.019	1.112*	0.985	0.997	1.006	1.058	0.965	1.009	1.026
	mixBART cons	1.073	1.000	1.001	1.020**	1.105**	1.004	0.999	1.011	1.052	0.982	0.997	1.008
	errorBART cons	1.071	1.005	1.010	1.020**	1.099**	1.013	1.007	1.004	1.041	0.988	0.992	0.993
	fullBART cons	1.073	0.985	1.001	1.019	1.114*	0.989	0.998	1.006	1.058	0.967	1.009	1.024
	BVAR SV	1.023	0.989	0.996	1.003	1.042	0.992	0.996	0.998	1.040	0.994	0.999	1.004
	BART SV	1.078	0.978	0.993	1.009	1.114*	0.983	0.994	1.002	1.078	0.966	1.002	1.015
	mixBART SV	1.072	0.991	0.996	1.013*	1.103*	0.996	1.000	1.008	1.069	0.976	0.993	0.999
	errorBART SV	1.035	0.985	0.996	1.005	1.059	0.991	0.994	0.993	1.050	0.990	1.002	1.008
	fullBART SV	1.086*	0.994	1.017	1.026**	1.152***	1.011	1.029	1.028	1.056	0.970	1.013	1.014
	BVAR heteroBART	1.020	0.973	0.979	0.983	1.041	0.977	0.975	0.970**	1.048	0.988	0.993	0.995
	BART heteroBART	1.095	0.976	0.985	1.001	1.131*	0.974	0.977	0.990	1.105	0.969	0.997	1.013
	mixBART heteroBART	1.082	0.984	0.991	1.002	1.115*	0.983	0.986	0.990	1.092	0.975	0.998	1.001
	errorBART heteroBART	1.044	0.989	0.995	1.008	1.068	0.993	0.989	0.991	1.036	0.984	0.990	0.995
	fullBART heteroBART	1.093	0.982	0.995	1.000	1.135*	0.978	0.984	0.988	1.105	0.974	1.004	1.007
LARGE													
	BVAR cons	1.144***	1.078**	1.107***	1.126***	1.268***	1.132**	1.164***	1.183***	1.121**	1.036	1.067	1.082
	BART cons	1.044	0.971	0.987	0.998	1.066*	0.974	0.982	0.983***	1.047	0.967	1.002	1.003
	mixBART cons	1.042*	0.999	1.051*	1.087***	1.045*	1.002	1.041	1.078**	1.050*	0.996	1.038	1.052
	errorBART cons	1.023	1.015	1.031	1.026	1.041	1.012	1.017	1.005	0.995	0.998	1.012	0.997
	fullBART cons	1.048	0.969	0.988	1.004	1.072*	0.972	0.982	0.989***	1.054	0.965	1.004	1.008
	BVAR SV	1.168	2.063	2.074	2.098	0.118	0.227	0.230	0.236	0.177	0.320	0.319	0.323
	BART SV	1.059	0.972	0.982	0.994	1.074*	0.975	0.981	0.982**	1.068	0.975	1.003	1.009
	mixBART SV	1.037	1.001	1.019	1.042***	1.042*	1.005	1.018	1.037*	1.045*	0.998	1.015	1.027
	errorBART SV	0.998	1.008	1.009	1.004	1.006	1.006	1.006	0.996	1.004	1.011	1.011	1.004
	fullBART SV	1.068*	0.986	1.012	1.022***	1.131***	1.003	1.021	1.021	1.052	0.972	1.014	1.014
	BVAR heteroBART	1.004	0.985**	0.986*	0.984**	1.007	0.982*	0.980	0.971***	1.022	0.994	0.992	0.986**
	BART heteroBART	1.048	0.973	0.971**	0.981**	1.069	0.973	0.967***	0.968***	1.063	0.976	0.990	0.995
	mixBART heteroBART	1.040	0.996	1.008	1.025*	1.040	0.995	1.002	1.014	1.056*	0.997	1.007	1.012
	errorBART heteroBART	1.006	1.003	1.017	1.017	1.015	0.999	1.004	0.996	0.993	0.997	1.009	1.000
	fullBART heteroBART	1.051	0.966	0.972**	0.988	1.074	0.967	0.968***	0.972**	1.068	0.970	0.992	1.000

Table 4: Cumulative ranked probability score (CRPS) and quantile weighted CRPSs for GDPC1.

Notes: CRPSs are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Quantile weights 'tail' indicate a weighting scheme capturing both tails; 'left' captures performance for downside risks. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

Model	CRPS				qwCRPS-tails				qwCRPS-left			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
SMALL												
BVAR cons	1.145***	1.078	1.051	1.047	1.135***	1.148**	1.156**	1.123	1.130**	0.970	0.924	0.909
BART cons	1.047	0.953	0.775***	0.673***	1.034	0.941	0.775***	0.675***	1.041	0.968	0.796***	0.692***
mixBART cons	1.068	0.936	0.809***	0.747***	1.044	0.917*	0.823***	0.790***	1.077	0.946	0.823**	0.775***
errorBART cons	1.123**	0.984	0.872**	0.838***	1.084*	0.957	0.886**	0.868***	1.146**	0.972	0.876*	0.863**
fullBART cons	1.046	0.950	0.778***	0.674***	1.033	0.940	0.779***	0.675***	1.043	0.964	0.798***	0.694***
BVAR SV	1.123**	0.984	0.921	0.911	1.096**	0.982	0.947	0.945	1.142**	0.972	0.896	0.878
BART SV	1.060	0.958	0.780***	0.677***	1.052	0.943	0.776***	0.681***	1.044	0.956	0.786***	0.689***
mixBART SV	1.069	0.939	0.805***	0.741***	1.046	0.928	0.821***	0.786***	1.063	0.933	0.810**	0.760***
errorBART SV	1.118**	0.946	0.789**	0.711***	1.085*	0.927	0.809***	0.759***	1.137**	0.954	0.815*	0.765**
fullBART SV	1.045	0.940	0.761***	0.663***	1.033	0.928	0.759***	0.668***	1.042	0.950	0.774***	0.681***
BVAR heteroBART	1.153***	0.993	0.919	0.880***	1.140**	0.980	0.926	0.906***	1.163***	1.003	0.938	0.907**
BART heteroBART	1.075	0.966	0.781***	0.674***	1.074	0.965	0.784***	0.673***	1.060	0.979	0.803**	0.698***
mixBART heteroBART	1.095*	0.947	0.810***	0.737***	1.077	0.937	0.813***	0.769***	1.085	0.957	0.827**	0.770***
errorBART heteroBART	1.128**	0.976	0.842**	0.798***	1.096*	0.950	0.849***	0.826***	1.146**	0.978	0.868*	0.850***
fullBART heteroBART	1.063	0.960	0.779***	0.678***	1.057	0.957	0.781***	0.677***	1.050	0.974	0.796***	0.701***
MEDIUM												
BVAR cons	1.141***	1.099	1.073	1.072	1.140***	1.151*	1.154**	1.129	1.125**	0.976	0.948	0.935
BART cons	1.009	0.912	0.817***	0.779***	1.002	0.912	0.818***	0.774***	1.029	0.930	0.824***	0.785***
mixBART cons	1.053	0.933	0.828***	0.806***	1.031	0.909*	0.839***	0.841***	1.081	0.944	0.829***	0.826***
errorBART cons	1.101*	1.007	0.898*	0.852**	1.063	0.976	0.899*	0.870**	1.114**	0.979	0.904	0.890*
fullBART cons	1.009	0.916	0.823***	0.786***	1.000	0.921	0.821***	0.779***	1.024	0.935	0.830***	0.792***
BVAR SV	1.097**	0.997	0.937	0.929	1.067	0.988	0.950	0.953	1.117**	0.988	0.925	0.918
BART SV	0.994	0.900*	0.785***	0.746***	0.977	0.898**	0.783***	0.745***	0.999	0.906	0.778***	0.747***
mixBART SV	1.025	0.923	0.812***	0.777***	1.005	0.908*	0.826***	0.818***	1.042	0.931	0.807***	0.793***
errorBART SV	1.063	0.942	0.795**	0.727***	1.032	0.917	0.805**	0.757***	1.079	0.945	0.823**	0.785**
fullBART SV	1.039	0.912	0.808***	0.765***	1.031	0.915	0.803***	0.765***	1.043	0.925	0.806***	0.772***
BVAR heteroBART	1.119**	0.990	0.929	0.912**	1.104**	0.976	0.932	0.938*	1.131**	0.998	0.954	0.966
BART heteroBART	1.006	0.928	0.789***	0.740***	0.997	0.931	0.786***	0.739***	1.009	0.949	0.803***	0.755***
mixBART heteroBART	1.035	0.929	0.811***	0.778***	1.020	0.916*	0.811***	0.803***	1.056	0.951	0.823***	0.812***
errorBART heteroBART	1.095*	0.985	0.865**	0.811***	1.061	0.957	0.866**	0.830***	1.105*	0.967	0.885*	0.864**
fullBART heteroBART	1.025	0.920	0.785***	0.733***	1.014	0.922	0.786***	0.736***	1.031	0.937	0.799***	0.750***
LARGE												
BVAR cons	1.051*	1.089**	1.082	1.085	1.094**	1.140**	1.153*	1.152	1.027	1.003	0.969	0.958
BART cons	1.064	0.866**	0.754***	0.709***	1.062	0.858***	0.744***	0.703***	1.064	0.876**	0.766***	0.738***
mixBART cons	0.969	0.954	0.823**	0.788***	0.952	0.925	0.828**	0.848*	0.962	0.980	0.842	0.824*
errorBART cons	0.986	1.029	0.927	0.839**	0.956	0.991	0.910	0.834***	0.983	1.018	0.941	0.896**
fullBART cons	1.032	0.866***	0.757***	0.714***	1.032	0.859***	0.748***	0.710***	1.035	0.882**	0.772***	0.748***
BVAR SV	0.510	0.695	0.892	1.035	0.051	0.070	0.089	0.105	0.078	0.106	0.128	0.144
BART SV	1.027	0.857***	0.776***	0.739***	1.013	0.854***	0.771***	0.729***	1.024	0.858***	0.776***	0.758***
mixBART SV	0.969	0.924	0.819**	0.781***	0.961	0.904**	0.824**	0.826**	0.959	0.945	0.835*	0.814*
errorBART SV	0.986	0.957	0.849	0.769**	0.964	0.928	0.835**	0.778***	0.984	0.966	0.880	0.830*
fullBART SV	1.180**	0.871**	0.744***	0.710***	1.211**	0.870**	0.739***	0.705***	1.203**	0.881*	0.758***	0.739***
BVAR heteroBART	0.993	0.978	0.973*	0.958**	0.990	0.979	0.975	0.975	0.986	1.002	1.016	1.014
BART heteroBART	1.144*	0.874**	0.772***	0.730***	1.169**	0.872***	0.763***	0.720***	1.139*	0.898*	0.792***	0.773***
mixBART heteroBART	0.973	0.913*	0.808***	0.778***	0.965	0.890**	0.805***	0.813***	0.958	0.942	0.832**	0.818**
errorBART heteroBART	0.988	1.027	0.927	0.848**	0.960	0.988	0.908	0.841***	0.985	1.020	0.951	0.920
fullBART heteroBART	1.135*	0.868***	0.753***	0.713***	1.133	0.866***	0.750***	0.707***	1.137*	0.881**	0.776***	0.754***

Table 5: Cumulative ranked probability score (CRPS) and quantile weighted CRPSs for GDPCTPI.

