Received: 21 October 2021 | Revised: 16 May 2024 | Accepted: 21 July 2024

DOI: 10.1002/for.3182

RESEARCH ARTICLE

WILEY

Macroeconomic real-time forecasts of univariate models with flexible error structures

Kelly Trinh[1,2] | **Bo Zhang**[3,4] | **Chenghan Hou**[5]

[1]Data61, The Commonwealth Scientific and Industrial Research Organisation, Clayton, Victoria, Australia

[2]College of Science and Engineering, James Cook University, Townsville, Queensland, Australia

[3]Business School, Wenzhou University, Wenzhou, Zhejiang, China

[4]Centre for Applied Macroeconomic Analysis (CAMA), Australian National University, Canberra, Australian Capital Territory, Australia

[5]Center for Economics, Finance and Management Studies, Hunan University, Changsha, Hunan, China

Correspondence
Bo Zhang, Business School, Wenzhou University, Wenzhou, Zhejiang 325035, China.
Email: bozhangyc@gmail.com

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 72250410372, 72003064; Wenzhou Philosophy and Social Sciences Key Research Base Project, Grant/Award Number: 22jd30

Abstract

This paper investigates the importance of flexible error structure specifications in two widely used univariate models, namely, autoregressive and unobserved component models, in fitting and forecasting 20 significant US macroeconomic variables. The in-sample estimation reveals that the models with flexible error structures provide better in-sample fit than the univariate models with homoscedastic errors. Furthermore, the density forecast analysis suggests that accommodating heavy tail, stochastic volatility, and serial correlation in error structures leads to significant improvements in short-term forecasts. For most macroeconomic variables, the univariate models tend to yield more accurate one-step-ahead forecasts than the multivariate (vector autoregressive) models in terms of both point and density forecasts.

KEYWORDS

autoregressive moving average, real-time forecasting, stochastic volatility

1 | INTRODUCTION

A large number of studies have emphasized the importance of considering time-varying variance in macroeconomic and financial data for estimation and prediction. A popular approach to incorporate this variability is through a stochastic volatility (SV) model (e.g., Kim et al., 1998; Taylor, 1994).[1] This approach has gained in popularity, as evidenced by the numerous extensions of the basic SV model, such as SV models with jump and Student's *t* error (Chib et al., 2002), SV models with leverage effect (Omori et al., 2007), SV models with moving average (MA) error (Chan, 2013), SV models with autoregressive moving average (ARMA) error (Zhang et al., 2020), and others. More recently, coefficient instability in macroeconomic time series models has been widely acknowledged (e.g., Chan et al., 2012; Cross & Poon, 2016; Koop & Potter, 2007; Stock & Watson, 1996). For instance, Stock and Watson (2007) show that the US CPI inflation is best modeled by an unobserved component (UC) model that includes time-varying volatility in both the transitory and trend equations. This paper aims to investigate the role of structural instability in both error process and mean process in fitting and forecasting a real-time dataset of 20 important US macroeconomic variables.

Several studies have compared the performance of various models for forecasting macroeconomic time series. For univariate models, Bauwens et al. (2015) investigate the forecasting performance of structural break models and autoregressive (AR) models for 60 quarterly and monthly macroeconomic series, while Marcellino et al. (2006) use 170 monthly time series to compare iterated forecast and direct forecast results for an AR model. For multivariate models, Swanson and White (1997) compare the forecasting performance of linear and nonlinear models using nine macroeconomic variables. Athanasopoulos and Vahid (2008) explore five multivariate models including both vector autoregressive (VAR) with homoscedastic errors and VAR with MA errors (VAR-MA) for macroeconomic forecasting. A recent paper by Hou et al. (2023) investigates the impacts of model size and error structures on AR models and VAR models for the Australian economy.

We contribute to the literature on macroeconomic forecasting by investigating the impacts of flexible error specifications on the forecast performance of both univariate and multivariate models. Our analysis employs real-time data comprised of 20 macroeconomic variables representing the US economy. The two classes of univariate models considered in this study are the AR model group and the UC group. The flexible error structures considered in this paper include homoscedastic errors, SV, MA, ARMA structures, and their variants with a heavy-tailed distribution. Furthermore, we compare the forecast performance of the univariate models with the multivariate models, namely, VAR models.

Unlike previous studies, we comprehensively evaluate forecast performance by using a larger range of forecast measures, including root mean squared forecast errors (RMSFEs) and mean absolute forecast errors (MAFEs) for the point forecast measures, as well as log-predictive likelihood, continuous ranked probability scores (CRPS), and quantile-weighted CRPS (qwCRPS) for the density forecast measures. By including the latter measures, we can assess not only the entire predictive distribution but also the tail forecast accuracy, resulting in a thorough evaluation of model performance.

The rest of the paper is structured as follows. Section 2 presents the univariate and multivariate models with flexible error structures. Section 3 describes the priors and estimation of the models. Section 4 reports the full-sample estimation results, while Section 5 presents the out-of-sample forecasting results based on the point and density forecast measures. Section 6 concludes the paper.

2 | MODELS

In this section, we present the univariate and multivariate models with flexible error specifications considered in our forecasting exercise.

2.1 | Univariate models

The univariate models we consider are within a general state space framework, where a time series y_t is decomposed into a conditional mean process μ_t and a stationary error process ϵ_t^y, which can be expressed as:

$$y_t = \mu_t + \varepsilon_t^y, \tag{1}$$

where ε_t^y is often assumed to be a Gaussian distributed with a zero mean and a constant variance σ_y^2, that is, $\varepsilon_t^y \sim \mathcal{N}(0, \sigma_y^2)$. Equation (1) is commonly called the measurement equation or observation equation.

2.1.1 | The mean process μ_t

One popular approach to modeling stochastic changes in the trend of a macroeconomic time series assumes that the conditional mean follows a random walk process. This is known as the unobserved component (UC) model. The conditional mean is specified as

$$\mu_t = \tau_t, \tag{2}$$

$$\tau_t = \tau_{t-1} + \varepsilon_t^\tau,\ \varepsilon_t^\tau \sim \mathcal{N}(0, \sigma_\tau^2), \tag{3}$$

for $t = 1, ..., T$. Equation (3) is the state equation describing the evolution of the state parameter τ_t, which is modeled by a random walk process. We assume the initial value τ_1 to be Gaussian, that is, $\tau_1 \sim \mathcal{N}(0, V_\tau)$.

Another popular mean process is the AR process that characterizes a linear dynamic relationship between y_t and its own lags, which assumes the conditional mean of y_t to be

$$\mu_t = \beta_0 + \beta_1 y_{t-1} + ... + \beta_m y_{t-m},$$

where β_0 is the intercept and $\beta_1,\ldots,$ and β_m are the AR coefficients associated with the lags. The lag order m of the AR model is selected using information criteria such as Akaike information criterion (AIC), Hannan–Quinn information criterion (HQC), and the Bayesian information criterion (BIC). More details are provided in Appendix S1.

Many studies in macroeconomic forecasting have shown that the assumption of error terms being homoscedastic Gaussian is too restrictive as it does not take into account important features exhibited in many macroeconomic time series such as heteroscedasticity, heavy tails, and serial correlation. In the following section, we consider different modeling approaches that incorporate these features.

2.1.2 | The error process ε_t^y

Heteroscedastic innovations

It has been shown that conditional variances of many macroeconomic data series substantially change over time; therefore, allowing time variations in the conditional variances is important for improving macroeconomic forecasting accuracy (Carriero et al., 2016; Clark & Ravazzolo, 2015). One prominent approach for modeling this feature is to allow the conditional variances to evolve as the following SV process:

$$\begin{aligned} \varepsilon_t^y &\sim \mathcal{N}(0, e^{h_t}), \\ h_t &= \phi_h h_{t-1} + \varepsilon_t^h,\ \varepsilon_t^h \sim \mathcal{N}(0, \sigma_h^2), \end{aligned} \tag{4}$$

where the log volatility h_t is specified as a stationary AR(1) process, that is, $|\phi_h| < 1$, and its initial state is assumed to follow the corresponding AR(1) stationary distribution, $h_1 \sim \mathcal{N}\left(0, \frac{\sigma_h^2}{1-\phi_h^2}\right)$.

Serially dependent innovations

The second variant of the error process allows for serial correlation, which has been considered in several studies, including Stock and Watson (2007), Chan (2013), and Zhang et al. (2020). To be specific, the error term ε_t^y is assumed to be evolving as a MA(1) process or an ARMA(1,1) process. For the MA(1) error structure, ε_t^y has the following specification:

$$\varepsilon_t^y = u_t + \psi u_{t-j},\ u_t \sim \mathcal{N}(0, \sigma_y^2), \tag{5}$$

where ψ is the MA coefficient, and its absolute value is restricted to be less than 1, that is, $|\psi| < 1$, to ensure the invertibility condition of the MA process. This specification is motivated by a theoretical justification of the Wold decomposition theorem, which suggests that any zero-mean covariance-stationary time series has an infinite MA representation.

The ARMA(1,1) error structure is given by

$$\varepsilon_t^y = \phi \epsilon_{t-1}^y + u_t + \psi u_{t-1},\ u_t \sim \mathcal{N}(0, \sigma_y^2), \tag{6}$$

where the AR coefficient ϕ and the MA coefficient ψ are restricted to be less than 1 to ensure the invertibility and stability conditions (Chib & Greenberg, 1994), that is, $|\phi| < 1$ and $|\psi| < 1$.

Non-Gaussian innovations

The conventional Gaussian error distribution assumption often places little mass on extreme values. To alleviate the effects of extreme values, heavy-tailed distributions are often used. A heavy-tailed distribution can be written as a scale of mixtures of Gaussian distributions which simplifies the model estimation via data augmentation. A popular heavy-tailed distribution used in macroeconomics is the *t*-distributed distribution, which can be represented as

$$\varepsilon_t^y \sim \mathcal{N}(0, \lambda_t \sigma_y^2),\ \lambda_t \sim \mathcal{IG}(\nu/2, \nu/2).$$

Geweke (1993) shows that by integrating out the scale mixture parameter λ_t, the resulting marginal distribution of the error term ϵ_t^y is a *t*-distribution with degrees of freedom ν.

Heteroscedastic with/without non-Gaussian and/or serially dependent innovations

Various flexible specifications for the error process can be obtained further by combining different specifications presented in the aforementioned sections. For example,

- MASV: $\varepsilon_t^y = u_t + \psi u_{t-j},\ u_t \sim \mathcal{N}(0, e^{h_t})$;
- ARMASV: $\varepsilon_t^y = \phi \epsilon_{t-1}^y + u_t + \psi u_{t-1},\ u_t \sim \mathcal{N}(0, e^{h_t})$;
- t-SV: $\varepsilon_t^y \sim \mathcal{N}(0, \lambda_t e^{h_t}),\ \lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$;
- t-MASV: $\varepsilon_t^y = u_t + \psi u_{t-j},\ u_t \sim \mathcal{N}(0, \lambda_t e^{h_t}),\ \lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$;
- t-ARMASV: $\varepsilon_t^y = \phi \epsilon_{t-1}^y + u_t + \psi u_{t-1},\ u_t \sim \mathcal{N}(0, \lambda_t e^{h_t})$, $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$.

2.2 | Multivariate models

The Bayesian VAR is a widely used multivariate model for macroeconomic structural analysis and forecasting.

For this reason, we also discuss the VAR model and some of its variants.

A conventional VAR with a constant covariance matrix has the following specification:

$$\mathbf{y}_t = \mathbf{b}_0 + \mathbf{B}_1\mathbf{y}_{t-1} + \ldots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t^y,\ \boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{7}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of endogenous variables in the VAR system; $\mathbf{b}_0$ is an $n \times 1$ vector of intercepts; $\mathbf{B}_1, \ldots,$ and $\mathbf{B}_p$ are $n \times n$ coefficient matrices; and $\boldsymbol{\Sigma}$ is an $n \times n$ covariance matrix of error term $\boldsymbol{\varepsilon}_t^y$, which is assumed to be a Gaussian.

