Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Forecasting Oil Prices: Can Large BVARs Help?

CAMA Working Paper 65/2022
October 2022

Bao H. Nguyen
University of Tasmania
CAMA, Crawford, ANU, Australia

Bo Zhang
Wenzhou University
CAMA, Crawford, ANU, Australia

Abstract

Large Bayesian Vector Autoregressions (BVARs) have been a successful tool in the forecasting literature and most of this work has focused on macroeconomic variables. In this paper, we examine the ability of large BVARs to forecast the real price of crude oil using a large dataset with over 100 variables. We find consistent results that the large BVARs do not beat the BVARs with small and medium sizes for short forecast horizons but offer better forecasts at long horizons. In line with the forecasting macroeconomic literature, we also find that the forecast ability of the large models further improves upon the competing standard BVARs once endowed with flexible error structures.

Keywords

forecasting, non-Gaussian, stochastic volatility, oil prices, big data

JEL Classification

C11, C32, C52, Q41, Q47

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

The Centre for Applied Macroeconomic Analysis in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Forecasting Oil Prices: Can Large BVARs Help?*

Bao H. Nguyen[†] Bo Zhang[‡]

October 2022

Abstract

Large Bayesian Vector Autoregressions (BVARs) have been a successful tool in the forecasting literature and most of this work has focused on macroeconomic variables. In this paper, we examine the ability of large BVARs to forecast the real price of crude oil using a large dataset with over 100 variables. We find consistent results that the large BVARs do not beat the BVARs with small and medium sizes for short forecast horizons but offer better forecasts at long horizons. In line with the forecasting macroeconomic literature, we also find that the forecast ability of the large models further improves upon the competing standard BVARs once endowed with flexible error structures.

JEL-codes: C11, C32, C52, Q41, Q47

Keywords: forecasting, non-Gaussian, stochastic volatility, oil prices, big data

*We thank Christiane Baumeister, Joshua Chan, Jamie Cross, Chenghan Hou and Francesco Ravazzolo for constructive comments and suggestions.

[†]Tasmanian School of Business and Economics, University of Tasmania and Centre for Applied Macroeconomic Analysis (CAMA), Australian National University, Australia. Email: b.nguyen@utas.edu.au

[‡]*Corresponding author.* Business School, Wenzhou University, Wenzhou, Zhejiang Province, 325035, P. R. China, and Centre for Applied Macroeconomic Analysis (CAMA), Australian National University, Australia. Email: bozhangyc@gmail.com

1 Introduction

In recent years, macroeconomists have greatly benefited from using large datasets as having more information generally allows for better prediction of economic phenomena and improves causal inference. To employ large datasets, two main alternative approaches have been proposed in the literature. The first and the earlier approach primarily relies on the use of a factor model as means of dimension reduction (see e.g. Geweke (1977); Stock and Watson (2002); Bernanke, Boivin, and Eliasz (2005); McCracken and Ng (2016), and many others).[1] The second approach is to apply shrinkage priors in a Bayesian Vector Autoregression (BVAR) to handle large dynamic systems (e.g. Bańbura, Giannone, and Reichlin (2010); Koop (2013); Carriero, Clark, and Marcellino (2019); Chan (2020a), and among others).[2]

Although the literature on the aforementioned approaches has rapidly expanded and their benefits have been well established in forecasting macroeconomic and financial variables, we have observed little empirical work focusing on other important variables that are also of interest to academic researchers, business practitioners, and government planners, such as the price of crude oil. The most notable study is Baumeister, Korobilis, and Lee (2020) (henceforth BKL). This paper investigates the usefulness of using large datasets to forecast oil prices and global petroleum consumption. More specifically, to capture important information that relates to energy demand the paper considers a broad range of indicators and compiles two sets of data. The first and medium dataset includes 16 variables and the second and large one contains 256 variables. Based on the factor approach, they find that the first principle component extracted from the first dataset is useful for forecasting oil prices while the corresponding one obtained from the larger dataset is not. In this paper, we revisit this exercise. We question whether the usefulness of the large BVARs, as a valid alternative to the factor model, can still hold under the large dataset constructed by BKL. We believe that this investigation is important, as it shows how large BVARs can be used to overcome the problem of cross-correlated idiosyncratic errors, which often arises in the factor model when the number of variables is very large (Boivin and Ng, 2006). Indeed, as BKL mentions, when adding more series from the same data category, the cross-correlation in the idiosyncratic errors tends to be too large and thus the extracted factor is less useful for forecasting.

Our work not only complements the work by BKL but also contributes to the small but growing literature, which explores the benefit of using large multivariate models in forecasting energy and commodity prices. To that end, we obtain the forecasting

[1]For surveys and the usefulness of the factor model in economics, see Bai and Ng (2008) and Stock and Watson (2016).

[2]A recent survey of the extensions of large BVARs can be found in Chan (2020b).

performance for the real price of oil by firstly beginning with a set of small BVARs that includes four core variables of the global oil market: global crude oil production, a proxy of global oil inventories, the real price of crude oil and an index of global real economic activity. The last variable is the variable of interest. While BKL respectively replaces this variable with alternative proxies that are commonly used in the literature, as well as factors extracted from their proposed datasets, we directly include their datasets in our proposed models. We then consider BVARs with a set of selected 16 series as the medium-sized specifications and with the second dataset as the large ones. To remain the same sample period as in BKL, e.g. 1973.2-2018.8, we eliminate some series those only start later and come up with a balanced panel of 108 variables in the large specifications. This large system makes our paper relevant to recent studies also using large information to forecast energy prices, for example, Ferrari, Ravazzolo, and Vespignani (2021) and Gianfreda, Ravazzolo, and Rossini (2020). The former study uses a large dataset that contains around 200 series for 33 countries and forecasts quarterly energy prices, including the price of crude oil. Different from our approach, to deal with the large cross-sectional dimension, they utilize a penalized maximum likelihood method to extract latent factors and evaluate the forecasting performance through various factor model modifications. The latter work evaluates the forecast ability of large BVARs but focuses on electricity markets in Italy and Germany. Similar to our analysis, this study also considers some forms of the BVAR model with flexible errors, which are discussed as follows.

In addition to employing large datasets, this paper also contributes to the fast-growing literature on large BVARs by considering different model specifications that allow for more flexible error covariance structures. Traditionally, the BVAR model is embedded within standard error assumptions, e.g. homoscedastic, Gaussian, and serially independent. However, recent extensions, such as heteroscedastic, non-Gaussian, and serially dependent innovations, are found to be crucial features that enhance the forecasting power of the BVAR (see, for example, Clark and Ravazzolo (2015); Carriero, Clark, and Marcellino (2016); Chan (2020a); and Hou, Nguyen, and Zhang (2022)). For example, BKL considers a specification that allows for stochastic volatility, one of the mentioned extensions, and finds that the inclusion of this element can improve the forecast ability of oil prices for long horizons. Similar conclusions are also found in forecasting electricity and natural gas prices, see Gianfreda, Ravazzolo, and Rossini (2020) and Gao, Hou, and Nguyen (2021), respectively. We extend BKL's analysis further by taking other forms of flexible disturbances, including common stochastic volatility, heavy tails, and dependent errors, as well as the combinations of these features. It must be emphasized that, to our best knowledge, this paper is the first to provide a multivariate analysis of crude oil prices that takes into account almost all possible combinations of non-standard errors existing in the current literature.

Our forecasting exercises yield several intriguing results, highlighting the importance of model sizes and error specifications. The findings are robust under alternative measures of crude oil, e.g. the U.S. refiner acquisition cost (RAC) of crude imports and the Brent price, and also hold when oil production is replaced by a measure of petroleum consumption. Overall, the main results indicate that the inclusion of flexible errors and the use of large datasets improve point and density forecasts of real oil prices. Moreover, their forecasting performance can be further enhanced by using these ingredients jointly, especially with long-horizon forecasts. These results match those observed in forecasting macroeconomic and financial variables, see Carriero, Clark, and Marcellino (2019) for example.