Notes: CRPSs are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Quantile weights 'tail' indicate a weighting scheme capturing both tails; 'left' captures performance for downside risks. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

Model	CRPS				qwCRPS-tails				qwCRPS-left			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
SMALL												
BVAR cons	1.299***	1.073***	1.036	1.030	1.409***	1.099***	1.056	1.047	1.273***	1.049	1.013	1.014
BART cons	0.999	0.993	1.030	1.047***	0.956	0.987	1.019	1.038***	0.941	0.974	1.020	1.047***
mixBART cons	0.970	0.996	1.011	1.015	0.941	0.995	1.007	1.012	0.913	0.975	0.997	1.012***
errorBART cons	1.077***	1.030	0.993	0.994	1.083***	1.029	0.994	0.992	1.052**	1.018	0.975**	0.984
fullBART cons	1.000	0.996	1.029	1.048***	0.957	0.989	1.017	1.038***	0.940	0.978	1.019	1.050***
BVAR SV	1.107***	1.008	0.997	0.996	1.098***	1.013	1.006	1.004	1.118***	1.014	1.001	1.002
BART SV	1.008	1.002	1.026	1.035***	0.971	0.998	1.018	1.027***	0.952	0.983	1.014	1.031**
mixBART SV	0.998	0.993	1.006	1.007	0.983	0.996	1.008	1.011	0.968	0.980	0.996	1.005
errorBART SV	1.091***	1.008	0.991	0.988	1.084***	1.013	0.999	0.995	1.105***	1.014	0.993	0.991
fullBART SV	1.001	1.001	1.017	1.022**	0.960	0.996	1.008	1.015*	0.941	0.980	1.005	1.018
BVAR heteroBART	1.094***	0.992	0.983	0.983	1.074**	0.994	0.987	0.987	1.093***	0.992	0.983	0.986
BART heteroBART	1.018	1.001	1.018	1.026	0.981	0.993	1.008	1.023	0.976	0.988	1.009	1.028
mixBART heteroBART	0.993	0.984	1.002	1.005	0.962	0.981	0.998	1.006	0.958	0.972	0.992	1.003
errorBART heteroBART	1.076***	1.013	0.989	0.990	1.068***	1.011	0.990	0.989	1.058***	1.005	0.975*	0.982
fullBART heteroBART	1.019	0.997	1.014	1.026	0.979	0.991	1.005	1.023	0.977	0.985	1.006	1.027
MEDIUM												
BVAR cons	1.279***	1.072***	1.033	1.026	1.382***	1.096***	1.052	1.043	1.265***	1.051	1.015	1.013
BART cons	1.032	1.022	1.042	1.058***	0.992	1.011	1.027	1.041**	0.995	0.997	1.029	1.055**
mixBART cons	0.995	1.022	1.024	1.025**	0.971	1.015	1.015	1.014	0.963	1.002	1.011	1.018
errorBART cons	1.082***	1.029	0.993	0.990	1.086***	1.025	0.991	0.987	1.072***	1.020	0.979**	0.982
fullBART cons	1.025	1.023	1.040	1.060***	0.981	1.010	1.025	1.043**	0.988	0.998	1.026	1.056***
BVAR SV	1.079***	1.010	0.989	0.986	1.079***	1.011	0.994	0.993	1.109***	1.016	0.992	0.991
BART SV	1.021	1.023	1.029	1.042***	0.982	1.012	1.018	1.029***	0.987	0.999	1.015	1.039***
mixBART SV	1.014	1.009	1.005	1.004	1.000	1.008	1.004	1.003	1.003	0.999	0.997	1.002
errorBART SV	1.081***	1.011	0.989	0.984	1.083***	1.012	0.993	0.989	1.110***	1.016	0.990	0.987
fullBART SV	1.025	1.024	1.030	1.037***	0.977	1.011	1.017	1.024***	0.982	0.999	1.015	1.033***
BVAR heteroBART	1.064**	0.996	0.980	0.981	1.055	0.994	0.981	0.984	1.071*	0.994	0.977	0.980
BART heteroBART	1.011	1.021	1.024	1.036***	0.973	1.007	1.012	1.028**	0.983	1.000	1.010	1.032***
mixBART heteroBART	1.001	1.004	1.005	1.005	0.970	0.996	1.000	1.001	0.980	0.988	0.994	1.003
errorBART heteroBART	1.080***	1.020	0.991	0.988	1.081***	1.016	0.990	0.987	1.076***	1.012	0.979*	0.980
fullBART heteroBART	1.027	1.026	1.027	1.038***	0.979	1.011	1.013	1.028**	0.996	1.003	1.011	1.034***
LARGE												
BVAR cons	1.219***	1.072***	1.046	1.041	1.338***	1.097**	1.067	1.066	1.186***	1.044	1.021	1.026
BART cons	0.933	0.991	1.024	1.029*	0.900**	0.979	1.009	1.011	0.887**	0.967	1.009	1.021
mixBART cons	0.945	1.009	1.040	1.049*	0.922*	0.997	1.025	1.038	0.905**	0.989	1.022	1.038
errorBART cons	0.992	1.010	1.000	0.989	0.993	1.003	0.994	0.986	0.964*	0.997	0.986	0.980
fullBART cons	0.938	0.989	1.027	1.031*	0.895**	0.977	1.012	1.014	0.892*	0.965	1.010	1.023
BVAR SV	0.109	0.283	0.320	0.334	0.011	0.032	0.037	0.038	0.017	0.047	0.052	0.054
BART SV	0.947	1.001	1.028	1.036	0.917*	0.990	1.013	1.016	0.897**	0.981	1.014	1.029
mixBART SV	0.959*	0.999	1.020	1.023	0.953*	0.995	1.014	1.019	0.934**	0.986	1.009	1.020
errorBART SV	1.004	0.999	0.993	0.985	1.009*	1.000	0.998	0.991	1.008	1.000	0.996	0.987
fullBART SV	0.986	0.996	1.027	1.036*	0.941	0.981	1.010	1.019	0.916	0.975	1.012	1.027
BVAR heteroBART	0.967***	0.978***	0.985**	0.992	0.951***	0.974***	0.985	0.990	0.935***	0.967***	0.976	0.985
BART heteroBART	0.967	0.989	1.011	1.019***	0.930	0.978	1.002	1.008	0.929	0.968	0.997	1.012***
mixBART heteroBART	0.947*	0.986	1.010	1.015	0.926**	0.978	1.005	1.012	0.906***	0.967	0.995	1.008
errorBART heteroBART	1.003	1.005	0.999	0.992	1.008	1.000	0.994	0.990	0.982	0.992	0.986	0.983
fullBART heteroBART	0.935	0.995	1.014	1.024**	0.902*	0.977	1.002	1.012	0.895*	0.970	0.997	1.017*

Table 6: Cumulative ranked probability score (CRPS) and quantile weighted CRPSs for UNRATE.

Notes: CRPSs are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Quantile weights 'tail' indicate a weighting scheme capturing both tails; 'left' captures performance for downside risks. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

B.2 Additional tail forecast metrics

Model	QS5				QS10				QS25			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
SMALL												
BVAR cons	1.317***	0.970	0.959	0.934	1.263***	1.035	1.030	1.009	1.078	1.026	1.043	1.053
BART cons	1.058	0.968	0.998	0.973	1.074	0.976	1.021	1.004	0.986	0.953	1.030	1.042***
mixBART cons	1.043	0.964	0.976	0.950	1.021	0.976	1.004	0.978	0.973	0.966	1.019	1.015
errorBART cons	0.983	0.969	0.948***	0.936**	1.004	1.001	0.964***	0.952***	0.987	0.992	0.990	0.994
fullBART cons	1.061	0.969	0.996	0.968	1.070	0.975	1.026	1.002	0.992	0.957	1.030	1.045**
BVAR SV	1.024	1.050	1.073	1.032	1.024	1.021	1.046	1.025	1.004	0.995	1.001	1.008
BART SV	1.081	0.997	1.034	0.995	1.083	0.990	1.037	1.008	0.998	0.960	1.027	1.021
mixBART SV	1.039	0.999	1.029	0.982	1.032	0.991	1.033	0.994	0.985	0.958	1.013	1.005
errorBART SV	1.066	1.052	1.062	1.023	1.073	1.028	1.026	1.017	1.019	0.999	1.010	1.010
fullBART SV	1.113	0.965	1.000	0.965	1.126	1.000	1.020	0.993	1.019	0.972	1.026	1.025
BVAR heteroBART	1.122	1.036	1.052	1.001	1.089	1.016	1.020	0.993	1.022	0.989	1.025	1.026
BART heteroBART	1.126	1.002	1.026	0.995	1.097	0.994	1.022	1.004	1.012	0.973	1.026	1.016
mixBART heteroBART	1.088	0.999	1.034	0.983*	1.049	0.991	1.027	0.993	1.006	0.974	1.021	1.014
errorBART heteroBART	0.997	0.996	0.976	0.963	1.003	1.007	0.976	0.974*	1.000	0.989	0.999	1.001
fullBART heteroBART	1.125	1.002	1.028	0.993	1.104	0.990	1.022	1.004	1.022	0.975	1.021	1.015
MEDIUM												
BVAR cons	1.355***	0.972	0.959	0.944	1.297***	1.043	1.040	1.022	1.131*	1.035	1.059	1.063
BART cons	1.115	0.959	0.972	0.948	1.127	0.964	1.005	0.993	1.049	0.947	1.026	1.046
mixBART cons	1.029	0.967	0.946	0.938	1.092	0.983	0.983	0.968	1.046	0.969	1.008	1.015
errorBART cons	1.031	0.968	0.948**	0.937**	1.075	0.990	0.970**	0.947***	1.032	0.981	0.988	0.988
fullBART cons	1.109	0.966	0.972	0.948	1.131	0.970	1.004	0.989	1.048	0.946*	1.025	1.043
BVAR SV	1.112	1.026	1.033	1.005	1.101	1.012	1.014	1.004	1.048	0.991	0.993	1.004
BART SV	1.164	0.984	0.993	0.972	1.158	0.982	1.010	1.002	1.077	0.947*	1.010	1.029*
mixBART SV	1.113	0.989	0.986	0.960	1.139*	0.992	0.997	0.976	1.066	0.955*	0.995	1.000
errorBART SV	1.154	1.027	1.032	1.005	1.123	1.011	1.005	1.006	1.054	0.986	0.998	1.006
fullBART SV	1.121**	0.939	0.966	0.941	1.135*	0.987	1.011	0.984	1.046	0.959	1.026	1.032
BVAR heteroBART	1.159	1.024	1.012	0.969	1.143	1.013	0.992	0.979	1.053	0.985	0.998	1.006
BART heteroBART	1.268*	0.985	0.991	0.972	1.221*	0.975	0.997	1.000	1.098	0.952	1.002	1.021
mixBART heteroBART	1.195	0.986	0.998	0.959**	1.186*	0.985	0.998	0.981	1.095	0.956	0.999	1.006
errorBART heteroBART	1.083	0.991	0.971	0.951**	1.086	0.998	0.977	0.966**	1.033	0.979	0.991	0.996
fullBART heteroBART	1.275*	0.990	0.995	0.965*	1.243**	0.978	0.998	0.994	1.095	0.960	1.004	1.012
LARGE												
BVAR cons	1.333***	0.968	0.976	0.977	1.304***	1.049	1.072	1.080	1.104	1.051	1.096	1.120
BART cons	1.021	0.925	0.950	0.932*	1.097*	0.961	0.986	0.964	1.056	0.975	1.030	1.022
mixBART cons	1.046	0.919	0.915	0.942	1.076*	0.966	0.987	0.994	1.057	1.018	1.048	1.051
errorBART cons	0.990	0.952	0.953**	0.937***	0.993	0.975	0.973*	0.943***	0.996	0.997	1.006	0.992***
fullBART cons	1.048	0.921	0.945	0.937*	1.106*	0.962	0.987	0.971	1.065	0.972	1.033	1.025
BVAR SV	0.259	0.779	0.796	0.829	0.405	0.939	0.956	0.993	0.686	1.195	1.177	1.193
BART SV	1.079	0.954	0.960	0.948	1.115	0.977	0.999	0.979	1.071	0.983	1.029	1.035
mixBART SV	1.040	0.957	0.949	0.946	1.066*	0.992	1.000	0.991	1.050	1.013	1.024	1.037
errorBART SV	1.006	1.018	1.032	1.004	1.009	1.017	1.011	0.998	1.016	1.009	1.012	1.004
fullBART SV	1.133**	0.915	0.943	0.933	1.123	0.972	0.997	0.978	1.057	0.977	1.041	1.038
BVAR heteroBART	1.042	0.993	0.995	0.962*	1.046	0.997	0.987	0.963***	1.027	0.997	0.993	0.990
BART heteroBART	1.118	0.964	0.971	0.947**	1.122	0.975	0.984	0.970***	1.065	0.984	1.011	1.010
mixBART heteroBART	1.044	0.962	0.960	0.942	1.077**	0.984	0.989	0.975	1.067*	1.004	1.015	1.017
errorBART heteroBART	1.001	0.974	0.968	0.954***	0.986	0.985	0.982	0.954***	1.001	1.000	1.008	0.999
fullBART heteroBART	1.150*	0.959	0.975	0.950**	1.126	0.968	0.984	0.975**	1.073	0.980	1.016	1.012

Table 7: Quantile scores (QS) for GDPC1 with the holdout ending in 2019:Q4.

