Structural macroeconomic forecasting using a text-based measure of policy uncertainty

Evidence from a Bayesian VAR Analysis

Renato Vassallo

Forecasting and Nowcasting with Text as Data

Term Paper

Barcelona School of Economics

Data Science for Decision Making

June 08, 2023

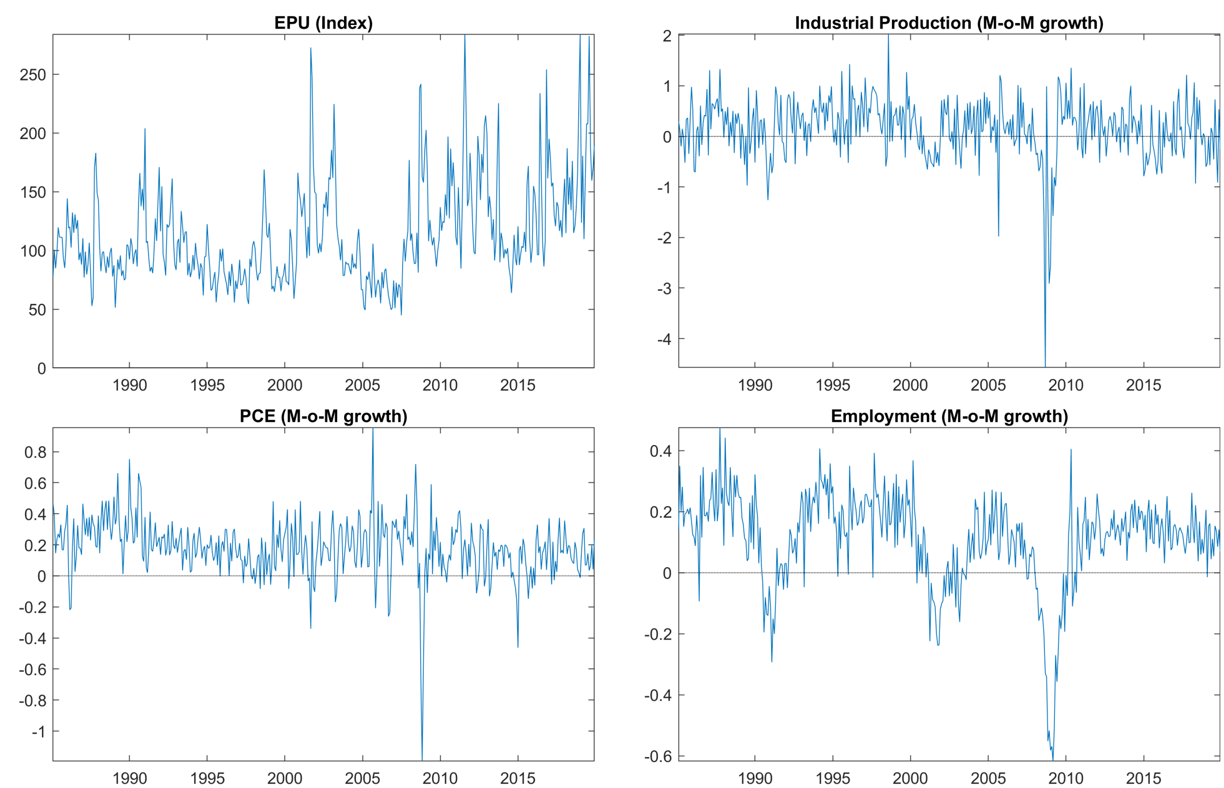
1 Introduction

The study by Baker, Bloom and Davis (2016) demonstrated the empirical usefulness of text-based measures for the analysis of micro and macroeconomic fluctuations. In their paper, the authors investigate the relationship of policy uncertainty with firm-level stock price volatility, investment rates and employment growth, as well as in macroeconomic aggregates such as investment and output, finding that uncertainty shocks do indeed have sizable negative effects on economic fluctuations.

In this document we extend this analysis to incorporate monthly data up to 2019, using a Bayesian approach that will allow us to account for the uncertainty behind the model parameters. In particular, we consider a VAR(1) model with four variables: (i) Economic Policy Uncertainty Index (EPU_t); and the month-on-month growth rates for (ii) Industrial Production (IP_t); (iii) Personal Consumption Expenditures Index (PCE_t); and, (iv) Non-Farm Employment (EMP_t). We obtain this data from `www.policyuncertainty.com` and the Federal Reserve Bank of St. Louis. The evolution of this variables are shown in Figure 1.

