ISSN 1178-2293 (Online)



BUSINESS SCHOOL Te Kura Pakihi University of Otago Economics Discussion Papers No. 2301

JANUARY 2023

Bayesian VARs of the U.S. Economy Before and During the Pandemic

Anna Sznajderska Warsaw School of Economics, al. Niepodległości 162, 02-554 Warsaw, Poland Email: asznajd@sgh.waw.pl Alfred A. Haug Department of Economics, University of Otago, Dunedin 9054, New Zealand Email: alfred.haug@otago.ac.nz

Address for correspondence:

Alfred Haug Department of Economics University of Otago PO Box 56 Dunedin NEW ZEALAND Email: alfred.haug@otago.ac.nz Telephone: 64 3 479 5636

Bayesian VARs of the U.S. Economy Before and During the Pandemic

Anna Sznajderska Warsaw School of Economics, al. Niepodległości 162, 02-554 Warsaw, Poland Email: asznajd@sgh.waw.pl Alfred A. Haug Department of Economics, University of Otago, Dunedin 9054, New Zealand Email: alfred.haug@otago.ac.nz

Abstract

We compare the forecasting performance of small and large Bayesian vector-autoregressive (BVAR) models for the United States. We do the forecast evaluation of the competing models for the sample that ends before the pandemic and for the sample that contains the pandemic period. The findings document that these models can be used for structural analysis and generate credible impulse response functions. Furthermore, the results indicate that there are only small gains from the application of a large BVAR model compared to a small BVAR model.

Keywords: COVID-19, Bayesian VAR models, impulse response functions, forecasting.

JEL codes: C32, E32, F10, O50

Vector Autoregressions (VARs) are one of the most popular tool within central banks and academia to analyse and forecast economic developments. The VAR models provide a very general representation of a statistical model, often in linear form. VARs allow the capture of complex dynamic data relationships, because in contrast to other models, such as structural economic models, they do not impose economic restrictions on the parameters.¹ Broadly speaking, the variables in the VAR model depend on their own lags and other model variables' lags.

The problem is that the high level of generality implies a large number of parameters to be estimated, whereas the typical sample size for macroeconomic applications is rather small. This entails a risk of overparameterization, resulting in parameter instability and a loss of precision for forecasting. On the other hand, reducing the number of parameters by cutting back on variables in a VAR may potentially lead to an omitted variable bias with adverse consequences both for structural analysis and forecasting. Furthermore, frequently too many lags are included in a VAR model to improve the in-sample fit, which results in a significant loss of degrees of freedom and poor out-of-sample forecast performance.²

The recent literature has proposed a number of ways to overcome these problems. There are three main econometric recommendations. The first is to apply factor models, which are based on the assumption that interrelations within a large dataset can be explained by a few common factors. The second is to use high frequency data, which may be used to identify the shocks from outside the model. The third is to take the Bayesian point of view, because the supply of prior information limits the overparameterization issue.³

The aim of this paper is to assess the performance of the Bayesian VAR (BVAR) for monetary models during the COVID-19 pandemic. The idea is to model the pandemic shocks and capture the adjustment process of the economy through the BVAR model parameters, without imposing a structural change in those parameters. Assuming in contrast a structural break would imply that the model parameters have changed due to the pandemic and that economic relationships are different after the start of the pandemic. A serious drawback of such conventional breaks is that they are treated as having zero probability of occurring again in the future and are thus ignored in forecasting (Hamilton, 2016). Instead, we use relatively flexible and not very restrictive formulations of the statistical distributions for the regression parameters and errors to capture the pandemic shocks, as well as the shocks due to the recent global financial crisis. The statistical tool we use for that purpose is Bayesian modelling.

This paper formulates and estimates small and large BVAR models for the U.S. economy before and during the pandemic. It contributes to the recent literature on BVARs that proposes ways to deal with the COVID-19 episode in macro-econometric VAR models. Schorfheide and Song (2021) drop observations during the early part of the pandemic (March – June 2020) in a

¹ An advantage over so-called dynamic stochastic general equilibrium (DSGE) macroeconomic models is that VARs impose less economic theory. See Pagan and Wickens (2022) on the advantages and disadvantages of each type of modelling approach and on how VARs relate to DSGE models.

² See, for example, Hamilton (1994).

³ See, for example, Kilian and Lütkepohl (2017) on the various approaches and their advantages and disadvantages.

mixed frequency BVAR with monthly and quarterly observations. Lenza and Primiceri (2022), Carriero et al. (2022), and Bobeica and Hartwig (2022) account in alternative ways for the excess variability of different macroeconomic series during the pandemic period. Lenza and Primiceri (2022) inversely weigh the observations during the pandemic period based on the innovation variances, using a standard homoskedastic VAR. On the other hand, Bobeica and Hartwig (2022) use a multivariate t-distribution for errors in a small-scale BVAR. Carriero et al. (2022) use a mid-sized BAVR with variable-specific outlier-adjusted stochastic volatility and t-distributed innovations. The idea behind using a distribution with fatter tails, i.e., the t-distribution, is to capture large shocks, such as those occurring during the recent pandemic. Instead, we use a recently proposed flexible Bayesian methodology for small- and large-scale VARs, developed and successfully applied in the pre-pandemic period to forecasting U.S. macroeconomic time series by Giannone et al. (2015).⁴ In other words, we explore whether a BVAR model used for forecasting with good results before the pandemic is able to forecast well once the pandemic period is included, without dropping observations, outlier adjustments, or assuming a t-distribution.

Giannone et al. (2015) build on the approach in Bańbura et al. (2010) that formulates a BVAR based on the conjugate normal-inverse Wishart family for the variance-covariance matrix and the VAR coefficients. To estimate the model, our paper uses the BEAR toolbox of the European Central Bank (Dieppe et al. 2016, Dieppe and van Roye 2021). Our paper divides the variables into fast-moving financial variables and slow-moving real variables and prices, with the identifying assumption that the slow-moving variables do not respond contemporaneously to a monetary policy shock and the information set of the monetary policy contains only past values of the fast-moving variables. We consider a small and a large BVAR model. The small model includes industrial production, the consumer price index and a measure of the Federals Funds Rate. The large model includes employment, housing and other financial indicator data. Next, we conduct impulse response function and forecasting analyses with samples that include and exclude the pandemic.