We also consider a few variants of VAR models accounting for features of the data series such as the time-varying volatility, serial correlation, and heavy-tailed distribution in the error process. The first variant is the VAR with a common SV (VAR-SV), in which the time variation in the covariance matrix of the VAR is assumed to be driven by one common SV:

$$\boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, e^{h_t}\boldsymbol{\Sigma}),$$

where h_t is the common log volatility at time t and is assumed to be a stationary AR(1) process

$$h_t = \phi_h h_{t-1} + \varepsilon_t^h,\ \varepsilon_t^h \sim \mathcal{N}(0, \sigma_h^2),$$

where $|\phi_h| < 1$ and the log-volatility process is initialized by a stationary distribution, $h_1 \sim \mathcal{N}\left(0, \frac{\sigma_h^2}{1-\phi_h^2}\right)$.

The second variant we consider is a VAR model with its error terms featured by both SV and MA components (VAR-MASV), that is,

$$\boldsymbol{\varepsilon}_t^y = \mathbf{u}_t + \psi\mathbf{u}_{t-1},\ \mathbf{u}_t \sim \mathcal{N}(0, e^{h_t}\boldsymbol{\Sigma}),$$

where ψ is the MA coefficient and $|\psi| < 1$ to ensure the invertibility condition.

In addition, we also consider the variants of our VAR-SV and VAR-MASV with t-distributed error terms:

- t-SV: $\boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \lambda_t e^{h_t}\boldsymbol{\Sigma})$, $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$;
- t-MASV: $\boldsymbol{\varepsilon}_t^y = \mathbf{u}_t + \psi \quad \mathbf{u}_{t-1}$, $\mathbf{u}_t \sim \mathcal{N}(0, \lambda_t e^{h_t}\boldsymbol{\Sigma})$, $\lambda_t \sim \mathcal{IG}$ $(\nu/2, \nu/2)$.

Overall, our forecasting exercise includes 19 univariate and multivariate models with flexible error structure. We summarize all our competing models in Table 1.

TABLE 1 A list of competing models.

Model	Description
UC	Unobserved component model with homoscedastic errors
UC-SV	UC model with stochastic volatility errors
UC-MASV	UC model with moving average MA(1) and stochastic volatility errors
UC-ARMASV	UC model with autoregressive moving average ARMA(1,1) and stochastic volatility errors
UCt-SV	UC-SV with t-distributed errors
UCt-MASV	UC-MASV with t-distributed errors
UCt-ARMASV	UC-ARMASV with t-distributed errors
AR(m)	Autoregressive model with homoscedastic errors
AR(m)-SV	AR model with stochastic volatility errors
AR(m)-MASV	AR model with moving average and stochastic volatility errors
AR(m)-ARMASV	AR model with autoregressive moving average and stochastic volatility errors
ARt(m)-SV	AR(m)-SV with t-distributed errors
ARt(m)-MASV	AR(m)-MASV with t-distributed errors
ARt(m)-ARMASV	AR(m)-ARMASV with t-distributed errors
VAR	Vector autoregressive model
VAR-SV	VAR with stochastic volatility errors with homoscedastic errors
VAR-MASV	VAR with moving average and stochastic volatility errors
VARt-SV	VAR-SV with t-distributed errors
VARt-MASV	VAR-MASV with t-distributed errors

3 | PRIORS AND ESTIMATION

To complete the model specifications, this section presents the prior choices for the model parameters discussed in the previous section. For an easy comparison, we use the same priors for the common parameters across the models.

For the AR models, we use a relatively non-informative Gaussian priors for the AR coefficients $\boldsymbol{\beta} \sim \mathcal{N}(0, \mathbf{V}_{\boldsymbol{\beta}})$ with $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_m)'$ and $\mathbf{V}_{\boldsymbol{\beta}} = 10\mathbf{I}_{m+1}$. For the UC models, the priors of the initial condition τ_1 and the variance σ_τ^2 are

$$\tau_1 \sim \mathcal{N}(0, V_\tau), \; \sigma_\tau^2 \sim \mathcal{IG}(\alpha_\tau, \eta_\tau).$$

We set $V_\tau = 10$, the shape parameter $\alpha_\tau = 10$, and the scale parameter $\eta_\tau = 0.2(\alpha_\tau - 1)$, which implies that the prior mean of σ_τ^2 is 0.2. The chosen prior mean reflects the desired smoothness of the corresponding state τ_t associated with the conditional mean process.

The variance σ_y^2 in Equation (1) is assumed to have an inverse-Gamma prior, that is, $\sigma_y^2 \sim \mathcal{IG}(\alpha_y, \eta_y)$ with $\alpha_y = 10$ and $\eta_y = 9$, which implies that the prior mean of σ_y^2 is equal to 1. This choice is comparable for the models with SV where the prior mean for μ_h, the unconditional mean of log volatility, is $\mu_h = 0$ ($\exp(\mu_h) = 1$).

For the SV component described in Equation (4), we use the following priors for the parameters in the log-volatility process:

$$\phi_h \sim \mathcal{N}(\phi_{h0}, V_{\phi_h})\mathbf{1}(|\phi_h| < 1), \; \sigma_h^2 \sim \mathcal{IG}(\alpha_h, \eta_h),$$

and we set $\phi_h = 0.95, V_{\phi_h} = 0.4^2, \alpha_h = 10$ and $\eta_y = 0.01(\alpha_h - 1)$. With this setup, the variance of innovations, that is, σ_h^2, in the log-volatility process is assumed to have a prior mean of 0.01.

For the MA component stated in Equation (5), we impose a truncated Gaussian prior for the MA coefficient ψ:

$$\psi \sim \mathcal{N}(0, V_\psi)\mathbf{1}(|\psi| < 1)$$

with the prior variance set to $V_\psi = 10$.

Similarly, we impose a truncated Gaussian prior for the AR coefficient ϕ for the ARMA component in Equation (6)

$$\phi \sim \mathcal{N}(0, V_\phi)\mathbf{1}(|\phi| < 1)$$

and set $V_\phi = 10$.

The degrees of freedom in the t-distributed error are treated as an unknown parameter and assumed to follow a uniform prior:

$$\nu \sim \mathcal{U}(2, 100),$$

where ν is restricted to be greater than 2 to ensure that the first and second moments of the t-distribution exists.

Next, we present the priors for the VAR models. To facilitate our discussion, let $\boldsymbol{\theta} = \text{vec}([\mathbf{b}, \mathbf{B}_1, ...\mathbf{B}_p]')$. We impose a natural conjugate prior, that is, Gaussian-inverse-Wishart prior, for the VAR coefficients and the covariance matrix:

$$(\boldsymbol{\theta}|\boldsymbol{\Sigma}) \sim \mathcal{N}(\text{vec}(\boldsymbol{\Phi}_0), \boldsymbol{\Sigma} \otimes \mathbf{V}_{\boldsymbol{\theta}}), \; \boldsymbol{\Sigma} \sim \mathcal{IW}(\nu_0, \mathbf{S}_0),$$

where $\boldsymbol{\Phi}_0 = \mathbf{0}, \nu_0 = n + 4$, and $\mathbf{S}_0 = \mathbf{I}_n$. The covariance matrix $\mathbf{V}_{\boldsymbol{\theta}}$ is assumed to be diagonal in which we set 100 for the intercepts and the elements associated with the lags l $(l = 1..., p)$ of variable i $(i = 1, ..., n)$ to be $\kappa / l^2 s_i^2$ where s_i^2 is the residual variance of AR(p) model and the shrinkage $\kappa = 0.2^2$. These hyperparameter choices are similar to those used in Carriero et al. (2016) and Chan (2020).

All models considered in this paper are estimated via the use of Markov chain Monte Carlo (MCMC) methods. As the univariate UCt-ARMASV is the most general model considered in this paper, we provide the details of the MCMC sampler for estimating the UCt-ARMASV model in Appendix S3. For the VAR models, we implement the algorithm developed by Chan (2020) to efficiently sample the VAR parameters. The algorithm is built upon the precision sampler of Chan and Jeliazkov (2009) and exploits the Kronecker covariance and sparse matrix structure in the error terms. We refer readers to Chan (2020) for further details.

4 | APPLICATION TO MACROECONOMIC SERIES

4.1 | Data

In this paper, we consider 20 macroeconomic variables that serve as important indicators for both nominal and real economic activities and that are commonly studied by macroeconomists (Bauwens et al., 2015). The data are obtained from the Federal Reserve Bank of Philadelphia and the Federal Reserve Bank of St. Louis. The descriptions of the macroeconomic variables, along with their transformations, are displayed in Table 2. Our forecasting analysis includes macroeconomic series belonging to the following five categories:

A. *Income, output, sales, and capacity utilization*;
B. *Employment and unemployment*;
C. *Construction, inventories, and orders*;

TABLE 2 Description of the data.

Acronym	Trans	Vintage	Series	Description
A. Income, output, sales, and capacity utilization				
1.RGDP	400ΔLog	Quarterly	Quarterly	Real Gross Domestic Product
2.RCON	400ΔLog	Quarterly	Quarterly	Real Personal Consumption Expenditures: Total
3.NPI	400ΔLog	Quarterly	Quarterly	Nominal Personal Income
4.IPT	400ΔLog	Monthly	Monthly	Industrial Production Index: Total
5.OPH	400ΔLog	Quarterly	Quarterly	Business Sector: Output Per Hour of All Persons
B. Employment and unemployment				
6.RUC	No	Quarterly	Monthly	Real Unemployment Rate
7.EMPL	400ΔLog	Monthly	Monthly	Nonfarm Payroll Employment
8.HOURS	400ΔLog	Monthly	Monthly	Indexes of Aggregate Weekly Hours: Total
C. Construction, inventories, and orders				
9.RINVN	400ΔLog	Quarterly	Quarterly	Real Gross Private Domestic Investment: Nonresidential
10.RINVR	400ΔLog	Quarterly	Quarterly	Real Gross Private Domestic Investment: Residential
11.HSTAR	400ΔLog	Monthly	Monthly	Housing Starts
D. Nominal prices, wages, and money				
12.PCON	400ΔLog	Quarterly	Quarterly	Price Index for Personal Consumption Expenditures, Constructed
13.PIMP	400ΔLog	Quarterly	Quarterly	Price Index for Imports of Goods and Services
14.M1	400ΔLog	Quarterly	Monthly	M1 Money Stock
15.M2	400ΔLog	Quarterly	Monthly	M2 Money Stock
16.PPPI	400ΔLog	Monthly	Monthly	Producer Price Index: Finished Consumer Goods
E. Interest rates and asset prices				
17.FFUND	No	-	Monthly	Effective Federal Funds Rate
18.GS1	No	-	Monthly	1-Year Treasury Constant Maturity Rate
19.GS10	No	-	Monthly	10-Year Treasury Constant Maturity Rate
20.BAAFR	No	-	Monthly	Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate

Note: The second column refers to the transformation methods (Trans): 400ΔLog = the growth rate after the first difference of logged variables.

D. *Nominal prices, wages, and money*;
E. *Interest rates and asset prices*.

The real-time dataset for the first four categories A–D compromises 16 macroeconomic variables that are revised on a monthly or a quarterly basis. It is worth noting that the real-time data released at vintage time t contain observations up to time $t-1$ due to reporting lag issues. For example, the 2022Q4 vintage covers estimates and revisions from 1964Q1 to 2022Q3. The last category E, containing four variables, is not subject to revisions. All monthly data are converted to quarterly data by calculating their monthly average within the given quarters. For instance, we calculate the quarterly data of 2021Q1 as an average of monthly data in 2021M1–2021M3 and the quarterly data of 2021Q2 as an average of monthly data in 2021M4–2021M6 and so forth. To obtain the quarterly vintage for the time series with the monthly data vintage, we use the final month vintage of a quarter as the quarterly vintage. For instance, observations in the 2022Q4 vintage are obtained from those in the 2022M12 vintage. To this end, this paper contains the vintage ranging from 1991Q1 to 2022Q4, with each vintage including observations dating back to 1964Q1. The vintage range from 1991Q1 to 2022Q4 is chosen based on their availability. We also summarize the data treatment process in the graphical presentation in Figure 1.