Finally, a forecasting exercise is carried out, both for the unconditional case and for a structural forecast conditional to a specific pattern of the EPU.

Figure 1: Variables in the model



Note: Variables considered in the model. The Economic Policy Uncertainty (EPU) index enters in levels; however, the series for Industrial Production (IP), Personal Consumption Expenditures (PCE) and Employment are transformed into month-on-month growth rates. Monthly data between 1985 and 2019.

2 Econometric Strategy

We consider the estimation of a $VAR(1)$ model using monthly data on Economic Policy Uncertainty Index (EPU_t), Industrial Production Growth (IP_t), Consumption (PCE_t), and Employment (Emp_t) for the US from 1985m02 to 2019m12:

$$\begin{pmatrix} EPU_t \\ IP_t \\ PCE_t \\ Emp_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} \end{pmatrix} \begin{pmatrix} EPU_{t-1} \\ IP_{t-1} \\ PCE_{t-1} \\ Emp_{t-1} \end{pmatrix} + \begin{pmatrix} \upsilon_{1t} \\ \upsilon_{2t} \\ \upsilon_{3t} \\ \upsilon_{4t} \end{pmatrix} \tag{1}$$

where:

$$var \begin{pmatrix} \upsilon_{1t} \\ \upsilon_{2t} \\ \upsilon_{3t} \\ \upsilon_{4t} \end{pmatrix} = \Sigma$$

The VAR can be written compactly as:

$$Y_t = X_t \Phi + \upsilon_t$$

with $X_t = \{c_i, Y_{it-1}\}$. Note that as each equation in the VAR has identical regressors, it can be rewritten as:

$$y = (I_N \otimes X)\phi + V$$

where $y = vec(Y_t)$, $\phi = vec(\Phi)$, and $V = vec(\upsilon_t)$.

2.1 The Independent Normal-Wishart Prior

As the name suggests, this prior involves setting the prior for the VAR coefficients and the error covariance independently (unlike the natural conjugate prior):

$$p(\phi) \sim \mathrm{N}(\tilde{\phi}_0, H)$$

$$p(\Sigma) \sim \mathrm{IW}(\bar{S}, \alpha)$$

where $\bar{S}$ is the prior scale matrix and α the prior degrees of freedom. It can be shown that the *posterior distribution of the VAR coefficients* conditional on Σ is normal (see Kadiyala and Karlsson (1997)). Similarly, the posterior for Σ conditional on ϕ is also inverse Wishart.

We incorporate the prior belief that the endogenous variables included in the VAR follow a random walk process or an AR(1) process (Minnesota-type prior.). In other words, the prior mean for the

VAR coefficients in equation 1 implies the following form for the VAR:

$$
(2) \qquad \begin{pmatrix} EPU_t \\ IP_t \\ PCE_t \\ EMP_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_{11}^0 & 0 & 0 & 0 \\ 0 & \phi_{22}^0 & 0 & 0 \\ 0 & 0 & \phi_{33}^0 & 0 \\ 0 & 0 & 0 & \phi_{44}^0 \end{pmatrix} \begin{pmatrix} EPU_{t-1} \\ IP_{t-1} \\ PCE_{t-1} \\ EMP_{t-1} \end{pmatrix} + \begin{pmatrix} \upsilon_{1t} \\ \upsilon_{2t} \\ \upsilon_{3t} \\ \upsilon_{4t} \end{pmatrix}
$$

After performing the Augmented Dickey-Fuller test, we find that all our variables reject the null hypothesis of presence of Unit Root at a 5% level. Therefore we consider more realistic to incorporate the prior that they follow an AR(1) process. For this application, we set:

$$\phi_{11}^0 = \phi_{22}^0 = \phi_{33}^0 = \phi_{44}^0 = 0.75$$

The variance of the prior H is set in a more structured manner and is given by the following relations for the VAR coefficients ϕ_{ij}:

$$
\begin{array}{rl}
\left(\dfrac{\lambda_1}{l^{\lambda_3}}\right)^2 & \text{if} \quad i = j \\
\left(\dfrac{\sigma_i \lambda_1 \lambda_2}{\sigma_j l^{\lambda_3}}\right)^2 & \text{if} \quad i \neq j \\
(\sigma_i \lambda_4)^2 & \text{for the constant}
\end{array}
$$