Concerning the model sizes, we find that both point and density forecasts consistently agree that the small BVAR models which include the four key variables outperform the medium and large models at short-horizon forecasts for up to six months. However, at longer horizons, the medium and large-sized BVAR models tend to forecast the real price of oil better than their small counterparts. This finding is new in the oil forecasting literature but reflects those of Ferrari, Ravazzolo, and Vespignani (2021) and BKL, although they employ different models. Ferrari, Ravazzolo, and Vespignani (2021), for example, they find that their factor models only do well when predicting oil prices one quarter ahead. Similarly, KBL also observes that the model based on the factor extracted from their large dataset does not beat the corresponding models based on a small dataset and even leads to a deterioration of the forecasting performance at long horizons. That said, at long-horizon forecasts, the large-sized BVAR model becomes more helpful as compared to the factor model in predicting oil prices.

In terms of model specifications, we find clear evidence that the models with flexible structures of the error term can further improve forecast accuracy when compared to the models with conventional error assumptions. In line with findings in BKL and Gianfreda, Ravazzolo, and Rossini (2020), the results show that stochastic volatility is an important ingredient, especially for long-horizon forecasts. In many cases, introducing the combination of stochastic volatility, dependent errors, and heavy tails can even further improve forecast accuracy, confirming evidence established in the forecasting macroeconomic literature, such as Carriero, Clark, and Marcellino (2019), Chan (2020a) and Hou, Nguyen, and Zhang (2022).

The rest of the paper is organized as follows. In Section 2 we describe the data used. We then introduce in Section 3 alternative covariance structures that incorporate into BVARs. Detailed forecast results and discussion are presented in Section 4. We report the sensitivity analysis in Section 5 and Section 6 concludes the paper.

2 Data

In this section, we briefly introduce the selected data used in this paper. Detailed descriptions of the data, sources, and respective transformation methods can be found in BKL and its online appendix. As discussed, the core variables of the global crude oil markets entering our small BVARs consist of the percent change in global oil production, an estimate of the change in global crude oil inventories, the log of the real price of oil and a proxy of real global economic activity. In the main analysis, we evaluate the forecasting performance of the real price of Brent and the forecasts for the U.S. refiner acquisition cost (RAC) of crude oil imports are reported in the robustness check. We also forecast the oil prices using the production- and consumption-based models, as in BKL. The latter model is the specification that replaces oil production with a measure of global petroleum consumption.

With regard to the proxy of global economic activity, BKL evaluates alternative indicators that have been used in the oil literature, including a measure of dry-cargo shipping rates developed by Kilian (2009), a real shipping cost factor constructed by Hamilton (2021), a real commodity price factor suggested by Alquist, Bhattarai, and Coibion (2020) and Delle Chiaie, Ferrara, and Giannone (2022), a global steel production factor proposed by Ravazzolo and Vespignani (2020), and a world industrial production index developed by Baumeister and Hamilton (2019). One of their key findings is that the index of world industrial production is the most useful candidate for forecasting oil prices. In this paper, therefore, we use this index as the main proxy of global real economic activity.

To investigate the forecast ability under different model sizes, we firstly consider a medium size by replacing the measure of global energy demand with a set of series that BKL use to construct their indicator of global economic conditions. To find a new indicator that can better capture global economic conditions and thus relates to global demand for energy, BKL carefully selects 16 series that span from 1973M1 to 2018M8 and cover eight dimensions of the global economy: real economic activity, commodity prices, financial indicators, transportation, uncertainty, expectations, weather, and energy-related measures. Note that, the first category still remains world industrial production that is used in the small dataset. To obtain a balanced panel, we discard 5 series that only start later: OECD consumer confidence index, long-run world oil price uncertainty, US consumer expectation, and world spread between long-run, and short-run oil price expectations. Therefore, our medium specification contains 14 variables. For the large specification, we utilize the set of 256 variables considered in BKL. This large dataset incorporates all possible disaggregated data from the eight categories above. After eliminating some series that do not start in 1973M1, we obtain a balanced panel containing 108 variables.

3 Competing Bayesian VARs

Table 1 summarizes the specifications considered in our analysis. In the following sections, we first introduce the BVAR with conventional error assumptions and then the model with idiosyncratic stochastic volatility (SV), common stochastic volatility (CSV), heavy tailedness (e.g., Student's t distribution), and serial dependence moving average error (MA).

3.1 Standard VARs with Conventional Error Assumptions

We start from an expression of the standard VAR model, which can be written in a reduced form of VAR with order p as below:

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

where $\mathbf{y}_t = (y_{1t}, \cdots, y_{nt})'$ denote an $n \times 1$ vector of endogenous variables in a BVAR, $\mathbf{b}$ is an $n \times 1$ vector of intercepts, and $\mathbf{B}_1, \cdots, \mathbf{B}_p$ are $n \times n$ coefficient matrices, and $\mathbf{\Sigma}$ is an $n \times n$ cross-sectional covariance matrix of VAR. In a standard VAR, $\boldsymbol{\varepsilon}_t^y$ can be assumed to be independent and identically Gaussian distributed (iid). In practice, Equation 1 can be rewritten as below for parameter estimation:

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t^y, \tag{2}$$

where $\mathbf{X}_t = \mathbf{I}_n \otimes [1, y'_{t-1}, \cdots, y'_{t-p}]$ in which notation $\otimes$ denotes the Kronecker product, and $\boldsymbol{\beta}$ is stacked by rows of $[\mathbf{b}, \mathbf{B}_1, ..., \mathbf{B}_p]'$ with the size of $(1 + np)n \times 1$.

Let $x'_t = (1, y'_{t-1}, \cdots, y'_{t-p})$ be a $1 \times (1 + np)$ vector, when stacking the observations over time T, we get $\mathbf{X}$ which is a $T \times (1 + np)$ matrix. Then we have:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{3}$$

where $\mathbf{Y}$ is $\mathbf{y}_t$ stacked over time T, $\mathbf{B} = (\mathbf{b}, \mathbf{B}_1, ..., \mathbf{B}_p)'$ with a size of $(1 + np) \times n$, $\mathbf{E} = (\boldsymbol{\varepsilon}_1^y, \cdots, \boldsymbol{\varepsilon}_T^y)'$, so that

$$\text{vec}(\mathbf{E}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{\Omega}), \tag{4}$$

where $\mathbf{\Omega}$ is the serial covariance matrix of the VAR model.

As mentioned, taking into account of improving model fitness and forecastability, the standard BVAR model with iid Gaussian innovations can be extended in different ways to capture important features of macroeconomic time series. In what follows, we introduce these extensions in detail and consider those proposed models as flexible BVARs.

3.2 VARs with Idiosyncratic Stochastic Volatility

As in BKL, we also consider VAR models with idiosyncratic stochastic volatility (VAR-SV). The framework of the VAR-SV model is obtained by extending equation (1) as below:

$$\mathbf{y}_t = \mathbf{b} + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t^y, \quad \boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{5}$$

where a Minnesota type of prior is used for the VAR coefficients $(\mathbf{b}, \mathbf{B}_1, \ldots, \mathbf{B}_p)$. The time-varying variance-covariance matrix $\boldsymbol{\Sigma}_t$ is assumed as $\boldsymbol{\Sigma}_t = \mathbf{B}_0^{-1}\text{diag}(e^{h_{1,t}}, \ldots, e^{h_{n,t}})\mathbf{B}_0^{-1'}$, where $\mathbf{B}_0$ is an unrestricted non-singular matrix. Each log-volatility process is specified as the following stationary AR(1) process:

$$h_{i,t} = \phi_{h,i}h_{i,t-1} + \varepsilon_{i,t}^h, \quad \varepsilon_{i,t}^h \sim \mathcal{N}(0, \sigma_{h,i}^2), \quad i = 1, \ldots, n,$$

with $h_{i,t} \sim \mathcal{N}(0, \sigma_{h,i}^2/(1-\phi_{h,i}^2))$ and $|\phi_{h,i}| < 1$.

