Journal of the Operational Research Society

ISSN: 0160-5682 (Print) 1476-9360 (Online) Journal homepage: www.tandfonline.com/journals/tjor20

Modeling the distribution of key economic indicators in a data-rich environment: new empirical evidence

Iason Kynigakis & Ekaterini Panopoulou

To cite this article: Iason Kynigakis & Ekaterini Panopoulou (2025) Modeling the distribution of key economic indicators in a data-rich environment: new empirical evidence, Journal of the Operational Research Society, 76:10, 2071-2090, DOI: 10.1080/01605682.2025.2457645

To link to this article: https://doi.org/10.1080/01605682.2025.2457645

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
2025, VOL. 76, NO. 10, 2071–2090
https://doi.org/10.1080/01605682.2025.2457645

RESEARCH ARTICLE

OPEN ACCESS

Check for updates

Modeling the distribution of key economic indicators in a data-rich environment: new empirical evidence

Iason Kynigakis[a] and Ekaterini Panopoulou[b]

[a]University College Dublin, Dublin, Ireland; [b]University of Essex, Colchester, UK

ABSTRACT

This study explores the ability of a large number of macroeconomic variables to forecast the mean, quantiles and density of key economic indicators. In the baseline case, we construct the forecasts using an autoregressive model. We then consider several general specifications that augment the time series model with macroeconomic information, either directly using the full set of predictors, through targeted-factors, targeted-predictors or forecast combinations. Our findings suggest that aggregating information across quantiles leads to improved estimates of the conditional mean. Overall, augmenting the autoregressive model with macroeconomic variables through methods that perform variable selection or account for non-linearities improves predictive performance. This increase in out-of-sample performance arises from the improved estimation of the lower and middle part of the distribution.

ARTICLE HISTORY
Received 6 February 2024
Accepted 19 January 2025

KEYWORDS
Distribution forecasting; economic forecasting; forecast combination; machine learning; quantile regression

JEL CLASSIFICATIONS
C22; C51; C52; C53; C55

1. Introduction

The prediction of the future evolution of key economic indicators is imperative to the conduct of economic policy, due to the delayed impact of a central bank's actions on economic activity. Therefore, accuracy is essential when forecasting the effects of various shocks on the future dynamics of key business cycle and inflation indicators. Typically, forecasting focuses on modeling the conditional mean using a large number of macroeconomic variables (see e.g., Kim & Swanson, 2014; Medeiros et al., 2021; Stock & Watson, 2002a, 2002b). However, central banks are increasingly concerned about the uncertainty around the point forecasts of economic indicators such as industrial production, inflation and employment. This has led to a growing number of recent papers that focus on modeling and forecasting the quantiles and density of economic indicators (see e.g., Amisano & Giacomini, 2007; Barbaglia et al., 2023; Carriero et al., 2024; De Gooijer & Zerom, 2019; Manzan, 2015; Pfarrhofer, 2022; Rossi & Sekhposyan, 2014). Forecasting the distribution of economic variables is important for several reasons. Density forecasts are able to fully capture the uncertain future behavior of an economic indicator, instead of measuring its central tendency similar to point forecasts. Furthermore, forecasting the distribution becomes important when a central bank evaluates the risks of a future increase or decrease of an economic indicator differently. Distribution forecasts can also lead to more accurate estimates of the conditional mean, which can be modeled as a function of individual quantiles (Lima & Meng, 2017).

The aim of this study is to forecast the mean, quantiles and distribution of the Industrial Production Index, the Consumer Price Index for all urban consumers and Non-farm Payroll Employment using models that include both a time series component and a large number of economic variables. In the baseline case, we forecast the conditional mean using an autoregressive model and to forecast the quantiles we employ the quantile autoregressive model, proposed by Koenker and Xiao (2006), where the forecasts depend only on the past values of the target variable. To enhance the forecasting models with a richer information set, we follow Kim and Swanson (2014) and Manzan (2015) by assuming that both the conditional mean and quantiles of the variables being forecast are functions of their own lags in addition to a large panel of macroeconomic and financial indicators based on the FRED-MD database by McCracken and Ng (2016).

We explore four general specifications to forecast the conditional mean or quantiles, all of which augment the baseline time series model with economic indicators. In the first approach, we directly augment the baseline model by incorporating the full set of

CONTACT Iason Kynigakis iason.kynigakis@ucd.ie University College Dublin, Dublin, Ireland
Supplemental data for this article can be accessed online at https://doi.org/10.1080/01605682.2025.2457645.

predictors through methods that perform dimensionality reduction, variable selection and account for non-linearities. The next two approaches exploit the ability of machine learning methods to uncover important variables and non-linear interactions. Specifically, the second approach directly includes the selected variables to the baseline model, while the third approach constructs latent factors based on the subset of targeted predictors before augmenting the time series model. The final approach extends the baseline model through forecast combinations of bivariate prediction models, based on either simple weighting schemes or with the weights estimated using machine learning methods. The proposed specifications allow for heterogeneous degrees of persistence of the variable we forecast and asymmetric dynamic responses of economic variables at different parts of the distribution. To take advantage of the varied information content in the forecasts of different quantiles, we follow Lima and Meng (2017) and Meligkotsidou et al. (2019) and construct forecasts of the conditional mean as the weighted average of a set of conditional quantile forecasts.

The analysis by Manzan (2015) is similar to the one considered in this paper, however, there are several differences between the two. First, our study considers a general specification of models based on forecast combinations, in addition to targeted predictors or factors. We also expand the scope of machine learning models beyond the lasso to encompass non-linear models such as random forests, gradient boosting, and neural networks. In addition to examining the accuracy of individual quantile forecasts, we also assess the models' ability to predict the entire distribution of economic indicators, similar to Manzan and Zerom (2013). However, unlike their study, we use a larger dataset of macroeconomic predictors, a wider selection of models and focus on industrial production and employment in addition to inflation. Finally, in line with Kim and Swanson (2014) and Medeiros et al. (2021), we add to the current literature on machine learning-based forecasting of the conditional mean of economic indicators in a data-rich environment. However, we enhance the forecasting accuracy of the center of the distribution by combining information across quantiles from several model specifications.

An alternative framework often used to forecast the quantiles and distribution of economic indicators is based on stochastic volatility models. Therefore, in addition to models within the four augmented autoregressive specifications, we consider univariate autoregressive stochastic volatility models with symmetric or asymmetric error distributions (see e.g., Jacquier et al., 1994; Kim et al., 1998; Omori et al., 2007; Taylor, 1982). Since cross-lags have been shown to improve predictive performance (Gruber & Kastner, 2022), we also employ vector autoregressions with stochastic volatility based on Bayesian estimation methods. Bayesian vector autoregressions (BVAR) are widely used for macroeconomic forecasting, as Bayesian shrinkage helps mitigate the curse of dimensionality from the large number of parameters and Bayesian estimation facilitates the efficient computation of time-varying volatility (see e.g., Bańbura et al., 2010; Carriero et al., 2019; Giannone et al., 2015; Huber & Feldkircher, 2019). The BVAR models differ according to the structure of the covariance matrix, with one case assuming that the errors follow a Cholesky stochastic volatility structure, and the other that the covariance matrix has a factor stochastic volatility structure.

Our findings suggest that combining information across quantiles improves the accuracy of conditional mean forecasts relative to generating point forecasts directly. Additionally, including economic variables to the baseline model through methods that perform variable selection or account for non-linearities can further improve predictive performance. The results also show that models combining information or forecasts generate performance equivalent to or surpassing that of stochastic volatility models. The out-of-sample analysis indicates that augmenting the autoregressive model with macroeconomic and financial information can increase the accuracy of conditional quantile forecasts. This improved forecasting performance is observed especially in the lower and middle quantiles. For the density forecast evaluation, the majority of the models significantly outperform the benchmark, which can be attributed to greater forecasting accuracy in the left tail and center of the distribution. Models incorporating a large number of predictors offer the highest accuracy for industrial production and inflation, while stochastic volatility models yield comparable out-of-sample performance for employment.

Our findings show that the global financial crisis represents a break for the evolution of model performance over time. During the crisis, cumulative performance is more volatile for industrial production and inflation than it is for employment, while after the crisis, performance of the models over the benchmark accumulates higher, especially for industrial production and inflation. Furthermore, extending the sample to cover the turbulent COVID-19 pandemic period leads to a decline in out-of-sample performance, with inflation forecasts being less affected compared to those for industrial production and employment. To account for the heterogeneity in forecasting accuracy across different models, we construct an amalgamation of all individual forecasts. Amalgam forecasts significantly outperform the benchmark for the conditional mean, although their performance is equivalent to that of the best-performing individual model. The amalgamation approach results in improved density forecasts and higher predictive accuracy for the lower and middle quantiles. Finally, since the contribution of a specific predictor to the formation of a forecast of the conditional mean can be different to that of the conditional quantiles, we conduct a variable importance analysis to explore which variables influence the conditional distribution far from its center. We find that variable importance not only differs across the conditional mean and quantiles, but also based on whether the predictor set is comprised of macroeconomic variables, or their individual forecasts based on simpler models.

The article is organized as follows. Section 2 discusses the models used to construct point, quantile and density forecasts. Section 3 describes the data and sample splitting. Section 4 presents the empirical results, and Section 5 concludes.

2. Methodology

In this Section we introduce the models used to forecast the conditional mean and quantiles and the benchmarks used to evaluate their predictive performance.

2.1. Autoregressive models

Let y_t, for $t = 1, 2, \ldots, T$, be the macroeconomic variable we are interested in forecasting h-step ahead that we assume is stationary. We focus on one period ahead forecasts and set $h = 1$. The baseline model we use to forecast the conditional mean is the autoregressive (AR) model of order p, where p is determined by the Bayesian information criterion (BIC), with the maximum value of p set to 12 lags. The model can be written as:

$$y_{t+h|t} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i+1} + \varepsilon_{t+h}, \tag{1}$$

where β_i, for $i = 0, 1, \ldots, p$, are the model parameters and ε_t is the error of the regression. The estimates of the parameter vector $\boldsymbol{\beta}$ are obtained by minimizing the least squares loss function:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{\beta}) \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \left(y_{t+h} - \beta_0 - \sum_{i=1}^{p} \beta_i y_{t-i+1} \right)^2.\end{aligned} \tag{2}$$

Once the parameters have been estimated the forecast for the conditional mean can be obtained as $\hat{y}_{t+h|t} = \hat{\beta}_0 + \sum_{i=1}^{p} \hat{\beta}_i y_{t-i+1}$.

However, the above loss function is affected by the presence of extreme observations and can be restrictive since it focuses only on one aspect of the distribution of y_t. These potential limitations led to the development of quantile regression, introduced by Koenker and Bassett (1978), who generalize ordinary sample quantiles to a regression setting, thus providing a more complete approximation for the distribution of y_t.

In this setting, the baseline approach we use to model the quantiles of y_t is the quantile autoregressive (QAR) specification considered in Koenker and Xiao (2006):

$$q_{t+h|t}(\tau) = \beta_0(\tau) + \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1} + \varepsilon_{t+h}(\tau), \tag{3}$$

where $q_t(\tau)$ indicates the $\tau \in (0, 1)$ conditional quantile of y_t, $\beta_i(\tau)$, for $i = 0, 1, \ldots, p_\tau$, are the model parameters depending on τ, p_τ is the lag order used to model $q_t(\tau)$ and $\varepsilon_t(\tau)$ is the error for quantile τ. The QAR model extends the AR model used for the conditional mean to a quantile regression setting. This model allows the lag order to vary at different parts of the distribution. To select the lag order at each quantile we follow Manzan (2015) and minimize a BIC-type criterion based on the quantile loss function. The parameters of the model are estimated by minimizing the following function:

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\tau) &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \mathcal{Q}_\tau(\boldsymbol{\beta}(\tau)) \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \mathcal{Q}_\tau \left(y_{t+h} - \beta_0(\tau) - \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1} \right),\end{aligned} \tag{4}$$

where $\mathcal{Q}_\tau(\cdot)$ is the quantile loss function defined as:

$$\mathcal{Q}_\tau(\varepsilon_{t+h}(\tau)) = \begin{cases} \tau \varepsilon_{t+h}(\tau), & \text{if } \varepsilon_{t+h}(\tau) \geq 0 \\ (\tau - 1)\varepsilon_{t+h}(\tau), & \text{if } \varepsilon_{t+h}(\tau) < 0 \end{cases} \tag{5}$$

and $\varepsilon_{t+h}(\tau) = y_{t+h} - \beta_0(\tau) - \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1}$. Similarly, the forecast for the τth quantile can be obtained as $\hat{q}_{t+h|t}(\tau) = \hat{\beta}_0(\tau) + \sum_{i=1}^{p_\tau} \hat{\beta}_i(\tau) y_{t-i+1}$. For all models, we rearrange the quantiles to avoid quantile crossing as proposed in Chernozhukov et al. (2010).