where i refers to the dependent variable in the i^{th} equation and j to the independent variables in that equation. Therefore, if $i = j$ then we are referring to the coefficients on the own lags of variable i. σ_i and σ_j are variances of error terms from AR regressions estimated via OLS using the variables in the VAR. The ratio of σ_i and σ_j in the formulas above controls for the possibility that variable i and j may have different scales. Note that l is the lag length. The λ's are parameters set by the researcher that control the tightness of the prior.

For example, λ_1 controls the standard deviation of the prior on own lags. As $\lambda_1 \to 0$, then $\phi_{11}, \phi_{22}, \phi_{33}, \phi_{44} \to \phi_{11}^0, \phi_{22}^0, \phi_{33}^0, \phi_{44}^0$ respectively, and all other lags go to zero in our setting in equation 2. Following Canova (2007), we set:

$$\lambda_1 = 0.2, \qquad \lambda_2 = 0.5, \qquad \lambda_3 = 1, \qquad \lambda_4 = 10^5$$

In relation to the prior of the covariance matrix $p(\Sigma) \sim \mathrm{IW}(\bar{S}, \alpha)$, we define:

$$\bar{S} = I_N, \qquad \alpha = N + 1$$

2.2 Identification Strategy

We are interested in estimating the response of the variables in our system in 1 to an increase in the Economic Policy Uncertainty Index. We assume that uncertainty shocks affects contemporaneously the other variables, but the other variables will only affect uncertainty with lags. The standard way to impose this restriction on the contemporaneous period is to identify the EPU shock using a Cholesky decomposition of Σ:

$$\Sigma = A_0 A_0'$$

where A_0 is a lower triangular matrix. Therefore, the relationship between reduced-form and structural errors will be defined by:

$$\begin{pmatrix} \upsilon_{1t} \\ \upsilon_{2t} \\ \upsilon_{3t} \\ \upsilon_{4t} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{pmatrix}$$

Note that we are also imposing that industrial production shocks simultaneously affect inflation and employment, but these do not affect industrial production in the same period. In this context, the last variable of the system (EMP_t) is considered as the most endogenous variable.

2.3 The Gibbs Sampling

Under the independent Normal-Wishart prior, analytical expressions for the marginal posterior distributions are not available. Therefore, we use the Gibbs sampling algorithm.

Bayesian simulation methods such as Gibbs Sampling provide an efficient way not only to obtain point estimates but also to characterise the uncertainty around those point estimates. The Gibbs Sampling algorithm for the VAR model consists of the following steps:

Step 1. Set the initial value for $\Sigma^{(0)}$. Set $s = 1$ and S.

Step 2. Draw $\phi^{(s)}$ from its conditional posterior distribution $p(\phi^{(s)}|\Sigma^{(s-1)}, Y_t)$.

Step 3. Draw $\Sigma^{(s)}$ from its conditional posterior distribution $p(\Sigma^{(s)}|\phi^{(s)}, Y_t)$

Step 4. Obtain $A_0^{(s)} = (P^{-1})$ through the Cholesky decomposition:

$$P'P = \Sigma^{(s)}$$

Step 5. If $s < S$, then set $s = s + 1$, and return to step 2. Otherwise, stop the simulation.

Note that the draws of the model parameters (after the burn-in period) are typically used to calculate forecasts or impulse response functions and build the distribution for these statistics.