Section 2 presents a short review of related Bayesian research. Section 3 introduces the basic features of our BVAR model and the data employed. Section 4 discusses the results and Section 5 concludes.

2. A brief review of recent related literature

While BVAR models started out with small models, recent research has developed ways to deal with large BVAR models (Bańbura et al. 2010, Kapetanios et al. 2012, Bobeica and Jarociński 2019, and Crump et al. 2021). Our econometric approach is based primarily on the work of Bańbura et al. (2010). They describe a large BVAR model with Bayesian shrinkage as an appropriate tool for forecasting and structural analysis. Their dataset covers the period 1959:01 to 2003:12 and hence does not include developments in regards to the pandemic. Also, new shrinkage methods,

⁴ See also Crump et al. (2021) for a recent use of this methodology, though not dealing with the pandemic.

such as the procedure of Giannone, Lenza and Primiceri that is developed in Giannone et al. (2015), were introduced after the publication of their article.

A contemporary paper on large BVAR models is Crump et al. (2021). The paper is centred around the concept of conditional forecasting. The authors use quarterly data from 1973:Q1 to 2018:Q3. It is worth noting that our paper considers unconditional forecasting instead and uses monthly data between 1990:01 and 2021:11. Crump et al. (2021) define the impulse response functions as the difference between conditional and unconditional forecasts. They compare their forecasts with those of professional forecasters and Greenbook forecasts. The authors conclude that a large-scale BVAR model produces reliable predictions of the joint distribution for a large set of macroeconomic and financial indicators monitored by the Federal Reserve staff and professional forecasters.

We briefly describe a few papers closely related to our research. Carriero et al. (2011) compare the forecasting performance of numerous BVAR models. Their general finding is that there are only very small losses from the adoption of BVAR modelling choices that make forecast computation quick and easy. An approach that works well is to specify a normal-inverted Wishart prior along the lines of Sims and Zha (1998) for the VAR in levels, preferably optimizing its tightness and lag length. Also they argue that specifications in levels benefit substantially from the imposition of the sum-of-coefficients and dummy initial observation priors.

Kapetanios et al. (2012) apply a large BVAR model (with 43 variables), as one of three different alternative models, to examine the effects of unconventional monetary policy in the UK.⁵ The BVAR model is estimated over rolling windows to allow for structural change. Their work is based on the assumption that quantitative easing (QE) reduced medium- to long-term government bond yields by about 100 basis points in the UK. They estimate the effects of QE on real GDP and inflation using counterfactual scenarios. The results show that without QE real GDP and inflation would have fallen even more during 2009.

Several recent papers deal with the COVID-19 pandemic in BVAR models, such as Lenza and Primiceri (2022) and Bobeica and Hartwig (2022). Lenza and Primiceri (2022) apply the BVAR model with observations weighed inversely proportional to their innovation variance. In other words, less weight is assigned to observations from the pandemic period.⁶ Their reasoning is that the COVID-19 pandemic caused unprecedented variation in many key macroeconomic variables. The U.S. unemployment rate, for instance, increased by approximately 10 percentage points from March to April 2020, which is two orders of magnitude more than its typical monthly change. Their VAR model includes six variables for the U.S. economy: unemployment, employment and four price indices. They come to the following two conclusions. The first is that the strategy of dropping the COVID-19 observations may be acceptable for the purpose of

⁵ They also use a change-point structural VAR (SVAR) model and a time-varying parameter SVAR model with sign restrictions. We prefer not to use sign restrictions for the impulse response functions and let instead the data speak for themselves. Sign restrictions require specifying periods affected by the restrictions, along with other assumptions (see, e.g., Fry and Pagan, 2011).

⁶ Along similar lines, Carriero et al. (2022) down-weigh instead the variance of the residuals of outlier observations.

parameter estimation. The second is that the strategy of dropping the COVID-19 observations is inappropriate for forecasting the future evolution of the economy.

An alternative possibility to deal with the COVID-19 pandemic is to relax the assumption of normal errors and allow them to follow a fat-tailed distribution as recommended by Bobeica and Hartwig (2022). They show that this strategy, assuming a multivariate t-distribution for the BVAR errors, ensures more stable parameters and forecasts for the model that includes the COVID-19 pandemic. Moreover, they demonstrate that a large standard BVAR with normal errors, but more prior shrinkage than usual, ensures stable unconditional forecasts. They emphasize the necessity to check how the COVID-19 observations affect the parameter estimates and the implied forecasts. Additionally, they point out that as new observations on the variables become available, the distortion of traditionally estimated parameters diminishes. However, they find that a simple BVAR model with normal and homoskedastic errors cannot cope with the outlier observations in COVID-19 period.

3. Research methodology

In this section, we briefly outline our research methodology. First, we discuss BVAR models. Second, we describe the data used in the paper. Our monthly U.S. data cover the period from 1990:01 to 2021:11. The start date is chosen based on data availability. The end date is the most recent month available when this research was started.

3.1 The Bayesian VAR model

Bayesian VARs deal with the problem of overparameterization (the curse of dimensionality) of VAR models by incorporating in the estimation and forecasting process prior information. It is information beyond that contained in the data itself and takes the form of inexact prior restrictions. VAR coefficients are treated as being random variables around their prior means. The tightness of the distribution is controlled for with so-called hyper-parameters. Therefore, one must specify the form of the prior distribution and also the covariance matrix of the regression errors.⁷ Similarly, the degree of uncertainty about the long-run stochastic trends (unit roots) enters explicitly when model parameters are estimated.⁸ The Bayesian VAR literature has suggested various alternative priors over the years. Recent research has developed algorithms that do not require the VAR error covariance matrix to be fixed or diagonal (Robertson and Tallman 1999). In our approach we use a prior distributions that belong to the conjugate normal-inverse Wishart family and we apply the shrinkage methods of Giannone et al. (2015). They advocate using prior information that shrinks large BVAR models towards parsimonious representations.