2.2. Augmented AR models

To examine whether augmenting the AR and QAR models with information from a large number of macroeconomic variables can improve forecast accuracy, we consider several model specifications. In the first specification we augment the autoregressive models with the direct inclusion of macroeconomic variables modelled using a variety of flexible functions from machine learning that induce sparsity or introduce non-linearities. The second approach allows a subset of predictors chosen by machine learning methods to enter the model directly in a linear fashion. An alternative approach consists of extracting principal components of the macroeconomic variables chosen by machine learning models. The final approach combines the forecasts of bivariate prediction models to construct forecasts for the conditional mean and quantiles, similar to Huang and Lee (2010) and Rossi and Sekhposyan (2014). We consider both simple forecast combination schemes and approaches based on machine learning.

The augmented AR models build upon the residuals of the autoregressive model similarly to Kim and Swanson (2014). For all models described below we start by fitting an autoregressive model to the dependent variable, excluding predictor variables, using least squares for the AR model or quantile regression for the QAR model, and retain the corresponding residual series. For all augmented model specifications, we use the same lag orders, p or p_τ, selected for the autoregressive part of the model as discussed above. The objective then becomes to map the macroeconomic variables to the residuals using different functions, in order to improve upon the fit of the autoregressive model.

Let $\mathbf{z}_t$ be the N-vector of stationary predictors at time t, the augmented AR model for the conditional mean is:

$$y_{t+h|t} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i+1} + g(\mathbf{z}_t;\boldsymbol{\theta}) + u_{t+h}, \tag{6}$$

where $g(\cdot)$ is a flexible function with parameters $\boldsymbol{\theta}$ that maps the predictors to the macroeconomic variable y_t through the residuals of the AR model and u_t are the errors of the augmented model. The parameters are derived by minimizing the least squares loss:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} \left(\varepsilon_{t+h} - g(\mathbf{z}_t;\boldsymbol{\theta})\right)^2. \tag{7}$$

The forecast for the conditional mean is $\hat{y}_{t+h|t} = \hat{\beta}_0 + \sum_{i=1}^{p} \hat{\beta}_i y_{t-i+1} + \hat{g}\left(\mathbf{z}_t;\hat{\boldsymbol{\theta}}\right)$, where $\hat{g}$ is the estimated function based on data up to time t.

The augmented QAR model is given by:

$$\begin{aligned} q_{t+h|t}(\tau) = \beta_0(\tau) + \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1} + g_\tau(\mathbf{z}_t;\boldsymbol{\theta}(\tau)) \\ + u_{t+h}(\tau), \end{aligned} \tag{8}$$

where the parameters $\boldsymbol{\theta}(\tau)$ of function $g_\tau(\cdot)$ now depend on the conditional quantile τ and are estimated by minimizing the following function:

$$\begin{aligned} \hat{\boldsymbol{\theta}}(\tau) &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{Q}_\tau(\boldsymbol{\theta}(\tau)) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} \mathcal{Q}_\tau(\varepsilon_{t+h}(\tau) - g_\tau(\mathbf{z}_t;\boldsymbol{\theta}(\tau))), \end{aligned} \tag{9}$$

where $\mathcal{Q}_\tau(\cdot)$ is defined as:

$$\mathcal{Q}_\tau(u_{t+h}(\tau)) = \begin{cases} \tau u_{t+h}(\tau), & \text{if } u_{t+h}(\tau) \geq 0 \\ (\tau-1) u_{t+h}(\tau), & \text{if } u_{t+h}(\tau) < 0 \end{cases} \tag{10}$$

and $u_{t+h}(\tau) = \varepsilon_{t+h}(\tau) - g_\tau(\mathbf{z}_t;\boldsymbol{\theta}(\tau))$. The forecast for the τth conditional quantile in the case of the augmented AR models is $\hat{q}_{t+h|t}(\tau) = \hat{\beta}_0(\tau) + \sum_{i=1}^{p_\tau} \hat{\beta}_i(\tau) y_{t-i+1} + \hat{g}_\tau\left(\mathbf{z}_t;\hat{\boldsymbol{\theta}}(\tau)\right)$.

We consider four general model specifications for the conditional mean and quantiles that vary based on the choice of flexible function, g and the type of inputs, $\mathbf{z}$.

2.2.1. Predictor-augmented AR

In this specification the models incorporate information from the full set of predictors. The predictor-augmented (PA) AR and QAR models can respectively be rewritten as:

$$y_{t+h|t} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i+1} + g(\mathbf{x}_t;\boldsymbol{\theta}) + \varepsilon_{t+h}, \tag{11}$$

$$\begin{aligned} q_{t+h|t}(\tau) = \beta_0(\tau) + \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1} + g_\tau(\mathbf{x}_t;\boldsymbol{\theta}(\tau)) \\ + \varepsilon_{t+h}(\tau), \end{aligned} \tag{12}$$

where $\mathbf{x} \equiv \mathbf{z}$ is a large panel of macroeconomic variables and the function g is approximated using several fitting methods. We denote and outline below the five predictor-augmented AR approaches as PCA-PA, LASSO-PA, GB-PA, RF-PA and NN-PA.

In the first approach, we augment the autoregressive model with a linear combination of principal components extracted from $\mathbf{x}_t$ (PCA) in order to forecast y_t. These factor-augmented autoregressions have been proposed, among others, by Stock and Watson (2002a, 2002b) and Forni et al. (2000) for the conditional mean, while Manzan (2015) explores the ability of these models to forecast the conditional quantiles. The advantage of this approach is that it reduces the dimensionality of the initial predictor set by concentrating the informational content of N macroeconomic indicators in a small number $K \ll N$ principal components. The number of factors to be included in the model is selected using BIC, while the maximum number of factors is set to $K = 5$.

Another popular approach to reduce the dimensionality of a large panel of macroeconomic variables is through variable selection using shrinkage methods. In this study we employ the lasso

(Tibshirani, 1996), which adds a penalty term to the objective function based on the l_1 norm of the model parameters. This way, the parameters are shrunk towards zero and depending on the strength of regularization, they may be set to zero, thus performing variable selection. The penalized linear model in high-dimensional settings has been extended to the estimation of conditional quantiles by Belloni and Chernozhukov (2011) and by Yi and Huang (2017). Lasso has been employed by De Mol et al. (2008) and Medeiros et al. (2021) along with other methods for the conditional mean case, while Manzan (2015) considers it to forecast conditional quantiles.

The methods described above assume a linear relationship between the variable of interest and the macroeconomic variables. We also consider ensembles of regression trees and artificial neural networks, which connect y_t to the predictor set in a non-linear way. The first ensemble approach is gradient boosting (GB), proposed by Friedman (2001) for the regression framework, which combines a large number of shallow trees, to form an ensemble with greater stability than a single more complex regression tree. The trees are sequentially combined by refitting shallow trees to the residuals from previous iterations. This process is repeated until a certain number of iterations is reached. The objective function to be minimized is the least squares or quantile loss. The second ensemble method we consider is based on bootstrap aggregating or bagging (Breiman, 1996), which combines forecasts from a large number of trees estimated for different bootstrap subsamples to obtain a single low-variance model. The bagging-based approach we use is random forests (RF), proposed by Breiman (2001), which aims to reduce the variance of the forecast relative to bagging by combining a large set of decorrelated trees based only on a randomly drawn subset of predictors. Random forests have also been extended to a quantile regression setting by Meinshausen (2006).

Finally, we consider is artificial neural networks (NN) and specifically feed-forward neural networks. These models are comprised of a number of layers with multiple nodes in each layer. They consist of an input layer of the predictors, one or more hidden layers, with nodes that transform the predictors using non-linear activation functions and an output layer that allows a final transformation of the outcome of the hidden layers to form a prediction. We consider a shallow neural network, which minimizes the least squares loss in the case of the conditional mean and the quantile loss in the case of the conditional quantiles.

2.2.2. Targeted predictor-augmented AR

This specification allows us to exploit the ability of the machine learning models to uncover important variables and non-linear interactions to forecast y_t. The targeted predictor-augmented (TP) autoregressions re-estimate the least squares or quantile regression model using the most important variables as selected by the lasso, the two ensemble methods or the neural network. The models are given by:

$$y_{t+h|t} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i+1} + \sum_{j=1}^{n} \theta_j x_{t,j} + \varepsilon_{t+h}, \tag{13}$$

$$q_{t+h|t}(\tau) = \beta_0(\tau) + \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1} + \sum_{j=1}^{n_\tau} \theta_j(\tau) x_{t,j} + \varepsilon_{t+h}(\tau), \tag{14}$$

where $n \ll N$ denotes the number of variables selected by one of the four machine learning approaches. In the case of the conditional quantiles, n and the variables $\mathbf{x}_t$ vary according to τ. The top ten most influential variables for each model are considered and the rankings are constructed based on the absolute change in the mean squared error (MSE) or quantile loss for setting one of the predictors to zero over the validation sample. We refer the four targeted predictor-augmented AR approaches as LASSO-TP, GB-TP, RF-TP and NN-TP.

2.2.3. Targeted factor-augmented AR

We follow Bai and Ng (2008), and construct targeted factors (TF), namely factors based on a subset of the macroeconomic variables selected with the specific target of forecasting y_t. Manzan (2015) extends this approach to a quantile regression setting, where the variables used to construct factors differ for each quantile. Specifically, this approach involves building factors by extracting information from the subset of macroeconomic variables selected by the machine learning models. The conditional mean and quantile models are given respectively by:

$$y_{t+h|t} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i+1} + \sum_{j=1}^{K} \theta_j f_{t,j} + \varepsilon_{t+h}, \tag{15}$$

$$q_{t+h|t}(\tau) = \beta_0(\tau) + \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1} + \sum_{j=1}^{K_\tau} \theta_j(\tau) f_{t,j} + \varepsilon_{t+h}(\tau), \tag{16}$$

where f_j denotes the jth factor derived using PCA and K the number of factors selected *via* BIC. The targeted factors and the number of factors vary based on the conditional quantile. We define the subset of variables as those that when set to zero the validation MSE or quantile loss will increase.

In the case that no variables are selected, the model is reduced to an AR specification. The four targeted factor-augmented AR approaches are denoted as LASSO-TF, GB-TF, RF-TF and NN-TF.

2.2.4. Forecast combination-augmented AR

In the fourth specification we augment the baseline AR model by combining forecasts generated from simple models. Forecast combinations (FC), originally proposed by Bates and Granger (1969), may be preferred over individual models that combine information, since they reduce model instability and parameter uncertainty. The models for the conditional mean and quantiles are:

$$y_{t+h|t} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i+1} + g\left(\hat{\mathbf{y}}_{t+h|t}; \boldsymbol{\theta}\right) + \varepsilon_{t+h}, \tag{17}$$

$$q_{t+h|t}(\tau) = \beta_0(\tau) + \sum_{i=1}^{p_\tau} \beta_i(\tau) y_{t-i+1} + g_\tau\left(\hat{\mathbf{q}}_{t+h|t}(\tau); \boldsymbol{\theta}(\tau)\right) + \varepsilon_{t+h}(\tau), \tag{18}$$

where $\hat{\mathbf{y}}_{t+h|t}$ and $\hat{\mathbf{q}}_{t+h|t}(\tau)$ are N-vectors of individual forecasts for the mean and quantiles respectively, derived from bivariate prediction models using least squares or quantile regression. The forecast combinations vary depending on estimation method of the combining weights, $\boldsymbol{\theta}$. We consider simple forecast combinations such as the mean (MN-FC) and median (MD-FC) forecast, and approaches where the combining weights are computed based on the historical forecasting performance of the individual models over an initial holdout period, such as the rank (RANK-FC) and cluster (CL-FC) combinations by Aiolfi and Timmermann (2006) or the discounted forecast error (DFE-FC) proposed by Stock and Watson (2004). Finally, we employ PCA, the lasso, ensemble methods and neural networks as another approach to combine individual forecasts. These forecast combination methods are referred to as PCA-FC, LASSO-FC, GB-FC, RF-FC and NN-FC.