3 Results

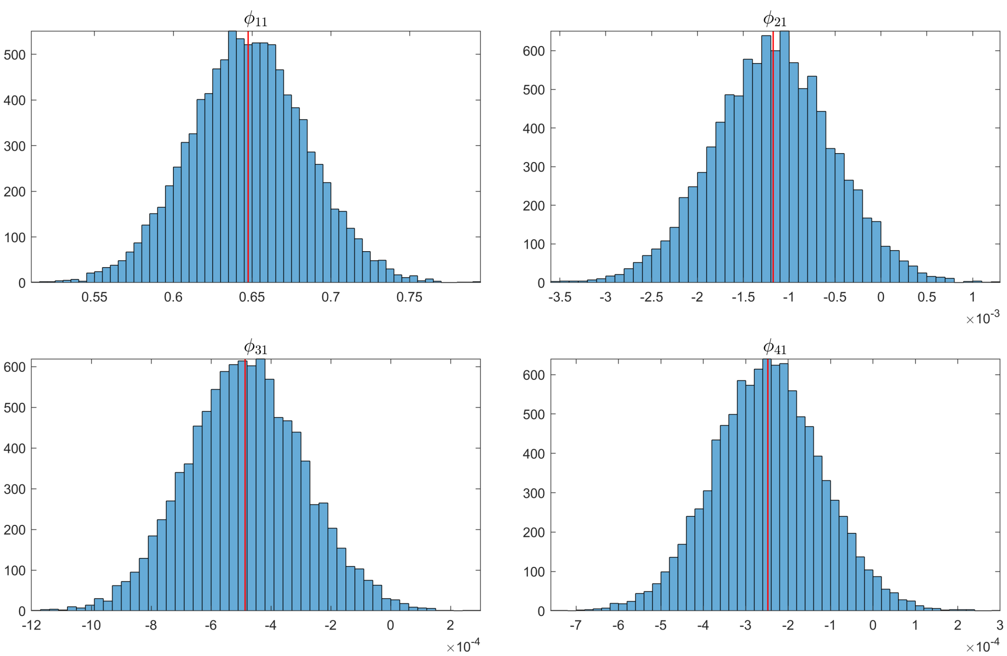
3.1 Estimation Results

The Gibbs sampling was performed for a total of 40,000 draws, of which the first 30,000 were discarded (burn-in) to avoid initial condition problems. We use the 10,000 draws of the model parameters to build the marginal posterior distribution for these coefficients.

Figure 2 shows the distributions for the coefficients associated with the EPU lag in the four equations of the model. That is, it gives us an idea of the relationship between the EPU in $t-1$ and the endogenous variables in t. As can be seen, the distributions of ϕ_{21}, ϕ_{31} and ϕ_{41} are strongly centered on negative values (median values of -0.0012, -0.00048 and -0.00025, respectively). This would be a first indicator of the negative relationship between EPU and real variables in US economy.

It is worth noting that although our a priori belief was that $\phi_{21}^0 = \phi_{31}^0 = \phi_{41}^0 = 0$, the interaction with information contained in the data through the *likelihood* allowed us to update our beliefs, resulting in significantly negative posterior estimates for these coefficients.

Figure 2: Posterior Distributions for selected VAR coefficients



Note: Marginal posterior distributions for the coefficients associated with the EPU lag in the four equations of the model in 1. The red lines represents the median values. The distributions are obtained using the Gibbs Sampling algorithm, from which 10,000 final posterior draws are obtained, after discarding the first 30,000 draws (burn-in).

According to Baker et al. (2016), the EPU captures quite well the effects of presidential elections, wars in the Gulf region and political battles over taxes and government spending-events. This is consistent with the idea that policy uncertainty shocks deteriorate private confidence, and consequently paralyse investment decisions, affecting real aggregates such as consumption and employment.