We apply the following baseline BVAR model with n endogenous variables, p lags, and m exogenous variables.

⁷ See, for example, Robertson and Tallman (1999) for a concise introduction to forecasting with BVARs.

⁸ This means that exact unit roots are not imposed on the VAR (Robertson and Tallman 1999)

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + C x_t + \varepsilon_t,$$
$$\varepsilon_t \sim N(0, \Sigma),$$

where y_t is a $n \times 1$ vector of endogenous variables, $A_1, \dots A_p$ are $p \ n \times n$ matrices of parameters, *C* is a $n \times m$ matrix, x_t is a $m \times 1$ vector of exogenous variables (it may contain constant terms), and ε_t is a $n \times 1$ vector of errors, following a multivariate normal distribution with zero mean and covariance matrix Σ .

We define the vectors $\beta = vec([A_1, A_2, ..., A_p, C]')$, $y = vec([y_1, y_2, ..., y_T]')$, and $\varepsilon = vec([\varepsilon_1, \varepsilon_2, ..., \varepsilon_T]')$ for any sample size *T*. Further,

$$X = \begin{pmatrix} y'_0 & y'_{-1} & \cdots & y'_{1-p} & x'_1 \\ y'_1 & y'_0 & \cdots & y'_{2-p} & x'_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y'_{T-1} & y'_{T-2} & \cdots & y'_{T-p} & x'_T \end{pmatrix}$$

$$\overline{X} = I_n \otimes X$$

Now, the previous equation (following Dieppe et al. 2016) can be reformulated as:

$$y = X\beta + \varepsilon$$
,

with β a $q \times 1$ vector, where $q = n(np + m) \times 1$. $E(\varepsilon_t \varepsilon'_t) = \Sigma$, while $E(\varepsilon_t \varepsilon'_s) = 0$ for $t \neq s$. Σ is a $n \times n$ symmetric positive definite covariance matrix. Taking the Bayesian perspective, one needs to specify the priors for β and for the covariance matrix Σ . The procedure is based on a modified version of Litterman's (1986) approach. The equations are centred around a random walk with drift (cf. equation (2) in Bańbura et al. 2010). The prior specification incorporates the belief that the more recent lags should provide more reliable information than the more distant ones. Also, own lags should explain more of the variation of the given variable than the lags of other variables in the equation. It is assumed that little is known about exogenous variables, so that the variance of these terms should be large. To account for correlation among the residuals of different variables, we use the inverse Wishart prior for Σ . Thus, in the baseline specification, we focus on prior distributions that belong to the conjugate normal-inverse Wishart family:

$$\Sigma \sim \mathrm{IW}(\Psi, \mathrm{d}),$$

$$\beta \sim N(b, \Sigma \otimes \Omega)$$

We want to compare the forecasting performance and impulse responses from small and large BVAR models. Technically a large model with about 40 endogenous variables is not possible to estimate using the proposed prior structure (Minnesota structure). Dieppe et al. (2016) calculate that in the case of a model with 40 endogenous variables, 5 exogenous variables and 15 lags, q = 24200, which implies that each iteration of the Gibbs sampler requires the inversion of a 24200 ×

24200 matrix.⁹ This makes the process very slow and practically intractable. To partly solve this problem we use the dummy observation prior with two extensions, the sum of coefficients prior and the dummy initial observations. If we write the VAR model in its error-correction form we get:

$$y_{t} = (I_{n} - A_{1} - \dots - A_{p})y_{t-1} + B_{1}\Delta y_{t-1} + \dots + B_{p-1}\Delta y_{t-p+1} + Cx_{t} + \varepsilon_{t}$$

The sum of coefficient prior shrinks $I_n - A_1 - \dots - A_p$ to 0. The procedure is implemented by adding the specific dummy observations (Bańbura et al. 2010).

We use for the majority of our calculations the European Central Bank's (ECB's) toolbox *Bayesian Estimation, Analysis And Regression* (BEAR) version 5.1.3.¹⁰ We apply variable specific priors for the autoregressive coefficients. We follow Giannone et al.'s (2015) approach for the construction of the prior. If λ controls the overall tightness of the prior distribution, then for $\lambda = 0$ the posterior equals the prior and the data do not influence the estimates and for $\lambda = \infty$ the posterior coincides with ordinary least squares estimates. Bańbura et al. (2010) argue that the more variables we include in the VAR model, the more λ should shrink in order to avoid overfitting.

3.2 Data

Monthly data for the U.S. economy are used to set up a small and a large BVAR model. The list of the variables with their descriptions, sources and transformations is in the Appendix. We consider the following two VAR models:

- 1) The *small model* that includes: Industrial Production: Total Index (INDPRO), Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL), and the Wu-Xia Shadow Federal Funds Rate.
- 2) The *large model* that extends the small model with the following variables:
 - [°] All Employees, Total Nonfarm (LOG_PAYEMS),
 - ° Unemployment Rate (UNRATE),
 - ° Capacity Utilization: Manufacturing (MCUMFN),
 - ° New Privately-Owned Housing Units Started: Total Units (LOG_HOUST),
 - ° University of Michigan: Consumer Sentiment (UMCSENT),
 - ° Personal Consumption Expenditures Chain-Type Price Index (LOG_PCEPI),
 - [°] Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average (LOG_CPILFESL),
 - ° Personal Consumption Expenditures Excluding Food and Energy (LOG_PCEPILFE),
 - [°] Compensation of Employees, Received: Wage and Salary Disbursements (LOG_COMPENSATION),
 - ° Crude Oil Prices: West Texas Intermediate (WTI) Cushing, Oklahoma (LOG_OIL),

⁹ The Gibbs sampling method draws randomly from the unconditional posterior distribution of the parameters. It is part of the class of so-called *Markov Chain Monte Carlo* Bayesian simulation methods. For an outline of the Gibbs procedure see Dieppe et al. (2016).

¹⁰ Available at <u>https://www.ecb.europa.eu/pub/research/working-papers/html/bear-toolbox.en.html</u>, last accessed 20 July 2022. The methodology for BEAR is developed in Dieppe et al. (2016).