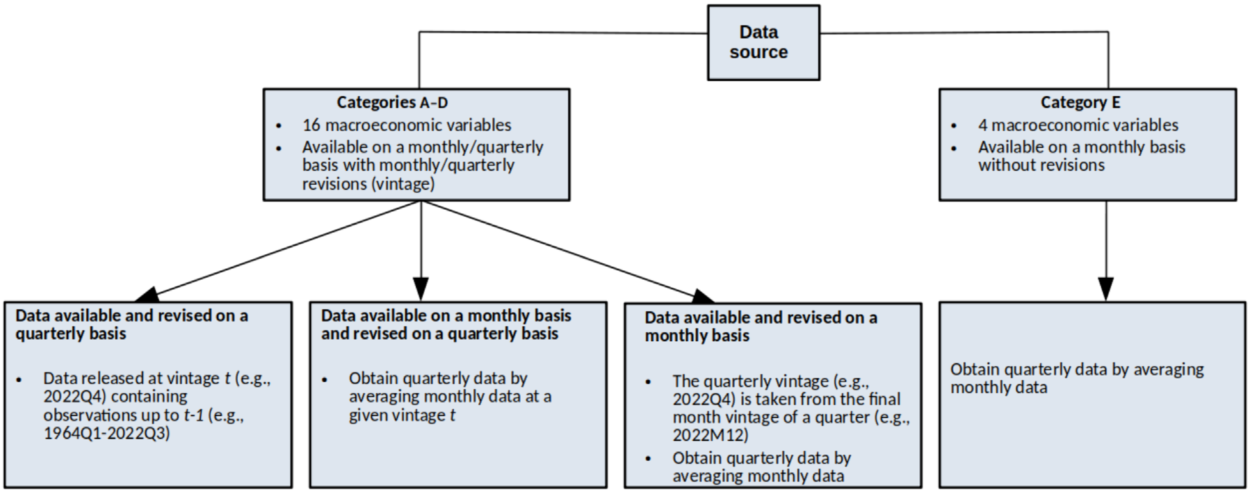

FIGURE 1 Graphical presentation of data treatment process.

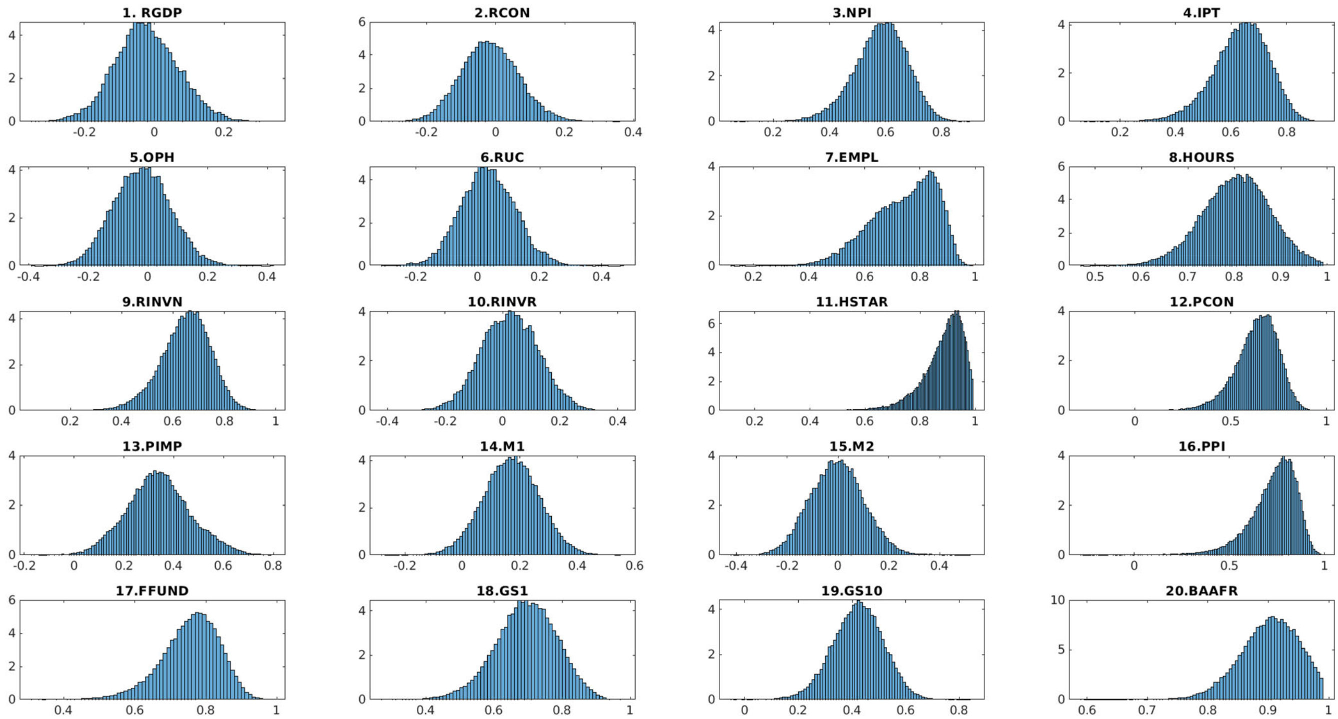

FIGURE 2 Posterior distributions for ϕ under the UCt-ARMASV model.

4.2 | Full-sample estimation results

Before we discuss the forecasting analysis, in this section, we present the full-sample estimation results of the UCt-ARMASV model for highlighting the importance of allowing SV and serial correlation in an error process. All estimates reported in this section are based on 50,000 draws from the posterior distribution after a burn-in period of 10,000 using data in the 2022Q4 vintage.

The posterior distributions displayed in Figures 2 and 3 reveal that most of the probability mass of the posterior distributions for the AR parameter ϕ and/or the MA parameter ψ are concentrated around regions that are far from zero, with the exception of real gross domestic product (**1.RGDP**), real personal consumption expenditures (**2.RCON**), output per hour of all persons in business sectors (**5.OPH**), real unemployment rate (**6.RUC**), real gross private domestic investment–residential (**10.**

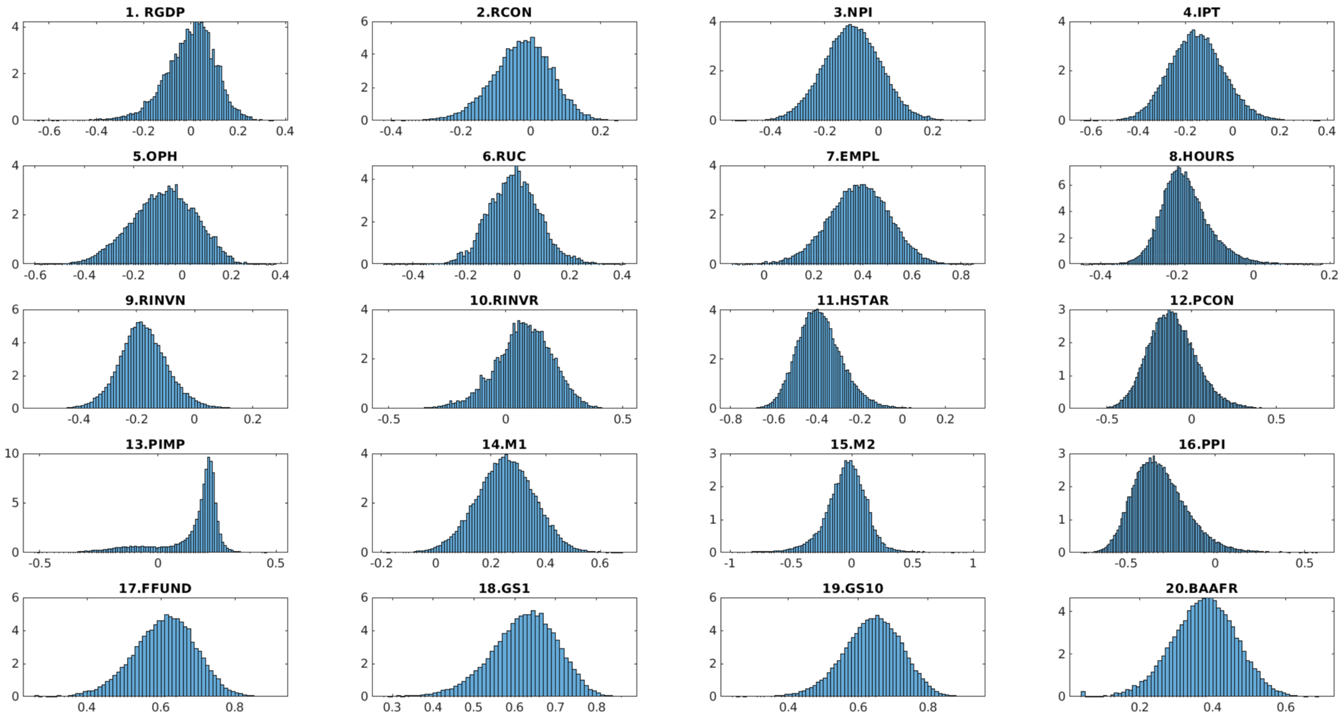

FIGURE 3 Posterior distributions for ψ under the UCt-ARMASV model.

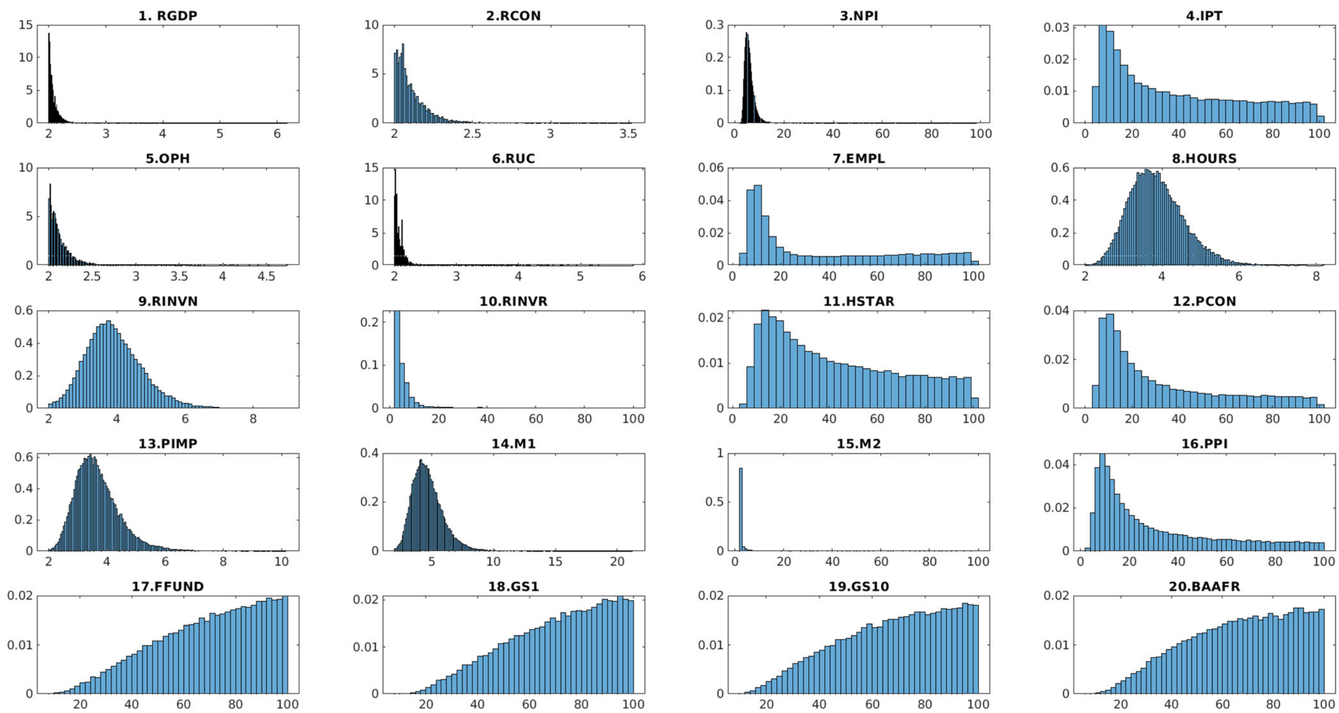

FIGURE 4 Posterior distributions for degrees of freedom ν under the UCt-ARMASV model.