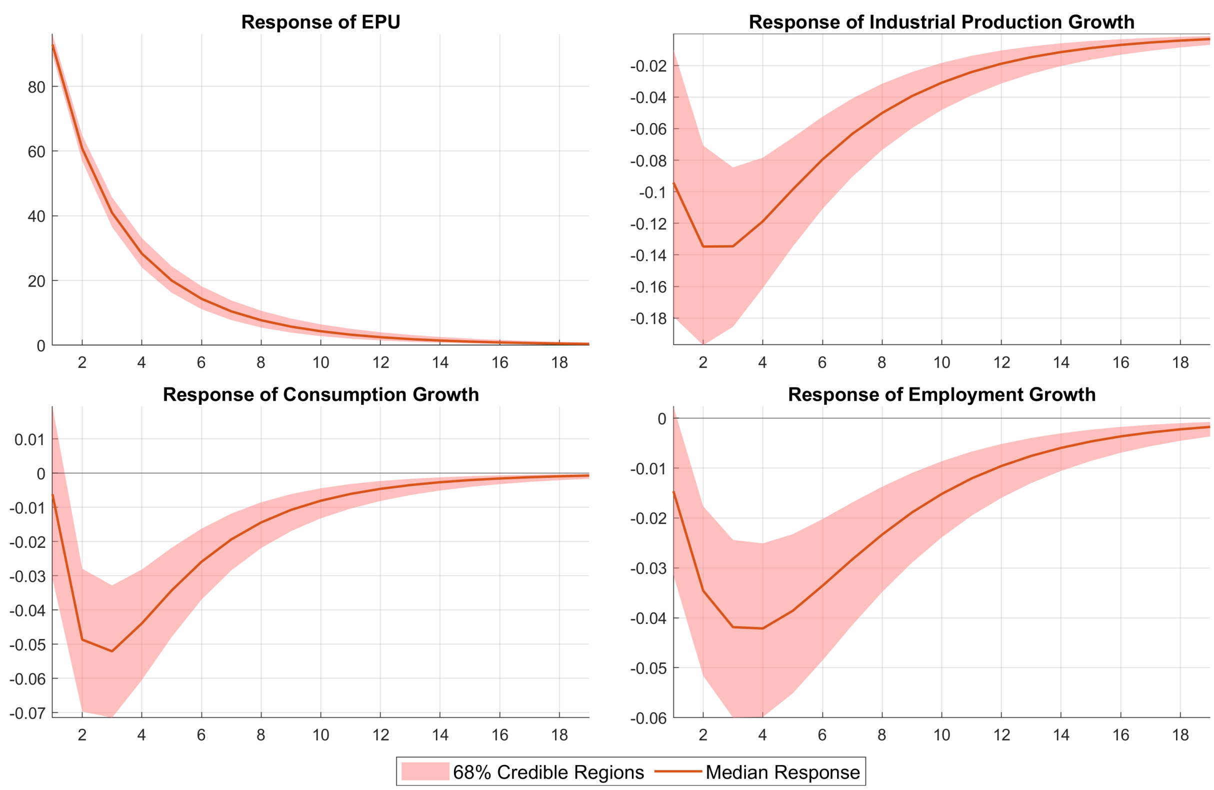
3.2 Impulse-Response Functions

Figure 3 depicts the model-implied responses of our variables to a 90-point upward EPU innovation, equal in size to the EPU change from its average value in 2005-06 (before the financial crisis) to its average value in 2011-12 (a period with major fiscal policy battles and high EPU levels). We also show the associated 68% credibility region.

The results would indicate that in the face of this unexpected increase in EPU, the three real variables in our model would show not only a significant decline (16th and 84th percentiles are below zero), but also persistent (the shock would dissipate by month 20).

In particular, the growth rate of industrial production would fall by around 0.14% in the second and third period, and then gradually dissipate by month 20, while consumption and employment growth would fall to a lesser but significant extent (-0.05% and -0.04% drop in month 3, respectively).

Figure 3: Impulse responses to an Economic Policy Uncertainty Shock



Note: BVAR-estimated impulse response functions to a 90-point upward EPU shock with 68% credible regions. Identification based on one lag and a Cholesky decomposition with the following ordering: EPU Index, Industrial Production Growth, Personal Consumption Expenditures Growth and Employment Growth. Fit to monthly data from 1985 to 2019.

The results reported here are consistent with those found by Baker et al. (2016). For the same increase in EPU shock, the authors documented maximum estimated drops of 1.2% in the level of industrial production and 0.35% in the employment rate (they estimate a VAR in log-levels).

4 Structural Forecasting

4.1 Analysis Framework

In this Section, our goal is to forecast industrial production, consumption and employment growth rates assuming that EPU index follow fixed future paths. Waggoner and Zha (1999) provide a convenient framework to calculate not only the conditional forecast but also the forecast distribution using Gibbs Sampling algorithm. Consider our simple VAR(1) model:

$$Y_t = c + \Phi Y_{t-1} + A_0 \epsilon_t$$

where ϵ_t are the uncorrelated structural shocks. Iterating forward K times we obtain:

$$Y_{t+K} = c \sum_{j=0}^{K} \Phi^j + \Phi^j Y_{t-1} + A_0 \sum_{j=0}^{K} \Phi^j \epsilon_{t+K-j} \tag{3}$$

The key point to note is that if a restriction is placed on the future path of the J^{th} variable in Y_t, this implies restrictions on the future shocks to the other variables in the system. Waggoner and Zha (1999) express these constraints on future innovations as:

$$R\epsilon = r \tag{4}$$

where r is a $(M \times k) \times 1$ vector where M are the number of constrained variables and k denotes the number of periods the constraint is applied. The elements of the vector r are the path for the constrained variables minus the unconditional forecast of the constrained variables. R is a matrix with dimensions $(M \times k) \times (N \times k)$. The elements of this matrix are the impulse responses of the constrained variables to the structural shocks ϵ at horizon $1, 2, ..., k$. The $(N \times k) \times 1$ vector ϵ contains the constrained future shocks.