- ° Moody's Seasoned Aaa Corporate Bond Yield (DAAA),
- ° Moody's Seasoned Baa Corporate Bond Yield (DBAA),
- ° Market Yield on U.S. Treasury Securities at 2-Year Constant Maturity (DGS2),
- ° Market Yield on U.S. Treasury Securities at 5-Year Constant Maturity (DGS5),
- ^o Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity (DGS10),
- ° Nominal narrow BIS effective exchange rate (NEER2),
- ° S&P 500 index (SP_CLOSE).

We choose the variables guided by previous large BVAR studies discussed in Section 2 and the availability of data. We note that when the Wu-Xia shadow interest rate is above 1/4 percent, it is exactly equal to the Wu-Xia model implied one-month interest rate by construction. Also, we decided to omit the real personal consumption expenditures series (PCEC96), because the series is available only from 2002:01. For the same reason we omit the real broad Bank for International Settlements (BIS) effective exchange rate and the nominal broad BIS effective exchange rate, which starts in 1994:01.

We follow Bańbura et al. (2010) and divide the variables into slow and fast moving. This means that the former group contains real variables and prices, while the latter consists of financial variables. The assumption that is made here is that slow-moving variables do not respond contemporaneously to a monetary policy shock and that the information set of the monetary authority contains only past values of the fast-moving variables.

4. Results

4.1 Unit root test results

We apply standard Augmented Dickey-Fuller unit root tests only to get an indication about the dynamic persistence (memory) of shocks to variables. This helps sorting variables into slow and fast moving processes. Table 1 reveals that most of the variables are nonstationary in levels, with covariance-stationary first differences, i.e., with unit roots in levels. There are only two series which appear to be covariance stationary in levels, namely these are the unemployment rate (UNRATE) and the University of Michigan consumer sentiment (UMCSENT). We set the autoregressive *prior* coefficient δ_i equal to 1 for nonstationary variables and equal to 0.8 for stationary variables. This is based on the assumption that each nonstationary variable follows an independent random walk process, potentially with drift.

4.2 Impulse response functions

First we estimate the small VAR model with only 3 variables (industrial production, the consumer price index, and the interest rate) and 13 lags. We test whether it is better to use a federal funds effective rate (Figure 1) or a shadow rate (Figure 2). One may observe that the positive shadow rate impulse causes a statistically significant decrease in inflation, whereas the positive federal funds rate impulse leads to statistically insignificant reactions of prices. Also, the positive

shadow rate impulse does not cause an increase in the level of industrial production as does the positive federal funds rate impulse.¹¹ If we restrict the sample to end before the pandemic we end with the same conclusion that the impulse response functions are more reasonable for the shadow rate than for the federal funds effective rate. This initial experiment recommends using the shadow rate in further VAR models as the appropriate monetary policy tool.

Next, we move to estimating BVAR models.¹² In the first step we redo the experiment with the smaller model. We obtain Figures 3 and 4. The results show that an increase in the Wu-Xia shadow rate and federal funds rate lead to a decrease in the CPI, but the response of industrial production is positive, contrary to economic theory. In the second step we estimate the large BVAR model consisting of 20 variables (Figure 5). In this case a monetary policy impulse leads to a short lived increase in industrial production $(2^{nd} - 3^{rd} \text{ month after the impact of the shock})$ and a decrease in the consumer price index $(3^{rd} - 11^{th} \text{ month})$. It is worth emphasizing that the monetary policy impulse becomes statistically insignificant after 37 months for the large BVAR model, and that this is not the case for the small BVAR model. Thus, we find the impulse response functions for the large BVAR model are the most reliable.

4.3 Forecasting

We compare first the forecasting performance of the small and large models. Lenza and Primiceri (2022) show that the BVAR model with rescaled variables produces forecasts with higher degrees of uncertainty than the BVAR model with the latest pandemic data being excluded. Next, we continue our research by doing an out-of-sample competition among the two models. We apply recursive forecast evaluation schemes, so that forecasting models are estimated on expanding windows. We calculate forecasts for 1 month, 6 months and 1 year horizons. We report 7 forecast evaluation criteria: the root mean squared error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), Theil's U coefficient, the continuous ranked probability score (CRPS), and two different versions of the log score.

4.3.1 Assessing the forecasting performance

The root mean squared error of the forecast for variable *i* in the VAR is defined as:

$$RMSE_i = \sqrt{\frac{1}{h}\sum_{i=1}^{h}(y_{T+i} - \tilde{y}_{T+i})^2},$$

where \tilde{y}_{T+i} denotes the predicted value of y_{T+i} . The mean absolute error of the forecast for variable i in the VAR is defined as:

¹¹ The short-lived positive initial impulse is a case of borderline significance only.

¹² It appears that the impulse response functions significantly change for the small model when the sum-of coefficient prior and the dummy observation prior are applied.

$$MAE_i = \frac{1}{h} \sum_{i=1}^{h} |y_{T+i} - \tilde{y}_{T+i}|.$$

For these two measures, a lower value indicates a better fit.

The mean absolute percentage error of the forecast for variable *i* in the VAR is defined as:

$$MAPE_{i} = \frac{100}{h} \sum_{i=1}^{h} \left| \frac{y_{T+i} - \tilde{y}_{T+i}}{y_{T+i}} \right|$$

The MAPE describes the size of the error relative to the value of the variable. The Theil inequality coefficient or Theil U statistic of the forecast for variable *i* in the VAR is defined as:

$$U_{i} = \frac{RMSE_{i}}{\sqrt{\frac{1}{h}\sum_{i=1}^{h}(y_{T+i})^{2}} + \sqrt{\frac{1}{h}\sum_{i=1}^{h}(\tilde{y}_{T+i})^{2}}}$$

This coefficient ranges between 0 and 1, with a lower value indicating a better forecast. A value of 0 indicates a perfect fit, while a value of 1 says that the forecast is no better than a naïve guess.

A further alternative measure of forecasting performance is the continuous ranked probability score, defined as:

$$CRPS(F, y_{T+h}^o) = \int_{-\infty}^{\infty} (F(x) - \mathbb{I}(x > y_{T+h}^o))^2 dx,$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, *F* is the cumulative distribution function corresponding to the marginal predictive density *f* for the forecast at period T + h, along with the realized value y_{T+h}^o for this period. The CRPS can be conceived as a penalty function sanctioning the overall distance between the distribution points and the realized values. The larger the value of the CRPS, the poorer the performance of the predictive distribution for the forecast at period T + h.