RINVR), and M2 money stock (**15.M2**). For these specific variables, the heavy-tailed distribution feature is more pronounced than the serial correlation. Figure 4 highlights that the posterior means of the degrees of freedom are small (less than 10) for the aforementioned variables. However, for some variables, such as aggregate weekly hours (**8.HOURS**), real gross private domestic investment–nonresidential (**9.RINVN**), and price index

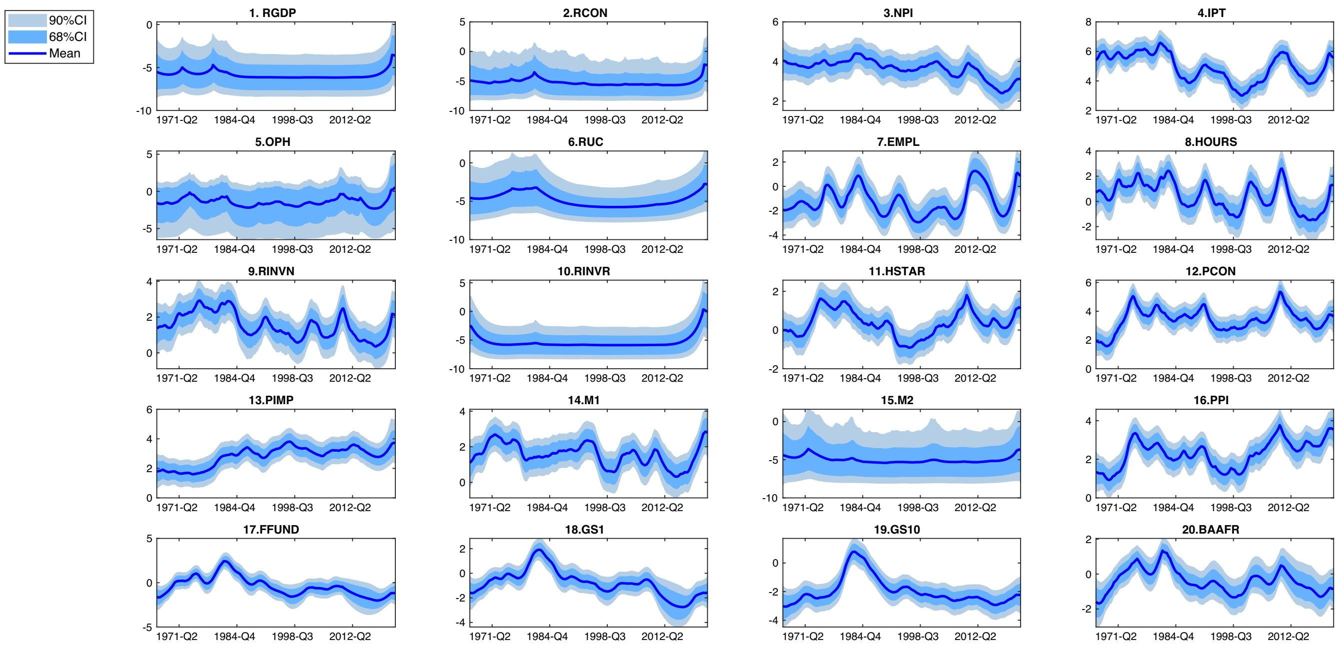

FIGURE 5 The estimated log volatilities under the UCt-ARMASV model. The solid lines are posterior means, while the shaded areas correspond to 68% and 90% posterior credible intervals.

for imports of goods and services (**13.PIMP**), the data indicate the presence of both serial dependent and non-Gaussian error processes.

Figure 5 presents the estimated posterior means and the 68% and 90% credible intervals for the log volatilities under the UCt-ARMASV model. The data strongly support the presence of time variation in volatilities across most macroeconomic variables. This is in line with numerous empirical macroeconomic studies including Cogley and Sargent (2005), Primiceri (2005), and Carriero et al. (2016). We note that some variables such as **1.RGDP**, **2.RCON**, **6.RUC**, **10.RINVR**, and **15.M2** exhibit less time variation in log volatilities than the others. An explanation for this is that the heavy-tailed distribution might capture most of the variations in these specific variables.

4.3 | Model comparison

We next conduct a formal Bayesian model comparison exercise by using the marginal likelihood as our model comparison criterion to examine if allowing for more flexible error structures improves in-sample fit of the data. The marginal likelihood of a model can be expressed as the product of one-step-ahead predictive likelihood over the full sample; thus, it is also closely related to the one-step-ahead density forecasting performance of the given model. Further discussion of the connection between marginal likelihood and predictive likelihood can be found in Geweke and Amisano (2010).

To save space, Table 3 presents the log marginal likelihoods of the competing models listed in Table 1 relative to the log marginal likelihoods of the benchmark AR models for some selected variables. Positive values of the relative marginal likelihoods indicate that the data are in favor of the competing models over the benchmark model. The results for the other variables are reported in Tables A2 and A3 in Appendix S1. Tables 3, A2, and A3 uncover overwhelming empirical evidence suggesting that the models with more general specifications in the error terms, such as heavy tail, SV, AR, and MA, substantially improve the in-sample fit over the models with constant variances, that is, **AR** and **UC**. For example, the UCt-ARMASV model provides the best in-sample fit for **1.RGDP**, **6.RUC**, **9.RINVN**, and **19.GS10**, while the UC-MASV model performs the best for **15.M2**.

5 | FORECASTING RESULTS

This section conducts a pseudo-out-of-sample forecasting exercise in which we compare both point and density forecasting performance of those candidate models listed in Table 1. In each exercise, we divide the data into three sub-samples. The first part is the *initialization period*, which contains the first $q = \max\{m,p\}$ observations, where m and p are the lag orders of the AR and VAR models. Following the standard practice, we set the lag orders of the VAR models equal to 4 for quarterly data. The lag orders for the AR models are selected based on BIC criterion (see Table A1 in Appendix S1 for more details)[2]. The first q observations, for example, $q=4$, 1964Q1–1964Q4, are therefore used to initialize the AR(m) and VAR(p) models. This is to ensure that all AR,

TABLE 3 Log marginal likelihoods of the competing models relative to log marginal likelihood of the benchmark model AR for some selected variables.

Model	1.RGDP	2.RCON	6.RUC	9.RINVN	15.M2	19.GS10
AR-SV	27.4	29.8	39.0	39.8	6.1	34.9
ARMASV	28.2	30.4	37.1	39.1	7.7	36.0
AR-ARMASV	24.0	26.5	28.7	30.9	5.8	37.5
ARt-SV	27.9	30.0	40.0	40.9	6.0	35.3
ARt-MASV	28.7	30.8	38.2	40.1	7.7	36.5
ARt-ARMASV	25.7	27.6	33.4	33.2	6.0	36.6
UC	−13.1	−12.1	−37.3	−26.8	1.7	−74.6
UC-SV	16.8	24.4	0.7	21.4	14.4	24.9
UC-MASV	28.4	31.1	28.3	24.7	**15.2**	33.8
UC-ARMASV	**34.7**	**36.5**	**42.0**	44.5	15.1	40.4
UCt-SV	16.6	23.8	−1.0	21.1	14.2	24.9
UCt-MASV	28.3	30.6	27.8	24.5	15.1	33.8
UCt-ARMASV	**34.7**	36.0	**42.0**	**44.8**	15.0	**41.0**

Note: Values in bold indicate the best relative log marginal likelihoods among all the competing models for each variable.

TABLE 4 Proportion of time a model achieves the most accurate forecast based on the relative RMSFE and relative MAFE measures for 20 macroeconomic variables given an evaluation period from 1999Q1 to 2022Q2.

	Relative RMSFE				Relative MAFE			
Model	$k=1$	$k=2$	$k=4$	$k=8$	$k=1$	$k=2$	$k=4$	$k=8$
AR	10%	0%	5%	15%	5%	0%	15%	15%
AR-SV	15%	15%	5%	0%	15%	15%	0%	0%
ARMASV	5%	0%	5%	0%	10%	0%	5%	0%
AR-ARMASV	15%	5%	5%	20%	0%	5%	0%	20%
ARt-SV	5%	10%	15%	0%	5%	5%	5%	0%
ARt-MASV	0%	0%	0%	0%	0%	5%	10%	10%
ARt-ARMASV	0%	20%	15%	5%	5%	20%	5%	5%
UC	0%	5%	0%	0%	0%	5%	0%	10%
UC-SV	15%	0%	10%	5%	30%	0%	5%	5%
UC-MASV	0%	5%	15%	20%	0%	5%	10%	5%
UC-ARMASV	10%	10%	0%	5%	10%	5%	0%	0%
UCt-SV	15%	0%	5%	5%	0%	10%	10%	5%
UCt-MASV	0%	0%	0%	0%	0%	0%	10%	0%
UCt-ARMASV	0%	10%	10%	10%	15%	10%	5%	0%
VAR	5%	15%	10%	5%	0%	10%	20%	5%
VAR-SV	0%	0%	0%	0%	0%	5%	0%	5%
VAR-MASV	0%	5%	0%	5%	0%	0%	0%	5%
VARt-SV	0%	0%	0%	0%	0%	0%	0%	0%
VARt-MASV	5%	0%	0%	5%	5%	0%	0%	10%
AR group	50%	50%	50%	40%	40%	50%	40%	50%
UC group	40%	30%	40%	45%	55%	35%	40%	25%
VAR group	10%	20%	10%	15%	5%	15%	20%	25%

UC, and VAR model variants have the same initial observations. The second part is the *training period*, denoted as $\mathbf{y}^{T_0}_{1965Q1:T_0-1} = \{y^{T_0}_{1965Q1}, y^{T_0}_{1965Q2} \ldots, y^{T_0}_{T_0-2}, y^{T_0}_{T_0-1}\}$, where the superscript T_0 refers to a vintage, and the subscript refers to data observations ranging from 1965Q1 to $T_0 - 1$. The data vector $\mathbf{y}^{T_0}_{1965Q1:T_0-1} = \{y^{T_0}_{1965Q1}, y^{T_0}_{1965Q2} \ldots, y^{T_0}_{T_0-2}, y^{T_0}_{T_0-1}\}$ comprises the first estimate of y_{T_0-1}, the first revision (i.e., the second estimate) of y_{T_0-2}. We start the training period from $\mathbf{y}^{1999Q1}_{1965Q1:1998Q4}$ and then recursively produce forecast estimates with the expanding training period. The parameter estimates obtained from the training period are used to conduct k-step-ahead forecasts, denoted as $\mathbb{E}(y_{T_0-1+k}|\mathbf{y}^{T_0}_{1965Q1:T_0-1})$. We report the forecasting results for the macroeconomic variables over horizons ranging from one step ahead ($k=1$) to eight steps ahead ($k=8$). The third part is the *evaluation period* (1999Q1–2022Q2). To evaluate the forecast performance of the models, we compare the predicted values, $\mathbb{E}(y_{T_0-1+k}|\mathbf{y}^{T_0}_{1965Q1:T_0-1})$, to the first revised value of y_{T_0-1+k} which is observed in vintage T_0+k+1, denoted as $y^{T_0+k+1}_{T_0+k-1}$. The forecasting setup is motivated by a number of studies that consider the presence of data revisions in forecasting exercises such as Croushore (2011), Clements and Galvão (2013), and Clements (2017). Clements and Galvão (2013) and Clements (2017) highlight that using dataset from a single vintage (e.g., the latest vintage) might lead to an inaccurate assessment. One reason for this is that most of the macroeconomics variables are subject to data revisions and these revisions are not small and random. To alleviate this potential problem, we consider a real-time database that includes all possible data vintages in the forecasting exercise. All forecast estimates