Doan et al. (1983) show that a least square solution for the constrained innovations is given as:

$$\hat{\epsilon} = R'(R'R)^{-1} r$$

With these constrained shocks $\hat{\epsilon}$ in hand, the conditional forecasts can be calculated by substituting these in equation 3.

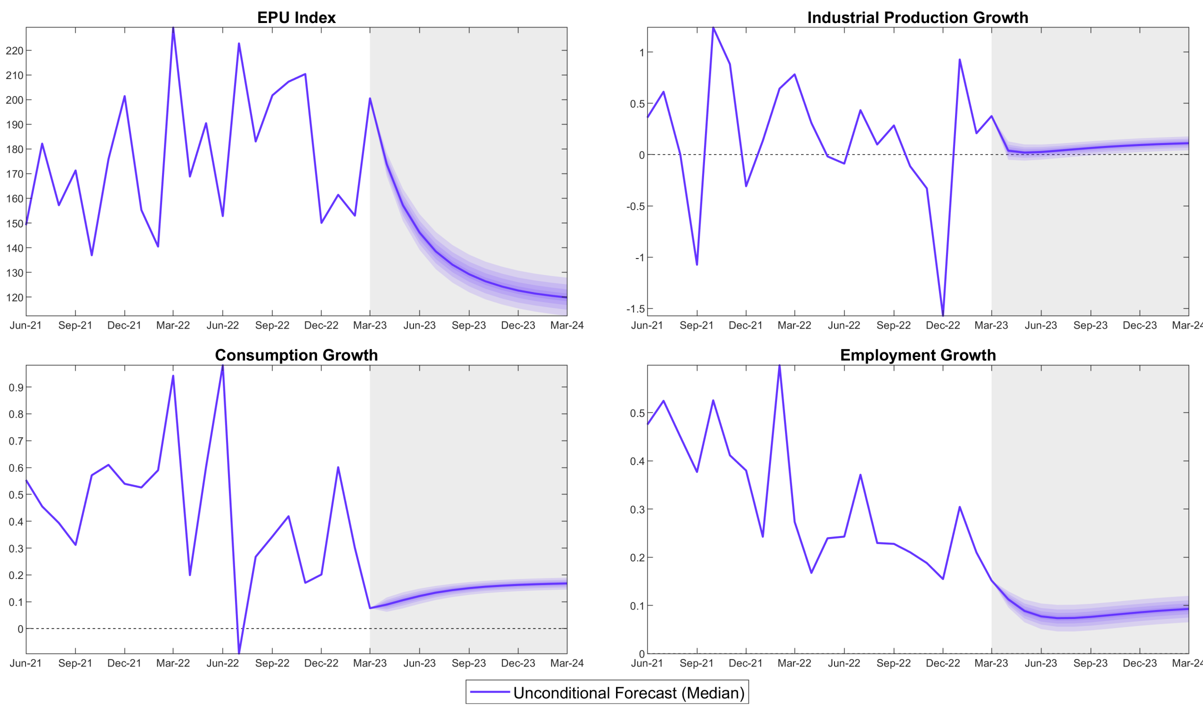
4.2 Unconditional Forecast

Using the estimated VAR in Section 2, Figure 4 shows the unconditional forecast twelve periods in the future (starting in March 2023). Since we have the complete probability distribution of the forecasts, we present the results as a Fan Chart, with density intervals at 80, 60, 40 and 20 percent, as well as the median value.

As can be seen, the unconditional forecast for the EPU shows a significant decline with respect to recent periods, which is consistent with a gradual increase in the growth rate of real variables in the system. This behaviour is due to the nature of the model and the underlying assumptions.

In particular, by specifying a stationary VAR process, the variables are assumed to have constant mean and variance over time. When generating an unconditional forecast, it means that we are projecting the variables into the future without considering any specific shocks or external factors. Since the model assumes that the variables revert to their mean over time, the forecast will naturally converge towards the mean values of the variables.