Last, we move to the family of log predictive scores for forecasts. The log scores compare the realised values with the whole posterior predictive density. The idea behind log scores is that the predictive distribution should be such that it takes a high density at the actual data value. The procedure of computing log predictive scores is described in the ECB's online BEAR "technical guide".¹³ Two base forecasting scenarios are considered. In the first case, what is evaluated is the performance of the forecasts produced by model A for variable *i* at respective periods T + 1, T + 2, ..., T + h. In the second case, what is evaluated is the overall performance of the forecast produced by model B for variable *i* again from period T + 1 up to period T + h. If model A has a higher average log predictive score than model B, it means that values close to the actual realizations of a time series were a priori more likely according to model A relative to model B (Giannone et al. 2015). In other words, a higher log predictive score indicates that the density forecasts produced by our proposed procedure are more accurate than those of the alternative models.

¹³ https://github.com/european-central-bank/BEAR-toolbox, last accessed 22 July 2022.

4.3.2 Discussion of the forecasting results

We compare the forecasting abilities of the two models before and during the pandemic. Tables 2-7 present forecast evaluations for the small model, which consists of 3 variables. Tables 8 - 13 present forecast evaluations for the large model, which consists of 20 variables. We present forecast evaluations for 3 variables, namely industrial production, consumer price index, and shadow interest rate. Tables 2-4 and Tables 8-10 show the results for the time period before the pandemic, that is before 2020:02. Tables 5-7 and Tables 11-13 show the results for forecasting during the pandemic, which is in the period between 2020:05 and 2021:05. In order to assess whether these models could predict a pandemic, we decided to perform the forecasting experiment for after the outbreak of the pandemic. Also due to the short time that has elapsed since the outbreak of the pandemic, the assessment of forecasts applies only to the period up to 6 months.

We test two sets of hyper-parameters, the first set is the standard values proposed in the literature (overall tightness equal to 0.1 and lag decay equal to 1), the second set is the values obtained using Giannone et al.'s (2015) procedure. For the large model, for the two samples, the one that includes the pandemic and the one that ends before the pandemic, much better results were obtained using Giannone et al.'s prior. However, for the small model for the sample that ends before the pandemic no strong evidence was found that their prior is better than the standard hyper-parameter values, with results very similar for the two type of priors. Moreover, for the sample that includes the pandemic the model with standard hyper-parameter values generated better forecasts than the small model with Giannone et al.'s prior.

Based on our results so far, we are not able to conclusively confirm that the large model outperforms the small model. The forecasting performance of the two models seems to be similar. Thus, taking into account the time needed to calculate the large model, the small model might be preferable.

The large model outperforms the small model for the forecasts for the shadow interest rate, according to all analysed measures. For the forecasts for industrial production the large model seems to outperform the small model when the horizons is larger than 1 month (except of now-casting), according to the RMSE, MAE, MAPE, CRPS, and Log score 1 criteria. Theil's U seems to show the same results for the two models, whereas Log score 2 suggests better results for the small model. For the forecasts for the consumer price index both the traditional point forecasting measures and the density-based scores point towards the small model producing better forecasts. However, the results differ little and in many cases the difference is only in third decimal place. In addition, if we compare the results for the forecasts for the shadow rate all error statistics are better for the small model. For the consumer price index, only log predictive scores seem to show better performance of the large model in a few cases (see Table 6 and 12, log scores). Whereas for the forecasts for the CRPS measure, indicate that the large model performs better. Thus, we conclude that in this specific time period the results are mixed.

If we compare the forecasting performance of the models before the pandemic and during the pandemic, we observe that the forecasting performance of the models is somewhat better before the pandemic. The result holds for all three time series (i.e. industrial production, consumer price index, and shadow interest rate). All forecasting evaluation criteria show that forecasts produced before the pandemic (Tables 2-4 and 8-10) are more accurate than forecasts produced during the pandemic (Tables 5-7 and 11-13).

5 Conclusion

We estimate the BVAR models with Giannone et al.'s (2015) prior for the United States for the recent time period from 1990:01 to 2021:11. We do not apply any data modification for the pandemic data. The impulse response functions that we obtain are very similar to the impulse response functions obtained for the sample that ends before the pandemic, which is in 2020:01. We test the performance of the small BVAR model (with three variables) and the large BVAR model (with 20 variables). The models seem to produce credible impulse responses. We compare their forecasting performance using traditional point forecasting measures and density-based scores in two time periods, namely before and after the pandemic.

Our results indicate that it is preferable to use the Wu-Xia shadow rate instead of the effective federal funds rate as the monetary policy tool in the VAR models. Also, our results indicate that the forecasting performance of the small BVAR model is quite similar to performance of the large VAR model. Therefore, it seems better to use small BVAR models than large BVAR models that are more difficult and time-consuming to estimate. We leave for future research a number of important issues: the application of conditional forecasts, optimizing for lag length, application of tests of forecast performance equality, as well as conducting a similar study for some other economies.

References

- Bańbura M., Giannone D., Reichlin L. (2010). Large Bayesian vector auto regressions. Journal of Applied Econometrics, 25, 71-92.
- Bobeica E., Hartwig B. (2022). The COVID-19 shock and challenges for time series models. International Journal of Forecasting, in press, <u>https://doi.org/10.1016/j.ijforecast.2022.01.002</u>, last accessed 28 November 2022.
- Bobeica E., Jarociński M. (2019). Missing disinflation and missing inflation: A VAR perspective. International Journal of Central Banking, 15, 199-252.
- Carriero A., Clark T., Marcellino M. (2011). Bayesian VARs: Specification choices and forecast accuracy. Working Paper No. 1112, Federal Reserve Bank of Cleveland.

Carriero A., Clark T., Marcellino M., Mertens E. (2022). Addressing COVID-19 outliers in BVARs with stochastic volatility. Review of Economics and Statistics, in press, <u>https://doi.org/10.1162/rest_a_01213</u>, last accessed 20 June 2022.