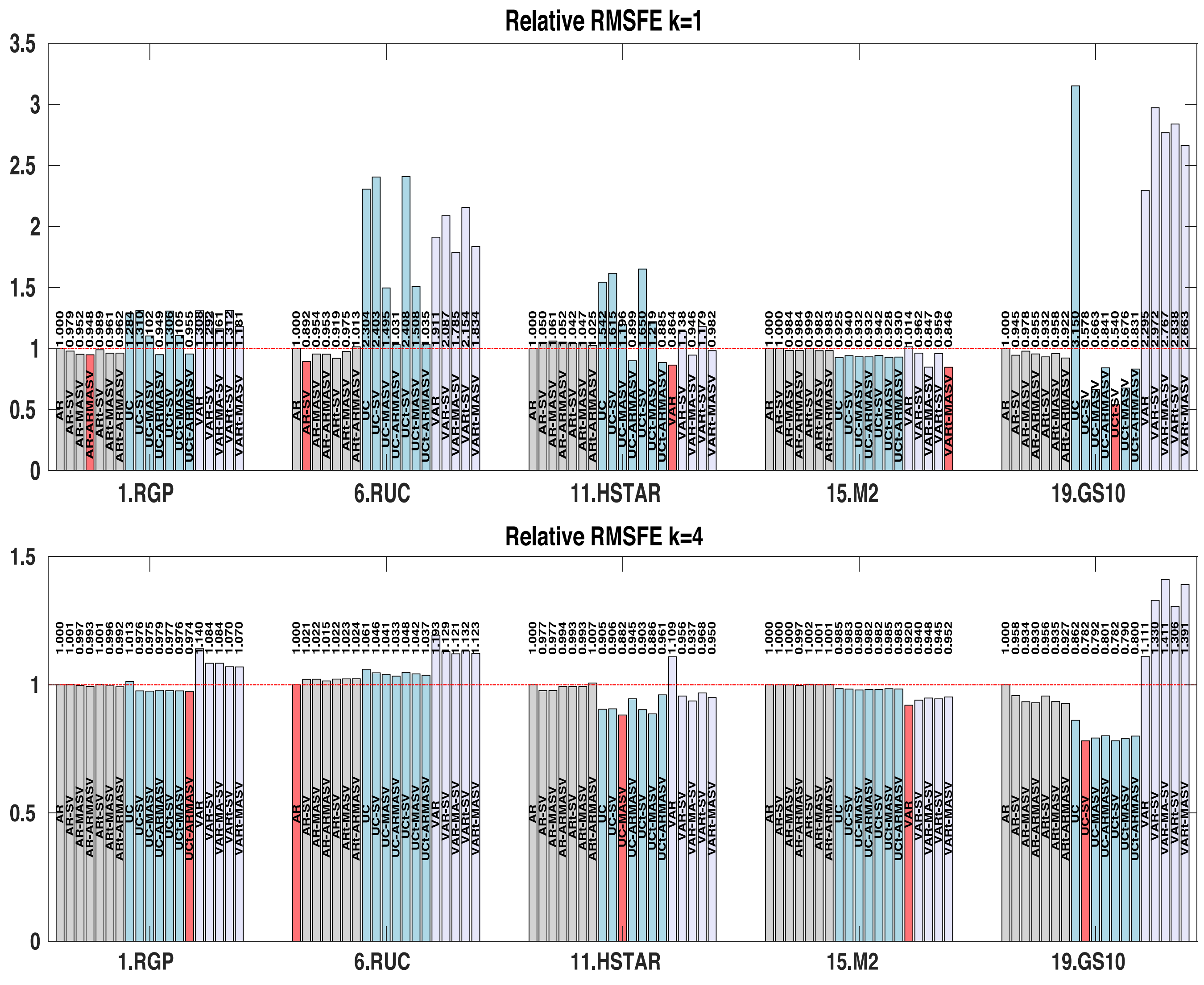

FIGURE 6 Relative RMSFE for selected variables at $k=1$ and $k=4$. The light gray, blue, and purple colors represent the model groups: light gray (AR group), light blue (UC group), and light purple (VAR group). The red color indicates the best model performance.

are based on 50,000 draws with a burn-in period of 10,000, which is the same as the full-sample study.

5.1 | Forecast evaluation methods

To compare the forecasting performance of the models listed in Table 1, we consider a range of forecast measures, including point forecast and density forecast measures. The first point forecast measure is the RMSFE. Specifically, RMSFE for a variable i at forecast horizon k is defined as

$$\mathrm{RMSFE}_{i,k} = \sqrt{\frac{1}{T-k-T_0}\sum_{t=T_0}^{T-k-1}\left(y_{i,t+k-1}^{o,t+k+1} - \mathbb{E}(y_{i,t+k-1}|\mathbf{y}_{1:t-1}^{t})\right)^2},$$

where T_0 denotes the starting period of the evaluation periods, $\mathbb{E}(y_{i,t+k-1}|\mathbf{y}_{1:t-1}^{t})$ is the posterior mean of k-step-ahead forecasts of variable i, given the data information up to time $t-1$ in vintage t, and $y_{i,t+k-1}^{o,t+k+1}$ is the first revised value of $y_{i,t+k-1}$, which is observed at vintage $t+k+1$.

Another widely used point forecast measure is the MAFE, which measures an absolute value between the actual values and the forecasts. MAFE for a variable i is defined as

$$\mathrm{MAFE}_{i,k} = \frac{1}{T-k-T_0}\sum_{t=T_0}^{T-k-1}\left|y_{i,t+k-1}^{o,t+k+1} - \mathbb{E}(y_{i,t+k-1}|\mathbf{y}_{1:t-1}^{t})\right|.$$

For an easy comparison, we report the RMSFE and MAFE of the competing models relative to the

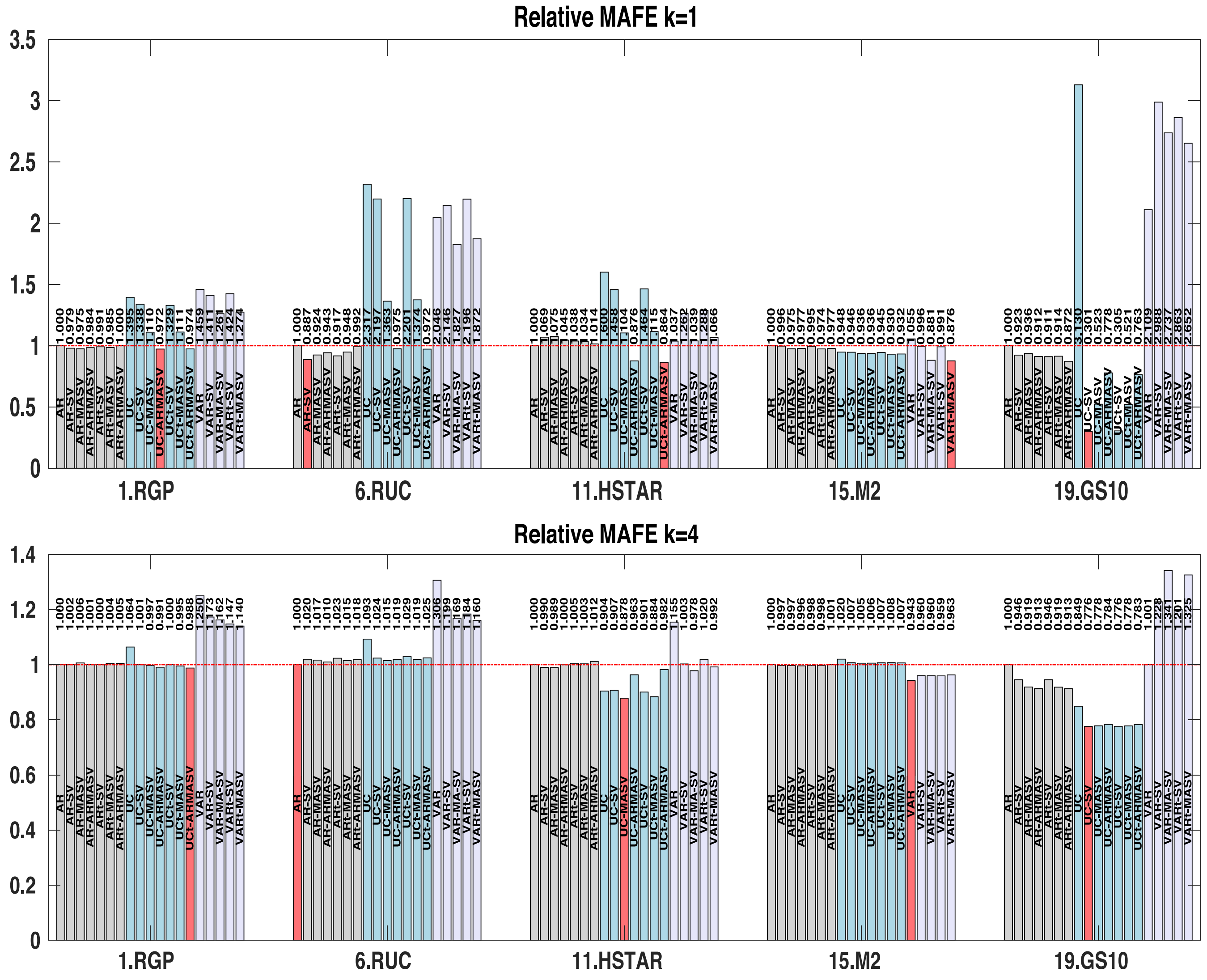

FIGURE 7 Relative MAFE for selected variables at $k=1$ and $k=4$. The light gray, blue, and purple colors represent the model groups: light gray (AR group), light blue (UC group), and light purple (VAR group). The red color indicates the best model performance.

benchmark model, that is, the AR model, to obtain the relative RMSFE and relative MAFE. If the relative RMSFE/MAFE measures are smaller than one, it implies that the competing models provide better point forecasts than the benchmark AR and vice versa.

To evaluate density forecasting performance, we consider the average of log-predictive likelihood (ALPL), which is defined as

$$\mathrm{ALPL}_{i,k} = \frac{1}{T-k-T_0}\sum_{t=T_0}^{T-k-1} \log p(y_{i,t+k-1} = y_{i,t+k-1}^{o,t+k+1} | \mathbf{y}_{1:t-1}^{t}),$$

where $p(y_{i,t+k-1} = y_{i,t+k-1}^{o,t+k+1} | \mathbf{y}_{1:t-1}^{t})$ is the k-step-ahead predictive likelihood of a variable i up to time $t-1$ evaluated at the first revised value $y_{i,t+k-1}^{o,t+k+1}$, which is observed at vintage $t+k+1$. For an easy comparison, we report the relative ALPLs, calculated as the difference of the log-predictive likelihoods of the given models and the benchmark AR model. Thus, a positive relative ALPL value indicates that the competing model has a better density forecasting performance than the AR model.