Notes: QSs for the 5th, 10th, and 25th quantile are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

		QS5				QS10				QS25			
Model		h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
SMALL													
	BVAR cons	1.133	1.078	1.256*	1.037	1.013	0.990	1.111	0.954	1.143**	0.906	0.797*	0.792**
	BART cons	1.075	1.082	1.028	0.786	0.998	0.998	0.880	0.748***	1.041	0.955	0.777***	0.677***
	mixBART cons	1.019	0.981	1.117	0.921	1.019	0.963	0.982	0.908	1.096	0.932	0.782**	0.779**
	errorBART cons	1.120	0.906	1.053	0.921	1.104	0.923	0.983	0.969	1.160***	0.962	0.843*	0.870*
	fullBART cons	1.064	1.061	1.017	0.797	1.004	0.994	0.881	0.752***	1.049	0.954	0.780**	0.682***
	BVAR SV	1.142	0.922	1.008	0.901	1.104	0.969	0.957	0.933	1.147**	0.957	0.868	0.842*
	BART SV	1.057	1.040	0.962	0.791**	0.999	0.954	0.846	0.740***	1.047	0.945	0.762**	0.666***
	mixBART SV	0.961	0.976	1.041	0.847	0.980	0.926	0.973	0.887	1.084	0.916	0.749**	0.761*
	errorBART SV	1.150	0.929	1.013	0.841	1.107	0.953	0.964	0.898	1.137**	0.948	0.797*	0.810
	fullBART SV	1.060	1.051	0.971	0.792*	0.996	0.965	0.848	0.742***	1.048	0.940	0.747***	0.660***
	BVAR heteroBART	1.214	1.051	1.161	1.081	1.146*	1.034	1.048	1.001	1.172***	0.978	0.930	0.895
	BART heteroBART	1.095	1.126	1.026	0.841	1.018	1.027	0.902	0.754***	1.064	0.969	0.786**	0.683***
	mixBART heteroBART	1.050	1.023	1.133	0.957	1.010	0.999	0.966	0.888	1.094	0.940	0.794**	0.771**
	errorBART heteroBART	1.180	0.951	1.067	0.975	1.115	0.960	1.000	0.980	1.149**	0.963	0.843*	0.874
	fullBART heteroBART	1.069	1.121	1.024	0.832	1.019	1.035	0.891	0.751***	1.049	0.956	0.771**	0.691***
MEDIUM													
	BVAR cons	1.140	1.056	1.245*	1.015	1.043	0.955	1.087	0.942	1.131**	0.897	0.810**	0.803**
	BART cons	1.188	1.039	1.003	0.852	1.057	0.981	0.860	0.788*	1.030	0.924	0.810**	0.781***
	mixBART cons	1.171	0.900	1.082	0.989	1.062	0.931	0.952	0.926	1.095	0.946	0.793***	0.826***
	errorBART cons	1.159	0.871	1.027	0.946	1.073	0.923	1.009	0.970	1.105*	0.957	0.871	0.916
	fullBART cons	1.159	1.079	0.999	0.861	1.034	0.993	0.867	0.787*	1.026	0.933	0.813**	0.785***
	BVAR SV	1.183	0.939	1.043	0.962	1.086	0.967	0.989	0.959	1.116**	0.980	0.897	0.893
	BART SV	1.034	0.999	0.929	0.797	0.953	0.920	0.813*	0.757***	1.017	0.899	0.753***	0.735***
	mixBART SV	1.077	0.941	1.029	0.940	1.003	0.936	0.935	0.901	1.059	0.931	0.764***	0.782***
	errorBART SV	1.176	0.914	0.994	0.889	1.052	0.936	0.953	0.902	1.065	0.935	0.802**	0.826**
	fullBART SV	1.177	1.044	0.949	0.847	1.061	0.975	0.836	0.791**	1.039	0.920	0.778***	0.765***
	BVAR heteroBART	1.224	1.035	1.197	1.163	1.124	1.030	1.070	1.092	1.128**	0.980	0.944	0.968
	BART heteroBART	1.138	1.123	0.960	0.860	1.024	1.028	0.845	0.772**	0.993	0.938	0.796***	0.754***
	mixBART heteroBART	1.151	1.037	1.091	1.026	1.046	0.995	0.924	0.915	1.071	0.947	0.802***	0.810***
	errorBART heteroBART	1.198	0.897	1.057	0.964	1.078	0.927	1.003	0.962	1.089	0.945	0.860*	0.901
	fullBART heteroBART	1.132	1.093	0.998	0.888	1.030	1.001	0.869	0.792**	1.028	0.925	0.779***	0.736***
LARGE													
	BVAR cons	1.129	1.061	1.257*	1.058	0.981	0.993	1.095	0.969	1.053	0.953	0.836	0.838
	BART cons	1.222	0.978	0.873	0.867	1.072	0.898*	0.796**	0.770**	1.059	0.873*	0.760***	0.746***
	mixBART cons	0.899**	0.946	0.996	0.895	0.891**	0.971	0.948	0.925	0.978	0.988	0.822	0.875
	errorBART cons	0.922	0.933	1.012	0.976	0.911	0.963	0.975	0.977	0.989	1.007	0.931	0.910**
	fullBART cons	1.210	0.967	0.894	0.874	1.057	0.907*	0.809*	0.794**	1.021	0.882*	0.761***	0.755***
	BVAR SV	0.100	0.144	0.149	0.199	0.182	0.241	0.254	0.318	0.304	0.421	0.495	0.529
	BART SV	1.165	0.943	0.872	0.852	1.014	0.865**	0.807*	0.763**	1.010	0.853**	0.765***	0.758***
	mixBART SV	0.907*	0.948	0.978	0.865	0.901*	0.956	0.949	0.904	0.971	0.947	0.808	0.851
	errorBART SV	0.959	0.975	1.001	0.931***	0.930	0.960	0.954	0.933**	0.987	0.956	0.877	0.846*
	fullBART SV	1.679**	0.997	0.875	0.844	1.341**	0.919	0.807*	0.759**	1.185**	0.867*	0.749***	0.745***
	BVAR heteroBART	1.030	1.088	1.119	1.156	0.961	1.039	1.109	1.088*	0.982	1.002	1.025	1.056
	BART heteroBART	1.459*	1.049	0.913	0.907	1.199	0.953	0.820*	0.805**	1.120	0.884*	0.798***	0.788***
	mixBART heteroBART	0.936	0.946	0.978	0.891	0.899**	0.950	0.928	0.888	0.965	0.946	0.813**	0.857
	errorBART heteroBART	0.932	0.931	1.048*	1.045	0.914	0.972	0.987	1.010	0.994	1.011	0.947	0.945
	fullBART heteroBART	1.386	1.025	0.933	0.874	1.147	0.931	0.847	0.792**	1.126	0.873**	0.767***	0.773***

Table 8: Quantile scores (QS) for GDPCTPI with the holdout ending in 2019:Q4.

Notes: QSs for the 5th, 10th, and 25th quantile are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

Model	QS5				QS10				QS25			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
SMALL												
BVAR cons	1.169	0.965	0.938	0.944	1.319***	1.032	0.995	1.003	1.332***	1.069	1.032	1.025
BART cons	0.739**	0.963**	1.006	1.012	0.790**	0.957**	1.010	1.035**	0.935	0.963	1.012	1.047***
mixBART cons	0.744**	0.962**	0.981	0.985	0.804**	0.963*	0.990	1.001	0.890*	0.965	0.988	1.014***
errorBART cons	0.991	0.995	0.968***	0.977	1.033	1.006	0.971***	0.980	1.038	1.017	0.968***	0.979
fullBART cons	0.742**	0.965**	1.002	1.014	0.791**	0.961	1.007	1.038***	0.929	0.967	1.009	1.051***
BVAR SV	1.097***	1.024	1.019	1.015	1.130***	1.019	1.018	1.020	1.114***	1.017	1.002	1.003
BART SV	0.777**	0.968*	1.012	1.008	0.818**	0.964	1.008	1.023**	0.939	0.975	1.002	1.029
mixBART SV	0.851	0.973	0.993	0.993	0.906	0.974	0.999	1.005	0.962	0.974	0.993	1.006
errorBART SV	1.086***	1.025	1.010	1.009	1.116***	1.017	1.010	1.010	1.102***	1.018	0.993	0.990
fullBART SV	0.741**	0.965*	0.994	1.000	0.793**	0.960*	0.995	1.012	0.931	0.970	0.996	1.015
BVAR heteroBART	1.075	1.017	1.025	1.022	1.083	0.999	1.008	1.010	1.083**	0.988	0.973	0.978
BART heteroBART	0.806**	0.981	1.020	1.024	0.854**	0.972	1.008	1.028	0.975	0.983	0.993	1.030
mixBART heteroBART	0.813*	0.969***	1.007	0.997	0.855*	0.963***	0.999	1.006	0.951	0.963	0.979	1.003
errorBART heteroBART	0.992	0.995	0.979**	0.987	1.026	0.997	0.977*	0.984	1.049**	1.001	0.966**	0.975
fullBART heteroBART	0.801**	0.976	1.019	1.026	0.857*	0.971	1.007	1.026	0.969	0.980	0.989	1.026
MEDIUM												
BVAR cons	1.169	0.968	0.942	0.943	1.326***	1.032	0.998	0.999	1.321***	1.073	1.034	1.025
BART cons	0.876*	0.973	0.989	0.997	0.901	0.965*	1.003	1.024	0.970	0.985	1.025	1.058**
mixBART cons	0.892**	0.976	0.972	0.979	0.914**	0.982	0.988	0.993	0.931*	0.992	1.009	1.018
errorBART cons	1.085	0.995	0.972***	0.978	1.082*	1.002	0.971***	0.977	1.055*	1.023	0.972**	0.977
fullBART cons	0.861**	0.971	0.995	1.000	0.896*	0.962*	1.002	1.028	0.962	0.989	1.021	1.058**
BVAR SV	1.174**	1.023	1.005	1.005	1.145**	1.018	1.005	1.007	1.119***	1.019	0.991	0.989
BART SV	0.875*	0.972*	0.990	1.004	0.904*	0.967**	0.998	1.019*	0.969	0.991	1.012	1.039***
mixBART SV	0.986	0.985	0.980	0.984	0.991	0.992	0.988	0.992	0.980	0.993	0.997	1.005
errorBART SV	1.168**	1.024	1.006	1.005	1.156**	1.018	1.004	1.002	1.115***	1.022	0.989	0.984
fullBART SV	0.844**	0.970*	0.985	0.995	0.867**	0.963**	0.992	1.010	0.960	0.993	1.008	1.034***
BVAR heteroBART	1.087	1.010	1.010	1.007	1.070	0.992	0.996	0.998	1.063	0.992	0.966	0.970
BART heteroBART	0.896	0.978	0.995	1.007	0.902*	0.967*	0.993	1.020	0.962	0.992	1.001	1.032***
mixBART heteroBART	0.929	0.976**	0.988	0.988	0.924*	0.973**	0.988	0.998	0.959	0.980	0.986	1.002
errorBART heteroBART	1.098	0.995	0.977***	0.983	1.085*	0.999	0.977**	0.979	1.066**	1.013	0.971**	0.973
fullBART heteroBART	0.861**	0.977	0.996	1.009	0.901*	0.963*	0.992	1.018	0.973	0.998	1.000	1.035***
LARGE												
BVAR cons	1.097	0.953	0.939	0.951	1.238**	1.021	0.999	1.012	1.247***	1.059	1.039	1.045
BART cons	0.712***	0.928*	0.970*	0.984	0.785***	0.931**	0.983	0.993	0.869**	0.959	1.003	1.019
mixBART cons	0.790**	0.944	0.955	0.972	0.832***	0.956	0.988	1.006	0.878**	0.987	1.023	1.039
errorBART cons	0.924*	0.969*	0.974***	0.982	0.936*	0.971*	0.977**	0.978	0.951**	0.995	0.979*	0.972
fullBART cons	0.685***	0.920*	0.971*	0.986	0.770***	0.926*	0.984	0.995	0.880**	0.959	1.004	1.019
BVAR SV	0.032	0.146	0.154	0.159	0.046	0.160	0.172	0.177	0.067	0.177	0.196	0.204
BART SV	0.722***	0.940	0.971	0.991	0.796***	0.947	0.986	0.998	0.874***	0.975	1.012	1.025
mixBART SV	0.865*	0.955	0.967	0.977	0.900**	0.968	0.988	1.001	0.918**	0.983	1.011	1.025
errorBART SV	1.025*	1.006	1.016	1.006	1.025*	1.000	1.010	1.005	1.006	1.003	0.994	0.983
fullBART SV	0.673***	0.925*	0.973	0.985	0.762***	0.937	0.984	0.998	0.907	0.969	1.004	1.025
BVAR heteroBART	0.842***	0.972***	0.995	0.993	0.872***	0.962***	0.985	0.988	0.920***	0.957***	0.966*	0.976
BART heteroBART	0.773***	0.938**	0.980**	0.992	0.835**	0.939**	0.981**	0.994	0.914*	0.960	0.989	1.009
mixBART heteroBART	0.769**	0.942	0.970	0.983	0.835***	0.945	0.980	0.994	0.877***	0.958*	0.992	1.010
errorBART heteroBART	0.936	0.972*	0.982**	0.987	0.968	0.972*	0.979**	0.981	0.974	0.992	0.978*	0.974
fullBART heteroBART	0.695***	0.922*	0.976***	0.990	0.781***	0.926*	0.978	1.001	0.893*	0.963	0.988	1.013

Table 9: Quantile scores (QS) for UNRATE with the holdout ending in 2019:Q4.

Notes: QSs for the 5th, 10th, and 25th quantile are computed as the ratio with respect to the large-scale Bayesian VAR with SV. Asterisks indicate statistical significance of the Diebold-Mariano test for equal predictive performance at the 1, 5, and 10 percent level. The row associated with the benchmark shows absolute numbers.

B.3 Additional results for Kullback-Leibler distances

[Figure 18 about here.]

[Figure 19 about here.]

[Figure 20 about here.]