- Crump R. K., Eusepi S., Giannone D., Qian E., Sbordone A. M. (2021), A large Bayesian VAR of the United States economy. Staff Report No. 976, Federal Reserve Bank of New York.
- Dieppe A., Legrand R., van Roye B. (2016). The BEAR toolbox. Working Paper No. 1934, European Central Bank.
- Dieppe A., van Roye B. (2021). The Bayesian estimation, analysis and regression (BEAR) toolbox. User guide. Version 5.0. European Central Bank. <u>https://github.com/european-central-bank/BEAR-toolbox/tree/master/tbx/doc</u>, last accessed 2 December 2022.
- Fry, R., Pagan, A. (2011). Sign restrictions in structural vector autoregressions: A critical review. Journal of Economic Literature, 49, 938-60.
- Giannone D., Lenza M., Primiceri G. E. (2015). Prior selection for vector autoregressions. Review of Economics and Statistics, 27, 436–451.
- Hamilton, J. D. (2016). Macroeconomic Regimes and Regime Shifts. Working Paper No. 21863, National Bureau of Economic Research.
- Hamilton, J. D. (1994). Time series analysis. Princeton University Press, Princeton, NJ.
- Kapetanios G., Mumtaz H., Stevens I., Theodoridis K. (2012). Assessing the economy-wide effects of quantitative easing. Economic Journal, 122, F316-F347.
- Kilian, L., Lütkepohl, H. (2017). Structural vector autoregressive analysis. Cambridge University Press, Cambridge, UK.
- Lenza M., Primiceri G. E. (2022). How to estimate a VAR after March 2020. Journal of Applied Econometrics, 37, 688-699.
- Litterman R. B. (1986). Forecasting with Bayesian vector autoregressions Five years of experience. Journal of Business and Economic Statistics, 4, 25–38.
- Pagan, A., Wickens, M. (2022). Checking if the straitjacket fits. In: Chudik, A., Hsiao, C., Timmermann, A. (Eds.). Essays in Honor of M. Hashem Pesaran: Prediction and macro modeling (Advances in Econometrics, Vol. 43A). Emerald Publishing Limited, Bingley, pp. 269-290.
- Robertson, J. C., Tallman, E. W. (1999). Vector autoregressions: Forecasting and reality. Federal Reserve Bank of Atlanta Economic Review, first quarter, 4-18.
- Schorfheide, F., Song, D. (2021). Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic. Working Paper No. 29535, National Bureau of Economic Research.
- Sims C. A., Zha T. (1998): Bayesian methods for dynamic multivariate models. International Economic Review, 39, 949–968.

Series	p-value	Lags	Maximun	n lags Observations
LOG_INDPRO	0.2621	2	16	383
LOG_PAYEMS	0.5945	0	16	384
UNRATE	0.0178	0	16	384
MCUMFN	0.1099	1	16	383
LOG_PCEC96	0.8539	2	14	236
LOG_HOUST	0.6283	2	16	383
UMCSENT	0.0233	0	16	383
LOG_CPIAUCSL	0.4433	2	16	384
LOG_PCEPI	0.7102	1	16	383
LOG_CPILFESL	0.9413	12	16	384
LOG_PCEPILFE	0.7102	1	16	383
LOG_COMPENSATION	0.8235	0	16	383
LOG_OIL	0.2675	1	16	384
WU_XIA_RATE	0.2627	2	16	381
DFF	0.1579	3	16	384
DAAA	0.7950	13	16	384
DBAA	0.7319	12	16	384
DGS2	0.3353	13	16	384
DGS5	0.5544	13	16	384
DGS10	0.6759	13	16	384
REER	0.4910	2	16	332
NEER	0.3822	2	16	332
NEER2	0.2101	1	16	383
SP CLOSE	1.0000	0	16	384

Table 1. Unit root tests

Note: We applied the standard augmented Dickey-Fuller (ADF) unit root test, with lags selected according to Akaike's information criterion.

Table 2. Forecast evaluation for industrial	production (LOG	_INDPRO), small n	nodel, time p	period
before the pandemic				

	1 month	2 month	3 month	6 month	12 month
RMSE:	0.005	0.006	0.007	0.010	0.015
MAE:	0.005	0.006	0.007	0.008	0.013
MAPE:	0.100	0.121	0.140	0.180	0.280
Theil's U:	0.000	0.001	0.001	0.001	0.002
CRPS:	0.001	0.002	0.002	0.004	0.008
Log score 1:	3.779	3.401	3.236	2.783	2.139
Log score 2:	3.779	7.560	11.366	22.774	45.966

Note: Forecasting models are estimated on rolling windows, the first set of forecasts is generated using the sample 1990:01 – 2017:01 and the last set is generated using the sample 1992:01 – 2019:01. The optimal lambda parameters were found to be the following: $\lambda_1 = 0.0645$, $\lambda_3 = 1.021$, $\lambda_6 = 0.02022$, and $\lambda_7 = 0.6763$.

	1 month	2 month	3 month	6 month	12 month
RMSE:	0.002	0.002	0.003	0.003	0.004
MAE:	0.002	0.002	0.002	0.003	0.003
MAPE:	0.030	0.038	0.045	0.051	0.060
Theil's U:	0.000	0.000	0.000	0.000	0.000
CRPS:	0.001	0.001	0.001	0.002	0.002
Log score 1:	4.772	4.306	4.064	3.856	3.522
Log score 2:	4.772	9.540	14.320	28.754	58.233

Table 3. Forecast evaluation for consumer price index (CPI), small model, time period before the pandemic

Note: See Table 2.

Table 4. Forecast evaluation for Wu-Xia shadow federal funds rate (WU_XIA_RATE), small model, time period before the pandemic

	1 month	2 month	3 month	6 month	12 month
RMSE:	0.100	0.131	0.156	0.222	0.412
MAE:	0.100	0.124	0.142	0.200	0.350
MAPE:	7.983	9.463	10.378	12.935	19.733
Theil's U:	0.044	0.057	0.065	0.083	0.133
CRPS:	0.047	0.074	0.099	0.168	0.305
Log score 1:	0.512	0.045	-0.221	-0.721	-1.365
Log score 2:	0.512	1.065	1.633	3.360	6.656

Note: See Table 2.