Following Gneiting and Raftery (2007), we evaluate the density forecast by considering the average CRPS (ACRPS), which is given by

$$\mathrm{ACRPS}_{i,k} = \frac{1}{T-k-T_0}\sum_{t=T_0}^{T-k-1} \mathrm{CRPS}(F_{t+k-1, y_{i,t+k-1}^{o,t+k+1}}),$$

where $\mathrm{CRPS}(F_{t+k-1, y_{t+k-1}^{o,t+k+1}}) = \int_{-\infty}^{\infty} (F_{t+k-1}(z) - \mathbf{1}(y_{i,t+k-1}^{o,t+k+1} < z))^2 dz = \mathbb{E}|y'_{i,t+k-1} - y_{i,t+k-1}^{o,t+k+1}| - 0.5\mathbb{E}|y'_{i,t+k-1} - y_{i,t+k-1}^{o,t+k+1}|$ with $F_{t-1+k}(.)$ is the cumulative distribution function of the predictive density, $y'_{i,t+k-1}$ and $y''_{i,t+k-1}$ are independent draws from the posterior predictive distribution.

We also evaluate the tail forecast accuracy using the qwCRPS developed by Gneiting and Ranjan (2011) as a density forecast measure. The average qwCRPS (AqwCRPS) is computed as a weighted sum of quantile scores at a range of $J-1$ quantiles:

$$\begin{aligned}
\mathrm{AqwCRPS}_{i,k} &= \frac{1}{T-T_0-k}\sum_{t=T_0}^{T-k-1} \frac{2}{J-1}\sum_{j=1}^{J-1} \omega(\tau_j) \mathrm{QS}_{i,\tau_j,t-1+k},\\
\mathrm{QS}_{i,\tau_j,t-1+k} &= (\mathbf{1}(y_{i,t-1+k}^{o} \le \mathcal{Q}_{i,\tau,t-1+k}) - \tau_j)(\mathcal{Q}_{i,\tau,t} - y_{i,t-1+k}^{o}),
\end{aligned}$$

TABLE 5 Proportion of time a model achieves the most accurate forecast based on the relative ALPL and relative ACRPS measures for 20 macroeconomic variables given an evaluation period from 1999Q1 to 2022Q2.

	Relative ALPL				Relative ACRPS			
Model	$k=1$	$k=2$	$k=4$	$k=8$	$k=1$	$k=2$	$k=4$	$k=8$
AR	5%	0%	0%	10%	5%	0%	0%	0%
AR-SV	0%	0%	5%	0%	10%	15%	5%	0%
ARMASV	0%	0%	5%	0%	0%	0%	0%	0%
AR-ARMASV	0%	0%	0%	0%	0%	5%	35%	30%
ARt-SV	5%	25%	10%	0%	20%	10%	0%	0%
ARt-MASV	10%	10%	0%	0%	0%	15%	5%	0%
ARt-ARMASV	0%	0%	15%	40%	10%	20%	10%	15%
UC	0%	0%	0%	5%	0%	0%	0%	5%
UC-SV	0%	5%	5%	10%	0%	5%	10%	5%
UC-MASV	0%	5%	10%	5%	0%	0%	0%	5%
UC-ARMASV	5%	10%	5%	0%	10%	5%	10%	15%
UCt-SV	0%	5%	10%	0%	0%	0%	5%	5%
UCt-MASV	0%	5%	10%	5%	0%	0%	0%	0%
UCt-ARMASV	70%	30%	5%	10%	35%	15%	10%	10%
VAR	0%	5%	15%	10%	0%	5%	10%	0%
VAR-SV	0%	0%	0%	0%	0%	0%	0%	0%
VAR-MASV	0%	0%	0%	0%	5%	0%	0%	0%
VARt-SV	0%	0%	0%	0%	0%	5%	0%	0%
VARt-MASV	5%	0%	5%	5%	5%	0%	0%	10%
AR group	20%	35%	35%	50%	45%	65%	55%	45%
UC group	75%	60%	45%	35%	45%	25%	35%	45%
VAR group	5%	5%	20%	15%	10%	10%	10%	10%

with $\tau_j = j/J$. We rely on a grid of $J-1=19$ quantiles $\tau_i \in \{0.05, 0.1, \ldots, 0.90, 0.95\}$ to compute these weighted scores. $\mathcal{Q}_{i,\tau_j,t-1+k}$ is the forecast quantile at quantile τ_jth, and $\mathbf{1}(y^o_{i,t-1+k} \leq \mathcal{Q}_{i,\tau,t-1+k})$ is an indicator function that has a value of one if the observed value is at or below the forecast quantile and zero otherwise. We set the weight $\omega(\tau_j) = \tau_j(1-\tau_j)$ to target the center of the predictive distribution (denoted as AqwCRPS-center) and $\omega(\tau_j) = (2\tau_j - 1)^2$ to target the right and left tails of the predictive distribution (denoted as AqwCRPS-tail). A model with lower ACRPS and AqwCRPS indicates better forecasting performance. We report the density forecast measures of ACRPS and AqwCRPS relative to that of the benchmark AR model. If a value of the relative ACRPS/AqwCRPS is less than one, this indicates that the competing model has better forecast accuracy than the benchmark.

The equal predictive accuracy of forecasting comparison is also conducted by a one-sided sign test introduced by Diebold and Mariano (2002). The Diebold and Mariano (2002) test compares a competing model with the benchmark AR model for a given forecasting horizon at the level of significance of 5%, corresponding with one asterisk. We also conduct the model confidence set (MCS) test of Hansen et al. (2011). The results contain 90% MCSs. The Diebold and Mariano (2002) and Hansen et al. (2011) results are presented in Tables A10–A21 in Appendix S1.

5.2 | Point and density forecast results

Considering the significant number of variables (20) and 19 candidates models involved in the study, we select a representative subset of variables from groups A (**1.GDP**), B (**6.RUC**), C (**11.HSTAR**), D (**15.M2**), and E (**19.GS10**) to illustrate the point and density forecast results. More detailed forecast performance for the other variables can be found in Appendix S2.

5.2.1 | Point forecast results

Table 4 provides a comprehensive view of the performance of all the models across different forecast

TABLE 6 Proportion of time a model achieves the most accurate forecast based on the relative AqwCRPS-center and relative AqwCRPS-tail measures for 20 macroeconomic variables given an evaluation period from 1999Q1 to 2022Q2.

	Relative AqwCRPS-center				Relative AqwCRPS-tail			
Model	$k=1$	$k=2$	$k=4$	$k=8$	$k=1$	$k=2$	$k=4$	$k=8$
AR	5%	0%	0%	5%	5%	0%	0%	0%
AR-SV	5%	5%	0%	0%	0%	0%	0%	0%
ARMASV	0%	0%	5%	0%	0%	0%	0%	0%
AR-ARMASV	0%	10%	0%	30%	0%	0%	0%	15%
ARt-SV	10%	20%	15%	0%	25%	35%	15%	0%
ARt-MASV	0%	0%	5%	0%	0%	10%	0%	0%
ARt-ARMASV	10%	5%	15%	10%	5%	5%	10%	20%
UC	0%	5%	0%	5%	0%	0%	5%	5%
UC-SV	5%	0%	5%	5%	0%	0%	0%	0%
UC-MASV	0%	5%	10%	10%	0%	10%	5%	5%
UC-ARMASV	25%	15%	0%	0%	15%	10%	10%	10%
UCt-SV	0%	0%	5%	5%	0%	0%	5%	0%
UCt-MASV	0%	0%	0%	5%	0%	0%	15%	10%
UCt-ARMASV	35%	25%	20%	0%	40%	15%	10%	0%
VAR	0%	10%	20%	5%	0%	10%	25%	10%
VAR-SV	0%	0%	0%	0%	0%	0%	0%	0%
VAR-MASV	0%	0%	0%	5%	5%	5%	0%	5%
VARt-SV	0%	0%	0%	0%	0%	0%	0%	5%
VARt-MASV	5%	0%	0%	15%	5%	0%	0%	15%
AR group	30%	40%	40%	45%	35%	50%	25%	35%
UC group	65%	50%	40%	30%	55%	35%	50%	30%
VAR group	5%	10%	20%	25%	10%	15%	25%	35%

horizons, based on their relative RMSFE and MAFE. It summarizes the percentage of time that a model achieves the most accurate forecast. The results suggest that there is no clear winner among the models for the macroeconomic variables. However, the univariate models (the AR and UC groups) with flexible error structures provide more accurate forecasts than those with homoscedastic errors. For example, based on the relative RMSFE measures, the univariate models with flexible error structures, that is, SV, MASV, ARMASV, and t-SV, provide the best forecast for 16 out of 20 variables at $k=1$, equivalent to a success rate of 80%. On the other hand, the benchmark (AR) with a homoscedastic error performs best in 10% of the considered variables.

When looking more closely at Figures 6 and 7 and Figures B1–B6 in Appendix S2, we gain additional insights into the forecasting performance of each competing model relative to the benchmark, the AR model. We refer the reader to Appendix S1 for the Diebold and Mariano (2002) and Hansen et al. (2011) tests. The main conclusions obtained from the figures are as follows. For certain variables, the best forecasting model remains consistent across the point forecast measures. However, for some other variables, different point forecast measures suggest different models as optimal. For example, both the relative RMSFE and MAFE measures indicate that the UCt-ARMASV model is superior in forecasting **1. RGP** at $k=4$. In contrast, the relative MAFE measure suggests the UCt-ARMASV model as the preferred forecasting model, whereas the relative RMSFE recommends the VAR model when forecasting the-one-step-ahead for **11.HSTAR**. These findings are also reflected in Table 4.

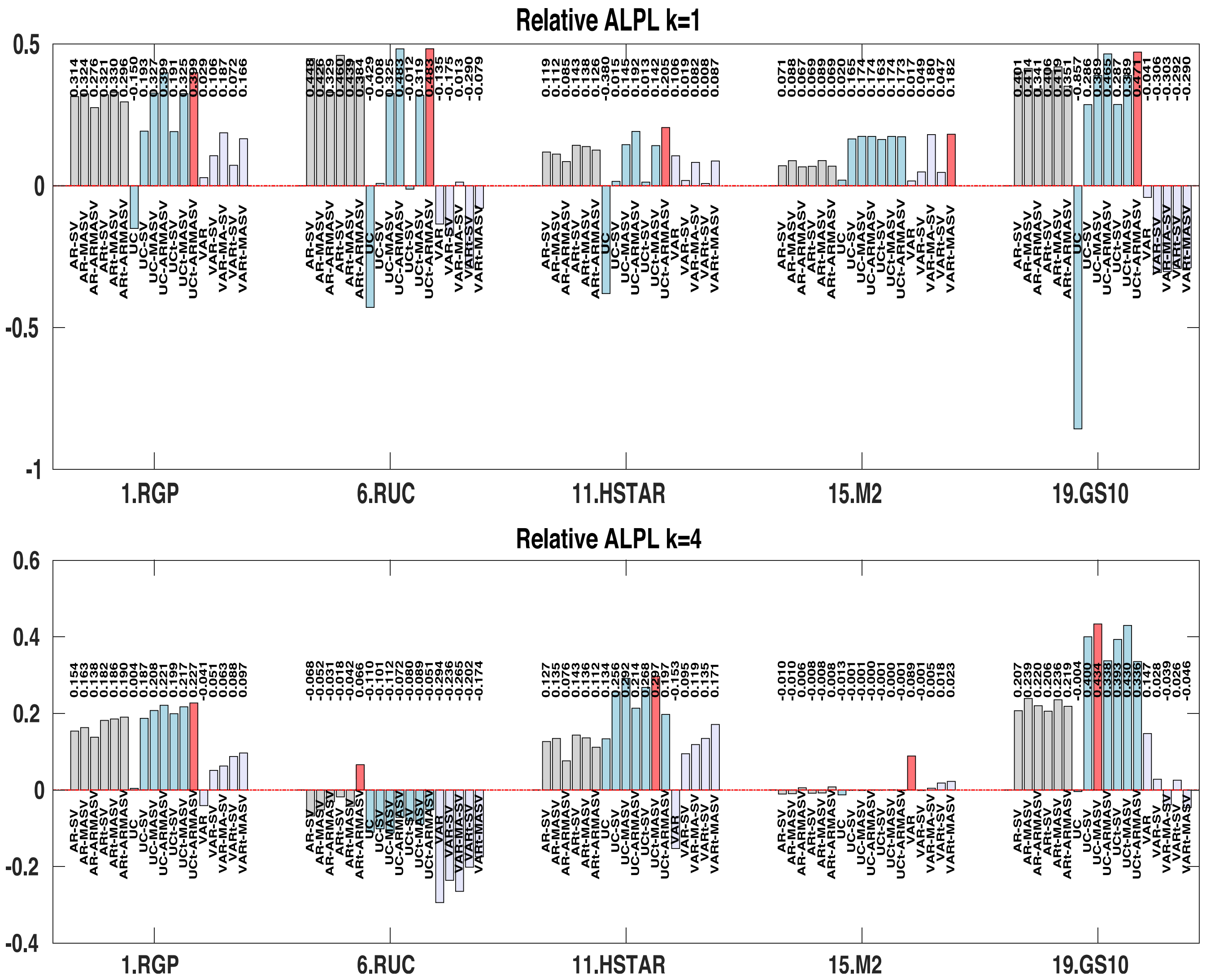

FIGURE 8 Relative ALPL for selected variables at $k=1$ and $k=4$. The light gray, blue, and purple colors represent the model groups: light gray (AR group), light blue (UC group), and light purple (VAR group). The red color indicates the best model performance.