List of Figures

1	CRPS over time.	48
2	qwCRPS-tails over time.	49
3	qwCRPS-left over time.	50
4	QS5 over time.	51
5	QS10 over time.	52
6	QS25 over time.	53
7	KLD between exact and approximate predictive distribution, GDPC1.	54
8	KLD between exact and approximate predictive distribution, GDPCTPI.	55
9	KLD between exact and approximate predictive distribution, UNRATE.	56
10	Predictive densities for GDPC1.	57
11	Predictive densities for GDPCTPI.	58
12	Predictive densities for UNRATE.	59
13	Volatility predictions for GDPC1 (posterior median).	60
14	Volatility predictions for GDPCTPI (posterior median).	61
15	Volatility predictions for UNRATE (posterior median).	62
16	Percentiles of the posterior predictive distributions for different quantiles of the NFCI.	63
17	One-step-ahead predictive distributions based on different values of the NFCI . . .	64
18	KLD between exact and approximate predictive distribution, GDPC1.	65
19	KLD between exact and approximate predictive distribution, GDPCTPI.	66
20	KLD between exact and approximate predictive distribution, UNRATE.	67

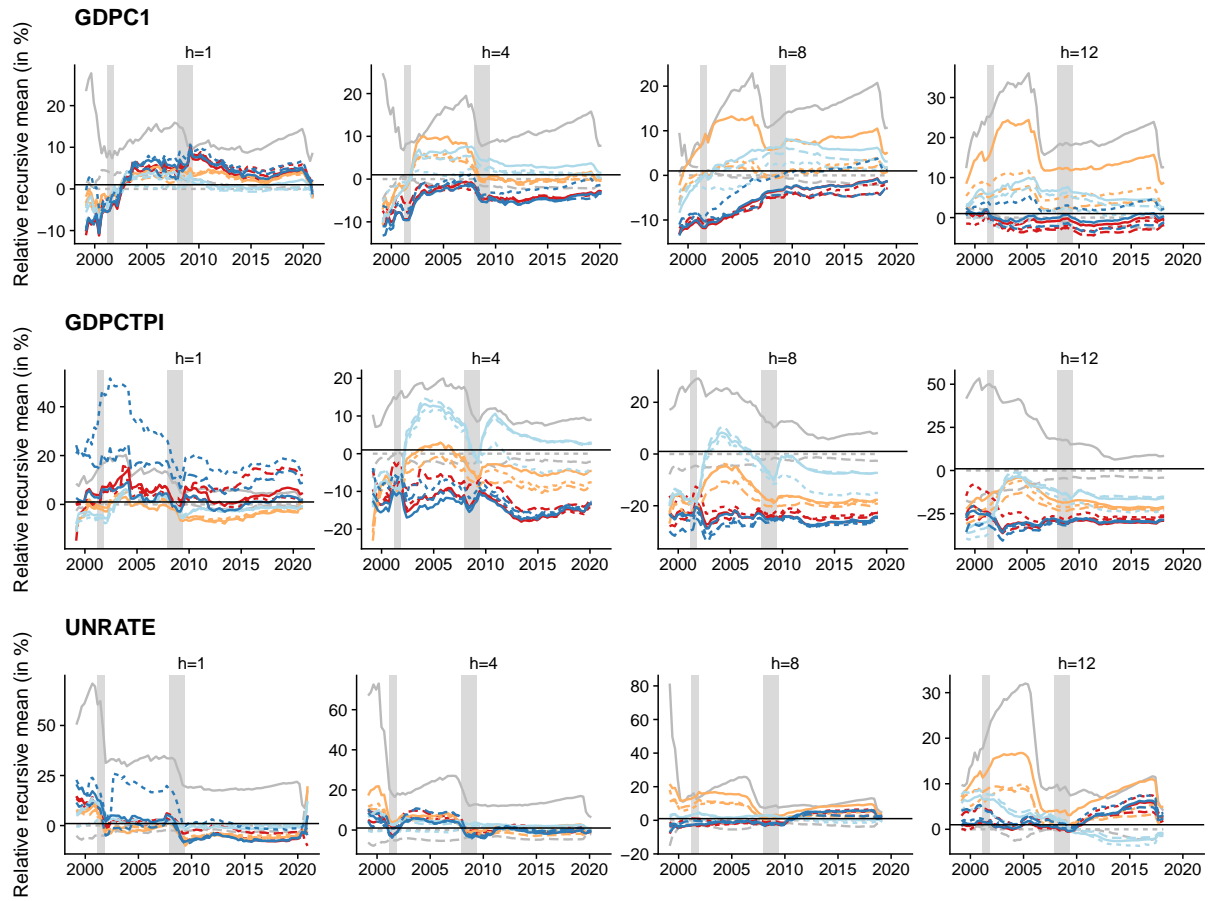


Figure 1: CRPS over time.

Notes: Recursive mean in percent relative to the benchmark model (large BVAR SV). The figure is based on the large information set. Models: BVAR (—), BART (—), mixBART (—), errorBART (---) and fullBART (---). Solid line indicates homoskedastic variances, short dashes mark SV, and long dashes are heteroBART.

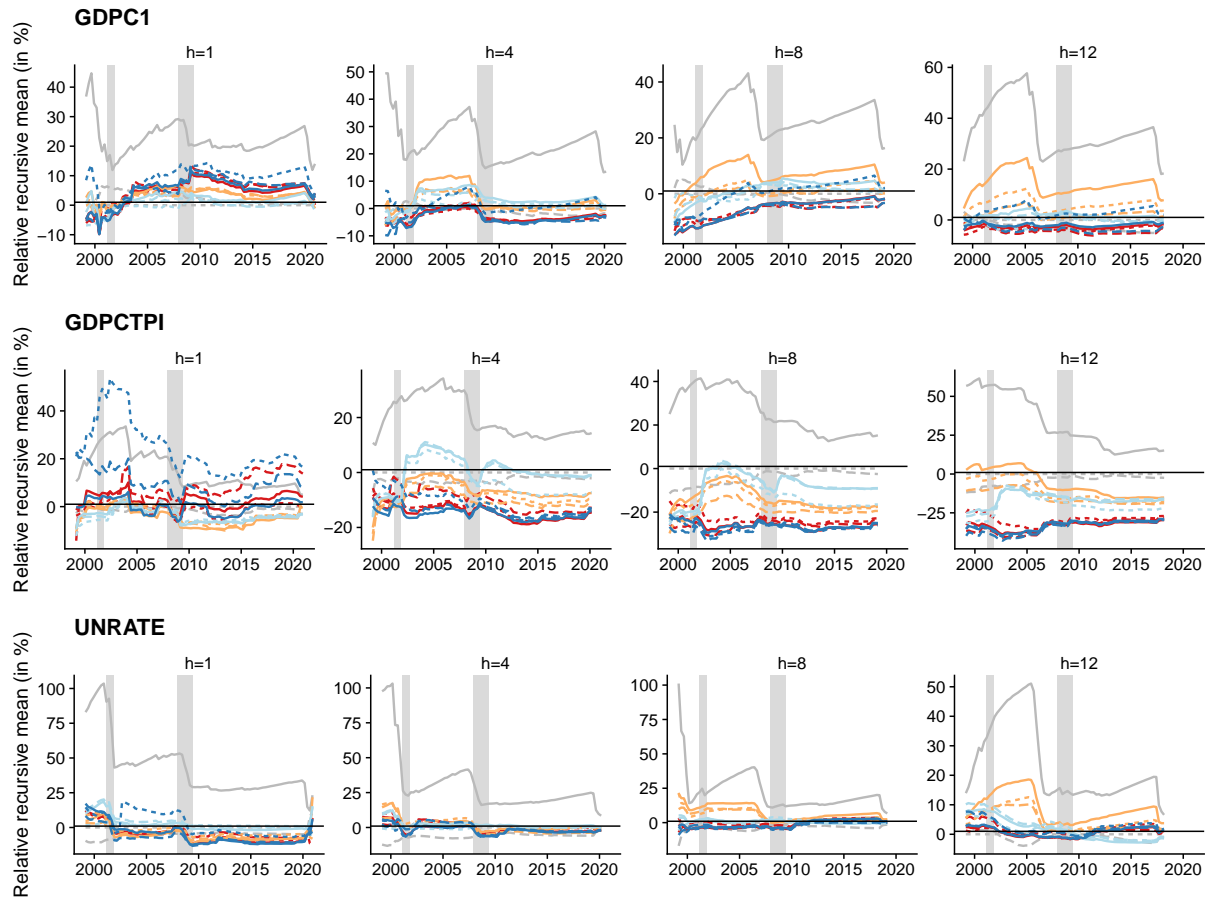


Figure 2: qwCRPS-tails over time.

Notes: Recursive mean in percent relative to the benchmark model (large BVAR SV). The figure is based on the large information set. Models: BVAR (—), BART (—), mixBART (—), errorBART (---) and fullBART (---). Solid line indicates homoskedastic variances, short dashes mark SV, and long dashes are heteroBART.

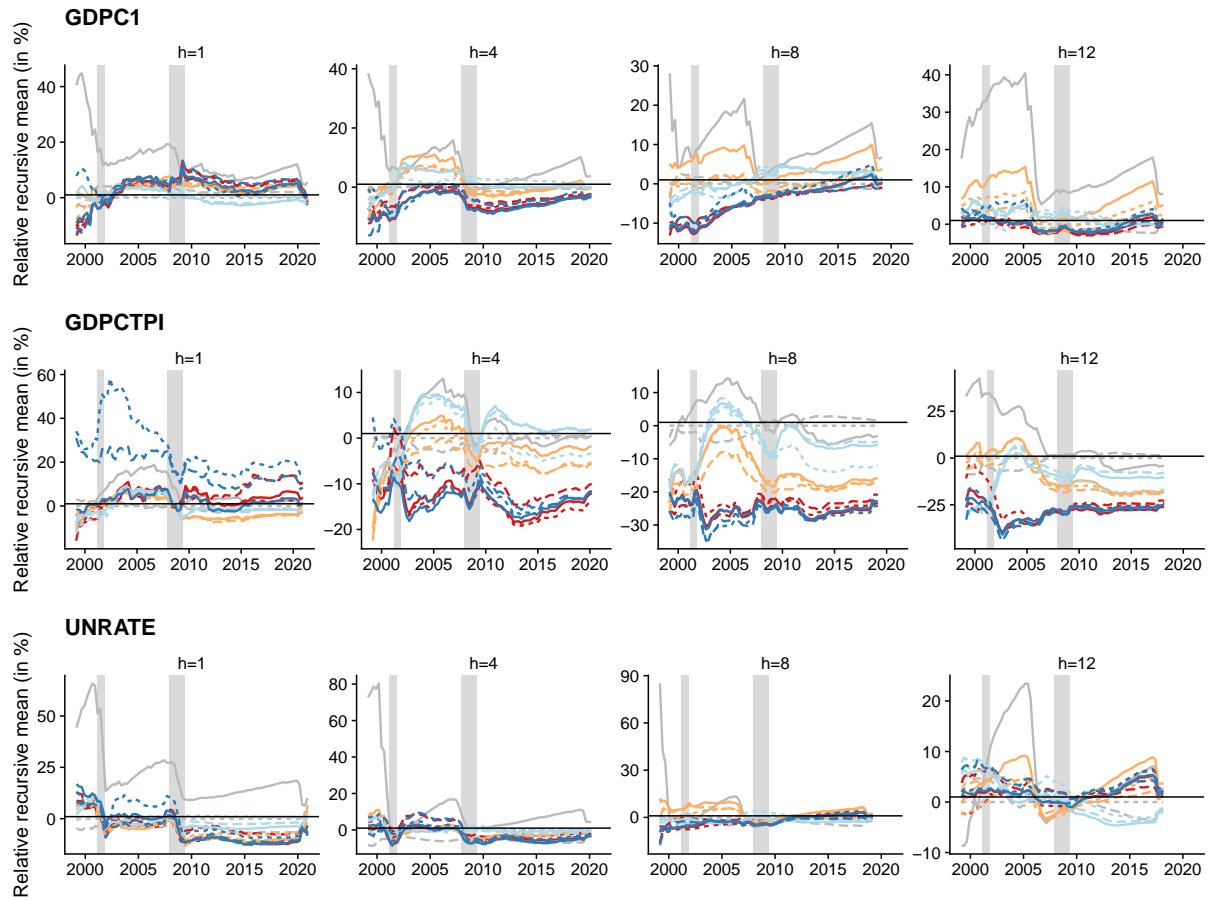


Figure 3: qwCRPS-left over time.

Notes: Recursive mean in percent relative to the benchmark model (large BVAR SV). The figure is based on the large information set. Models: BVAR (—), BART (—), mixBART (—), errorBART (---) and fullBART (---). Solid line indicates homoskedastic variances, short dashes mark SV, and long dashes are heteroBART.