3.3 VARs with a Common Stochastic Volatility

One of the most useful extensions of VARs is the adoption of a common stochastic volatility (CSV) factor. There has been recognized that the volatilities of a wide range of macroeconomic variables are time-varying and tend to move together (Carriero, Clark, and Marcellino, 2016; Mumtaz and Theodoridis, 2018; Poon, 2018). However, standard VARs with homoscedastic error, would not be able to capture this feature. The inclusion of CVS error specification allows VARs to capture any common structural shifts in the macroeconomic time series. In the modeling framework of VARs with CSV, we first consider time-varying volatility. Suppose $\boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, e^{h_t}\boldsymbol{\Sigma})$, where h is the stochastic volatility parameter and e^{h_t} is the common stochastic volatility (Carriero, Clark, and Marcellino, 2016). More specifically, h follows an AR(1) process:

$$h_t = \phi_h h_{t-1} + \boldsymbol{\varepsilon}_t^h, \quad \boldsymbol{\varepsilon}_t^h \sim \mathcal{N}(\mathbf{0}, \sigma_h^2), \tag{6}$$

where $|\phi_h| < 1$. In this assumption, the variances of all the variables share the same stochastic volatility parameter which is a restrictive assumption. There is empirical evidence that the volatilities of macroeconomic time series have a comovement (Carriero, Clark, and Marcellino, 2016), thus, it is also a parsimonious assumption for parameter estimation.

3.4 VARs with a CSV and t Errors

Recent empirical studies also show that the forecasting performance of macroeconomic variables can be improved when a normal distribution is replaced by heavy-tailed distribution, e.g. Student's t distribution, in the covariance matrix of VARs. The importance of this extension is that when the model accounts for t-disturbances, this specification of heavy-tailed innovations turns out to present good features, such as reducing the variation of estimates, dealing well with outliers, such as the Great Recession, and thus providing good model fitness (e.g., Clark and Ravazzolo, 2015; Cross and Poon, 2016; Chiu, Mumtaz, and Pinter, 2017). For VARs which can capture such fat tail events, the distribution of error terms $\boldsymbol{\varepsilon}_t^y$ has one more hyperparameter λ_t for t innovations:

$$\boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \lambda_t e^{h_t} \boldsymbol{\Sigma}), \tag{7}$$

where $\lambda_t \sim \mathcal{IG}(\boldsymbol{\nu}_\lambda/2, \boldsymbol{\nu}_\lambda/2)$ following an inverse-gamma distribution with degree of freedom parameter $\boldsymbol{\nu}_\lambda$, and $\lambda_1, ..., \lambda_T$ are independent from each other.

3.5 VARs with a CSV and MA(1) t Errors

Another property of macroeconomic variables that has been recognized is serially dependent (Chan, 2013). To handle this property, the conventional assumption of serially independent innovations can be replaced by a moving average of error terms. Following Chan (2020b), for the serial dependence of covariance matrix over time, suppose the error term $\boldsymbol{\varepsilon}_t^y$ follows a heteroscedastic moving average innovation process. More precisely, we assume $\boldsymbol{\varepsilon}_t^y$ has an MA(1) stochastic volatility process:

$$\boldsymbol{\varepsilon}_t^y = u_t + \boldsymbol{\psi}_{\boldsymbol{\varepsilon}} u_{t-1}, \quad u_t \sim \mathcal{N}(\mathbf{0}, \lambda_t e^{h_t} \boldsymbol{\Sigma}). \tag{8}$$

Here, the covariance matrix $\boldsymbol{\Omega}$ in Equation 4 has $((1 + \boldsymbol{\psi}_{\boldsymbol{\varepsilon}}^2)\lambda_1 e^{h_1}, ..., (1 + \boldsymbol{\psi}_{\boldsymbol{\varepsilon}}^2)\lambda_T e^{h_T})$ along its main diagonal, $(\boldsymbol{\psi}_{\boldsymbol{\varepsilon}} \lambda_1 e^{h_1}, ..., \boldsymbol{\psi}_{\boldsymbol{\varepsilon}} \lambda_{T-1} e^{h_{T-1}})$ above and below the main diagonal, and 0 elsewhere.

For our forecasting exercise, we start with a BVAR model with conventional error assumptions and then include the aforementioned features of the covariance structure into this standard BVAR. We also consider possible combinations of these error structures. All models are estimated using Markov chain Monte Carlo method (MCMC), see Appendix A.1 for details on simulation. The estimation results in our empirical studies are all based on 10000 posterior samples obtained after a burn-in period of 5000. With regard to priors, for comparison purposes, whenever possible we choose the same priors for the common parameters across models. In particular, the Minnesota prior and the natural conjugate prior is used for the standard VARs and flexible VARs respectively. Details of values of the hyperparameter of these priors are reported in Appendix A.2.

4 Forecast Results

In this section, we perform a recursive out-of-sample forecasting exercise to evaluate the performance of the proposed VARs in terms of point and density forecasts. For expository purposes, in the analysis below, we focus on the performance of the models listed in Table 1. It is worth noting that, with large systems as we do, the computational cost of some settings, such as a model with idiosyncratic stochastic volatility across all variables, is huge. Therefore, we only consider this feature in the small and medium BVARs and employ fast algorithms that have been recently proposed in Chan (2020a) and Chan (2021). Alternatively, we consider the model with *common* stochastic volatility. Following BKL, we use the no-change forecast model as the benchmark for evaluating the forecast ability of our proposed models. The lag length is set to p = 12 and the parameters are estimated using data from 1973.2 to 1991.12 to forecast the log of oil price for one-month-ahead, 3-month-ahead, and so on. The evaluation period is from 1992.1 to 2018.8 based on the level of real oil prices.

4.1 Forecast Evaluation Metrics

The recursive exercise will involve using data available at t from T_0 up to time T to forecast at time $t+k$ for $k = 1, 3, 6, 12$ and 24. Thus, the forecast horizons are one-month-ahead, one-quarter-ahead, half-year-ahead, one-year-ahead, and two-year-ahead. The accuracy of the point forecast is assessed by root mean squared forecast error (RMSFE).[3] RMSFE is a commonly used scale-dependent measure for each time series with the same unit. The value of RMSFE for the target variable i, e.g. the real price of oil, at forecast horizon k ($k = 1, 3, 6, 12$ and 24) is calculated as:

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

where T_0 is the start of the evaluation period, $y^{\text{o}}_{i,t+k}$ is the observed value of the interested variable, and $\mathbb{E}(y_{i,t+k}\,|\,\mathbf{y}_{1:t})$ is the sample mean of forecasts given information of the variable up to time T. For RMSFE, a smaller value comes from a smaller forecast error and stands for a better forecast performance.

As point forecast ignores the predictive distribution of forecast results, we also evaluate the forecast performance from the predictive distribution of density forecast by the average of log predictive likelihood (ALPL). For the estimation $y_{i,t+k}$, the predictive likelihood is obtained by the predictive density evaluated at the observation $y^{\text{o}}_{i,t+k}$. More

[3] Note that, BKL use the mean squared forecast error to evaluate the point forecasting performance. Our findings remain unchanged under this alternative measure.

specifically, the ALPL is defined as:

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

where $p(y_{i,t+k} = y^{o}_{i,t+k} \,|\, \mathbf{y}_{1:t})$ is the predictive likelihood with information of the interested variable up to time T. Given the predictive distribution, a larger value of predictive likelihood means that the observation $y^{o}_{i,t+k}$ is more likely under the predicted density forecast. In other words, a larger value of ALPL indicates better forecast performance.