Details on the machine learning and forecast combination methods used to approximate g and g_τ throughout the four general specifications can be found in Supplementary Appendix B.

2.3. Quantile combinations

We also examine whether combining information from different quantiles can yield improved estimates for the conditional mean. We follow Lima and Meng (2017) and Meligkotsidou et al. (2019) and forecast the conditional mean of the macroeconomic variables as the weighted average of a set of quantiles. For a given model specification, the point forecast is derived from the estimates of the conditional quantiles in the following way:

$$\hat{y}_{t+h|t} = \sum_{\tau \in S} w_\tau \hat{q}_{t+h|t}(\tau), \ \sum_{\tau \in S} w_\tau = 1, \tag{19}$$

where w_τ denotes the weight associated with the τth quantile forecast and S is the set of quantiles being aggregated. We consider three different quantile combinations with time-invariant weights. Specifically, we consider three-quantile estimators similar to Tukey (1977) and Gastwirth (1966), given respectively by the following formulae:

$$\text{QC1: } \hat{y}_{t+h|t} = 0.25\hat{q}_{t+h|t}(0.2) + 0.50\hat{q}_{t+h|t}(0.5) + 0.25\hat{q}_{t+h|t}(0.8), \tag{20}$$

$$\text{QC2: } \hat{y}_{t+h|t} = 0.30\hat{q}_{t+h|t}(0.3) + 0.40\hat{q}_{t+h|t}(0.5) + 0.30\hat{q}_{t+h|t}(0.7). \tag{21}$$

In order to attach more weight on extreme events, we also employ the five-quantile estimator, suggested by Judge et al. (1988):

$$\text{QC3: } \hat{y}_{t+h|t} = 0.05\hat{q}_{t+h|t}(0.1) + 0.25\hat{q}_{t+h|t}(0.2) + 0.40\hat{q}_{t+h|t}(0.5) + 0.25\hat{q}_{t+h|t}(0.8) + 0.05\hat{q}_{t+h|t}(0.9). \tag{22}$$

2.4. Alternative models

We consider several alternative forecasting models that are employed in the literature of macroeconomic forecasting in addition to the AR(1) and AR(12) models described in Section 2.1.

The first alternative model is the random walk (RW), where the point forecasts are computed as $\hat{y}_{t+h|t} = y_t$, while the quantile forecasts are derived from the quantiles of a normal distribution with a mean and standard deviation estimated based on the random walk model.

Another popular approach in the forecasting literature is the stochastic volatility (SV) model, where the variance specification is stochastic and time varying. We consider an AR(p) model augmented by stochastic volatility and estimated by Markov chain Monte Carlo (MCMC) methods. The model is given by:

$$\begin{aligned} y_{t+h|t} &= \beta_0 + \beta_i y_t + \sigma_{t+h}\varepsilon_{t+h} \\ \log\left(\sigma^2_{t+h}\right) &= \mu + \varphi \log\left(\sigma^2_t\right) + \eta_{t+h}, \end{aligned} \tag{23}$$

where ε_t and η_t are the error terms that are assumed to be independent of each other and distributed as $N(0,1)$ and $N\left(0,\sigma^2_\eta\right)$ respectively. We report the results for the basic SV model with one lag and twelve lags (ARSV1 and ARSV12). A restriction of the "vanilla" stochastic volatility model is the assumption that the error distribution is symmetric. We allow for asymmetries in the error distribution by extending the basic SV model to incorporate the leverage effect, which is introduced to Equation (23) by allowing ε_{t+h} and η_{t+h} to be correlated with:

$$\begin{pmatrix} \varepsilon_{t+h} \\ \eta_{t+h} \end{pmatrix} = N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}, \tag{24}$$

where ρ is the corelation that captures the leverage effect. We estimate asymmetric ARSV with one and twelve lags (ARSVLEV1 and ARSVLEV12).

The alternative models considered thus far rely solely on lags of the response variable. While own-lags are important in forecasting macroeconomic variables, cross-lags have also been shown to enhance predictive performance (Gruber & Kastner, 2022). Therefore, we also employ Bayesian vector autoregressions (VARs), which are widely adopted in forecasting macroeconomic time series (see e.g., Bańbura et al., 2010; Clark, 2011; Koop, 2013; Korobilis, 2013; Huber & Feldkircher, 2019; Pfarrhofer, 2024).

A VAR model of order p, VAR(p), can be given by the following equation[1]:

$$\mathbf{y}_{t+h|t} = \sum_{i=1}^{p} \mathbf{B}_i \mathbf{y}_{t-i+1} + \boldsymbol{\epsilon}_{t+h}, \boldsymbol{\epsilon}_{t+h} \sim N\left(0, \boldsymbol{\Sigma}_{t+h|t}\right), \tag{25}$$

where $\mathbf{y}_t$ is an $m \times 1$ vector of endogenous variables, $\mathbf{B}_i$, for $i = 1, ..., p$, the $m \times m$ coefficient matrix, $\boldsymbol{\epsilon}_t$ is an $m \times 1$ vector of exogenous shocks and $\boldsymbol{\Sigma}_t$ corresponds to the $m \times m$ covariance matrix. For the VAR coefficients we use the Horseshoe (HS) shrinkage prior (Carvalho et al., 2010), which leads to sparse models and has the advantage that no tuning parameters need to be specified. We consider two types of VAR models that differ in the way the covariance matrix is decomposed. In the first case we assume that the errors follow a Cholesky stochastic volatility structure (see Cogley & Sargent, 2005; Feldkircher et al., 2024), where the covariance matrix can be decomposed to:

$$\boldsymbol{\Sigma}_t = \mathbf{U}^{'} \mathbf{H}_t \mathbf{U}, \tag{26}$$

where $\mathbf{U}$ is an $m \times m$ upper triangular matrix with ones on the diagonal, whose off-diagonal elements are distributed based on the HS prior and $\mathbf{H}_t$ is an $m \times m$ diagonal matrix, whose elements are assumed to follow independent, univariate AR(1) processes. In the second case we assume that the covariance matrix has a factor stochastic volatility structure (see Kastner & Huber, 2020). The covariance matrix is given by:

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}^{'} \mathbf{S}_t \boldsymbol{\Lambda} + \mathbf{Z}_t, \tag{27}$$

where $\boldsymbol{\Lambda}$ is an $m \times k$ matrix of factor loadings, $\mathbf{S}_t$ is a $k \times k$ diagonal matrix containing the variances of the k latent factors, while $\mathbf{Z}_t$ is an $m \times m$ diagonal matrix that contains the idiosyncratic variances. The logarithms of the elements in the factor and idiosyncratic components of the covariance matrix follow independent AR(1) processes. Further details on the estimation of the elements of $\boldsymbol{\Sigma}_t$ and prior specifications can be found in Kastner et al. (2017) and Kastner (2019). Here we consider the case of a single latent factor ($k = 1$). We estimate VAR models of lag order $p = 1, 2$ for both Cholesky SV (VARCSV1 and VARCSV2) and factor SV (VARFSV1 and VARFSV2).

3. Data and sample splitting

Our dataset consists of the FRED-MD database by McCracken and Ng (2016), which is a large monthly macroeconomic dataset ideally suited for empirical analysis in high-dimensional settings. We obtain the dataset from Michael McCraken's webpage.[2] The number of variables in the FRED-MD database is 127. Details on the variables included in the FRED-MD database can be found in Tables A1–A8 in Supplementary Appendix A. We forecast the h-month growth ($h = 1$) of Industrial Production Index (INDPRO), the Consumer Price Index for all urban consumers (CPIAUCSL) and Non-farm Payroll Employment (PAYEMS). The first twelve lags of each respective response variable are accounted as candidates in the autoregressive part of the model, leaving $N = 126$ candidate predictors for the part of the model that combines information or forecasts. The three response variables are transformed using log differences, while for the remaining variables we use the same transformations as McCracken and Ng (2016). The full sample period is from December 1964 to December 2019, for a total of $T = 661$ monthly observations.

3.1. Sample splitting and hyperparameter tuning

The forecasts from the four model specifications described in Section 2.2 are generated using a rolling window scheme. The first $t_0 = 60$ observations of the rolling window are only used in the estimation of the simple forecasts based on individual predictors used in forecast combinations, while the remaining $T_0 = 240$ observations are used to estimate the models that combine either information or simple forecasts. The initial rolling window is from December 1964 to November 1989 (or 300 monthly observations), which leaves a total of $T_{OOS} = 361$ observations, from December 1989 to December 2019, that can be used for forecast evaluation.

To choose the hyperparameters of the machine learning models, we adopt the validation sample approach. In each iteration, the rolling window, T_0, is split into two disjoint periods, the training subsample, consisting of 90% of the observations, with the remaining observations belonging to the validation subsample. The predictors are standardized for all methods using the mean and standard deviation calculated from observations from the training subsample. In the training subsample the model is estimated for several sets of hyperparameters. The second subsample is used to select the optimal set of tuning parameters, by constructing forecasts, using the model estimates from the training sample for the respective hyperparameter set, for the observations in the validation sample. The optimal set of hyperparameters is chosen to minimize the mean squared error or the quantile loss function over the validation subsample, for the mean and quantile forecasts respectively. Once the optimal set of hyperparameters is chosen, the model is refitted using all data from the rolling window, T_0, and the estimates of the model parameters are kept to construct the forecasts.

4. Empirical results

In this section, we first examine the forecasting performance of the proposed models that either forecast the conditional mean directly or indirectly by combining information across different quantiles. We then proceed to evaluate the out-of-sample performance of individual quantile forecasts and finally the ability of the models to approximate different parts of the distribution.

4.1. Point forecast evaluation

First, we examine the accuracy of the point forecasts of the three macroeconomic variables, which is evaluated based on the out-of-sample MSE computed as

$$\mathrm{MSE}_i = \frac{1}{T_{OOS}} \sum_{t=1}^{T_{OOS}} \hat{e}_{i,t}^2, \tag{28}$$

where $\hat{e}_{i,t} = y_t - \hat{y}_{i,t}$ and $\hat{y}_{i,t}$ is the forecast of the macroeconomic variable of model i. Specifically, we follow Medeiros et al. (2021) and report the MSE ratio of model i with respect to the random walk (RW) benchmark, with a smaller ratio indicating greater outperformance from the benchmark model. To evaluate the statistical significance of our conditional mean forecasts relative to the benchmark, we employ the Diebold and Mariano (1995) (DM) test for predictive accuracy. Table 1 reports the MSE ratio of the alternative model with respect to the random walk benchmark and its significance through the p-values of the DM test, for industrial production, inflation and employment.

Overall, the results indicate that the majority of the models outperform the RW benchmark, as revealed by the MSE ratios of the target variables. In addition, the majority of the models exhibit statistically significant outperformance over the RW model at the 1% level, with just a few exceptions at the 5% and 10% levels, while only three models, in the case of employment, fail to significantly outperform the benchmark. Furthermore, the results show that models that combine information or forecasts generate MSE ratios equivalent to or lower than those of univariate or multivariate stochastic volatility models. More importantly, combining information across quantiles considerably improves the out-of-sample performance of the models. This improved performance is more prominent in the four specifications that incorporate a large number of predictors to the models.

The results for industrial production show that direct forecasts of the conditional mean (LS) would lead to outperformance of the RW benchmark, with MSE ratios between 0.563 and 0.787. Approximately 40% of the models (9 models) conditioned on economic variables have MSE ratios lower than the AR(12) model, while 22% of the conditional forecasts (5 models) outperform VAR-SV2, which yields the lowest MSE ratio among the ten autoregressive models. The models with the lowest MSE ratio in each of the four specifications are the factor-augmented AR (PCA-PA), the targeted factor- and targeted predictor-augmented AR with predictors selected using the random forests algorithm (RF-TP and RF-TF) and forecast combinations based on PCA (PCA-FC). Combining information across quantiles has a positive effect on the predictive accuracy for the majority of the models, particularly those incorporating a large number of predictors. The percentage of models that outperform the AR(12) model in each quantile combination specification increases to over 90% and the percentage of the quantile combinations that outperform VAR-SV2 varies between 48% to 57%. Through quantile combinations the MSE ratio is reduced to the range of 0.545 to 0.675, with targeted predictor augmented regression based on the neural network and factor augmented regressions based on the two ensemble methods having the best performance.