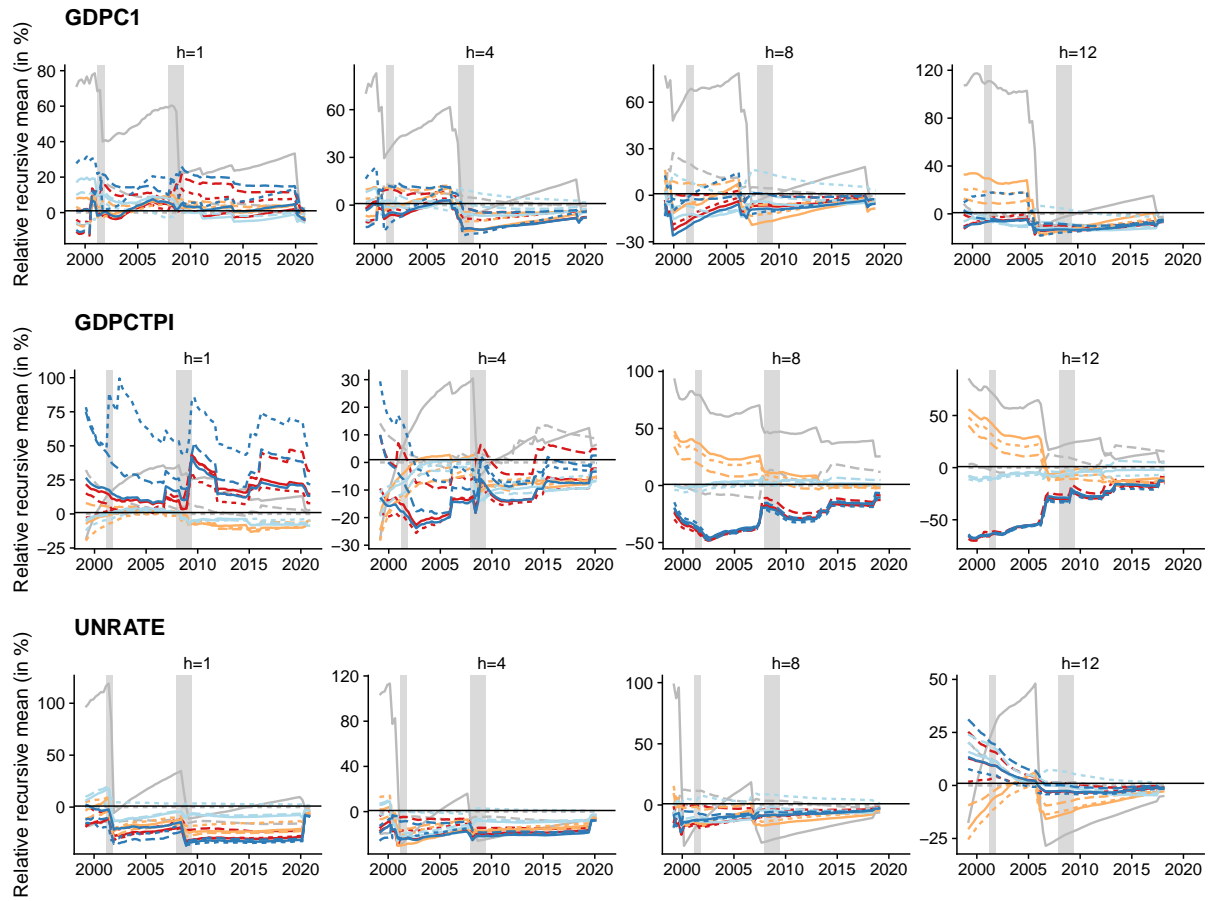


Figure 4: QS5 over time.

Notes: Recursive mean in percent relative to the benchmark model (large BVAR SV). Figure is based on the large information set. Models: BVAR (—), BART (—), mixBART (—), errorBART (---) and fullBART (---). Solid line indicates homoskedastic variances, short dashes mark SV, and long dashes are heteroBART.

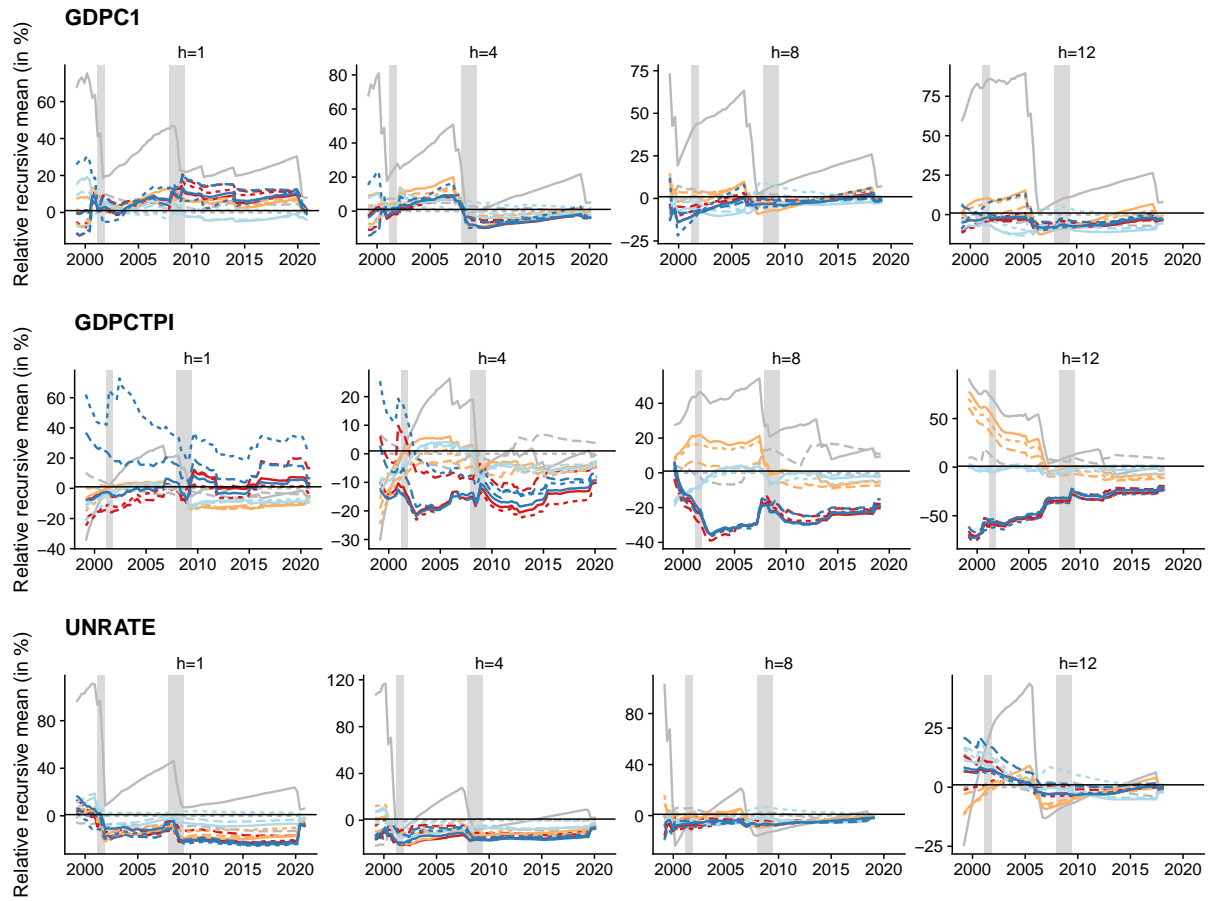


Figure 5: QS10 over time.

Notes: Recursive mean in percent relative to the benchmark model (large BVAR SV). Figure is based on the large information set. Models: BVAR (—), BART (—), mixBART (—), errorBART (---) and fullBART (—). Solid line indicates homoskedastic variances, short dashes mark SV, and long dashes are heteroBART.

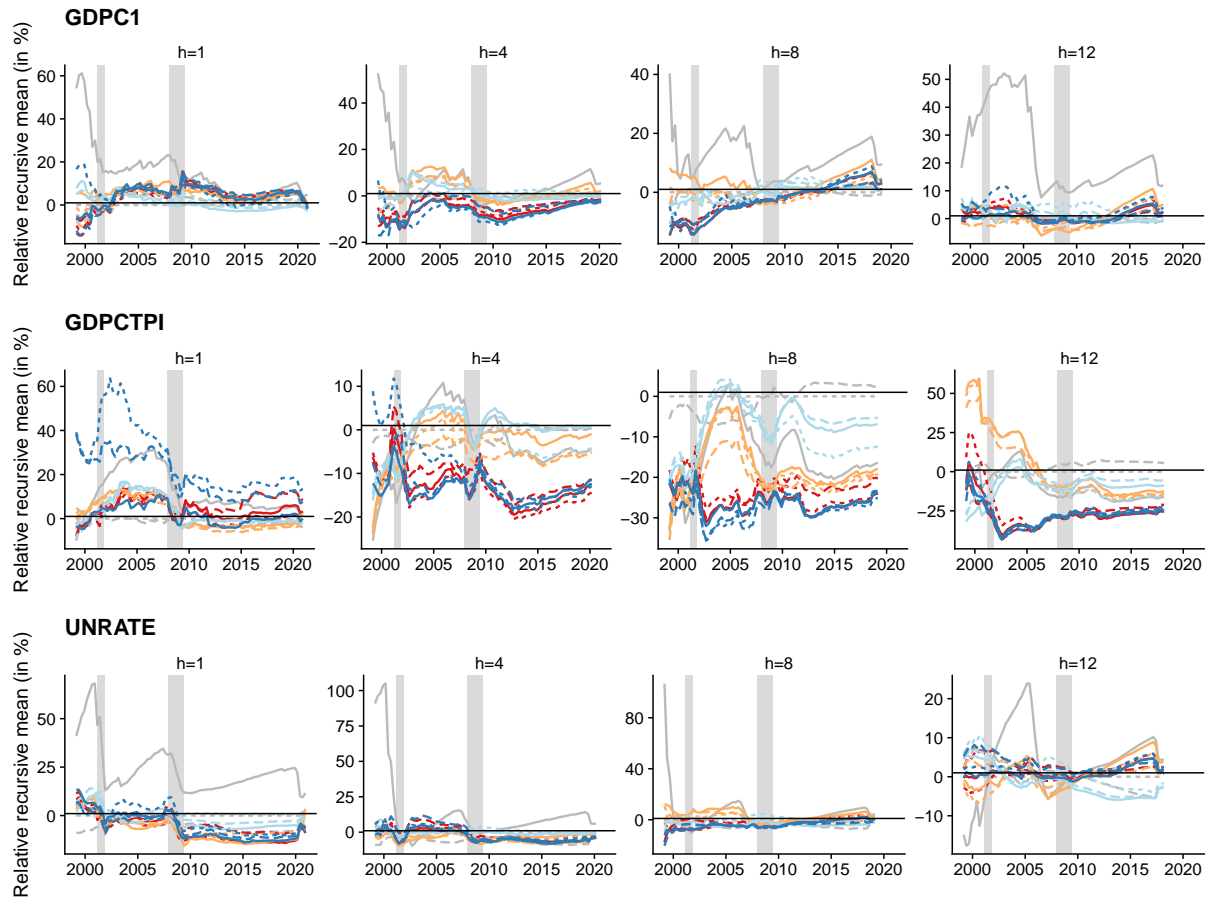


Figure 6: QS25 over time.

Notes: Recursive mean in percent relative to the benchmark model (large BVAR SV). Figure is based on the large information set. Models: BVAR (—), BART (—), mixBART (—), errorBART (---) and fullBART (—). Solid line indicates homoskedastic variances, short dashes mark SV, and long dashes are heteroBART.

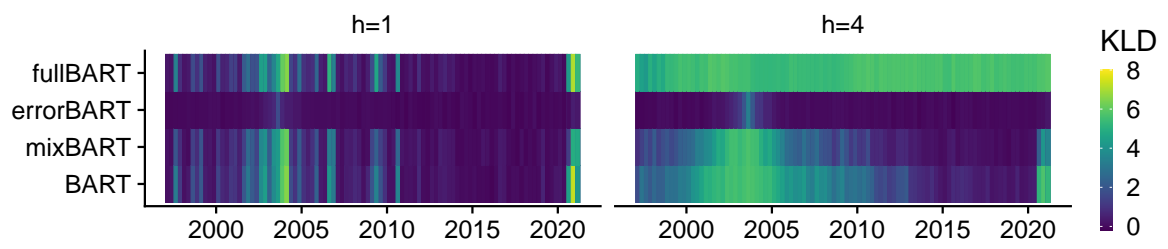


Figure 7: KLD between exact and approximate predictive distribution, GDPC1.

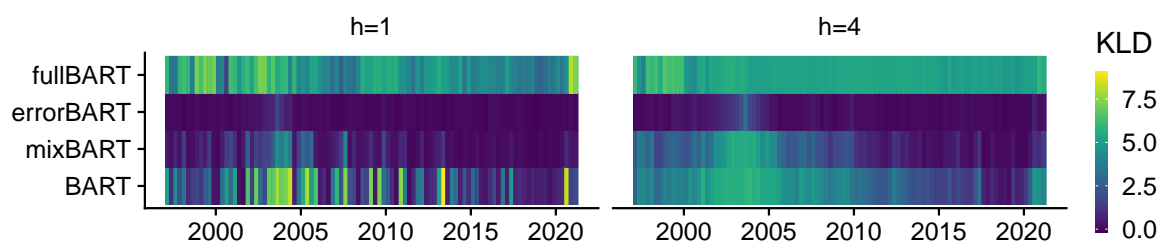


Figure 8: KLD between exact and approximate predictive distribution, GDPCTPI.

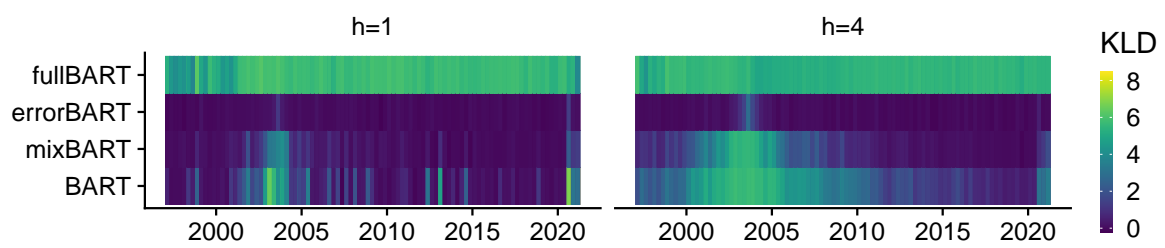


Figure 9: KLD between exact and approximate predictive distribution, UNRATE.