4.2 Forecasting Results

In this section, we report and discuss the forecasting performance of our proposed models listed in Table 1. We consider the performance for the Brent price as the main results and the results for RAC are reported in the following section as the robustness analysis although the results under these alternative measures of the global oil price are quantitatively similar. To judge the superiority between the benchmark and the competing model for a given horizon, we implement the Diebold-Mariano (DM) test (Diebold and Mariano, 1995). In the tables, dark gray cells and light grey cells indicate the significant difference in the predictive accuracy between alternative models and the benchmark model at 1% and 5% levels of significance, respectively. All RMSFE values are normalized relative to the no-change forecast so that a ratio below 1 indicates that the model does better than the benchmark model. For ALPL, we report the difference between the value obtained from a competing model and the benchmark, therefore, a larger value of relative ALPL stands for better forecasting performance.

Overall, we find that the small-sized models outperform the medium and large-sized models at short-horizon forecasts, however, at longer horizons, the models with medium and large sizes tend to forecast the Brent price better than the small counterparts. We also find that the models endowed with flexible error structures help to improve forecast accuracy, and the gains from these features increase with the system size for long-horizon forecasts. Our results not only further support findings in BKL, but also confirm what has been found in the macroeconomic literature. Carriero, Clark, and Marcellino (2019), for example, highlight the benefit of jointly using a large dataset and heteroskedastic model and also find that the benefit is larger at longer forecast horizons.

The following parts of this paper move on to describe in greater detail of these findings by first focusing on the role of model sizes and then examining the additional benefits of using flexible error structures.

4.2.1 The role of model size

Table 2 and 3 compare the forecasting performance of our competing models for the real Brent price according to the production-based and consumption-based models, respectively. As can be seen clearly, the results are similar under the two considered models. First, focusing solely on the small size, forecast results based on RMSFE show that the standard VAR model outperforms its counterpart modifications at short forecast horizons, e.g. 1 and 3-month ahead. For longer horizons, the small model with time-varying volatilities turns out to be the best for predicting the Brent price. This is consistent with what BKL finds. In terms of density forecasts, the small VARs with either idiosyncratic stochastic volatility or common stochastic volatility still largely dominate other models for short-horizon forecasts although the standard VAR model slightly does better for long horizons.

If we now turn to the medium-sized models, as we can see from the table, the heteroscedastic model provides superior accuracy across horizons based on point forecasts. However, adding a few more variables seems not necessarily sufficient to increase the accuracy of point forecasts relative to the small system with four variables. The forecasting performance of the VAR-SV is quantitatively similar between the small and medium systems. However, based on the density forecasts, we observe that forecasts using the medium models are a little better for long horizons. More specifically, we find that the homoskedastic BVAR model is the best model for density forecasts at 24-month-ahead forecasts. This finding reflects the interaction effect between the use of a large dataset and the inclusion of flexible error structures documented in Carriero, Clark, and Marcellino (2019).

Compared to the small and medium models, the forecasting performance of the Brent price based on the large models is more striking. With regard to point forecasts, we find that the model with heavy tails provides superior accuracy for 12- and 24-month-ahead forecasts. The forecast ability of this modification can be further improved when allowing for the errors being dependent, as we can see the reduction in the relative RMSFE values at 3- and 6-month-ahead forecasts. While adding more information helps sustainably improve the point forecasts at long horizons, using a larger dataset does not seem to be important under density forecasts. We observe some model specifications, such as VAR-CSV-MA or VAR-CSV, still outperform the benchmark model but the improvement of these specifications is not significant as compared to the small and medium models. In some stances, under the density metric, the results are not always statistically significant when comparing the competing models and the benchmark, especially with longer horizon forecasts.

4.2.2 The role of flexible error structures

Generally, we observe that the inclusion of different forms of the error structure is increasingly important when the model size is expanded and this feature is especially crucial for long forecast horizons. These patterns can be found from Table 2 and 3.

In the case of point forecasts, both the production-based and consumption-based models suggest that the VAR model with conventional error assumption is the best for predicting the Brent price at short horizons. Interestingly, the forecasting performance can be further enhanced by allowing for stochastic volatility and the benefit of having this feature is specifically pronounced at long-horizon forecasts under the small system. This finding is consistent with BKL who also highlight that stochastic volatility is an important ingredient for long-horizon forecasts of the real Brent price. As expected, turning to the medium VAR, this system still supports this model specification and its forecasting performance still stands out as compared with the other models. We also find clear evidence that the forecasting performance of our proposed models changes remarkably when moving from the small to larger data. We observe that, at 1-month-ahead forecast, none of these proposed models can beat the no-change forecast model when the large data is used, including the standard VAR-the best model under the small system. However, when increasing the forecast horizon beyond 1-month ahead, while the standard VAR still performs poorly, the models with flexible error structures increasingly become important and substantially improve the forecast accuracy. Interestingly, the fat tail error and the MA component are found to be the most important features in predicting the Brent price.

Turning to results for density forecasts, the results are likely in favour of the models with flexible errors, even under the small system. This is, not only because the model incorporated with SV can improve the forecast accuracy, but other specifications of the variant, such as CSV, MA or Student's-t error are also often found important. VAR-CSV, for example, improves upon the standard VAR across many model sizes and forecast horizons. Our results broadly in line with Gianfreda, Ravazzolo, and Rossini (2020) and Gao, Hou, and Nguyen (2021). These studies, respectively, also find that Student-t stochastic volatility is an important ingredient for forecasting electricity prices and gas prices. Although the diversity, the main results under this forecast metric are still in line with the point forecasts. They all agree that at short forecast horizons, the small-sized VARs are sufficient to produce the best forecast accuracy and the inclusion of the unconventional error structures further improves the performance of the small model. This finding is in line with KBL and also with the forecasting macroeconomic literature (Carriero, Clark, and Marcellino, 2019; Chan, 2020a; Hou, Nguyen, and Zhang, 2022).

5 Robustness

In this section, we discuss the forecasting performance of our proposed models for the alternative measure of the global oil price, e.g. RAC. Table 4 and 5 present individual results under the production-based and consumption-based models. As we already mentioned, the overall results are in line with the findings under the Brent price reported in the main analysis. In this session, we highlight the most striking observations emerging from forecasting real RAC.

First, focusing on the small-sized models, the relative RMSFE ratios indicate the standard VAR model continues to beat the random walk for near-term forecasts of real RAC. This finding matches those observed in BKL and Baumeister and Kilian (2012). For example, using the real-time data to forecast real RAC, Baumeister and Kilian (2012) find that the BVAR model beat the no-change forecast only at short horizons. Consistent with the Brent price, we also find that the models with unconventional error assumptions also do better than the benchmark and the best error modification is the model with MA. This model even does slightly better than the standard VAR for 1 and 3-month ahead forecasts. The density forecasts of real RAC do not change much under the small system as compared to those of the Brent price. This is, the models with stochastic volatility still stand out. Next, turning to the medium-sized models, while results under the point forecasts are still consistent with what we find with the Brent price, the density forecasts of real RAC provide strong evidence in favour of the VAR model with idiosyncratic stochastic volatility. This type of model outperforms other specifications at almost all horizon forecasts, especially under the production-based specification, and becomes the best model for forecasting real RAC among the three considered sets of data. Finally, with the large models, compared to the forecasting performance of the Brent price, we observe similar results. They show that the large VAR model, especially when it is incorporated with flexible errors, is the most useful model for forecasting real RAC at long horizons.