For the one-step-ahead forecasts ($k = 1$), the relative MAFE measure indicates that the UC group (especially UCt-ARMASV, UC-ARMASV, and UC-MASV) outperforms the other models for 11 out of 20 variables, equivalent to 55% of the considered variables. However, based on the relative RMSFE measure, the UC group provides the best forecast for eight variables, representing 40% of the examined variables.

In addition, the point forecasting measures reveal that the UC-SV and UCt-SV models excel in forecasting the variables in Group E–*Interest rates and asset prices*, namely, **17.FFUND**, **18.GS1**, and **19.GS10**, for short- and medium-term forecasts. For instance, the UCt-SV model reduces the RMSFE of the benchmark model by almost 46% for forecasting **19.GS10** at $k = 1$ and nearly 25% RMSFE of the benchmark model at $k = 4$.

Finally, we observe that the univariate models appear to perform better than the VAR models. The VAR models seem to improve their forecast performance over medium- and long-term forecasts. Also, the VAR group outperforms the univariate models when it comes to forecasting **15.M2** money stock. This result is consistent across the forecast horizons and the forecast measures.

5.3 | Density forecast results

Tables 5 and 6 present the percentage of times that a model yields the best forecasting performance based on the relative ALPL, CRPS, AqwCRPS-center, and AqwCRPS-tail measures. Figures 8–11 present the

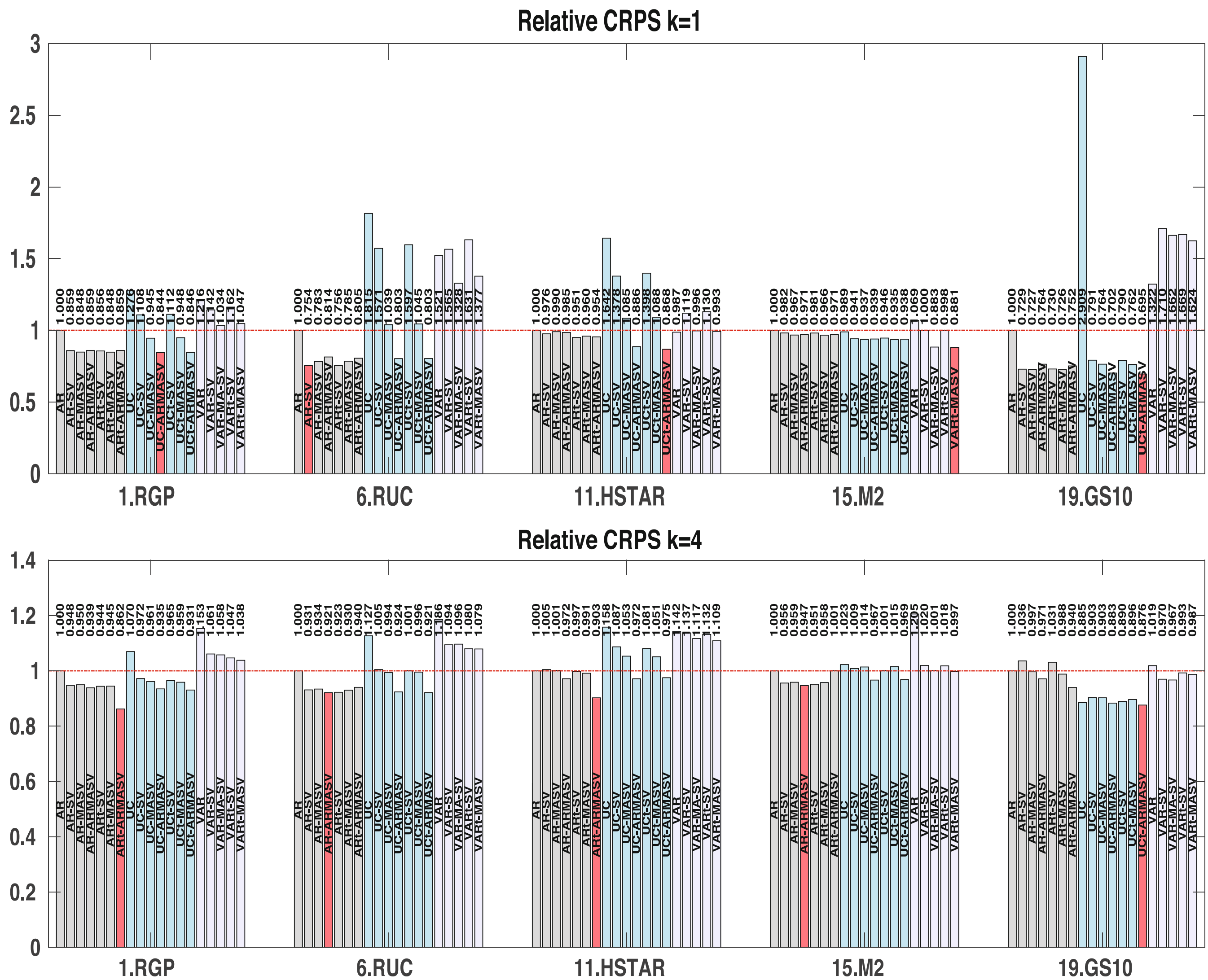

FIGURE 9 Relative average CRPS for selected variables at $k = 1$ and $k = 4$. The light gray, blue, and purple colors represent the model groups: light gray (AR group), light blue (UC group), and light purple (VAR group). The red color indicates the best model performance.

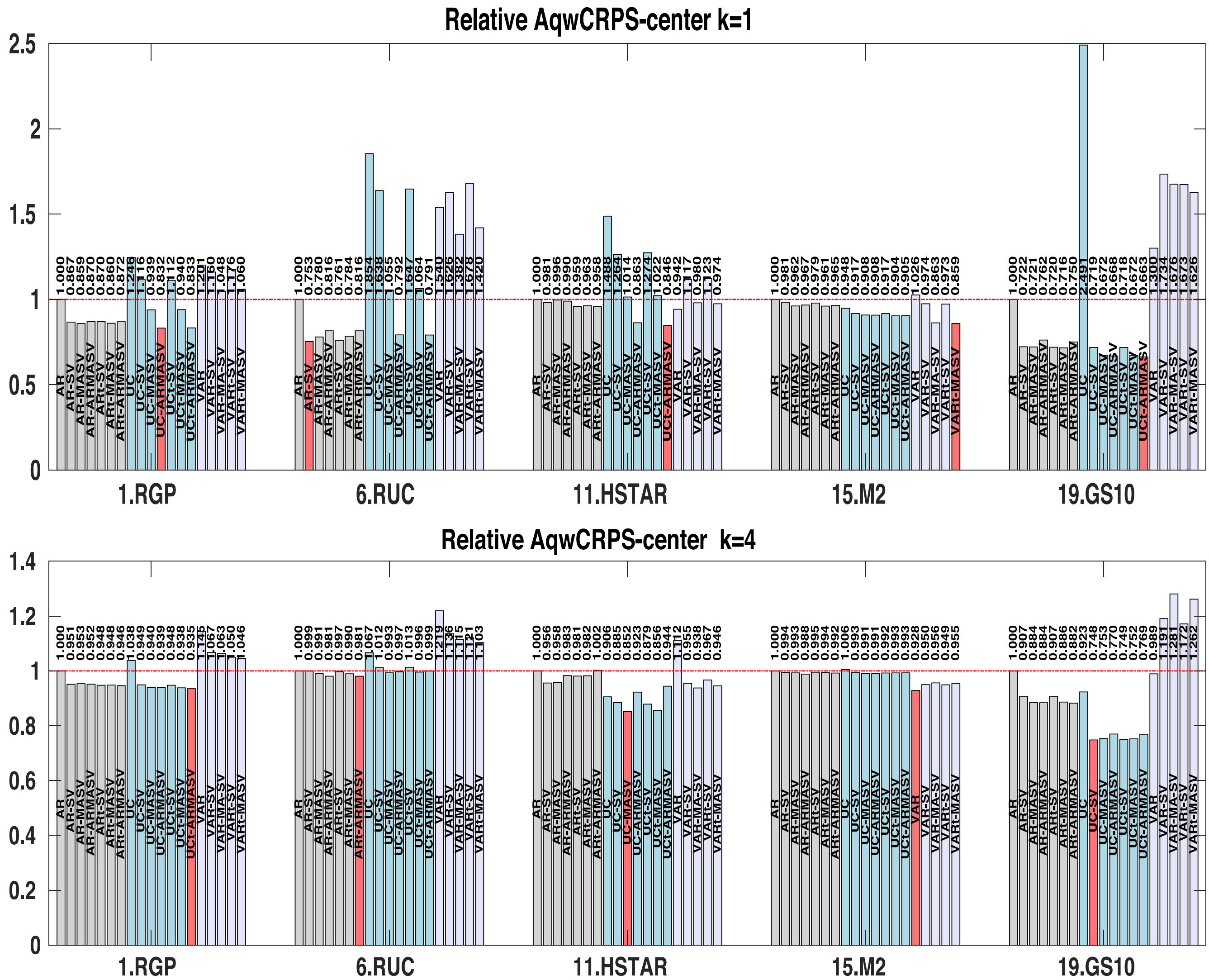

FIGURE 10 Relative AqwCRPS-center for selected variables at $k = 1$ and $k = 4$. The light gray, blue, and purple colors represent the model groups: light gray (AR group), light blue (UC group), and light purple (VAR group). The red color indicates the best model performance.

density forecast measures for some selected variables across 19 candidate models at $k = 1$ and $k = 4$.

There is strong evidence suggesting that the UC group, especially the UC-ARMASV and UCt-ARMASV models, produces the most accurate forecasts for the majority of macroeconomic variables at $k = 1$. Based on the relative ALPL results in Table 5, the UCt-ARMASV and UC-ARMASV models provide the best forecast for 15 out of 20 variables, representing about 75% of the examined variables. A similar pattern is also found in Table 6 when the relative AqwCRPS-center and AqwCRPS-tail measures are used to evaluate the forecasting performance. For medium- and long-term forecasts, other models within the AR and VAR groups demonstrate their superior forecast performance.