For inflation, we observe that methods minimizing the least squares loss outperform the benchmark with MSE ratios ranging from 0.693 to 1.012. Over half of the models conditioned on a large number of variables yield MSE ratios lower than the AR(12) baseline model, which is also the best performing autoregressive model. The models with the lowest MSE ratios for the three specifications that combine information are based on gradient boosting (GB-PA, GB-TP and GB-TF), while for the forecast combinations the neural network (NN-FC) has the lowest MSE ratio. Combining information across quantiles further reduces the MSE ratios to a range of 0.674 to

Table 1. Point forecast evaluation for the Industrial Production Index (INDPRO), the Consumer Price Index (CPIAUCSL) and Employment (PAYEMS).

	INDPRO				CPIAUCSL				PAYEMS			
	LS	QC1	QC2	QC3	LS	QC1	QC2	QC3	LS	QC1	QC2	QC3
AR1	0.643***	0.643***	0.632***	0.643***	0.762**	0.760**	0.749**	0.763**	0.953	0.915**	0.904**	0.920*
AR12	0.601***	0.583***	0.581***	0.583***	0.741***	0.732***	0.718***	0.732***	0.752***	0.723***	0.731***	0.727***
ARSV1	0.647***	0.647***	0.658***	0.647***	1.012	0.819***	0.829***	0.819***	0.912***	0.912***	0.937**	0.911***
ARSV12	0.583***	0.582***	0.593***	0.582***	0.808*	0.807*	0.807*	0.808*	0.741***	0.739***	0.754***	0.740***
ARSVLEV1	0.645***	0.645***	0.659***	0.645***	0.819***	0.820***	0.838***	0.820***	0.906***	0.905***	0.944*	0.904***
ARSVLEV12	0.581***	0.581***	0.592***	0.581***	0.805*	0.804*	0.806*	0.804*	0.734***	0.734***	0.750***	0.734***
VARCSV1	0.597***	0.595***	0.593***	0.595***	0.768**	0.769**	0.750**	0.769**	0.924	0.921*	0.898**	0.923*
VARCSV2	0.580***	0.580***	0.576***	0.580***	0.757**	0.760**	0.744**	0.759**	0.782***	0.772***	0.755***	0.771***
VARFSV1	0.600***	0.598***	0.597***	0.599***	0.773**	0.773**	0.752**	0.774**	0.940	0.939	0.910*	0.939
VARFSV2	0.575***	0.576***	0.574***	0.576***	0.761**	0.763**	0.744**	0.763**	0.762***	0.765***	0.753***	0.765***
A. Predictor-Augmented AR (PA)												
PCA-PA	0.568***	0.552***	0.552***	0.553***	0.769**	0.724***	0.710***	0.724***	0.713***	0.698***	0.696***	0.699***
LASSO-PA	0.673***	0.657***	0.675***	0.658***	0.721***	0.748***	0.745***	0.736***	0.793***	0.779***	0.776***	0.777***
GB-PA	0.569***	0.559***	0.548***	0.561***	0.706***	0.697***	0.690***	0.696***	0.738***	0.721***	0.729***	0.718***
RF-PA	0.614***	0.595***	0.587***	0.598***	0.725***	0.696***	0.687***	0.695***	0.718***	0.692***	0.699***	0.693***
NN-PA	0.572***	0.587***	0.589***	0.589***	0.720***	0.771**	0.775**	0.782**	0.735***	0.958	0.828***	1.107
B. Targeted Predictor-Augmented AR (TP)												
LASSO-TP	0.601***	0.560***	0.587***	0.557***	0.767***	0.695***	0.692***	0.692***	0.752***	0.735***	0.738***	0.733***
GB-TP	0.630***	0.559***	0.565***	0.559***	0.722***	0.677***	0.674***	0.674***	0.852**	0.717***	0.722***	0.719***
RF-TP	0.572***	0.564***	0.560***	0.564***	0.776***	0.718***	0.703***	0.717***	0.799***	0.714***	0.706***	0.713***
NN-TP	0.603***	0.545***	0.549***	0.547***	0.778**	0.737**	0.722***	0.735***	0.823**	0.724***	0.731***	0.726***
C. Targeted Factor-Augmented AR (TF)												
LASSO-TF	0.601***	0.563***	0.576***	0.561***	0.768***	0.695***	0.689***	0.694***	0.752***	0.725***	0.743***	0.718***
GB-TF	0.622***	0.553***	0.550***	0.553***	0.753***	0.695***	0.680***	0.691***	0.716***	0.691***	0.696***	0.691***
RF-TF	0.563***	0.548***	0.550***	0.547***	0.757**	0.714***	0.705***	0.717***	0.739***	0.688***	0.701***	0.687***
NN-TF	0.576***	0.559***	0.563***	0.559***	0.752**	0.729***	0.718***	0.729***	0.734***	0.704***	0.712***	0.703***
D. Forecast Combination-Augmented AR (FC)												
MN-FC	0.606***	0.587***	0.582***	0.587***	0.795**	0.774**	0.743***	0.777**	0.827**	0.795***	0.766***	0.800***
MD-FC	0.602***	0.584***	0.581***	0.584***	0.747**	0.732***	0.717***	0.733***	0.750***	0.728***	0.726***	0.732***
RANK-FC	0.596***	0.577***	0.575***	0.578***	0.732***	0.729***	0.713***	0.729***	0.746***	0.715***	0.719***	0.720***
CL-FC	0.595***	0.580***	0.578***	0.580***	0.734***	0.725***	0.712***	0.725***	0.738***	0.706***	0.714***	0.709***
DFE-FC	0.603***	0.584***	0.580***	0.584***	0.761**	0.756**	0.731***	0.758**	0.737***	0.729***	0.717***	0.734***
PCA-FC	0.587***	0.563***	0.568***	0.561***	0.739***	0.734***	0.722***	0.732***	0.721***	0.711***	0.716***	0.715***
LASSO-FC	0.601***	0.560***	0.565***	0.558***	0.740***	0.725***	0.712***	0.725***	0.752***	0.710***	0.717***	0.711***
GB-FC	0.605***	0.571***	0.574***	0.570***	0.731***	0.718***	0.709***	0.717***	0.764***	0.720***	0.728***	0.720***
RF-FC	0.618***	0.583***	0.578***	0.585***	0.696***	0.713***	0.696***	0.714***	0.817***	0.711***	0.723***	0.716***
NN-FC	0.787*	0.588***	0.595***	0.587***	0.693***	0.764**	0.751**	0.761**	0.835**	0.735***	0.757***	0.734***

This table reports the ratio of the mean squared error (MSE) of the alternative prediction model relative to the benchmark. The benchmark is the random walk (RW) model. A lower ratio indicates greater outperformance from the benchmark. The statistical significance of the forecasts is assessed using the Diebold and Mariano (1995) test. The significant outperformance of the alternative models from the benchmark is denoted by: *, **, and *** for significance at the 10%, 5%, and 1% level, respectively. The out-of-sample period is from December 1989 to December 2019. The results are reported for point forecasts from models that directly model the mean (LS) or combine information from different quantiles, using Tukey (1977) trimean (QC1), Gastwirth (1966) three-quantile estimator (QC2) and Judge et al. (1988) five-quantile estimator (QC3).

0.838, with approximately 80% of the models outperforming the AR(12) baseline across the three QC specifications. Models that combine information benefit more from quantile combinations, with the PA, TP and TF specifications based on gradient boosting consistently offering improved performance.

The results for employment reveal that directly forecasting the conditional mean can lead to good out-of-sample performance relative to the RW, with MSE ratios that are from 0.713 to 0.952. Furthermore, over 50% of the models outperform the AR(12) baseline, while 17% of the models conditioned on a large number of variables outperform the AR(12) augmented by stochastic volatility with leverage (ARSVLEV12), which has the lowest ratio among the ten alternative autoregressive models. Factor-augmented AR (PCA-PA), targeted predictor based on the lasso (LASSO-TP), targeted factor-augmented model based on gradient boosting (GB-TF) and forecast combinations based on PCA (PCA-FC) are the models with the lowest MSE ratios in each of the four specifications. Combining information across quantiles results to MSE ratios in the range of 0.696 to 1.107. The majority of the models benefit from quantile combinations, with over 80% of the models having a lower MSE ratio than the AR(12) and over 70% outperforming the ARSVLEV12 model. Predictor-augmented models based on the lasso and random forests, as well as targeted factor-augmented models based on the two ensemble methods have the lowest MSE ratios across all QC specifications. In contrast, the forecast based on NN-PA for QC3 is the only instance that the RW would be preferable to an alternative model.

4.2. Quantile forecast evaluation

We next examine the accuracy of individual quantile forecasts for each of the nine values of $\tau \in (0, 1)$. Gneiting and Raftery (2007) and Gneiting and Ranjan (2011) propose that the same loss function should be employed in both model estimation and forecast evaluation. Therefore, following Manzan and Zerom (2015) and Manzan (2015), we evaluate quantile forecasts using the quantile score (QS) function. The QS function focuses on a specific quantile τ and provides a local evaluation of the forecasts. The QS for the τth quantile forecast of model i is given by:

$$\mathrm{QS}_{i,\tau}(\hat{e}_{i,t}(\tau)) = \begin{cases} \tau\hat{e}_{i,t}(\tau) & \text{if } \hat{e}_{i,t}(\tau) \geq 0 \\ (\tau-1)\hat{e}_{i,t}(\tau), & \text{if } \hat{e}_{i,t}(\tau) < 0 \end{cases}, \tag{29}$$

where $\hat{e}_{i,t}(\tau) = y_t - \hat{q}_{i,t}(\tau)$ and $\hat{q}_{i,t}(\tau)$ is the forecast of the target variable of model i for the τth quantile. This scoring rule is negatively orientated coinciding with the notion of a loss function, so that when comparing two models, we prefer the one with the lowest score. To evaluate the hypothesis of equal predictive accuracy of the quantile forecasts, we follow Amisano and Giacomini (2007) and Giacomini and White (2006) and compare the quantile score of model i to that of the benchmark model using the test statistic:

$$t = \frac{\overline{\mathrm{QS}}_{i,\tau} - \overline{\mathrm{QS}}_{0,\tau}}{\hat{\sigma}}, \tag{30}$$

where $\overline{\mathrm{QS}}_{i,\tau}$ and $\overline{\mathrm{QS}}_{0,\tau}$ denote the averages over the out-of-sample period of the quantile scores for a given quantile τ for model i and the benchmark model respectively, while $\hat{\sigma}$ is the standard error estimator of the quantile score difference. Assuming suitable regularity conditions, the statistic t is asymptotically standard normal under the null hypothesis of vanishing expected score differentials. In the case of rejection, model i is preferred over the benchmark if t is negative, and the benchmark model is preferred if t is positive. We follow studies such as Manzan (2015), who consider an autoregressive model with stochastic volatility as the benchmark. Specifically, we set the autoregressive model augmented by stochastic volatility with leverage (ARSVLEV1) as the benchmark. In Sections 4.2.1–4.2.3, we describe the results for the quantile forecast accuracy for industrial production, inflation and employment. Tables 2–4 report the test statistics of the quantile score test for the null hypothesis of equal quantile forecast accuracy of a model relative to the stochastic volatility benchmark model for the three target economic variables.