6 Conclusion

Recent research has shown that having large data and allowing flexible error structures are key ingredients in a VAR model for forecasting macroeconomic and financial variables. So far, however, there has been little discussion about the benefit of using those specifications in forecasting crude oil prices. In this paper, therefore, we evaluated the usefulness of the VAR model with different sizes and various distributional assumptions regarding the innovation in terms of their forecast ability for the real price of oil. We found the large models with 108 variables do not beat the small-sized (4 variables) and medium-sized

(14 variables) VARs at short-term forecasts. However, at longer forecasting horizons, the large-sized (108 variables) models become the most useful, suggesting the large VAR model is a valid alternative to the factor model not only in forecasting macroeconomic literature but also for predicting oil prices. Especially, in line with the literature, the results reveal that the benefit of using a large dataset will increase along with the use of unconventional error terms.

References

Alquist, R., S. Bhattarai, and O. Coibion (2020): "Commodity-price comovement and global economic activity," *Journal of Monetary Economics*, 112, 41–56.

Bai, J., and S. Ng (2008): *Large dimensional factor analysis*. Now Publishers Inc.

Bańbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector autoregressions," *Journal of Applied Econometrics*, 25(1), 71–92.

Baumeister, C., and J. D. Hamilton (2019): "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks," *American Economic Review*, 109(5), 1873–1910.

Baumeister, C., and L. Kilian (2012): "Real-time forecasts of the real price of oil," *Journal of Business & Economic Statistics*, 30(2), 326–336.

Baumeister, C., D. Korobilis, and T. K. Lee (2020): "Energy markets and global economic conditions," *The Review of Economics and Statistics*, 104(4), 1–45.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly Journal of Economics*, 120(1), 387–422.

Boivin, J., and S. Ng (2006): "Are more data always better for factor analysis?," *Journal of Econometrics*, 132(1), 169–194.

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

Carriero, A., T. E. Clark, and M. Marcellino (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212(1), 137–154.

Carriero, A., G. Kapetanios, and M. Marcellino (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25, 400–417.

Chan, J. C. (2013): "Moving Average Stochastic Volatility Models with Application to Inflation Forecast," *Journal of Econometrics*, 176(2), 162–172.

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

——— (2020b): *Large Bayesian Vector Autoregressions*, Macroeconomic forecasting in the era of big data. Springer.

——— (2021): "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37(3), 1212–1226.

Chan, J. C., and C. Hsiao (2014): *Estimation of stochastic volatility models with heavy tails and serial dependence*, Bayesian Inference in the Social Sciences. John Wiley & Sons, Hoboken.

Chiu, C.-W. J., H. Mumtaz, and G. Pinter (2017): "Forecasting with VAR models: Fat tails and stochastic volatility," *International Journal of Forecasting*, 33(4), 1124–1143.

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

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

Delle Chiaie, S., L. Ferrara, and D. Giannone (2022): "Common factors of commodity prices," *Journal of Applied Econometrics*, 37(3), 461–476.

Diebold, F. X., and R. S. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Econometric Statistics*, 13(3), 134–144.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distribution," *Econometric Reviews*, 3(1), 1–100.

Ferrari, D., F. Ravazzolo, and J. Vespignani (2021): "Forecasting energy commodity prices: A large global dataset sparse approach," *Energy Economics*, 98, 105268.

Gao, S., C. Hou, and B. H. Nguyen (2021): "Forecasting natural gas prices using highly flexible time-varying parameter models," *Economic Modelling*, 105, 105652.

Geweke, J. (1977): "The dynamic factor analysis of economic time series," in *Latent variables in socio-economic models*, pp. 365–383. North-Holland Publishing Company.

Gianfreda, A., F. Ravazzolo, and L. Rossini (2020): "Large time-varying volatility models for electricity prices," *Centre for Applied Macro- and Petroleum Economics (CAMP)*, Working Papers, No 05/2020.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior selection for vector autoregressions," *The Review of Economics and Statistics*, 97(2), 436–451.

Hamilton, J. D. (2021): "Measuring global economic activity," *Journal of Applied Econometrics*, 36(3), 293–303.

Hou, C., B. H. Nguyen, and B. Zhang (2022): "Real-time forecasting of the Australian macroeconomy using flexible Bayesian VARs," *Journal of Forecasting*, forthcoming.

Kilian, L. (2009): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 99(3), 1053–69.

Koop, G., and D. Korobilis (2010): "Bayesian multivariate time series methods for empirical macroeconomics," *Foundations and Trends in Econometrics*, 3(4), 267–358.

Koop, G. M. (2013): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28(2), 177–203.

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

Mumtaz, H., and K. Theodoridis (2018): "The changing transmission of uncertainty shocks in the US," *Journal of Business & Economic Statistics*, 36(2), 239–252.

Poon, A. (2018): "Assessing the synchronicity and nature of Australian state business cycles," *Economic Record*, 94(307), 372–390.

Ravazzolo, F., and J. Vespignani (2020): "World steel production: A new monthly indicator of global real economic activity," *Canadian Journal of Economics/Revue canadienne d'économique*, 53(2), 743–766.

Stock, J. H., and M. W. Watson (2002): "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, 20(2), 147–162.

——— (2016): "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," in *Handbook of macroeconomics*, vol. 2, pp. 415–525. Elsevier.

Table 1: A list of competing models.

Abbreviation	Full description
VAR	VAR with conventional error assumptions
VAR-SV	VAR with idiosyncratic stochastic volatility
VAR-CSV	VAR with a common stochastic volatility
VAR-CSV-t	VAR with t errors and a common stochastic volatility
VAR-CSV-MA	VAR with a common stochastic volatility and moving average errors
VAR-CSV-t-MA	VAR with t, moving average errors and a common stochastic volatility
VAR-MA	VAR with moving average errors
VAR-t	VAR with t errors
VAR-t-MA	VAR with t, moving average errors

Table 2: RMSFE and ALPL of the **production-based models** relative to no-change forecast of the real Brent price.