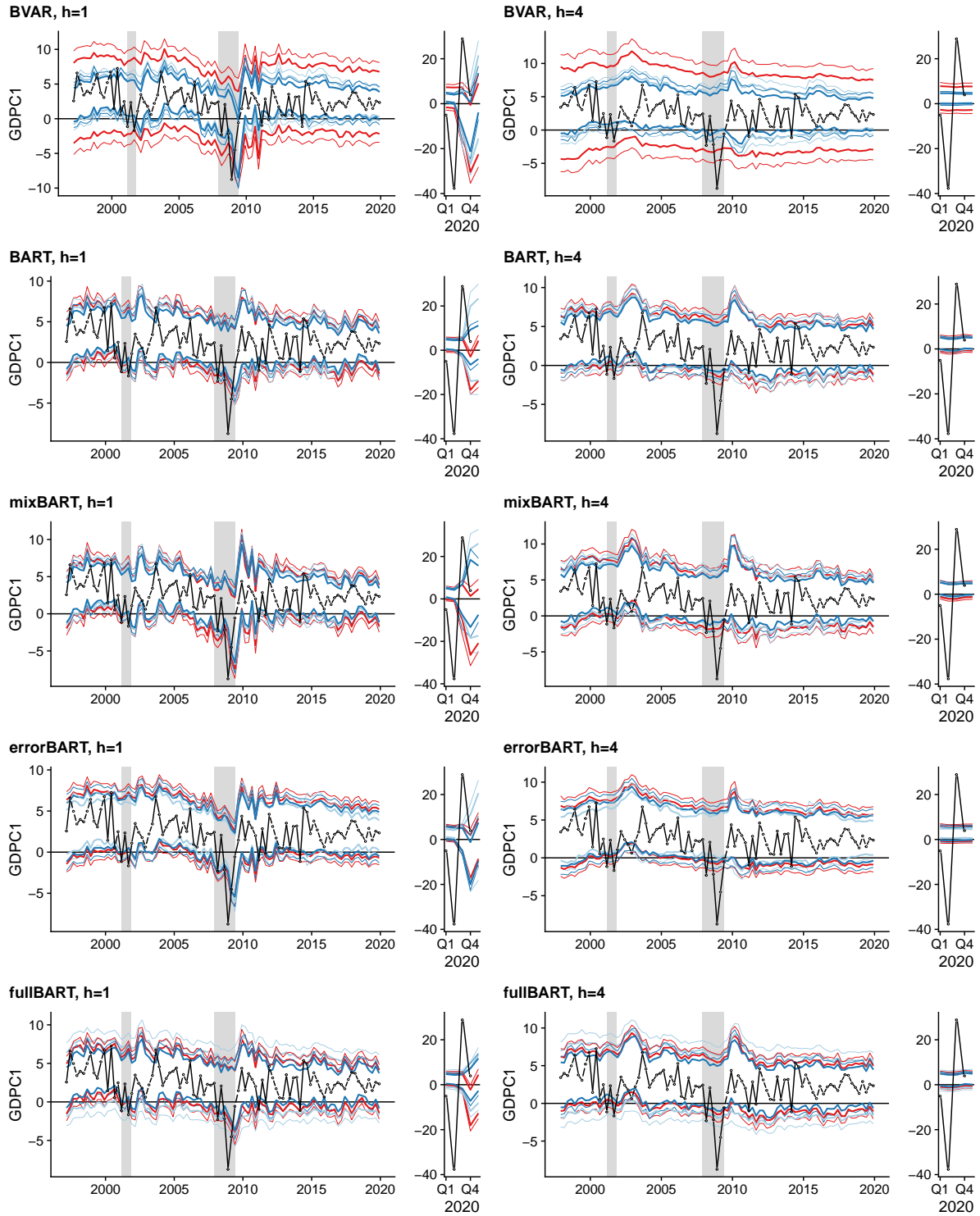


Figure 10: Predictive densities for GDPC1.

Notes: Constant volatility (—), SV (—), heteroBART (—). Thin colored lines mark the 5/95th percentile, thick lines the 10/90th percentile. Black lines and points are realizations of the final vintage.

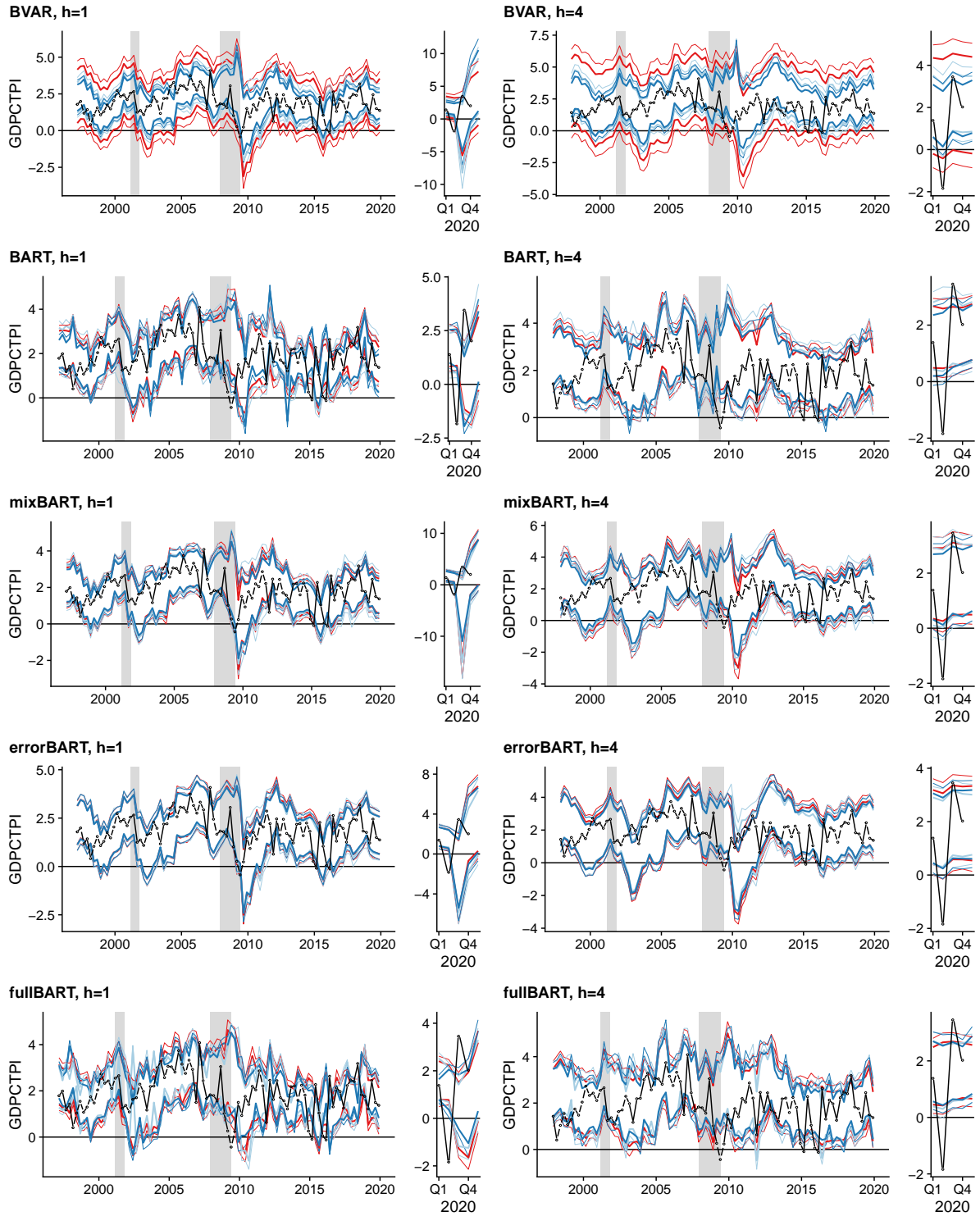


Figure 11: Predictive densities for GDPCTPI.

Notes: Constant volatility (—), SV (—), heteroBART (—). Thin colored lines mark the 5/95th percentile, thick lines the 10/90th percentile. Black lines and points are realizations of the final vintage.

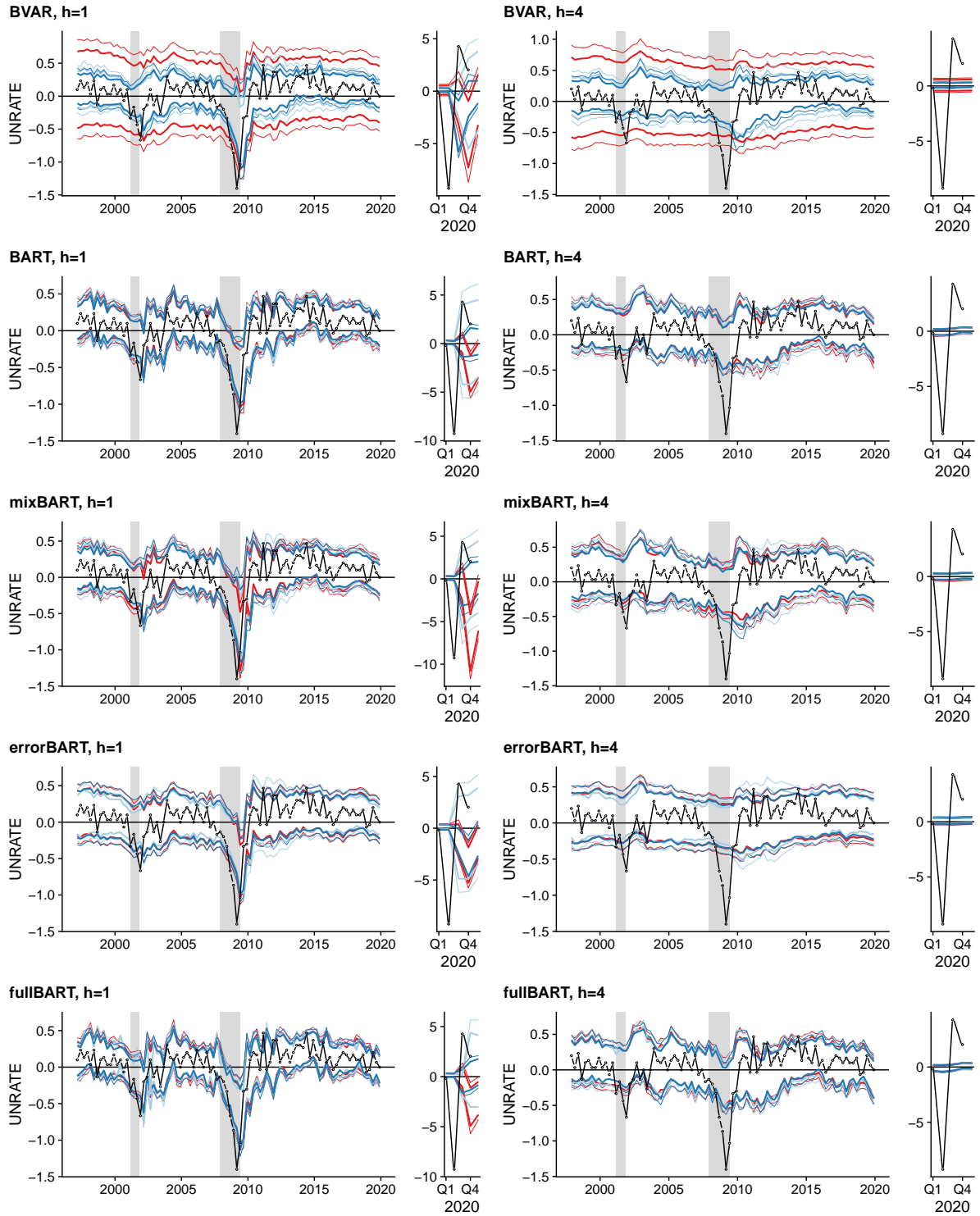


Figure 12: Predictive densities for UNRATE.

Notes: Constant volatility (—), SV (—), heteroBART (—). Thin colored lines mark the 5/95th percentile, thick lines the 10/90th percentile. Black lines and points are realizations of the final vintage.