4.2.1. Quantile forecast evaluation: Industrial Production Index

Starting with the results for industrial production, presented in Table 2, we observe that the majority of the models are more accurate than the benchmark. The forecast accuracy, based on the statistical significance of the QS tests, is more pronounced in the left tail ($\tau \in \{0.1, 0.2, 0.3\}$) and the center ($\tau \in \{0.4, 0.5, 0.6\}$) of the distribution. However, the performance of the models is weaker for the rightmost part of the distribution, where most models fail to significantly outperform the benchmark for $\tau = \{0.8, 0.9\}$. The autoregressive models that consistently outperform the benchmark across the lower and middle part of the distribution, are the AR(12) baseline and the AR(12) models augmented by stochastic volatility (ARSV12 and ARSVLEV12), as well as the VAR(2) with factor stochastic volatility (VARFSV2), with AR(12) yielding the lowest QS test statistic for $\tau \in [0.1, 0.5]$. The rankings of models that augment the QAR baseline by combining information or forecasts vary depending on the

Table 2. Quantile forecast evaluation for the Industrial Production Index (INDPRO).

	q(0.1)	q(0.2)	q(0.3)	q(0.4)	q(0.5)	q(0.6)	q(0.7)	q(0.8)	q(0.9)
RW	4.350	5.563	5.386	5.159	5.155	5.407	5.540	5.603	4.397
AR1	0.396	−0.199	−0.641	−0.439	0.007	−0.626	−0.938	−1.215	−0.753
AR12	−2.621	−2.632	−3.172	−2.785	−2.521	−2.074	−1.482	−1.457	−0.255
ARSV1	−1.546	−0.628	0.183	−0.317	1.461	2.239	1.949	2.583	2.824
ARSV12	−1.943	−1.905	−2.477	−2.382	−2.346	−2.168	−1.404	−0.529	0.325
ARSVLEV12	−2.591	−2.154	−2.433	−2.450	−2.418	−2.354	−1.880	−1.245	−0.765
VARCSV1	−1.788	−1.289	−0.762	−1.017	−1.723	−2.062	−1.617	−1.336	−0.106
VARCSV2	−2.103	−1.793	−1.398	−1.490	−2.173	−2.708	−2.376	−1.534	−0.156
VARFSV1	−2.016	−1.564	−1.076	−1.058	−1.631	−1.851	−1.427	−0.851	0.185
VARFSV2	−2.160	−2.417	−2.060	−1.858	−2.391	−2.714	−2.343	−1.311	0.068
A. Predictor-Augmented AR (PA)									
PCA-PA	−2.965	−3.015	−3.412	−3.352	−3.162	−2.018	−1.140	−0.962	0.064
LASSO-PA	0.603	−0.276	−0.099	−0.254	−0.223	0.143	1.061	1.531	1.867
GB-PA	−1.871	−1.902	−2.740	−3.319	−3.458	−2.656	−1.855	−1.077	0.299
RF-PA	−2.684	−3.240	−3.088	−2.267	−2.836	−2.303	−1.462	−1.622	−0.694
NN-PA	−1.025	−1.298	−2.632	−2.643	−2.292	−2.002	−1.039	−0.579	0.843
B. Targeted Predictor-Augmented AR (TP)									
LASSO-TP	−2.315	−3.575	−2.555	−2.379	−2.196	−1.441	−0.767	0.115	−0.032
GB-TP	−1.799	−2.152	−3.071	−2.800	−2.990	−2.313	−1.699	−0.335	0.415
RF-TP	−1.997	−2.163	−2.444	−2.646	−2.309	−2.338	−1.507	−0.567	1.713
NN-TP	−2.417	−3.329	−3.976	−2.852	−3.374	−3.338	−2.186	−1.244	1.160
C. Targeted Factor-Augmented AR (TF)									
LASSO-TF	−2.523	−3.067	−2.915	−2.443	−2.090	−1.425	−1.061	−0.204	0.507
GB-TF	−2.118	−2.692	−3.270	−3.211	−3.125	−2.585	−1.467	−1.201	0.131
RF-TF	−2.667	−3.555	−3.793	−3.502	−3.009	−2.291	−0.891	−0.512	0.613
NN-TF	−2.937	−3.056	−3.644	−3.082	−2.956	−2.241	−1.378	−1.155	−0.271
D. Forecast Combination-Augmented AR (FC)									
MN-FC	−2.191	−2.327	−3.065	−2.661	−2.040	−1.260	−0.197	0.363	1.511
MD-FC	−2.625	−2.568	−3.106	−2.758	−2.414	−1.998	−1.339	−1.208	−0.026
RANK-FC	−2.585	−2.686	−3.334	−3.099	−2.640	−2.031	−1.415	−1.245	0.122
CL-FC	−2.638	−2.681	−3.250	−2.881	−2.632	−2.175	−1.671	−1.631	−0.276
DFE-FC	−2.422	−2.407	−3.099	−2.689	−2.155	−1.492	−0.581	−0.172	0.756
PCA-FC	−2.578	−2.831	−3.438	−2.916	−2.582	−1.782	−1.610	−2.021	−0.118
LASSO-FC	−2.555	−3.029	−3.531	−3.143	−2.829	−2.636	−2.552	−2.099	−0.199
GB-FC	−1.854	−1.811	−2.589	−2.571	−2.835	−1.955	−1.047	−1.550	0.142
RF-FC	−1.473	−1.959	−3.131	−3.097	−2.660	−2.246	−2.177	−1.111	−0.301
NN-FC	−2.489	−2.179	−2.273	−2.320	−2.463	−2.112	−0.840	−0.749	0.332

This table reports the quantile score *t*-statistics for the null hypothesis of equal predictive ability of the alternative prediction model relative to the benchmark. The benchmark is the autoregressive model augmented by stochastic volatility with leverage (ARSVLEV1). A lower value indicates greater outperformance from the benchmark. Values less than −1.645 indicate that the alternative model outperforms the benchmark at the 5% level. The out-of-sample period is from December 1989 to December 2019.

quantile under examination. Targeted factor-augmented AR models based on random forests and the neural network (RF-TF and NN-TF) are consistent in outperforming the benchmark and improving upon the QAR baseline model compared to other specifications for $\tau \in [0.1, 0.6]$.

4.2.2. Quantile forecast evaluation: Consumer Price Index

The results for inflation (Table 3) indicate that most models significantly outperform the benchmark, especially in the left tail and middle part of the distribution. For $\tau = 0.9$, the AR models augmented by stochastic volatility tend to outperform those that combine information or forecasts from a large set of predictors, while for $\tau \in \{0.2, 0.3\}$, VAR models show improved out-of-sample performance compared to univariate AR models. The VARFSV2 model outperforms the remaining autoregressive models for the lower quantiles, while the AR(12) performs better for the middle quantiles. Specifications that augment the QAR model by combining information generate statistically significant outperformance over the benchmark for $\tau = 0.9$, compared to models that combine forecasts, where only the lasso yields significant outperformance. For $\tau \in [0.2, 0.7]$ the majority of the models that incorporate a large number of economic variables outperform the benchmark and improve upon the QAR baseline. Overall, specifications where variables are selected by the lasso, or the two ensemble methods demonstrate the best out-of-sample performance for the lower and middle quantiles. In contrast, when $\tau = \{0.8, 0.9\}$ the benchmark becomes increasingly difficult to significantly outperform for all models considered.

4.2.3. Quantile forecast evaluation: Employment

Turning to the results for employment, reported in Table 4, we observe that most models can significantly outperform the benchmark, however, out-of-sample performance diminishes for the upper quantiles. Univariate autoregressive models yield improved predictive performance relative to VAR models, with the AR(12) model augmented by stochastic volatility with leverage (ARSVLEV12) exhibiting strong performance across all quantiles. The majority of the

Table 3. Quantile forecast evaluation for the Consumer Price Index (CPIAUCSL).

	q(0.1)	q(0.2)	q(0.3)	q(0.4)	q(0.5)	q(0.6)	q(0.7)	q(0.8)	q(0.9)
RW	3.205	4.134	3.889	4.177	4.341	3.887	3.354	3.590	3.491
AR1	−0.726	−2.236	−2.660	−2.661	−2.538	−2.474	−2.297	−1.425	0.423
AR12	−1.293	−2.441	−3.022	−2.979	−3.050	−3.748	−3.445	−2.381	−0.367
ARSV1	1.398	−1.659	−1.403	−1.801	1.059	3.582	4.090	2.720	0.952
ARSV12	−2.966	−2.734	−2.009	−1.315	−1.306	−1.523	−1.231	−0.974	−0.547
ARSVLEV12	−2.584	−2.581	−1.862	−1.285	−1.424	−1.660	−1.382	−1.218	−0.808
VARCSV1	−3.570	−3.999	−3.299	−2.423	−2.212	−2.161	−1.774	−1.246	−0.752
VARCSV2	−3.534	−3.854	−3.120	−2.467	−2.421	−2.535	−2.153	−1.694	−1.040
VARFSV1	−3.310	−3.717	−3.117	−2.307	−2.171	−2.237	−1.902	−1.242	−0.557
VARFSV2	−3.375	−4.003	−3.336	−2.599	−2.385	−2.148	−1.790	−1.421	−0.635
A. Predictor-Augmented AR (PA)									
PCA-PA	−1.464	−2.558	−3.118	−3.058	−3.045	−3.729	−3.562	−2.584	−0.624
LASSO-PA	−1.952	−1.687	−1.789	−1.301	−1.504	−1.652	−1.339	−0.682	−0.252
GB-PA	−1.685	−2.784	−3.308	−3.455	−3.761	−4.101	−4.014	−3.125	−0.687
RF-PA	−2.478	−3.960	−3.571	−3.751	−3.995	−4.462	−3.169	−2.170	−0.506
NN-PA	2.968	0.109	−0.414	−1.226	−1.100	−0.193	0.967	1.242	2.493
B. Targeted Predictor-Augmented AR (TP)									
LASSO-TP	−1.343	−2.609	−3.079	−3.340	−3.589	−3.272	−2.863	−1.920	−0.849
GB-TP	−2.581	−4.001	−3.717	−3.713	−3.351	−3.665	−3.791	−2.586	−0.491
RF-TP	−2.211	−2.970	−2.800	−3.113	−3.188	−3.673	−2.768	−1.593	−0.170
NN-TP	−2.363	−2.482	−2.349	−2.753	−3.030	−3.710	−3.005	−1.383	−0.699
C. Targeted Factor-Augmented AR (TF)									
LASSO-TF	−1.672	−3.441	−3.287	−3.463	−3.593	−4.069	−3.705	−2.600	−0.732
GB-TF	−2.909	−3.718	−3.701	−3.655	−3.636	−4.366	−3.941	−2.846	−0.574
RF-TF	−1.289	−3.351	−3.093	−3.555	−3.237	−4.202	−3.939	−2.387	−0.687
NN-TF	−1.558	−2.748	−3.145	−3.212	−2.988	−3.798	−3.661	−2.339	−0.373
D. Forecast Combination-Augmented AR (FC)									
MN-FC	−1.420	−2.023	−1.652	−1.295	−1.248	−1.654	−1.468	−0.413	1.075
MD-FC	−1.317	−2.434	−2.824	−2.776	−2.747	−3.419	−3.200	−2.090	−0.276
RANK-FC	−1.448	−2.584	−2.843	−2.919	−3.066	−3.498	−3.201	−2.182	−0.324
CL-FC	−1.544	−2.699	−3.205	−3.141	−3.060	−3.671	−3.362	−2.230	−0.360
DFE-FC	−1.491	−2.290	−2.297	−1.949	−1.900	−2.533	−2.448	−1.257	0.192
PCA-FC	−1.373	−2.267	−2.846	−2.878	−2.852	−3.710	−3.562	−2.383	−0.649
LASSO-FC	−1.912	−3.572	−3.542	−3.060	−3.019	−3.308	−2.875	−1.634	−0.420
GB-FC	−1.478	−3.001	−2.742	−2.985	−2.993	−3.995	−3.757	−2.558	−0.583
RF-FC	−1.437	−3.068	−3.514	−3.208	−3.548	−4.413	−4.046	−2.237	−0.490
NN-FC	−1.162	−1.521	−2.186	−2.110	−2.447	−2.565	−2.699	−0.087	1.762

This table reports the quantile score *t*-statistics for the null hypothesis of equal predictive ability of the alternative prediction model relative to the benchmark. The benchmark is the autoregressive model augmented by stochastic volatility with leverage (ARSVLEV1). A lower value indicates greater outperformance from the benchmark. Values less than −1.645 indicate that the alternative model outperforms the benchmark at the 5% level. The out-of-sample period is from December 1989 to December 2019.

models that augment the QAR baseline with a large number of predictors outperform the benchmark in the lower and middle quantiles. Notably, the models that improve the most upon the QAR baseline are the predictor-augmented AR based on random forests (RF-PA), the targeted factor-augmented AR based on the neural network (NN-TF) and the cluster weighting scheme to combine forecasts (CL-FC).