	relative RMSFE					relative ALPL				
	Monthly horizon									
	1	3	6	12	24	1	3	6	12	24
Small size (4 variables)										
VAR	**0.944**	**0.953**	0.988	0.991	0.973	0.506	0.098	0.002	**0.221**	**0.814**
VAR-SV	0.947	0.962	**0.975**	**0.966**	**0.944**	**0.568**	0.092	0.024	0.163	0.667
VAR-CSV	0.956	0.962	0.991	1.016	1.037	0.516	**0.124**	**0.064**	0.214	0.784
VAR-CSV-*t*	0.957	0.964	0.992	1.014	1.034	0.519	0.123	0.056	0.205	0.771
VAR-CSV-MA	0.955	0.959	0.993	1.028	1.068	0.510	0.099	0.018	0.179	0.671
VAR-CSV-*t*-MA	1.035	0.991	0.997	1.010	1.035	0.487	0.045	-0.065	0.036	0.615
VAR-MA	0.945	0.954	0.990	0.996	0.999	0.495	0.066	-0.004	0.189	0.686
VAR-*t*	0.957	0.966	0.994	1.012	1.030	0.503	0.114	0.034	0.201	0.778
VAR-*t*-MA	0.957	0.961	0.994	1.023	1.063	0.496	0.089	0.020	0.171	0.658
Medium size (14 variables)										
VAR	0.977	1.005	1.040	1.009	0.956	0.478	0.060	-0.082	**0.229**	**0.842**
VAR-SV	**0.948**	**0.964**	**0.976**	**0.961**	**0.946**	**0.567**	0.093	0.021	0.164	0.689
VAR-CSV	0.984	0.992	1.009	1.010	0.994	0.524	**0.108**	**0.040**	0.166	0.723
VAR-CSV-*t*	0.984	0.995	1.015	1.018	1.001	0.529	0.081	-0.018	0.138	0.703
VAR-CSV-MA	0.970	0.975	1.028	1.094	1.222	0.524	0.017	-0.151	-0.071	0.244
VAR-CSV-*t*-MA	1.020	1.004	1.011	1.023	1.014	0.504	0.011	-0.142	-0.030	0.519
VAR-MA	0.977	1.004	1.074	1.125	1.253	0.451	-0.064	-0.277	-0.062	0.330
VAR-*t*	0.990	1.012	1.037	1.024	0.988	0.488	0.080	0.018	0.186	0.789
VAR-*t*-MA	0.985	1.004	1.061	1.116	1.228	0.470	-0.025	-0.175	-0.104	0.240
Large size (108 variables)										
VAR	1.175	1.109	1.096	1.024	0.957	0.293	-0.138	-0.275	-0.128	0.653
VAR-CSV	1.180	1.020	0.936	0.920	0.887	0.450	0.044	-0.015	**0.143**	**0.784**
VAR-CSV-*t*	1.181	1.024	0.939	0.920	0.884	0.456	0.033	-0.038	0.083	0.743
VAR-CSV-MA	1.169	1.006	0.932	0.924	0.904	**0.461**	**0.050**	-0.018	0.125	0.737
VAR-CSV-*t*-MA	1.238	1.050	0.952	0.932	0.903	0.427	-0.010	-0.089	0.023	0.661
VAR-MA	1.250	1.333	1.382	1.467	2.146	0.298	-0.219	-0.409	-0.319	0.249
VAR-*t*	1.154	1.009	0.931	**0.910**	**0.863**	0.127	-0.482	-0.593	-0.442	0.513
VAR-*t*-MA	1.139	**0.991**	**0.924**	0.913	0.880	0.108	-0.533	-0.698	-0.574	0.402

Note: Values in bold and blue indicate the best relative RMSFE and ALPL among all model sizes and specifications. Boldface indicates the best specification among each model size. Dark grey cells and light grey cells indicate the significant difference of the predictive accuracy between an alternative models and the benchmark no-change forecast at 1% and 5% level of significance respectively based on the the related asymptotic test introduced by Diebold and Mariano (1995).

Table 3: RMSFE and ALPL of the **consumption-based models** relative to no-change forecast of the real Brent price.

	relative RMSFE					relative ALPL				
	Monthly horizon									
	1	3	6	12	24	1	3	6	12	24
	Small size (4 variables)									
VAR	**0.934**	**0.944**	0.968	0.989	0.972	0.523	0.094	-0.001	0.221	**0.812**
VAR-SV	0.946	0.956	**0.968**	**0.963**	**0.938**	**0.567**	0.095	-0.006	0.169	0.675
VAR-CSV	0.949	0.959	0.988	1.031	1.059	0.546	**0.151**	0.079	**0.226**	0.787
VAR-CSV-t	0.945	0.957	0.984	1.029	1.057	0.550	0.145	0.080	0.212	0.782
VAR-CSV-MA	0.950	0.961	0.993	1.044	1.096	0.536	0.121	0.048	0.185	0.644
VAR-CSV-t-MA	0.967	0.966	0.985	1.024	1.057	0.549	0.114	0.033	0.160	0.742
VAR-MA	0.936	0.949	0.972	0.992	0.986	0.514	0.056	0.034	0.206	0.726
VAR-t	0.947	0.960	0.985	1.018	1.038	0.527	0.124	0.062	0.211	0.785
VAR-t-MA	0.946	0.958	0.989	1.032	1.078	0.518	0.093	0.004	0.151	0.620
	Medium size (14 variables)									
VAR	0.981	1.006	1.032	1.013	0.957	0.485	0.048	-0.093	**0.224**	**0.835**
VAR-SV	**0.947**	**0.961**	**0.972**	**0.958**	**0.948**	**0.568**	0.089	0.017	0.182	0.702
VAR-CSV	0.979	0.992	1.007	1.015	0.995	0.527	**0.091**	**0.026**	0.163	0.729
VAR-CSV-t	0.977	0.993	1.012	1.022	1.002	0.525	0.061	-0.025	0.142	0.720
VAR-CSV-MA	0.967	0.977	1.027	1.094	1.218	0.524	0.007	-0.160	-0.071	0.263
VAR-CSV-t-MA	1.005	0.999	1.008	1.027	1.019	0.507	0.016	-0.106	-0.011	0.554
VAR-MA	0.982	1.011	1.070	1.126	1.248	0.462	-0.072	-0.274	-0.065	0.323
VAR-t	0.991	1.015	1.042	1.036	0.999	0.494	0.078	0.009	0.196	0.812
VAR-t-MA	0.986	1.010	1.070	1.134	1.276	0.476	-0.026	-0.173	-0.094	0.274
	Large size (108 variables)									
VAR	1.176	1.124	1.099	1.027	0.957	0.294	-0.175	-0.287	-0.142	0.646
VAR-CSV	1.183	1.022	0.936	0.922	0.891	0.449	0.044	-0.011	**0.145**	**0.780**
VAR-CSV-t	1.184	1.024	0.937	0.920	0.886	0.454	0.034	-0.032	0.088	0.735
VAR-CSV-MA	1.170	1.006	0.929	0.924	0.906	**0.460**	**0.050**	-0.013	0.131	0.728
VAR-CSV-t-MA	1.238	1.049	0.949	0.932	0.905	0.426	-0.005	-0.077	0.030	0.663
VAR-MA	1.255	1.345	1.418	1.509	2.209	0.292	-0.226	-0.424	-0.335	0.245
VAR-t	1.157	1.008	0.927	**0.908**	**0.864**	0.124	-0.448	-0.547	-0.393	0.508
VAR-t-MA	1.142	**0.992**	**0.921**	0.912	0.883	0.110	-0.539	-0.667	-0.545	0.375

Note: See Table 2.

Table 4: RMSFE and ALPL of the **production-based models** relative to no-change forecast of real RAC.