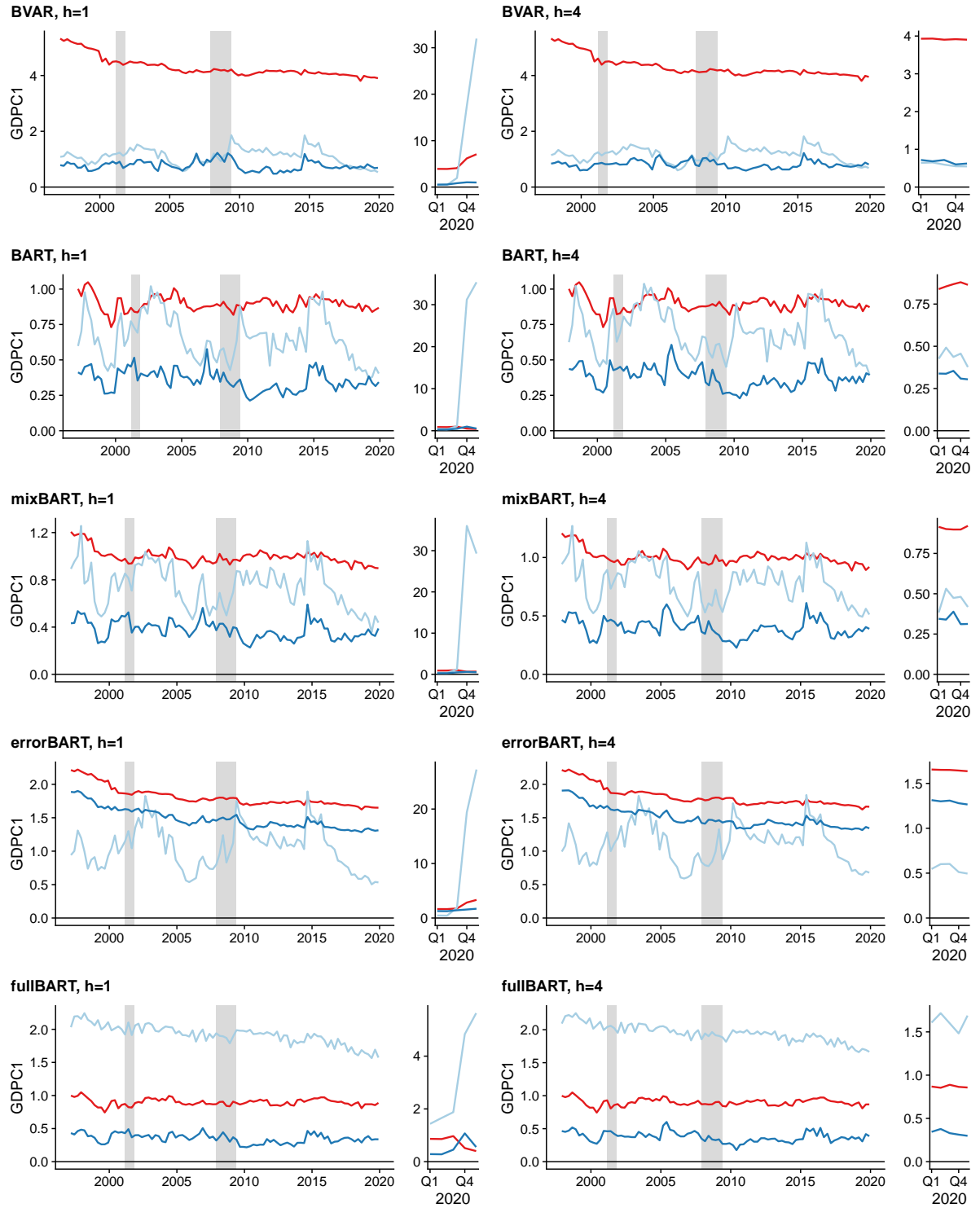


Figure 13: Volatility predictions for GDPC1 (posterior median).

Notes: Constant volatility (—), SV (—), heteroBART (—).

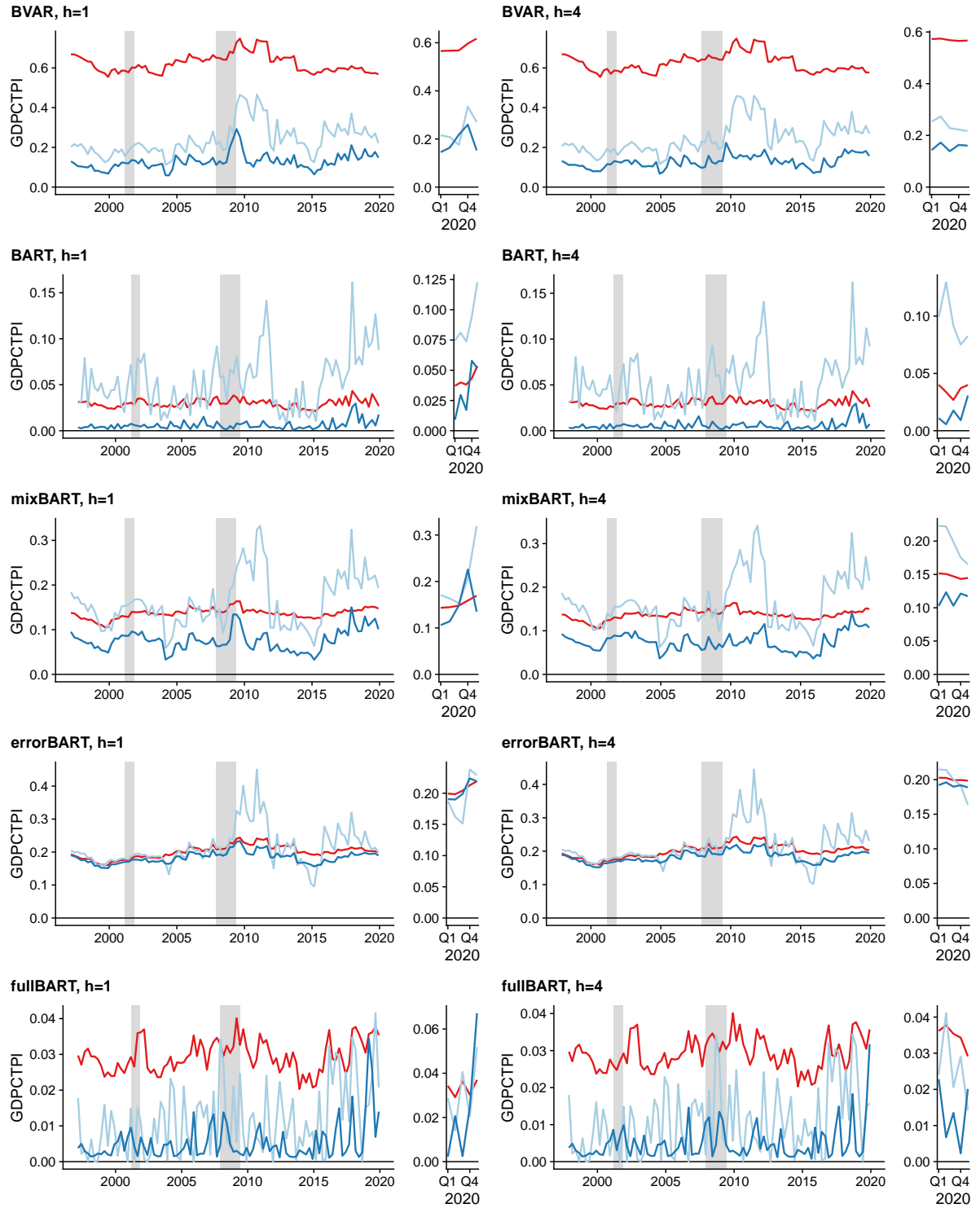


Figure 14: Volatility predictions for GDPCTPI (posterior median).

Notes: Constant volatility (—), SV (—), heteroBART (—).

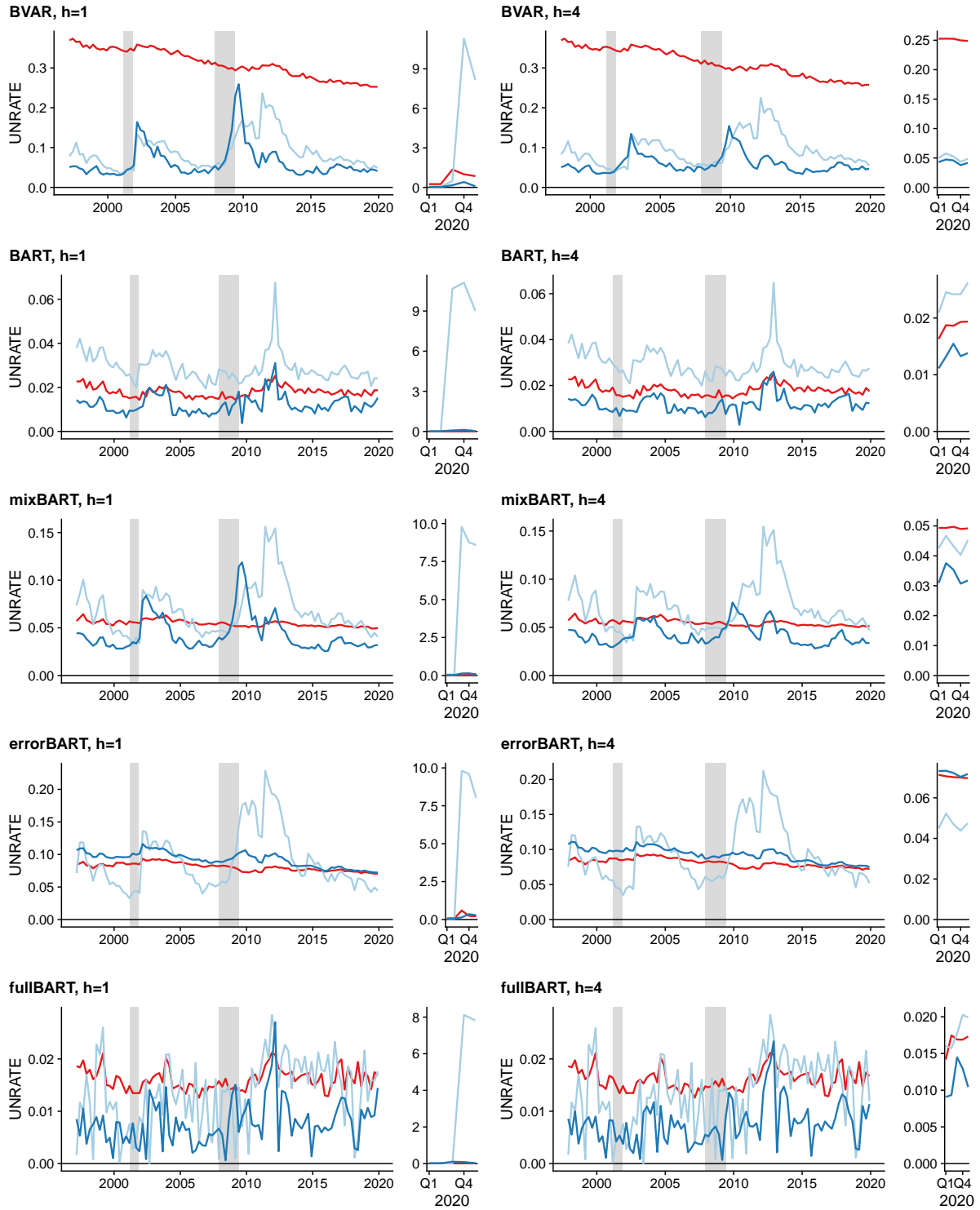
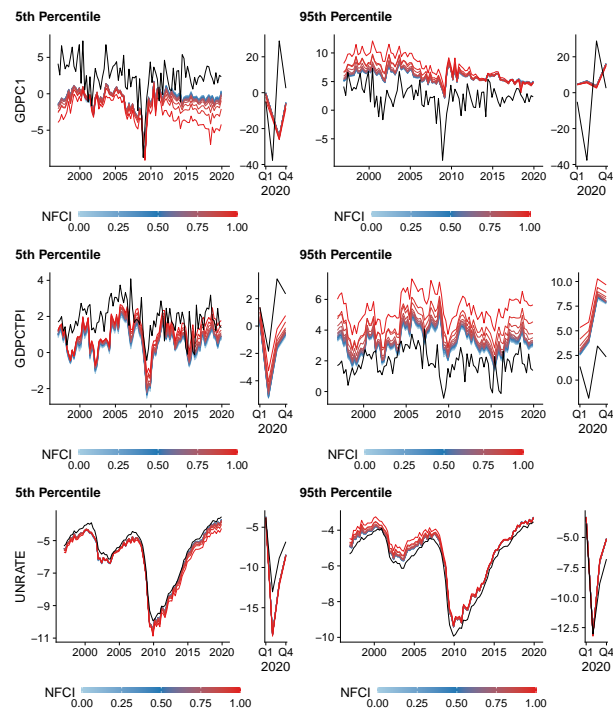


Figure 15: Volatility predictions for UNRATE (posterior median).

Notes: Constant volatility (—), SV (—), heteroBART (—).

(a) BVAR heteroBART



(b) BART heteroBART

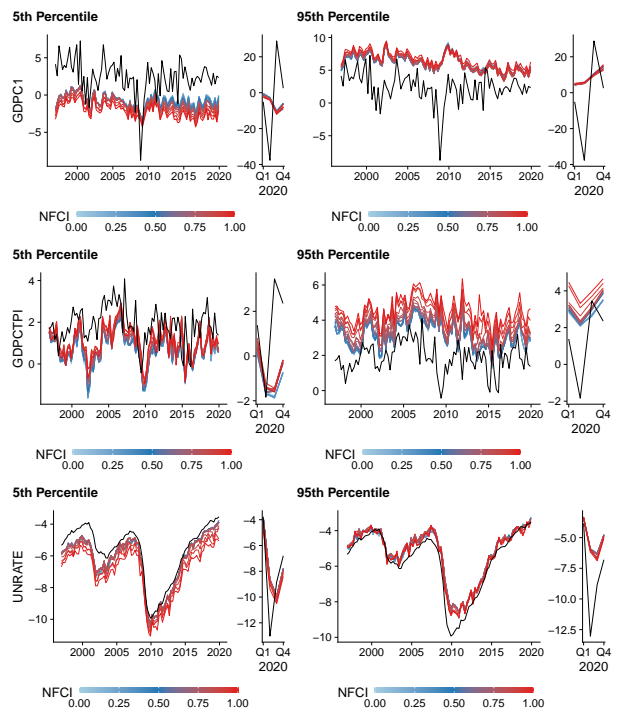
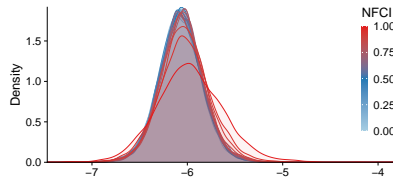


Figure 16: Percentiles of the posterior predictive distributions for different quantiles of the NFCI.