4.3. Density forecast evaluation

In this Section we examine the ability of the models to approximate the density of the three target economic variables. To evaluate the ability of a model to forecast an area of the distribution we follow Manzan and Zerom (2013) and Meligkotsidou et al. (2019) and use the weighted quantile score (WQS) function. The WQS is constructed by integrating the QS across a set of quantiles, with the score multiplied by a weight function that focuses on a specific part of the distribution. The WQS is defined as follows:

$$\mathrm{WQS}_{i,t} = \int_0^1 \mathrm{QS}_{i,\tau,t}\,\omega_\tau d\tau, \tag{31}$$

where ω_τ denotes a weight function in the unit interval. We replace the continuous version of WQS with a discrete version summing over the quantiles of interest, allowing us to evaluate specific areas of the distribution. We employ four different weighting functions ω_τ: 1. full: $\omega_\tau = 1$, which assigns uniform weights across the entire distribution; 2. mid: $\omega_\tau = \tau(1-\tau)$, places more weight in the middle of the distribution; 3. left: $\omega_\tau = (1-\tau)^2$, assigns more weight to the left tail of the distribution; 4. right: $\omega_\tau = \tau^2$, which focuses on the right tail of the distribution. Another measure we consider when evaluating density forecasts is the mean log predictive score (MLPS), where the log predictive score is derived as the logarithm of the predictive density generated by a model and evaluated at the realized value of the target variable (see e.g., Geweke & Amisano, 2010). If the alternative model generates a higher MLPS value than the benchmark model, then that model outperforms the benchmark. We evaluate the statistical significance of the WQS and MLPS measures using the Diebold and Mariano statistic. The test statistic is constructed to be negatively oriented, with model i outperforming the benchmark if the statistic is significantly negative.

Table 4. Quantile forecast evaluation for Employment (PAYEMS).

	q(0.1)	q(0.2)	q(0.3)	q(0.4)	q(0.5)	q(0.6)	q(0.7)	q(0.8)	q(0.9)
RW	5.561	5.687	3.422	2.347	2.558	3.272	4.451	6.063	6.772
AR1	−2.959	−2.438	0.366	0.205	0.305	0.566	1.533	3.244	4.325
AR12	−3.435	−3.041	−2.653	−3.351	−3.639	−3.637	−2.710	−1.619	1.109
ARSV1	−2.360	−4.225	−3.149	−1.556	1.215	3.566	3.761	2.911	3.105
ARSV12	−2.108	−2.046	−2.710	−3.241	−3.454	−3.601	−3.455	−2.981	−1.847
ARSVLEV12	−2.723	−2.548	−3.057	−3.509	−3.630	−3.686	−3.579	−3.207	−2.156
VARCSV1	−1.785	−1.043	−0.418	−0.307	−0.103	0.818	2.020	3.024	5.070
VARCSV2	−2.250	−2.913	−2.878	−2.536	−2.270	−1.987	−1.382	−0.964	−0.480
VARFSV1	−1.580	−0.942	−0.133	0.082	0.117	1.037	1.992	3.149	4.529
VARFSV2	−2.146	−2.898	−2.795	−2.647	−2.554	−2.146	−1.856	−1.771	−1.115
A. Predictor-Augmented AR (PA)									
PCA-PA	−3.805	−3.970	−3.786	−3.876	−4.063	−3.112	−2.344	−1.176	1.057
LASSO-PA	−1.697	−1.166	−1.241	−1.531	−1.240	−0.939	−0.351	0.929	1.527
GB-PA	−2.102	−2.505	−2.544	−3.014	−3.169	−2.548	−2.906	−2.299	−0.140
RF-PA	−3.481	−3.808	−3.412	−4.263	−4.319	−4.266	−3.131	−2.026	0.797
NN-PA	5.827	2.745	0.126	−0.640	−1.651	−1.427	−0.651	1.810	5.297
B. Targeted Predictor-Augmented AR (TP)									
LASSO-TP	−1.869	−1.915	−1.727	−2.084	−2.643	−2.117	−1.840	−1.556	0.600
GB-TP	−1.758	−2.355	−2.522	−3.301	−2.999	−2.820	−2.263	−1.322	2.050
RF-TP	−3.035	−2.048	−2.767	−3.132	−3.546	−3.823	−2.504	−1.396	1.078
NN-TP	−2.610	−2.581	−2.893	−3.258	−3.308	−3.429	−2.175	−1.503	0.796
C. Targeted Factor-Augmented AR (TF)									
LASSO-TF	−2.454	−2.782	−2.306	−2.633	−2.477	−2.812	−3.167	−1.810	−0.201
GB-TF	−2.813	−2.554	−3.260	−4.087	−3.703	−3.638	−2.639	−2.005	0.848
RF-TF	−3.245	−3.588	−3.233	−3.869	−4.058	−3.705	−2.781	−1.843	0.275
NN-TF	−3.474	−3.348	−3.101	−3.824	−3.949	−3.797	−2.304	−1.711	0.948
D. Forecast Combination-Augmented AR (FC)									
MN-FC	−2.657	−2.383	−1.982	−2.039	−1.698	−0.602	0.597	2.395	4.214
MD-FC	−3.665	−3.085	−2.673	−3.210	−3.372	−2.875	−1.889	−0.862	1.563
RANK-FC	−3.958	−3.397	−2.810	−3.391	−3.661	−3.346	−2.204	−0.904	1.677
CL-FC	−3.663	−3.340	−3.134	−3.602	−3.878	−3.800	−3.155	−1.959	1.022
DFE-FC	−3.160	−3.052	−2.858	−3.181	−3.024	−2.350	−1.308	0.143	2.637
PCA-FC	−3.073	−3.369	−3.076	−3.560	−4.090	−3.742	−2.768	−1.464	1.222
LASSO-FC	−2.822	−3.493	−3.228	−3.406	−3.273	−2.587	−1.992	−0.646	1.411
GB-FC	−2.279	−2.170	−1.889	−3.292	−3.220	−3.308	−2.667	−2.445	−0.422
RF-FC	−2.939	−3.027	−2.746	−3.210	−3.890	−3.837	−2.971	−1.883	1.109
NN-FC	−0.053	−1.583	−1.954	−1.707	−2.202	−2.366	−2.011	−1.723	6.382

This table reports the quantile score *t*-statistics for the null hypothesis of equal predictive ability of the alternative prediction model relative to the benchmark. The benchmark is the autoregressive model augmented by stochastic volatility with leverage (ARSVLEV1). A lower value indicates greater outperformance from the benchmark. Values less than −1.645 indicate that the alternative model outperforms the benchmark at the 5% level. The out-of-sample period is from December 1989 to December 2019.

Table 5 reports the WQS and MLPS *t*-statistics for the null hypothesis of equal predictive ability of the alternative prediction model relative to the AR(1) model augmented by stochastic volatility with leverage (ARSVLEV1) benchmark, for industrial production, inflation and employment.

Overall, the findings show that the majority of the models produce significantly superior density forecasts relative to the benchmark across the three variables of interest, with the augmented AR specifications improving upon the QAR baseline model. This outperformance can be attributed to greater forecasting accuracy in the left tail and center of the distribution. For industrial production and inflation, models from the four specifications that incorporate a large number of predictors offer the highest accuracy, while for employment, models augmented by stochastic volatility yield comparable out-of-sample performance.

For industrial production, the AR(12) outperforms the remaining autoregressive models according to the WQS metrics for the full distribution, middle and the left tail, however, based on the MLPS test statistic the AR(12) and VAR(2) models augmented by stochastic volatility exhibit stronger performance. In the case of models that use the predictors directly as inputs (Panel A), random forests (RF-PA) and PCA (PCA-PA) offer the lowest WQS test statistics, while gradient boosting (GB-PA) followed by PCA generate the lowest values for the MLPS *t*-statistic. In contrast, the lasso does not significantly outperform the benchmark in this specification. For the case of an AR augmented by targeted predictors (Panel B), the neural network (NN-TP) is the best performing model, while for the case of the targeted factor-augmented AR (Panel C), the neural network (NN-TF) and both ensemble methods show significant outperformance across all metrics. Most forecast combinations (Panel D) also significantly outperform the benchmark, with the lasso-based combination (LASSO-FC) having the strongest performance.

Turning to the results for inflation, the AR(12) outperforms the remaining autoregressive models according to the MLPS *t*-statistic and in terms of WQS for the full distribution, center and right tail of the distribution. However, VAR models exhibit better out-of-sample performance in the left tail. Overall, ensemble approaches are the best performing models across the four augmented AR

Table 5. Density forecast evaluation for the Industrial Production Index (INDPRO), the Consumer Price Index (CPIAUCSL) and Employment (PAYEMS).