	relative RMSFE					relative ALPL				
	Monthly horizon									
	1	3	6	12	24	1	3	6	12	24
	Small size (4 variables)									
VAR	0.867	0.912	**0.967**	0.996	0.978	0.654	0.055	-0.041	0.218	**0.815**
VAR-SV	0.875	0.936	0.968	**0.967**	**0.944**	**0.714**	**0.138**	**0.097**	**0.250**	0.748
VAR-CSV	0.880	0.927	0.980	1.025	1.040	0.656	0.116	0.072	0.225	0.791
VAR-CSV-*t*	0.881	0.927	0.980	1.021	1.034	0.660	0.128	0.056	0.221	0.778
VAR-CSV-MA	0.878	0.920	0.995	1.084	1.192	0.649	0.075	-0.050	0.053	0.471
VAR-CSV-*t*-MA	0.966	0.965	0.989	1.015	1.034	0.600	0.021	-0.056	0.045	0.638
VAR-MA	**0.866**	**0.909**	0.978	1.042	1.090	0.645	-0.001	-0.125	0.020	0.521
VAR-*t*	0.884	0.931	0.982	1.020	1.036	0.637	0.110	0.055	0.216	0.782
VAR-*t*-MA	0.881	0.924	0.998	1.078	1.182	0.633	0.055	-0.010	0.071	0.467
	Medium size (14 variables)									
VAR	0.914	0.962	1.023	1.022	0.966	0.618	0.008	-0.113	0.229	**0.834**
VAR-SV	**0.876**	**0.938**	**0.969**	**0.965**	**0.950**	**0.716**	**0.141**	**0.103**	**0.255**	0.749
VAR-CSV	0.939	0.973	1.011	1.041	1.044	0.645	0.069	0.009	0.165	0.707
VAR-CSV-*t*	0.936	0.972	1.013	1.041	1.041	0.645	0.039	-0.042	0.139	0.685
VAR-CSV-MA	0.917	0.949	1.035	1.148	1.372	0.658	-0.024	-0.218	-0.117	0.159
VAR-CSV-*t*-MA	0.983	0.987	1.009	1.045	1.056	0.594	-0.074	-0.211	-0.080	0.465
VAR-MA	0.913	0.965	1.067	1.179	1.393	0.602	-0.130	-0.468	-0.121	0.237
VAR-*t*	0.928	0.971	1.020	1.040	1.011	0.613	0.066	0.012	0.180	0.772
VAR-*t*-MA	0.921	0.966	1.054	1.161	1.351	0.603	-0.054	-0.202	-0.149	0.150
	Large size (108 variables)									
VAR	1.073	1.024	1.032	1.014	0.938	0.457	-0.116	-0.225	-0.083	0.684
VAR-CSV	1.123	0.960	0.881	0.897	0.881	0.555	0.091	**0.067**	**0.203**	**0.814**
VAR-CSV-*t*	1.124	0.961	0.884	0.898	0.878	0.559	0.067	0.032	0.139	0.773
VAR-CSV-MA	1.100	0.940	0.878	0.899	0.900	**0.577**	**0.096**	0.059	0.191	0.768
VAR-CSV-*t*-MA	1.170	0.979	0.890	0.905	0.893	0.535	0.029	-0.014	0.079	0.693
VAR-MA	1.151	1.181	1.266	1.423	1.976	0.436	-0.153	-0.320	-0.259	0.280
VAR-*t*	1.099	0.952	0.879	**0.890**	**0.860**	0.245	-0.498	-0.533	-0.423	0.586
VAR-*t*-MA	1.077	**0.933**	**0.876**	0.894	0.880	0.245	-0.571	-0.639	-0.504	0.424

Note: See Table 2.

Table 5: RMSFE and ALPL of the **consumption-based models** relative to no-change forecast of real RAC.

	relative RMSFE					relative ALPL				
	Monthly horizon									
	1	3	6	12	24	1	3	6	12	24
Small size (4 variables)										
VAR	0.863	0.912	**0.955**	0.996	0.980	0.664	0.010	-0.073	0.223	**0.809**
VAR-SV	0.873	0.929	0.963	**0.967**	**0.944**	**0.717**	0.142	0.100	0.245	0.748
VAR-CSV	0.882	0.937	0.987	1.052	1.081	0.685	**0.150**	0.093	**0.245**	0.789
VAR-CSV-*t*	0.874	0.926	0.975	1.043	1.073	0.684	0.148	**0.110**	0.224	0.776
VAR-CSV-MA	0.880	0.932	1.003	1.103	1.234	0.677	0.108	0.009	0.118	0.477
VAR-CSV-*t*-MA	0.920	0.950	0.983	1.039	1.072	0.653	0.073	0.015	0.137	0.722
VAR-MA	**0.862**	**0.912**	0.966	1.022	1.041	0.658	-0.029	-0.125	0.148	0.633
VAR-*t*	0.881	0.935	0.982	1.035	1.061	0.658	0.109	0.054	0.211	0.781
VAR-*t*-MA	0.877	0.930	0.999	1.087	1.204	0.651	0.055	-0.044	0.059	0.452
Medium size (14 variables)										
VAR	0.920	0.972	1.017	1.026	**0.968**	0.621	-0.007	-0.122	0.228	**0.833**
VAR-SV	**0.873**	**0.934**	**0.965**	**0.967**	0.979	**0.718**	**0.140**	**0.094**	**0.250**	0.748
VAR-CSV	0.935	0.975	1.010	1.051	1.056	0.650	0.061	0.005	0.164	0.711
VAR-CSV-*t*	0.931	0.972	1.010	1.052	1.058	0.641	0.015	-0.034	0.127	0.677
VAR-CSV-MA	0.914	0.954	1.032	1.146	1.356	0.659	-0.059	-0.247	-0.100	0.164
VAR-CSV-*t*-MA	0.969	0.984	1.008	1.053	1.073	0.597	-0.088	-0.203	-0.076	0.514
VAR-MA	0.920	0.974	1.063	1.170	1.363	0.610	-0.172	-0.433	-0.213	0.265
VAR-*t*	0.931	0.983	1.032	1.064	1.041	0.619	0.053	0.000	0.196	0.795
VAR-*t*-MA	0.924	0.977	1.066	1.182	1.401	0.610	-0.057	-0.209	-0.141	0.178
Large size (108 variables)										
VAR	1.081	1.045	1.038	1.017	0.938	0.444	-0.171	-0.219	-0.082	0.686
VAR-CSV	1.126	0.959	0.879	0.900	0.886	0.552	0.089	**0.070**	**0.200**	**0.806**
VAR-CSV-*t*	1.127	0.963	0.882	0.899	0.882	0.556	0.066	0.037	0.136	0.768
VAR-CSV-MA	1.102	0.941	0.876	0.901	0.903	**0.575**	**0.096**	0.066	0.189	0.760
VAR-CSV-*t*-MA	1.171	0.978	0.887	0.905	0.895	0.533	0.029	-0.006	0.079	0.696
VAR-MA	1.163	1.203	1.293	1.451	1.988	0.424	-0.166	-0.331	-0.275	0.275
VAR-*t*	1.104	0.954	0.876	**0.890**	**0.862**	0.230	-0.524	-0.530	-0.424	0.543
VAR-*t*-MA	1.081	**0.935**	**0.873**	0.893	0.881	0.251	-0.534	-0.599	-0.480	0.416

Note: See Table 2.

Appendix A Estimation and Priors

A.1 Estimation

The posterior estimation for parameters of the BVARs can be obtained by sampling sequentially by Markov chain Monte Carlo (MCMC) methods. Here, we take the estimation of parameters in VAR-CSV-t-MA as an example. There are seven steps in one loop of posterior draws for each parameter. Specifically, the posterior draws are obtained for the coefficients of VAR $\mathbf{B}$, the cross-sectional covariance matrix $\mathbf{\Sigma}$, the hyperparameter λ_t and ν of t distribution, the stochastic volatility parameter h and the related truncated normal parameter ρ and variance σ_h^2, and the moving average coefficient ψ. The simulation can be implemented as below:

1. $p(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{Y}, \lambda_t, \mathbf{h}, \sigma_h^2, \rho_h, \psi_{\boldsymbol{\varepsilon}}, \nu_\lambda)$;
2. $p(\lambda_t \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}, \mathbf{h}, \sigma_h^2, \rho_h, \psi_{\boldsymbol{\varepsilon}}, \nu_\lambda)$;
3. $p(\nu_\lambda \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}, \lambda_t, \mathbf{h}, \sigma_h^2, \rho_h, \psi_{\boldsymbol{\varepsilon}})$;
4. $p(\mathbf{h} \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}, \lambda_t, \sigma_h^2, \rho_h, \psi_{\boldsymbol{\varepsilon}}, \nu_\lambda)$;
5. $p(\sigma_h^2 \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}, \lambda_t, \mathbf{h}, \rho_h, \psi_{\boldsymbol{\varepsilon}}, \nu_\lambda)$;
6. $p(\rho_h \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}, \lambda_t, \mathbf{h}, \sigma_h^2, \psi_{\boldsymbol{\varepsilon}}, \nu_\lambda)$;
7. $p(\psi_{\boldsymbol{\varepsilon}} \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}, \lambda_t, \mathbf{h}, \sigma_h^2, \rho_h, \nu_\lambda)$;

In the first step, given that the coefficients and covariance matrix are the natural conjugate prior, the joint posterior distribution of $(\mathbf{B}, \mathbf{\Sigma})$ is a normal-inverse-Wishart distribution, so the posterior draws can be obtained from their posterior distribution directly.