Table 5. Forecast evaluation for industrial production (LOG_INDPRO), small model, during the pandemic (i.e., 2020:05 – 2021:05).

±		,		
	1 month	2 month	3 month	6 month
RMSE:	0.033	0.046	0.051	0.054
MAE:	0.033	0.043	0.047	0.050
MAPE:	0.734	0.944	1.039	1.087
Theil's U:	0.004	0.005	0.006	0.006
CRPS:	0.004	0.007	0.008	0.009
Log score 1:	0.621	-0.827	-1.283	0.455
Log score 2:	0.621	1.240	3.033	9.749

Note: Forecasting models are estimated on rolling windows, the first set of forecasts is generated using the sample 1990:01 – 2020:05 and the last set is generated using the sample 1991:01 – 2021:05. The optimal lambda parameters were found to be the following: $\lambda_1 = 0.05206$, $\lambda_3 = 1.068$, $\lambda_6 = 0.02024$, and $\lambda_7 = 0.5032$.

(1.0., 2020.05 2021.05).						
	1 month	2 month	3 month	6 month		
RMSE:	0.005	0.008	0.009	0.014		
MAE:	0.005	0.007	0.009	0.012		
MAPE:	0.083	0.127	0.153	0.216		
Theil's U:	0.000	0.001	0.001	0.001		
CRPS:	0.001	0.002	0.002	0.002		
Log score 1:	3.380	2.653	2.599	1.421		
Log score 2:	3.380	7.330	11.332	22.227		

Table 6. Forecast evaluation for consumer price index (CPI), small model, during the pandemic (i.e., 2020:05 - 2021:05).

Note: See Table 5.

Table 7. Forecast evaluation for Wu-Xia Shadow federal funds rate (WU_XIA_RATE), small model, during the pandemic (i.e., 2020:05 – 2021:05).

	1 month	2 month	3 month	6 month
RMSE:	0.358	0.493	0.600	0.850
MAE:	0.358	0.453	0.538	0.743
MAPE:	142.812	144.517	156.593	153.820
Theil's U:	0.486	0.507	0.533	0.573
CRPS:	0.081	0.136	0.167	0.224
Log score 1:	-0.762	-1.149	-1.399	-1.710
Log score 2:	-0.762	-1.568	-2.164	-3.553

Note: See Table 5.

Table 8. Forecast evaluation for industrial production (LOG_INDPRO), large model, time period before the pandemic

	1 month	2 month	3 month	6 month	12 month
RMSE:	0.005	0.006	0.007	0.009	0.014
MAE:	0.005	0.005	0.006	0.008	0.012
MAPE:	0.097	0.118	0.134	0.175	0.266
Theil's U:	0.000	0.001	0.001	0.001	0.002
CRPS:	0.001	0.002	0.002	0.004	0.007
Log score 1:	3.757	3.448	3.286	2.822	2.253
Log score 2:	3.757	7.546	11.344	22.716	42.049

Note: Forecasting models are estimated on rolling windows, the first set of forecasts is generated using the sample 1990:01 – 2017:01 and the last set is generated using the sample 1992:01 – 2019:01. The optimal lambda parameters were found to be the following: $\lambda_1 = 0.02144$, $\lambda_3 = 1.088$, $\lambda_6 = 0.02096$, and $\lambda_7 = 0.5027$.

	1 month	2 month	3 month	6 month	12 month	
RMSE:	0.002	0.002	0.003	0.004	0.005	
MAE:	0.002	0.002	0.003	0.003	0.004	
MAPE:	0.032	0.040	0.046	0.060	0.075	
Theil's U:	0.000	0.000	0.000	0.000	0.000	
CRPS:	0.000	0.001	0.001	0.002	0.002	
Log score 1:	4.708	4.271	4.054	3.814	3.486	
Log score 2:	4.708	9.422	14.163	28.501	53.016	

Table 9. Forecast evaluation for consumer price index (CPI), large model, time period before the pandemic

Note: See Table 8.

Table 10. Forecast evaluation for Wu-Xia Shadow federal funds rate (WU_XIA_RATE), large model, time period before the pandemic

	1 month	2 month	3 month	6 month	12 month
RMSE:	0.073	0.101	0.126	0.195	0.328
MAE:	0.073	0.095	0.116	0.173	0.282
MAPE:	6.359	7.716	8.818	11.685	16.701
Theil's U:	0.036	0.048	0.058	0.080	0.117
CRPS:	0.042	0.064	0.084	0.142	0.236
Log score 1:	0.644	0.168	-0.087	-0.565	-1.099
Log score 2:	0.644	1.286	1.943	3.794	6.619

Note: See Table 8.

Table 11. Forecast evaluation for industrial production (LOG_INDPRO), large model, during the pandemic (i.e., 2020:05 – 2021:05).

	1 month	2 month	3 month	6 month
RMSE:	0.024	0.035	0.044	0.058
MAE:	0.024	0.033	0.040	0.050
MAPE:	0.519	0.713	0.864	1.083
Theil's U:	0.003	0.004	0.005	0.006
CRPS:	0.010	0.016	0.017	0.021
Log score 1:	2.053	1.599	1.443	1.457
Log score 2:	2.053	3.991	6.036	10.983

Note: Forecasting models are estimated on rolling windows, the first set of forecasts is generated using the sample 1990:01 – 2020:05 and the last set is generated using the sample 1991:01 – 2021:05. The optimal lambda parameters were found to be the following: $\lambda_1 = 0.03509$, $\lambda_3 = 1.057$, $\lambda_6 = 0.03509$, and $\lambda_7 = 0.5033$.

(
	1 month	2 month	3 month	6 month
RMSE:	0.006	0.009	0.011	0.017
MAE:	0.006	0.008	0.010	0.015
MAPE:	0.099	0.145	0.175	0.266
Theil's U:	0.000	0.001	0.001	0.001
CRPS:	0.002	0.004	0.005	0.005
Log score 1:	3.463	2.779	2.543	1.748
Log score 2:	3.463	7.009	10.571	20.122

Table 12. Forecast evaluation for consumer price index (CPI), large model, during the pandemic (i.e., 2020:05 - 2021:05).