When examining Figures 8–11 (as well as Figures B7–B18 in Appendix S2), we find a similar pattern observed from the point forecast measures. That is, for some variables the same models consistently provide the best forecast performance across the density forecast measures. However, for other variables, the optimal models vary based on the forecast measures. For example, the UCt-ARMASV/UC-ARMASV model provides the most accurate forecast for **1.RGDP**, while the VARt-MASV model is superior in forecasting **15.M2** at $k = 1$ across the density forecast measures. For forecasting **6.RUC** at $k = 1$, the relative ALPL measure suggests the UCt-ARMASV model is the best, while the relative CRPS and the AqwCRPS-tail measures indicate the AR-SV and ARt-SV models, respectively.

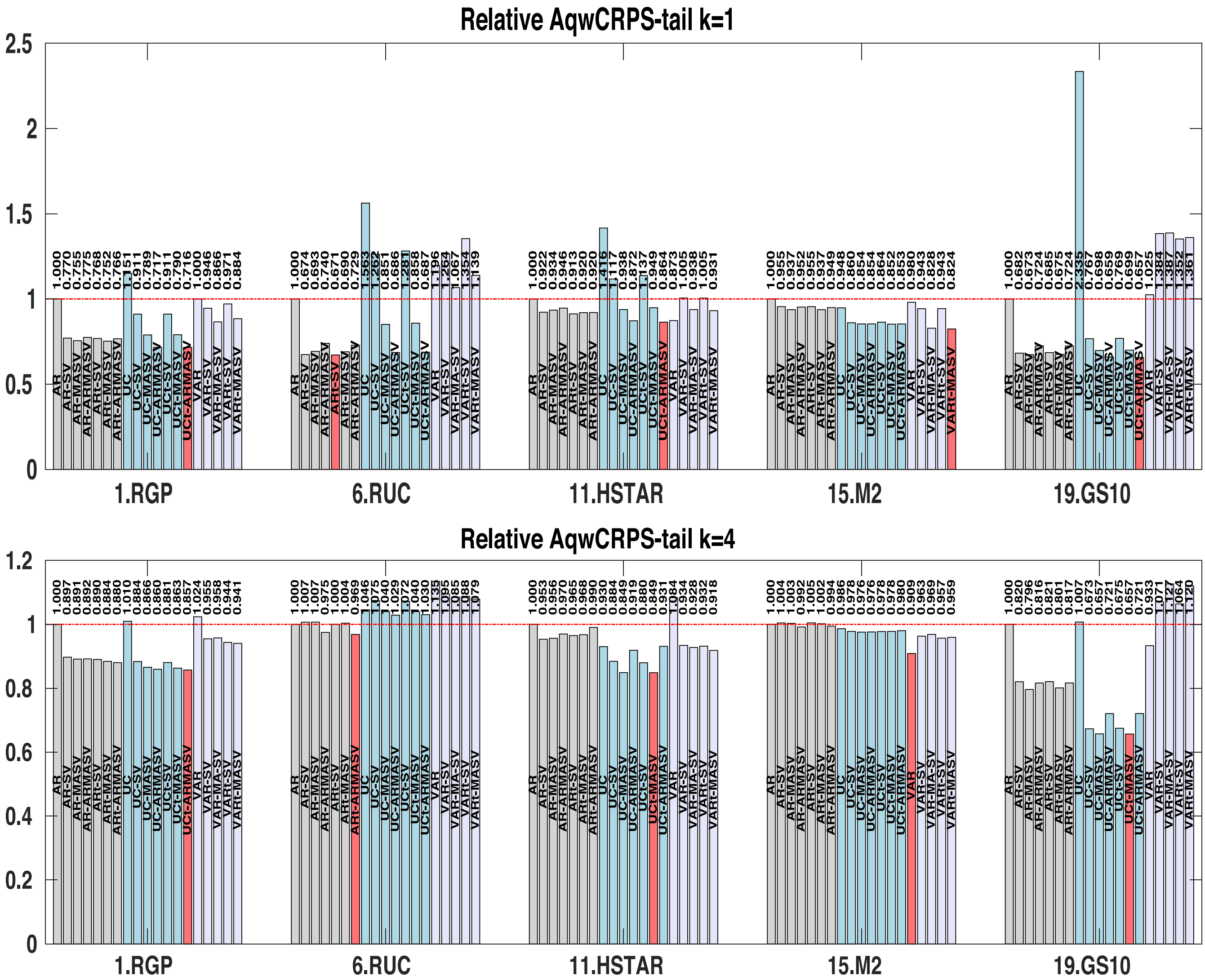

FIGURE 11 Relative AqwCRPS-tail for selected variables at $k=1$ and $k=4$. The light gray, blue, and purple colors represent the model groups: light gray (AR group), light blue (UC group), and light purple (VAR group). The red color indicates the best model performance.

The key takeaway from the density forecast analysis is that including heavy tails, SV, and serial correlation in error structures of the UC model enhances short-term forecasts for the majority of macroeconomic variables considered in this paper. For medium- and long-term forecasts, the AR group and UC groups are relatively compatible.

We further evaluate the forecast performance of the models during the COVID-19 pandemic, spanning from 2020Q1 to 2022Q2.[3] Tables A7–A9 in Appendix S1 suggest that the UC group, particularly UC-ARMASV and UCt-ARMASV, outperforms the other models for the one-step-ahead forecasts, according to both point and density forecast measures. For instance, the relative ALPL measure indicates that the UC-ARMASV and UCt-ARMASV models yield the most accurate forecast for 14 out of 20 variables, equivalent to a success rate of 70%. However, the VAR group demonstrates superior performance for medium- and long-term forecasts during the COVID-19 pandemic. For example, the VAR models with flexible error structures provide the best forecasts for 14 out of 20 variables at $k=8$, as revealed by the relative ALPL measure. Overall, the results from the pseudo-out-of-sample forecasting exercise confirm that the models with flexible error structures outperform those with homoscedastic errors in terms of forecast accuracy.

6 | CONCLUSION

In this paper, we have used two groups of univariate models, the AR and the UC models, with various error specifications to forecast 20 US macroeconomic series. Although the forecasting results do not indicate a clear winner, both point forecast and density forecast results show that the models with flexible error structures achieve better forecast accuracy for most variables than the models with homoscedastic errors. Specifically, accounting for heavy tail, SV, and serial correlation features in error processes significantly improves short-term forecast accuracy. While some variables exhibit consistent forecast accuracy across the point and/or the density measures, others show discrepancies between these measures in terms of the optimal model. Finally, the univariate models often perform better than the VAR models for most of the macroeconomic variables when it comes to the one-step-ahead forecasts.

ACKNOWLEDGEMENTS

We thank the editor Professor Derek Bunn and the anonymous reviewers for their constructive feedback throughout the review process of the journal. We thank Thomas Kim for his valuable input in developing an automatic process to summarize and present the comprehensive forecast results. Bo Zhang would like to acknowledge the financial support of the National Natural Science Foundation of China (72250410372) and Wenzhou Philosophy and Social Sciences Key Research Base Project (22jd30). Chenghan Hou would like to acknowledge the financial support of the National Natural Science Foundation of China (72003064). Open access publishing facilitated by Australian National University, as part of the Wiley - Australian National University agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

These data were derived from the following resources available in the public domain: the Federal Reserve Bank of Philadelphia at https://www.philadelphiafed.org/surveys-and-data and Federal Reserve Bank of St. Louis at https://fred.stlouisfed.org/.

ORCID

Kelly Trinh https://orcid.org/0000-0002-1814-6847
Bo Zhang https://orcid.org/0000-0003-0986-0018
Chenghan Hou https://orcid.org/0000-0002-2864-4831

ENDNOTES

[1] Another widely used approach to accommodate the time-varying variance is via the generalized autoregressive conditional heteroscedasticity (GARCH) model (Bollerslev, 1986).

[2] Table A1 in Appendix S1 presents the selected lag orders of the AR models for all variables using the AIC, BIC, and HQC criteria. Among these criteria, BIC chooses relatively lower orders for all variables. According to Bhansali (1997), selecting a low-order model may increase bias but could decrease estimation uncertainty and further improve forecasting performance. We therefore use the BIC criterion for selecting lags of the AR models in our forecasting analysis.

[3] We also examine the forecast performance of the models for periods excluding the COVID-19 pandemic. The results, presented in Tables A4–A6 in Appendix S1, align closely with those previously shown in Tables 4–6.

REFERENCES

Athanasopoulos, G., & Vahid, F. (2008). VARMA versus VAR for macroeconomic forecasting. *Journal of Business & Economic Statistics*, *26*(2), 237–252.

Bauwens, L., Koop, G., Korobilis, D., & Jeroen, J. V. K. (2015). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, *30*(4), 596–620.

Bhansali, R. J. (1997). Direct autoregressive predictors for multistep prediction: Order selection and performance relative to the plug in predictors. *Statistica Sinica*, *7*, 425–449.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://www.sciencedirect.com/science/article/pii/0304407686900631

Carriero, A., Clark, T. E., & Marcellino, M. (2016). Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics*, *34*(3), 375–390.

Chan, J. C. C. (2013). Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, *176*(2), 162–172.

Chan, J. C. C. (2020). Large Bayesian VARs: A flexible Kronecker error covariance structure. *Journal of Business & Economic Statistics*, *38*(1), 68–79.

Chan, J. C. C., & Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, *1*(1-2), 101–120.

Chan, J. C. C., Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2012). Time varying dimension models. *Journal of Business & Economic Statistics*, *30*(3), 358–367.

Chib, S., & Greenberg, E. (1994). Bayes inference in regression models with ARMA(p, q) errors. *Journal of Econometrics*, *64*(1–2), 183–206.

Chib, S., Nardari, F., & Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, *108*(2), 281–316.

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*(4), 551–575.

Clements, M. P. (2017). Assessing macro uncertainty in real-time when data are subject to revision. *Journal of Business & Economic Statistics*, *35*(3), 420–433.

Clements, M. P., & Galvão, A. B. (2013). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, *28*(3), 458–477.

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.

Cross, J., & Poon, A. (2016). Forecasting structural change and fat-tailed events in Australian macroeconomic variables. *Economic Modelling*, *58*, 34–51.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, *49*(1), 72–100.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, *8*(S1), S19–S40.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.

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

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

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497. https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA5771

Hou, C., Nguyen, B., & Zhang, B. (2023). Real-time forecasting of the Australian macroeconomy using flexible Bayesian VARs. *Journal of Forecasting*, *42*(2), 418–451.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, *65*(3), 361–393.

Koop, G., & Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies*, *74*(3), 763–789.

Marcellino, M., Stork, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1–2), 499–526.

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.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, *72*(3), 821–852.

Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, *14*(1), 11–30.

Stock, J. H., & Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*, 3–33.

Swanson, N. R., & White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, *13*(4), 439–461.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, *4*(2), 183–204.

Zhang, B., Chan, J. C. C., & Cross, J. L. (2020). Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts. *International Journal of Forecasting*, *36*(4), 1318–1328.

AUTHOR BIOGRAPHIES

Dr. Kelly Trinh is a research scientist at the Commonwealth Scientific and Industrial Research Organisation and an adjunct lecturer at James Cook University. She received her PhD from the University of Queensland. Her research focuses on Bayesian statistics, time series forecasting, large Bayesian VARs, and machine learning.

Dr. Bo Zhang is a professor at Business School, Wenzhou University, in China. She received her PhD from the Australian National University. Her research interests focus on macroeconomic forecasting, large Bayesian VARs, and energy and environmental economics.

Dr. Chenghan Hou is an associate professor at the Center for Economics, Finance and Management Studies, Hunan University. His research focuses on Bayesian econometrics, time series analysis, nonlinear state space models, and macroeconomic forecasting.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Trinh, K., Zhang, B., & Hou, C. (2025). Macroeconomic real-time forecasts of univariate models with flexible error structures. *Journal of Forecasting*, *44*(1), 59–78. https://doi.org/10.1002/for.3182