The second and third steps draw the parameter λ_t and ν_λ for t distribution which can be written as a scale mixture of Gaussian distribution. This multivariate t distribution has a mean vector $\mathbf{0}$, scale matrix $\mathbf{\Sigma}$ and degree of freedom ν, and $(\lambda_t \,|\, \nu_\lambda)$ follows an inverse-gamma distribution. Then, we have $\mathbf{\Omega} = \text{diag}(\lambda_1, ..., \lambda_T)$. The hyperparameter ν_λ in the inverse-gamma distribution of λ_t can be sampled by an independence-chain Metropolis-Hastings step described in Chan and Hsiao (2014).

The following three steps are related to the common stochastic volatility parameter $\mathbf{h}$ and its hyperparameter σ_h^2 and ρ_h. The simulation of common stochastic volatility can follow Carriero, Clark, and Marcellino (2016) and the models are assumed to have a stationary AR(1) stochastic volatility. We assume that σ_h^2 has an inverse-gamma prior and ρ_h has an independent truncated normal distribution. Then the posterior distribution of parameter $\mathbf{h}$ can be obtained by implementing the Newton-Raphson algorithm and the acceptance-rejection Metropolis-Hastings step.

Lastly, the posterior distribution of moving average parameter ψ_ε can be sampled by an independence-chain Metropolis-Hastings step, while the related estimation method and efficient algorithm are discussed in Chan (2013).

A.2 Priors

The selection of priors is a crucial step in the estimation of BVARs, as the number of coefficients that need to be estimated can be a great amount. This overparameterization problem can be eliminated by using informative priors or regularization. In the setup of coefficient prior, the Minnesota Prior is considered in the standard VARs and VARs with idiosyncratic stochastic volatility, and the natural conjugate prior is used in the VAR-CSV with various flexible covariance structures.

The Minnesota prior is firstly introduced with small VARs by Doan, Litterman, and Sims (1984). It uses an approximation $\widehat{\sigma^2}$ for error covariances in each VAR equation by OLS estimation, so it is not limited by the size of VARs and can be applied to large BVARs. In the prior distribution of the coefficients, the means and the variances imposed distributions associated with the lag length l of the variable's own lag and the lag of another variable. Specifically, a modified version is used which is discussed in Koop and Korobilis (2010):

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_{Minn}, \mathbf{V}_{Minn}), \tag{9}$$

$$\mathbf{V}_{Minn} = \begin{cases} \kappa_1 & \text{for intercept,} \\ \kappa_2/l^2 & \text{for own lags,} \\ \kappa_3\widehat{\sigma_i^2}/(l^2\widehat{\sigma_j^2}) & \text{otherwise,} \end{cases} \tag{10}$$

where $\boldsymbol{\beta}_{Minn} = 0$ indicates that growth rate data are used and they are stationary time series. $\mathbf{V}$ is the variance operator, κ_1, κ_2 and κ_3 are hyperparameters of $\mathbf{V}_{Minn}$.

With regards to $\mathbf{V}_{Minn}$, more reliable information is provided by more recent lags which should be given more weight in the estimation. In practice, the value of $\mathbf{V}_{Minn}$ is smaller when the lag length l turns larger. Besides, the value of $\mathbf{V}_{Minn}$ is also controlled by the ratio of prior variance from two variables. For the cross lags, it is supposed that the lags of other variables can not explain more variation of one variable than its own lags, so the $\mathbf{V}_{Minn}$ of cross lags should be smaller than that of its own lags.

In the application part, for the standard BVARs with the Minnesota prior, the hyperparameters of the variance operator are set to be $\kappa_1 = 10^2, \kappa_2 = 0.2^2$ and $\kappa_3 = 0.1^2$, where

κ_2 is bigger than κ_3 indicating that variables' own lags are more important than their cross lags. With the Minnesota prior, the BVARs are models with constant variances, then a two-step Gibbs sampler can be used to estimate the models. The VAR coefficients $\boldsymbol{\beta}$ are drawn from a conditional posterior distribution that is multivariate normal in the first step, and the covariance matrix $\boldsymbol{\Sigma}$ is simulated from an inverse Gamma distribution in the second step. The additional detail on algorithms and priors can be found in Koop and Korobilis (2010).

The setting of Minnesota prior provides a way of shrinkage for the standard VARs with a considerable amount of coefficient, but the parameters of Minnesota prior are restricted to be fixed and the covariance matrix is diagonal. To cover these concerns, alternative priors in the sensitivity analysis section introduce hyperparameters or other flexible specifications on the covariance matrix to the VAR models.

The natural conjugate prior is used as the prior of VAR-CSV with flexible covariance structures, which assumes that the error covariance matrix of BVARs is an unknown symmetric matrix. It can be considered as the Minnesota prior with a normal-inverted-Wishart assumption on the error covariance matrix $\boldsymbol{\Sigma}$ instead of a fixed diagonal matrix. This prior takes into account the uncertainty of the error covariance matrix. Moreover, it is computationally tractable and has a closed form of the marginal likelihood compared with the Minnesota prior. The normal-inverted-Wishart prior takes the following form:

$$\mathbf{B}|\boldsymbol{\Sigma} \sim \mathcal{N}(\mathbf{B}_0, \boldsymbol{\Sigma} \otimes \mathbf{V}_B), \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{\nu}_0, \mathbf{S}_0), \tag{11}$$

where $\mathbf{B}_0, \mathbf{V}_B, \boldsymbol{\nu}_0$ and $\mathbf{S}_0$ are prior hyperparameters of Normal distribution and inverted Wishart distribution, parameters with the subscript 0 stand for those of the prior distributions. Equation (11) can be written as:

$$(\mathbf{B}, \boldsymbol{\Sigma}) \sim \mathcal{NIW}(\mathbf{B}_0, \mathbf{V}_B, \boldsymbol{\nu}_0, \mathbf{S}_0). \tag{12}$$

For models with the natural conjugate prior, parameters of larger lag lengths are conducted higher degree of shrinkage. However, there is no difference between the prior variances of variables' own lags and other lags compared with the feature of Minnesota prior. The natural conjugate prior gives the same degree of shrinkage of variables' own lags and other lags, thus, they share the same tightness hyperparameter on variables' lags. The detailed algorithm for BVARs with the natural conjugate prior is described in Giannone, Lenza, and Primiceri (2015) and Carriero, Kapetanios, and Marcellino (2009).

In the application part, We set $\boldsymbol{\beta}_0 = \mathbf{0}$, the hyperparameters for the covariance matrix $\mathbf{V}_0$ are $\kappa_1 = 10^2, \kappa_2 = 0.2^2$ (so that the parameters with larger lag lengths are conducted higher degree of shrinkage, which is consist with the setting of the Minnesota prior), $\boldsymbol{\nu}_0 = n + 3$ and $\mathbf{S}_0 = diag(s_1^2, ..., s_n^2)$ (where $s_1^2, ..., s_n^2$ are obtained from the standard

OLS estimates of the error variance for each equation). To estimate the models with the natural conjugate prior, the Kronecker structure of the posterior covariance matrix can be considered for fast simulation. This approach is based on the algorithm of drawing posteriors from the matrix normal distribution. As the posterior distributions of the VAR coefficients $\boldsymbol{\beta}$ and the covariance matrix $\boldsymbol{\Sigma}$ have the same distributions of the priors, $\boldsymbol{\Sigma}$ can be drawn marginally from an inverse gamma distribution, then $\boldsymbol{\beta}$ can be simulated from a normal distribution.