Note: See Table 11.

Table 13. Forecast evaluation for Wu-Xia Shadow federal funds rate (WU_XIA_RATE), large model, during the pandemic (i.e., 2020:05 – 2021:05).

	1 month	2 month	3 month	6 month
RMSE:	0.787	1.199	1.569	1.903
MAE:	0.787	1.126	1.443	1.734
MAPE:	316.875	492.243	601.008	627.430
Theil's U:	0.632	0.651	0.684	0.645
CRPS:	0.225	0.357	0.412	0.498
Log score 1:	-1.290	-1.826	-2.134	-1.881
Log score 2:	-1.290	-2.451	-3.565	-7.397

Note: See Table 11.

Figure 1. Impulse response functions for the small VAR model for the whole sample when the effective federal funds rate is used



Figure 2. Impulse response functions for the small VAR model for the whole sample when the shadow interest rate is used



Figure 3. Impulse response functions for the small BVAR model with the effective federal funds rate (whole sample)



Figure 4. Impulse response functions for the small BVAR model with the shadow rate (whole sample)







Appendix

Table. Data description

	Carries	Describtion	Source	Unite	Transfor	
	Series	Describuon	Source	Units	mation	
Ind	icators of the level of	economic activity			mation	
1	IND PROD	Industrial Production: Total Index	Federal Reserve	SA 2017=100	ln	
1	nub_nubb	(INDPRO)	Board	511, 2017-100		
2	PAYEMS	All Employees Total Nonfarm	BLS	SA Thousands	ln	
2	1 TT LIVID	(PAYEMS)	DLS	of Persons		
3	UNEMP	Unemployment Rate (UNRATE)	BLS	SA. % (of labor	none	
5			DED	force)	none	
4	CAPACITY	Capacity Utilization:	Federal Reserve	SA, % (of	none	
		Manufacturing (NAICS)	Board	capacity)		
		(MCUMFN)		1 57		
5	R PCE	Real Personal Consumption	BEA	Billions of	ln	
	_	Expenditures (PCEC96)		Chained 2012		
		-		Dollars, SA		
6	HOUSING_STAR-	New Privately-Owned Housing	Census Bureau	SAAR,	ln	
	TS	Units Started: Total Units		Thousands of		
		(HOUST)		Units		
Ind	icators of the level of	prices				
7	CPI	Consumer Price Index for All	BLS	SA, 1982-84 =	ln	
		Urban Consumers: All Items in		100		
		U.S. City Average (CPIAUCSL)				
8	PCE	Personal Consumption	BEA	SA, 2012 = 100	ln	
		Expenditures: Chain-type Price				
		Index (PCEPI)				
9	CPI_LESS	Consumer Price Index for All	BLS	SA, 1982-84 =	ln	
		Urban Consumers: All Items Less		100		
		Food and Energy in U.S. City				
1.0		Average (CPILFESL)		<u> </u>	-	
10	PCE_LESS	Personal Consumption	BEA	SA, $2012 = 100$	ln	
		Expenditures Excluding Food and				
		Energy (Chain-Type Price Index)				
11		(PCEPILFE)		CA Dillions of	1	
11	COMPENSATION	Compensation of Employees,	BEA	SA, Billions of	In	
		Disburgements (A 576PC1)		Donars		
12	OII	Crude Oil Prices: West Taxes	EIA / W/SI	dollars	ln.	
12	UIL	Intermediate (WTI) Cushing	LIA / WSJ	uonais	111	
		Oklahoma (DCOII WTICO)				
Inte						
13	WII XIA RATE	Wu-Xia Shadow Federal Funds	https://www.atlanta	% n a	none	
15		Rate	fed org/coer/researc	70 p.a.	none	
		Rute	h/wu-xia-shadow-			
			federal-funds-rate			
14	FF EF RATE	Federal Funds Effective Rate	Federal Reserve	% p.a.	none	
		(DFF)	Board	,• p		
15	Moody_AAA	Moody's Seasoned Aaa Corporate	Moody's	% p.a.	none	
		Bond Yield (DAAA)				
16	Moody_BAA	Moody's Seasoned Baa Corporate	Moody's	% p.a.	none	
		Bond Yield (DBAA)	-			

17	2Y_yield	Market Yield on U.S. Treasury Securities at 2-Year Constant Maturity (DGS2)	Federal Reserve Board	% p.a.	none
18	5Y_yield	Market Yield on U.S. Treasury Securities at 5-Year Constant Maturity (DGS5)	Federal Reserve Board	% p.a.	none
19	10Y_yield	Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity (DGS10)Federal Reser Board		% p.a.	none
Exc	change rate				
20	USD_EUR	U.S. Dollars to Euro Spot Exchange Rate	Federal Reserve Board	U.S. Dollars to One Euro, Not Seasonally Adjusted	ln
22	REER	Real broad BIS effective exchange rate, Real (CPI-based), Broad Indices	BIS	Monthly averages; 2010=100	ln
23	NEER	Nominal broad BIS effective exchange rate	BIS	Monthly averages; 2010=100	ln
24	NEER2	Nominal narrow BIS effective exchange rate	BIS	Monthly averages; 2010=100	ln
Ade	ditional variables				
25	S&P 500	S&P 500 index, adjusted close price adjusted for splits and dividend and/or capital gain distributions.	Yahoo! Finance	dollars	ln
26	SENTIMENT	University of Michigan: Consumer Sentiment (UMCSENT) consumer confidence levels in the United States, before 1978 we use University of Michigan: Consumer Sentiment (DISCONTINUED) (UMCSENT1)	University of Michigan	Index 1966:Q1=100, Not Seasonally Adjusted	none

Note: BLS refers to the U.S. Bureau of Labor Statistics, BEA to the U.S. Bureau of Economic Analysis, WSJ to the Wall Street Journal, and BIS to the Bank for International Settlement. Also, "In" refers to taking the natural logarithm.