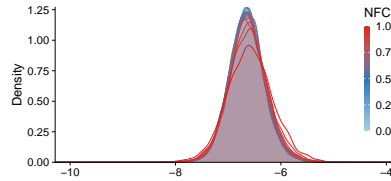
Notes: The quantiles range from 0 to 1 with step size of 0.05. The legend refers to the quantiles.

BVAR heteroBART

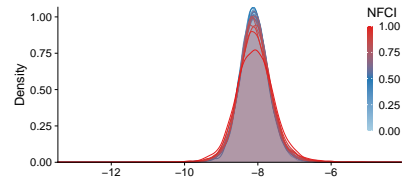
2008:Q3



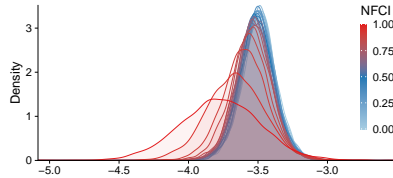
2008:Q4



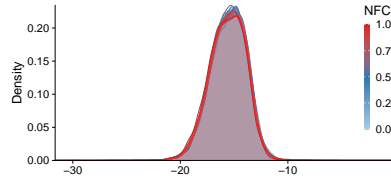
2009:Q1



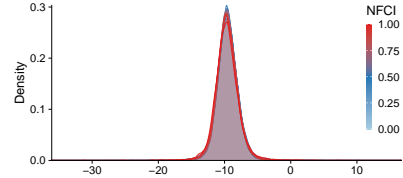
2020:Q1



2020:Q2

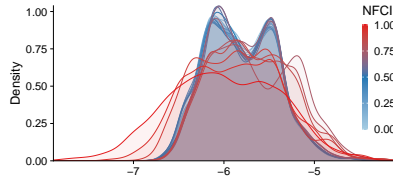


2020:Q3

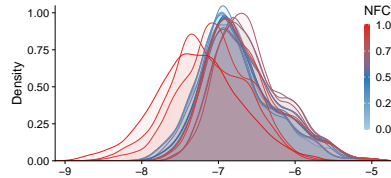


BART heteroBART

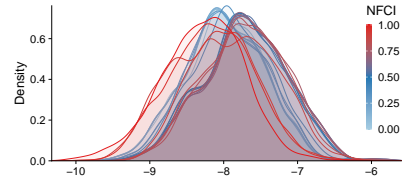
2008:Q3



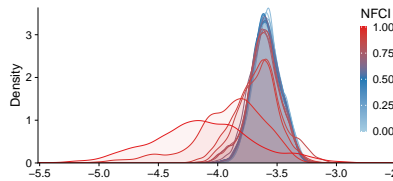
2008:Q4



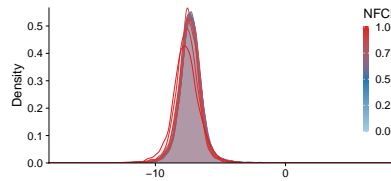
2009:Q1



2020:Q1



2020:Q2



2020:Q3

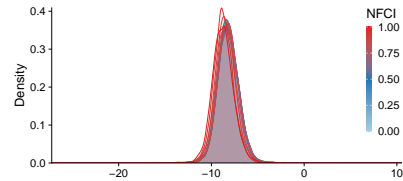


Figure 17: One-step-ahead predictive distributions based on different values of the NFCI

Notes: The quantiles range from 0 to 1 with step size of 0.05. The legend refers to the quantiles.

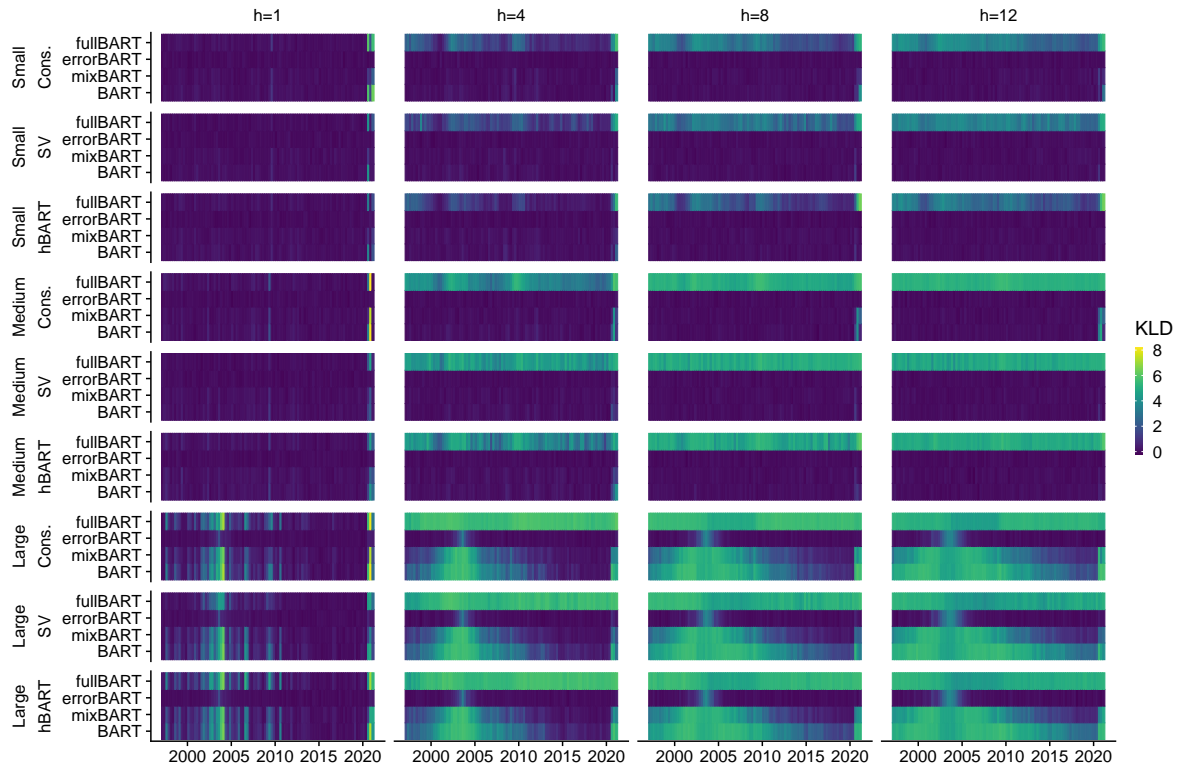


Figure 18: KLD between exact and approximate predictive distribution, GDPC1.

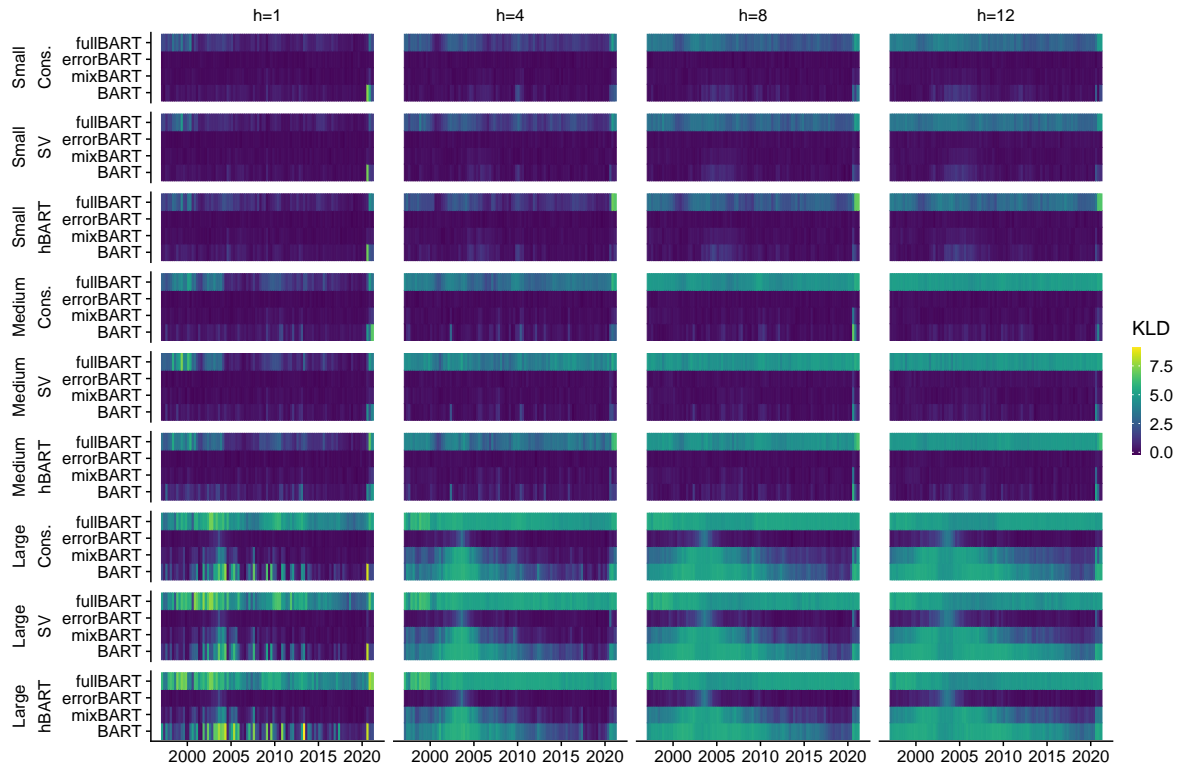


Figure 19: KLD between exact and approximate predictive distribution, GDPCTPI.

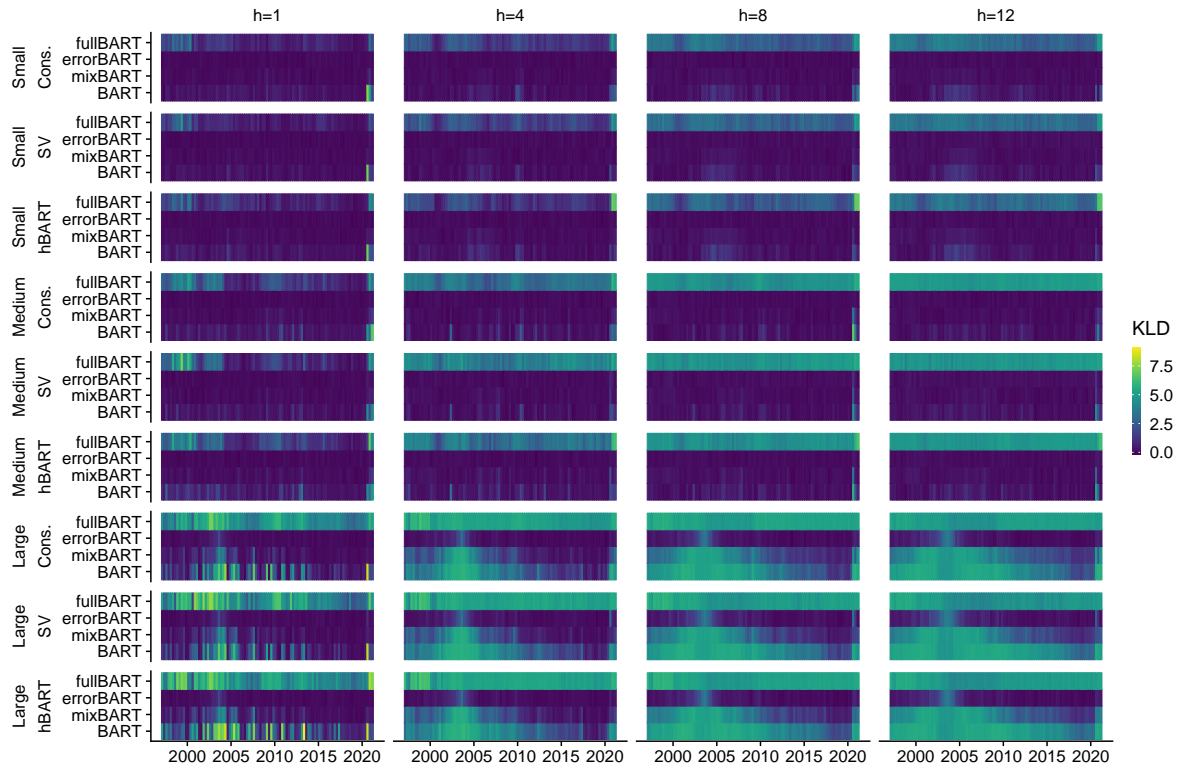


Figure 20: KLD between exact and approximate predictive distribution, UNRATE.