	INDPRO					CPIAUCSL					PAYEMS				
	Full	Mid	Left	Right	MLPS	Full	Mid	Left	Right	MLPS	Full	Mid	Left	Right	MLPS
RW	6.198	6.072	6.040	5.657	5.391	4.992	4.878	4.470	4.382	5.481	6.384	5.483	6.226	6.089	10.566
AR1	−0.656	−0.653	−0.191	−0.934	−0.320	−2.393	−2.588	−2.151	−1.935	−3.084	1.021	0.863	−0.963	2.741	−0.203
AR12	−2.746	−2.748	−3.131	−1.910	−2.321	−3.218	−3.382	−2.747	−2.884	−3.574	−3.343	−3.492	−3.760	−2.251	−2.386
ARSV1	2.225	2.391	−0.717	2.877	−1.536	3.514	3.471	0.908	3.896	12.140	1.344	1.415	−2.809	3.609	−2.661
ARSV12	−2.088	−2.196	−2.438	−1.329	−3.395	−1.953	−1.806	−2.454	−1.307	−0.287	−3.664	−3.659	−3.147	−3.564	−2.449
ARSVLEV12	−2.462	−2.474	−2.683	−1.917	−3.830	−1.996	−1.862	−2.313	−1.507	−0.066	−4.010	−3.954	−3.612	−3.793	−2.546
VARCSV1	−1.621	−1.614	−1.580	−1.392	−2.712	−2.986	−2.841	−3.643	−1.952	−1.042	0.564	0.483	−0.753	2.215	−0.756
VARCSV2	−2.170	−2.187	−2.100	−1.878	−3.859	−3.181	−3.027	−3.613	−2.336	−1.516	−2.576	−2.533	−2.918	−1.806	−1.836
VARFSV1	−1.571	−1.569	−1.755	−1.119	−2.622	−2.859	−2.741	−3.411	−1.926	−1.424	0.712	0.667	−0.487	2.197	−0.629
VARFSV2	−2.389	−2.440	−2.507	−1.843	−3.484	−3.024	−2.900	−3.601	−2.030	−1.409	−2.860	−2.794	−2.955	−2.340	−1.934
A. Predictor-Augmented AR (PA)															
PCA-PA	−2.928	−2.967	−3.520	−1.799	−3.481	−3.373	−3.502	−2.899	−3.058	−3.812	−3.539	−3.709	−4.506	−2.020	−2.487
LASSO-PA	0.566	0.360	0.087	1.151	0.650	−1.767	−1.700	−2.030	−1.220	−2.908	−0.878	−0.997	−1.529	0.010	−1.856
GB-PA	−2.842	−2.991	−2.946	−2.040	−4.752	−3.936	−4.054	−3.325	−3.607	−4.359	−3.174	−3.206	−3.055	−2.630	−3.057
RF-PA	−3.012	−2.941	−3.518	−2.157	−0.875	−4.280	−4.312	−4.287	−3.222	−3.753	−4.072	−4.233	−4.472	−2.872	−2.723
NN-PA	−2.047	−2.221	−2.231	−1.228	1.069	0.826	0.232	0.934	1.109	4.002	2.968	1.511	3.775	2.787	2.494
B. Targeted Predictor-Augmented AR (TP)															
LASSO-TP	−2.208	−2.192	−2.975	−1.112	−2.586	−3.342	−3.460	−3.073	−2.793	−3.680	−2.283	−2.344	−2.373	−1.722	−3.016
GB-TP	−2.526	−2.688	−2.919	−1.534	−2.375	−4.095	−4.100	−4.183	−3.223	−4.560	−2.712	−2.950	−3.069	−1.634	−2.366
RF-TP	−2.177	−2.361	−2.699	−1.104	−3.108	−3.223	−3.326	−3.306	−2.435	−2.971	−3.353	−3.467	−3.551	−2.393	−3.490
NN-TP	−3.203	−3.392	−3.766	−1.915	−3.319	−3.169	−3.194	−3.135	−2.558	−3.028	−3.122	−3.289	−3.476	−2.074	−2.397
C. Targeted Factor-Augmented AR (TF)															
LASSO-TF	−2.204	−2.238	−2.993	−1.078	−2.685	−3.857	−3.981	−3.482	−3.305	−3.896	−2.992	−3.007	−3.055	−2.445	−2.820
GB-TF	−2.914	−3.039	−3.318	−1.902	−2.959	−4.276	−4.302	−4.253	−3.493	−4.237	−3.573	−3.766	−3.923	−2.461	−2.859
RF-TF	−2.907	−3.006	−3.717	−1.514	−3.728	−3.673	−3.861	−3.200	−3.176	−3.449	−3.826	−3.928	−4.251	−2.705	−2.986
NN-TF	−2.992	−3.009	−3.556	−1.967	−3.090	−3.410	−3.557	−3.038	−2.907	−3.519	−3.574	−3.745	−4.114	−2.316	−2.500
D. Forecast Combination-Augmented AR (FC)															
MN-FC	−1.751	−1.905	−2.715	−0.407	1.307	−1.515	−1.564	−1.837	−0.791	−1.778	−0.508	−0.829	−2.199	1.296	−0.087
MD-FC	−2.619	−2.633	−3.077	−1.725	−1.769	−2.999	−3.135	−2.644	−2.610	−3.421	−2.858	−3.016	−3.623	−1.563	−1.926
RANK-FC	−2.767	−2.814	−3.252	−1.792	−1.835	−3.152	−3.283	−2.806	−2.725	−3.302	−3.145	−3.327	−3.952	−1.753	−2.299
CL-FC	−2.867	−2.872	−3.218	−2.050	−2.491	−3.329	−3.468	−2.980	−2.826	−3.671	−3.733	−3.868	−4.150	−2.589	−2.797
DFE-FC	−2.088	−2.172	−2.870	−0.927	0.191	−2.322	−2.379	−2.315	−1.780	−2.481	−2.327	−2.572	−3.438	−0.768	−1.371
PCA-FC	−2.831	−2.829	−3.261	−1.982	−3.941	−3.209	−3.331	−2.699	−2.966	−3.348	−3.555	−3.757	−4.020	−2.301	−2.676
LASSO-FC	−3.276	−3.302	−3.507	−2.464	−5.055	−3.404	−3.484	−3.392	−2.527	−3.516	−2.894	−3.061	−3.758	−1.607	−2.187
GB-FC	−2.391	−2.471	−2.574	−1.680	−4.630	−3.482	−3.589	−2.966	−3.194	−3.865	−3.254	−3.293	−3.008	−2.828	−3.197
RF-FC	−2.751	−2.848	−2.842	−2.076	−1.609	−3.747	−3.912	−3.253	−3.278	−3.681	−3.462	−3.631	−3.755	−2.426	−1.375
NN-FC	−2.255	−2.297	−2.672	−1.389	−2.216	−1.982	−2.283	−2.101	−1.034	0.553	−1.361	−1.953	−1.919	0.088	0.732

This table reports the weighted quantile score (WQS) and mean log predictive score (MLPS) *t*-statistics for the null hypothesis of equal predictive ability of the alternative prediction model relative to the benchmark. Four WQS functions are considered: full, provides an overall evaluation of the forecast distribution; mid, places more weight in the middle of the distribution; left, assigns more weight to the left tail of the distribution; right, which focuses on the right tail of the distribution. The benchmark is the autoregressive model augmented by stochastic volatility with leverage (ARSVLEV1). A lower value indicates greater outperformance from the benchmark. Values less than −1.645 indicate that the alternative model outperforms the benchmark at the 5% level. The out-of-sample period is from December 1989 to December 2019.

specifications. Specifically, for the predictor-augmented and forecast combination specifications, random forests (RF-PA and RF-FC) offer the lowest WQS t-statistics, while gradient boosting (GB-PA and GB-FC) has the lowest MLPS t-statistic. On the other hand, the neural network for the predictor-augmented AR (NN-PA) does not outperform the benchmark, and forecast combinations based on mean and neural network weighting schemes (MN-FC and NN-FC) show mixed results in terms of significant outperformance. For the remaining two specifications, all models significantly outperform the benchmark, with the targeted predictor- and the targeted factor-augmented AR based on gradient boosting (GB-TP and GB-TF) generating the best performance.

Finally, the results for employment indicate that among the autoregressive models, the AR(12) augmented by stochastic volatility with leverage (ARSVLEV12) exhibits the strongest performance across all metrics, except for the WQS t-statistic in the left tail, where the AR(12) baseline is better at outperforming the benchmark. In general, the majority of the models that augment the QAR baseline with a large number of predictors, exhibit improved forecasting performance in terms of the MLPS t-statistic and the WQS t-statistic in the left tail, while for the other metrics, the performance is comparable to that of the univariate autoregressive models such as the ARSVLEV12. In particular, random forests is better at significantly outperforming the benchmark and improving upon the QAR baseline in specifications that combine information (PA, TP and TF), while the cluster weighting scheme (CL-FC) and PCA-FC are among the top performing models in the forecast combination specification.

4.4. Further analysis

4.4.1. Performance over time

In this Section we examine how model performance evolves over time. For the point forecast evaluation, we plot in Figure A1 in Supplementary Appendix A the cumulative squared errors for the random walk benchmark relative to a selection of models over time. The chosen models are those that directly forecast the mean (LS) and exhibit the lowest MSE ratio in the five specifications of Table 1. The financial crisis of 2007–2008 is a break for the evolution of the performance of the models. Performance is more volatile for industrial production and inflation during the global financial crisis than it is for employment. After the crisis performance of the models over the benchmark accumulates higher, especially for industrial production and inflation. Focusing on the rightmost point in the respective figure, we observe that for industrial production a targeted factor-augmented AR based on random forests (RF-TF) closely followed by a factor-augmented AR based on PCA (PCA-PA) are the best performing models. For inflation, forecast combinations based on the neural network (NN-FC) is the best performing model especially after the global financial crisis, while before the crisis a predictor augmented AR based on gradient boosting (GB-PA), or targeted factor-augmented AR based on the neural network (NN-TF) outperform the remaining models. For employment, point forecasts generated by a factor-augmented AR (PCA-PA), targeted factor-augmented AR based on gradient boosting (GB-TF) and forecast combination based on PCA (PCA-FC) outperform the other models over the out-of-sample period.

For the density forecast evaluation, we plot the cumulative differences in the four WQS functions for the ARSVLEV1 benchmark relative to the chosen forecasting models. For each of the three variables of interest, the models selected are those that yield the lowest WQS t-statistics for the full distribution in each specification of Table 5. Figure A2 depicts how the different WQS metrics evolve over time for industrial production. The plots reveal that the global financial crisis constitutes a break for density forecasts as well, with cumulative performance being more volatile during the crisis, and increasing after the crisis with the trend being steeper for the left part of the distribution. Considering the rightmost point in the plots a forecast combination based on the lasso (LASSO-FC) is the best performing model for the full distribution, with a targeted predictor-augmented AR based on the neural network (NN-TP) offering equivalent performance in the middle part of the distribution and outperforming LASSO-FC in the left tail of the distribution. The results for inflation, presented in Figure A3, show that cumulative performance decreases during the crisis, then sharply increases and relatively plateaus until the end of the sample for all cases considered. The best performing models for inflation according to all WQS metrics are targeted predictor- and targeted factor-augmented AR based on gradient boosting (GB-TP and GB-TF). For employment (Figure A4), the results reveal that cumulative performance during and after the crisis is more stable in the left tail compared to the middle and right tail parts of the distribution. Predictor- and targeted factor-augmented AR models based on random forests (RF-PA and RF-TF) are the best performing models for the middle and left tail parts of the distribution. It is interesting to note that for the right tail of the distribution the AR(12) augmented by stochastic volatility with leverage outperforms the remaining models.

4.4.2. Variable importance analysis

In this Section, we provide a description of the results for the variable importance analysis of the predictor-augmented AR and forecast combination-augmented AR specifications for the cases of the conditional mean and the conditional quantiles. The out-of-sample variable importance is constructed for each of the eight groups of variables in FRED-MD.[3] The variable importance in each period is computed as the absolute change in MSE for the conditional mean forecasts, or the QS for the conditional quantile forecasts, by setting each one of the predictors to zero. Variable importance for a group is the average change in MSE or QS of the variables within that group. The variable importance measure is averaged throughout the out-of-sample period and is normalized for each model to sum to 100.

The results of the variable importance analysis for industrial production are presented in Figure A5 in Supplementary Appendix A for the conditional mean. In the case of models that combine information, the results indicate that labor market indicators are primarily chosen by the lasso, while variables from the money and credit group and output and income group are chosen by the GB and RF models respectively. On the other hand, bivariate forecasts based on variables from the money and credit or housing groups are those primarily selected by the ensemble methods. Overall, forecasts based on neural networks place approximately equal importance in each variable group.

The variable importance for industrial production for the conditional quantiles is reported in Figures A6 and A7 for models that combine information or combine forecasts respectively. The results show that the groups of important predictors vary greatly based on the value for τ and the forecasting model. In the case of the lower quantiles ($\tau \in \{0.1, 0.2, 0.3\}$), predictors from the output and income are better on average at explaining industrial production, however, when combining forecasts, the interest and exchange rate groups is most often selected across all models. For the middle part of the distribution ($\tau \in \{0.4, 0.5, 0.6\}$), the labor market, in addition to output and income groups are important regardless of whether predictors or simple forecasts are combined. For the right tail of the distribution ($\tau \in \{0.7, 0.8, 0.9\}$), output and income predictors are selected by most models, except for neural networks that focus on consumption, orders, and inventories. Simple forecasts of labor market variables are selected primarily by the forecast combination methods.

Figure A8 in the Supplementary Appendix presents the variable importance analysis for inflation in the case of point forecasts. A predictor-augmented AR based on the lasso would select variables primarily from the output and income group, while gradient boosting focuses on price indicators and random forests on labor market variables. In contrast, when the predictor set in comprised of individual forecasts, the lasso overall focuses on variables from the housing group. Both ensemble methods select bivariate forecasts from the price group, while random forests also selects simple forecasts based on money and credit variables. Neural networks again place approximately equal importance across groups.

Figures A9 and A10 report the variable importance for the quantile forecasts of inflation for models that use the predictors variables directly or combine forecasts respectively. For the left tail part of the distribution, models that combine information select predictors primarily from the output and income group and then from the price group. However, when modelling the middle and right tail part of the distribution the majority of the models place greater importance on the price group, with housing variables also being useful for modelling quantiles for $\tau \in \{0.4, 0.5, 0.6\}$. Bivariate forecasts based on consumption, orders, and inventories variables dominate all other groups when combining forecasts, especially for the middle to right tail part of the distribution. For the lower quantiles, simple forecasts based on output and income, or interest and exchange rate variables are also important in modelling inflation.

Lastly, the variable importance results for the conditional mean forecasts of employment are reported in Figure A11. Of particular note is that lasso and especially random forests select variables from the labor market group, while gradient boosting selects primarily housing indicators. Otherwise, when the predictor set is comprised of bivariate forecasts, the lasso selects variables from the output and income group, GB selects bivariate forecasts based on labor market variables and RF from those in the housing group.

The results for the variable importance of each quantile for employment are depicted in Figures A12 and A13 for the case of models that combine information or forecast combinations respectively. In terms of variable importance, regardless of whether information or forecasts are combined, the most important group to model employment are labor market indicators. When modelling the middle part of the distribution, the models that combine information also focus on interest and exchange rate variables. Forecast combinations also select bivariate forecasts based on variables from the consumption, order and inventories group to estimate the lower quantiles, stock market variables to model the lower tail and middle part of the distribution, while for the upper quantiles, forecasts based on housing variables are also important.