Figure 4: Unconditional Forecast



Note: The BVAR model is estimated with monthly data from 1985-2019 considering 1 lag. The grey area denotes the forecast period using March 2023 as the latest data to start the forecasts. The forecast is characterised as a Fan Chart with density intervals at 80%, 60%, 40% and 20%, as well as the median of the marginalized joint distribution.

4.3 Structural Forecast

We impose that the only shocks that are in the system are uncertainty shocks (EPU ordered first in the structural shock). This will allow us to interpret the forecasted variables in our model as driven by shocks to uncertainty.

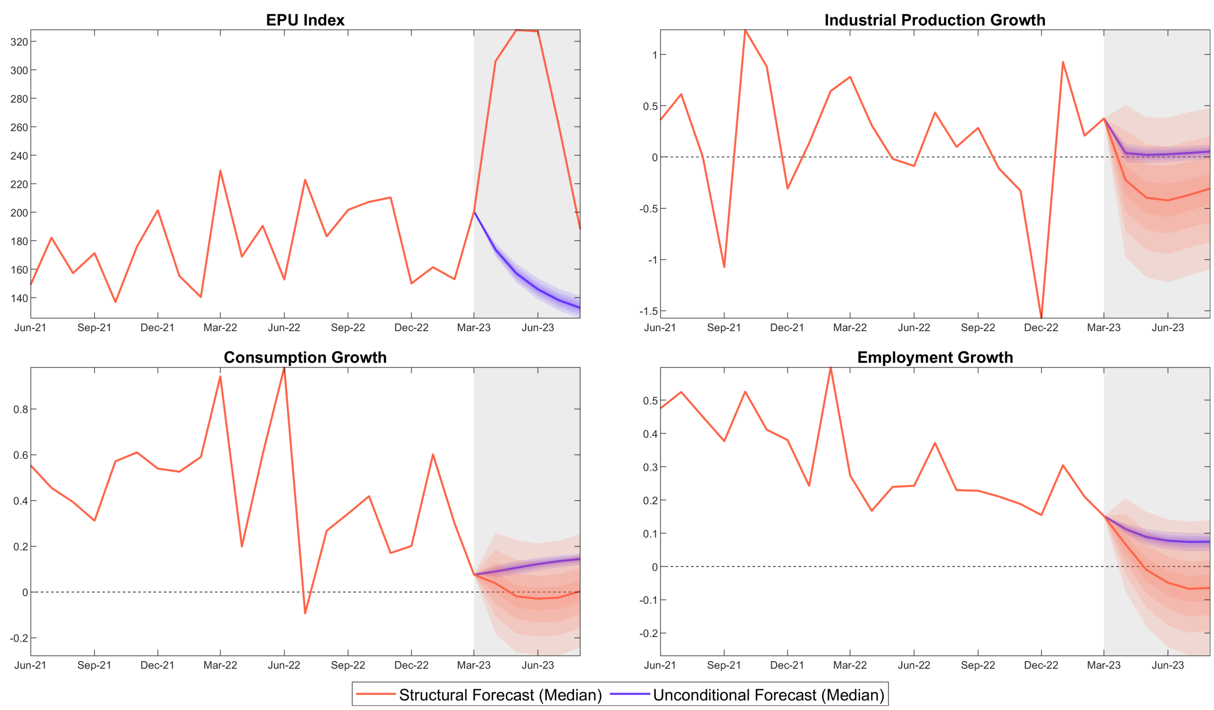
In particular, we consider forecasting Y_t five periods in the future (starting in March 2023). However we impose the condition that the paths from April to August 2023 of the US EPU is the same as during October 2020 to February 2021, i.e., during the last months before and the months during and after the last US national elections. Therefore, we asume that:

$$\begin{pmatrix} EPU_{t+1} \\ EPU_{t+2} \\ EPU_{t+3} \\ EPU_{t+4} \\ EPU_{t+5} \end{pmatrix} = \begin{pmatrix} 306 \\ 328 \\ 327 \\ 261 \\ 188 \end{pmatrix}$$

The results of the structural forecast are presented in Figure 5. We also include the unconditional forecast for comparative purposes. By forcing the EPU to take historically high values, we are introducing a structural shock that will contemporaneously affect the other variables in the system. With respect to the unconditional forecast, the dynamics of the three real variables deteriorate, even falling to negative growth rates.

The presented Fan Charts give evidence in favour of a probability of around 50% that industrial production would contract for all periods over the projection horizon. Similarly, although with a lower probability (20%), employment would contract from May 2023 onwards. In the case of consumption, its dynamics would be essentially zero for this period.

Figure 5: Unconditional and Structural Forecast



Note: The BVAR model is estimated with monthly data from 1985-2019 considering 1 lag. The grey area denotes the forecast period using March 2023 as the latest data to start the forecasts. The forecast is characterised as a Fan Chart with density intervals at 80%, 60%, 40% and 20%, as well as the median of the marginalized joint distribution.

4.4 Shortcomings and Limitations

- **Structural breaks and regime shifts:** our model assume that the relationships between variables remain stable over time. However, if there are structural breaks or regime shifts in the data-generating process (such as in COVID-19 period), the BVAR model may not capture these changes adequately, leading to poor conditional forecasts. Lenza and Primiceri (2022) and Álvarez and Odendahl (2022) provide potential ways to estimate VARs taking into account outliers. Other solution could be incorporating Stochastic Volatility and/or time-varying parameters as in Rodríguez, Vassallo and Castillo (2023).

- **The role of policy variables:** our specification does not incorporate the policy reaction that the monetary or fiscal authority might have. For example, the inclusion of the Fed Funds rate could capture additional information and improve forecasting accuracy. However, the selection and specification of these variables can be challenging, and their inclusion may introduce additional uncertainty and potential issues in the conditional forecasts.

- **Quality of the data:** The quality and availability of EPU can pose challenges in this framework. Periodic reviews of the indicator and constant assessment of its consistency with the stylised facts of the economy in question are required.

5 Conclusions

In this document we extend the model of Baker et al. (2016) to incorporate monthly observations of real variables up to 2019. Likewise, and to incorporate the uncertainty around the model parameters, we propose a Bayesian estimation with independent Normal-Wishart priors.

Despite the fact that *a priori* we assume values of zero for the coefficients associated with the impact of EPU first lag on the endogenous variables, our approach is capable of updating and modifying these beliefs, finding distributions strongly centered on negative values for these coefficients.

This first piece of evidence is put to the test by means of an impulse response analysis, for which the Cholesky decomposition is performed assuming a recursive ordering where the first variable is the EPU. Unexpected shocks of policy uncertainty are found to affect the real aggregates of the US economy significantly and persistently. In particular, it is found that in the face of a 90 point increase in the EPU shock, industrial production would fall around 0.14% in the second and third periods. The shock dissipates until month 20 after the shock is recorded.

Finally, an unconditional and structural forecast exercise is carried out assuming that the EPU follows a fixed future path. While the unconditional forecast presents a mean-reverting behaviour typical of VARs, our conditional forecast allows us to assess the dynamics of real variables driven by pure uncertainty shocks (*Rain dance-like*).

6 References

[1] Álvarez, L. J. and Odendahl F. (2022). Data outliers and Bayesian VARs in the Euro Area. *Bank of Spain Working Paper* **2239**.

[2] Baker, S. R., Bloom, N. and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, Oxford University Press, **131(4)**, 1593-1636.

[3] Canova, F. (2007). Methods for Applied Macroeconomic Research. *Princeton University Press*.

[4] Doan, T., Litterman, R. B. and Sims, C. A. (1983). Forecasting and Conditional Projection Using Realistic Prior Distributions. *NBER Working Papers* **1202**, National Bureau of Economic Research, Inc.

[5] Kadiyala, K. R. and Karlsson, S. (1997). Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics* **12(2)**, 99—132.

[6] Lenza, M and Primiceri, G E. (2022). How to estimate a vector autoregression after March 2020. *Journal of Applied Econometrics* **37**, 688-699.

[7] Rodríguez, G., Vassallo, R. and Castillo B., P. (2023). Effects of external shocks on macroeconomic fluctuations in Pacific Alliance countries. *Economic Modelling* **124**, 106302.

[8] Waggoner, D. F. and Zha, T. (1999). Conditional Forecasts In Dynamic Multivariate Models. *The Review of Economics and Statistics* **81(4)**, 639—651.