4.4.3. Amalgamation of forecasts

To account for the heterogeneity observed in the forecasting accuracy across different model specifications and performance metrics we follow Rapach and Strauss (2012) and construct an amalgamation of all individual models. The point, quantile and density performance evaluation of the amalgam forecasts is presented in Table A9 in Supplementary Appendix A.

Starting with the point forecast evaluation, aggregating forecasts across the LS and the three QS specifications leads to amalgam forecasts that significantly outperform the random walk benchmark for all variables of interest. However, pooling forecasts across all models does not lead to any significant gains compared to the best performing model, with the MSE ratios between the amalgam and the model with the strongest performance being equivalent.

The results for the quantile forecast evaluation indicate that aggregating forecasts across all models for each quantile leads to statistically significant outperformance over the benchmark for $\tau \in [0.1, 0.8]$ for all target variables. For industrial production, the amalgamation approach yields lower QS t-statistic than the best performing individual model for $\tau \in [0.1, 0.5]$. In the case of inflation, the amalgamation approach outperforms the top individual model for $\tau \in [0.2, 0.6]$. For employment, the amalgam forecast would be preferable over that of the individual model with the lowest QS t-statistic for $\tau \in [0.3, 0.6]$.

The amalgamation approach results to density forecasts that significantly outperform the benchmark in terms of all metrics for the three variables of interest. For industrial production, aggregating forecasts leads to lower WQS t-statistics than the individual models, while models such as GB-PA and LASSO-FC generate lower MLPS t-statistics. Similarly, for inflation the amalgamation approach yields improved performance according to the WQS t-statistics, although models that combine information using gradient boosting generate lower MLPS t-statistics. In the case of employment, the amalgam forecasts generate superior performance according to the WQS t-statistics for the full distribution, the left and middle parts of the distribution, while individual models based on ensemble approaches outperform the amalgam forecast in terms of MLPS.

4.4.4. Model performance including the COVID-19 pandemic period

So far, this paper has focused on a sample period ending in December 2019. In this Section, we explore the performance of the models for an extended sample up to March 2024, which includes the turbulent period of the COVID-19 pandemic and subsequent lockdowns. This poses a significant challenge for macroeconomic forecasting, as the values of the target variables deviate significantly from their historical range.[4]

The performance evaluation for the point forecasts is reported in Table A10 in Supplementary Appendix A. For inflation, the majority of the models provide significant outperformance over the benchmark. In contrast, although most models outperform the benchmark for industrial production, the results are not statistically significant. For employment, models that use a large number of predictors do not outperform the benchmark, while stochastic volatility models yield statistically insignificant outperformance.

The results for the quantile forecast evaluation can be found in Tables A11 to A13 for the respective variable of interest. For inflation, while performance diminishes after extending the sample, most models continue to significantly outperform the benchmark for $\tau \in [0.1, 0.8]$. For industrial production, we observe significant outperformance mainly in the lower quantiles ($\tau \in \{0.1, 0.2\}$), while for the remaining quantiles the results are not statistically significant. For employment, augmented AR models are most impacted by including the post-2019 period, while Bayesian VARs and univariate autoregressive models with a single lag show significant outperformance in the lower quantiles.

The density forecast evaluation is presented in Table A14. For inflation, most models continue to produce statistically significant results, with models incorporating a large number of predictors improving upon the QAR baseline, particularly in the left tail and middle parts of the distribution. For industrial production, performance is better in the left tail of the distribution, however, the results overall are statistically insignificant. In the case of employment, autoregressive models with a single lag tend to offer statistically significant outperformance.

5. Conclusion

In this study we examine the ability of a large number of economic variables to forecast key business cycle and inflation indicators. The aim of our analysis is to construct point, quantile, and density forecasts of industrial production, inflation and employment, by leveraging the benefits arising from the quantile regression framework, machine learning approaches and the information contained in the predictor set. Specifically, the variables of interest are analysed assuming that their future values depend on their own lags and a large set of macroeconomic and financial variables, through several specifications of an augmented time series model. These models allow for heterogeneous degrees of persistence of the target variable and asymmetric dynamic responses of economic variables at different parts of the distribution.

The results suggest that combining information across quantiles to forecast the conditional mean considerably improves the out-of-sample performance. In addition, augmenting an autoregressive model with economic variables though methods that perform variable selection or account for non-linearities can further increase the accuracy of the point forecasts relative to simpler models. Furthermore, our findings indicate that the improved forecasting performance arises from incorporating economic information into the quantile forecasts and is driven especially from the lower left and middle part of the distribution. The results also show that the proposed approaches and the information of the economic variables lead to more accurate distribution forecasts than the benchmark models.

Acknowledgements

We would like to thank two anonymous referees and the Editor, Zhe George Zhang, for helpful advice and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes

1. For simplicity of notation, we omit the intercept in the description of the VAR models.
2. https://research.stlouisfed.org/econ/mccracken/fred-databases/.
3. Following McCracken and Ng (2016) the groups are Consumption, orders, and inventories; Housing; Interest and exchange rates; Labor market; Money and credit; Output and income; Prices; Stock market. Details on which variables belong in each group can be found in Tables A1–A8 in Supplementary Appendix A.
4. For the univariate models, we follow Medeiros et al. (2021) and incorporate a dummy variable for March 2020 and April 2020 to account for the NBER-dated recession triggered by the COVID-19 pandemic. The dummy variable is included in the models only after those months are within the in-sample period, which ensures that there is no look-ahead bias in the analysis.

Funding

The authors gratefully acknowledge the support of the Irish Research Council under grant number GOIPD/2022/721.

References

Aiolfi, M., & Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, *135*(1–2), 31–53. https://doi.org/10.1016/j.jeconom.2005.07.015

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, *25*(2), 177–190. https://doi.org/10.1198/073500106000000332

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317. https://doi.org/10.1016/j.jeconom.2008.08.010

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92. https://doi.org/10.1002/jae.1137

Barbaglia, L., Consoli, S., & Manzan, S. (2023). Forecasting with economic news. *Journal of Business & Economic Statistics*, *41*(3), 708–719. https://doi.org/10.1080/07350015.2022.2060988

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468. https://doi.org/10.1057/jors.1969.103

Belloni, A., & Chernozhukov, V. (2011). ℓ1-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics*, *39*(1), 82–130. https://doi.org/10.1214/10-AOS827

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Carriero, A., Clark, T. E., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, *212*(1), 137–154. https://doi.org/10.1016/j.jeconom.2019.04.024

Carriero, A., Clark, T. E., & Marcellino, M. (2024). Capturing macro-economic tail risks with Bayesian vector autoregressions. *Journal of Money, Credit and Banking*, *56*(5), 1099–1127. https://doi.org/10.1111/jmcb.13121

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480. https://doi.org/10.1093/biomet/asq017

Chernozhukov, V., Fernández-Val, I., & Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, *78*(3), 1093–1125.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, *29*(3), 327–341. https://doi.org/10.1198/jbes.2010.09248

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*(2), 262–302. https://doi.org/10.1016/j.red.2004.10.009

De Gooijer, J. G., & Zerom, D. (2019). Semiparametric quantile averaging in the presence of high-dimensional predictors. *International Journal of Forecasting*, *35*(3), 891–909. https://doi.org/10.1016/j.ijforecast.2018.10.009

De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, *146*(2), 318–328. https://doi.org/10.1016/j.jeconom.2008.08.011

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Feldkircher, M., Gruber, L., Huber, F., & Kastner, G. (2024). Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian vector autoregressions? *Journal of Forecasting*, *43*(6), 2126–2145. https://doi.org/10.1002/for.3121

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, *82*(4), 540–554. https://doi.org/10.1162/003465300559037

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232.

Gastwirth, J. L. (1966). On robust procedures. *Journal of the American Statistical Association*, *61*(316), 929–948. https://doi.org/10.1080/01621459.1966.10482185

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230. https://doi.org/10.1016/j.ijforecast.2009.10.007

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578. https://doi.org/10.1111/j.1468-0262.2006.00718.x

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*(2), 436–451. https://doi.org/10.1162/REST_a_00483

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, *29*(3), 411–422. https://doi.org/10.1198/jbes.2010.08110

Gruber, L., & Kastner, G. (2022). *Forecasting macroeconomic data with Bayesian VARs: Sparse or dense? It Depends!*. arXiv Preprint, arXiv:2206.04902.

Huang, H., & Lee, T. H. (2010). To combine forecasts or to combine information? *Econometric Reviews*, *29*(5–6), 534–570. https://doi.org/10.1080/07474938.2010.481553

Huber, F., & Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics*, *37*(1), 27–39. https://doi.org/10.1080/07350015.2016.1256217

Jacquier, E., Polson, N. G., & Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *12*(4), 371–389. https://doi.org/10.1080/07350015.1994.10524553

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., & Lee, T.-C. (1988). *Introduction to the theory and practice of econometrics*. Wiley.

Kastner, G. (2019). Sparse Bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics*, *210*(1), 98–115. https://doi.org/10.1016/j.jeconom.2018.11.007

Kastner, G., & Huber, F. (2020). Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*, *39*(7), 1142–1165. https://doi.org/10.1002/for.2680

Kastner, G., Frühwirth-Schnatter, S., & Lopes, H. F. (2017). Efficient Bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics*, *26*(4), 905–917. https://doi.org/10.1080/10618600.2017.1322091

Kim, H. H., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, *178*, 352–367. https://doi.org/10.1016/j.jeconom.2013.08.033

Kim, S., Shepherd, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, *65*(3), 361–393. https://doi.org/10.1111/1467-937X.00050

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, *101*(475), 980–990. https://doi.org/10.1198/016214506000000672

Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, *28*(2), 177–203. https://doi.org/10.1002/jae.1270

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, *28*(2), 204–230. https://doi.org/10.1002/jae.1271

Lima, L. R., & Meng, F. (2017). Out-of-sample return predictability: A quantile combination approach. *Journal of Applied Econometrics*, *32*(4), 877–895. https://doi.org/10.1002/jae.2549

Manzan, S. (2015). Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business & Economic Statistics*, *33*(1), 144–164. https://doi.org/10.1080/07350015.2014.937436

Manzan, S., & Zerom, D. (2013). Are macroeconomic variables useful for forecasting the distribution of US inflation? *International Journal of Forecasting*, *29*(3), 469–478. https://doi.org/10.1016/j.ijforecast.2013.01.005

Manzan, S., & Zerom, D. (2015). Asymmetric quantile persistence and predictability: The case of US inflation. *Oxford Bulletin of Economics and Statistics*, *77*(2), 297–318. https://doi.org/10.1111/obes.12065

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589. https://doi.org/10.1080/07350015.2015.1086655

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*, 983–999.

Meligkotsidou, L., Panopoulou, E., Vrontos, I. D., & Vrontos, S. D. (2019). Quantile forecast combinations in realised volatility prediction. *Journal of the Operational Research Society*, *70*(10), 1720–1733. https://doi.org/10.1080/01605682.2018.1489354

Omori, Y., Chib, S., Shephard, N., & Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, *140*(2), 425–449. https://doi.org/10.1016/j.jeconom.2006.07.008

Pfarrhofer, M. (2022). Modeling tail risks of inflation using unobserved component quantile regressions. *Journal of Economic Dynamics and Control*, *143*, 104493. https://doi.org/10.1016/j.jedc.2022.104493

Pfarrhofer, M. (2024). Forecasts with Bayesian vector autoregressions under real time conditions. *Journal of Forecasting*, *43*(3), 771–801. https://doi.org/10.1002/for.3055

Rapach, D. E., & Strauss, J. K. (2012). Forecasting US state-level employment growth: An amalgamation approach. *International Journal of Forecasting*, *28*(2), 315–327. https://doi.org/10.1016/j.ijforecast.2011.08.004

Rossi, B., & Sekhposyan, T. (2014). Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, *30*(3), 662–682. https://doi.org/10.1016/j.ijforecast.2013.03.005

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. https://doi.org/10.1198/016214502388618960

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. https://doi.org/10.1198/073500102317351921

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430. https://doi.org/10.1002/for.928

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, *1*, 203–226.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

Yi, C., & Huang, J. (2017). Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, *26*(3), 547–557. https://doi.org/10.1080/10618600.2016.1256816