Anttonen, Jetro

Working Paper

Nowcasting the Unemployment Rate in the EU with Seasonal BVAR and Google Search Data

ETLA Working Papers, No. 62

Provided in Cooperation with:
The Research Institute of the Finnish Economy (ETLA), Helsinki

Suggested Citation: Anttonen, Jetro (2018) : Nowcasting the Unemployment Rate in the EU with Seasonal BVAR and Google Search Data, ETLA Working Papers, No. 62, The Research Institute of the Finnish Economy (ETLA), Helsinki

This Version is available at:
https://hdl.handle.net/10419/201277

ETLA Working Papers | 62 5.11.2018

Nowcasting the Unemployment Rate in the EU with Seasonal BVAR and Google Search Data

Jetro Anttonen
The Research Institute of the Finnish Economy
jetro.anttonen@etla.fi and
jetro.anttonen@helsinki.fi

Suggested citation:
Anttonen, Jetro (5.11.2018).
"Nowcasting the Unemployment Rate in the EU with Seasonal BVAR and Google Search Data".

ETLA Working Papers No 62.
http://pub.etla.fi/ETLA-Working-Papers-62.pdf

Abstract

In this paper a Bayesian vector autoregressive model for nowcasting the seasonally non-adjusted unemployment rate in EU-countries is developed. On top of the official statistical releases, the model utilizes Google search data and the effect of Google data on the forecasting performance of the model is assessed. The Google data is found to yield modest improvements in forecasting accuracy of the model. To the author's knowledge, this is the first time the forecasting performance of the Google search data has been studied in the context of Bayesian vector autoregressive model. This paper also adds to the empirical literature on the hyperparameter choice with Bayesian vector autoregressive models. The hyperparameters are set according to the mode of the posterior distribution of the hyperparameters, and this is found to improve the out-of-sample forecasting accuracy of the model significantly, compared to the rule-of-thumb values often used in the literature.

Tiivistelmä

Työttömyysasteen ennustaminen EU-maissa BVAR-mallilla ja Googlen hakudatalla

Tässä tutkimuksessa kehitetään kausiluontoinen bayesiläinen vektoriautoregressiivinen malli työttömyystason ennustamiseksi EU-maissa. Virallisten tilastojulkaisujen lisäksi malli käyttää ennustamiseen Googlen hakudataa ja tutkimuksessa tarkastellaan Googlen hakudatan käyttökelpoisuuttaa työttömyystason ennustamiseksi. Googlen hakudatan havaitaan parantavan mallin ennustetarkkuutta vain hieman. Kirjoittajan tiedossa ei ole aikaisempia tutkimuksia, jotka tarkastelisivat Googlen hakudatan käyttökelpoisuutta ennustamisessa käyttäen bayesiläisiä vektoriautoregressiivisiä malleja. Tässä tutkimuksessa tarkastellaan myös hyperparametrien valinnan merkitystä bayesiläisen vektoriautoregressiivisen mallin ennustetarkkuudelle. Hyperparametreille muodostetun marginaalisen posteriorijakauman moodin käyttämisen havaitaan parantavan mallin ennustetarkkuutta huomattavasti, verrattuna kirjallisuudelle tyypillisiin ennalta määrättyihin arvoihin.

I would like to thank Markku Lehmus, Jani Luoto and seminar participants at Etla for helpful comments and conversations.

Haluaisin kiittää Markku Lehmusta, Jani Luotoa ja Etlan seminaariosallistujia hyödyllisistä kommenteista ja keskusteluista.

Key words: Nowcasting, Forecasting, BVAR, Big Data, Unemployment

Asiasanat: Ennustaminen, Nykyhetken ennustaminen, BVAR, Big Data, Työttömyys

JEL: C32, C53, C55, C82, E27

1 Introduction

Nowcasting can be defined as predicting the present, the near future and sometimes the very recent past. In economics the need for predicting not only the future, but also the present and the past, arises from the publication lags and revisions of the economic variables. The distinction between nowcasting and short-term forecasting is not always clear, and in this paper the freedom of treating the two terms somewhat substitutively is indulged.

As policymakers rely on accurate information on the current state of the economy, great amount of literature in recent years has focused on making accurate and timely predictions of the quarterly published Gross Domestic Product (GDP) figures. For example, the Bank of Finland has recently published a new nowcasting model for predicting the quarterly growth rates of the GDP in Finland (Itkonen & Juvonen 2017) and Barnett et al. (2016) is just one example of recent publications, where nowcasting the GDP in the United States is addressed. Despite the GDP being the most widely studied economic variable in the context of nowcasting, there have been prominent work on predicting the other economic variables, such as inflation, as well. To name a couple, Stelmasiak & Szafranski (2016) predict the monthly headline inflation in the Polish economy with seasonal Bayesian vector autoregressive models, while Modugno (2011) nowcasts the inflation in the Euro area and in the United States using high frequency data.

Recently, the Bayesian vector autoregressive (BVAR) models have gained popularity especially in the short term macroeconomic forecasting literature. Their recent popularity stems from the fact that they have been found to produce more accurate forecasts than the popular factor augmented methods, even in the presence of a great number of variables (see e.g Banbura et al. 2010, Koop 2013). Apart from a few exceptions (e.g Stelmasiak & Szafranski 2016, Raynauld & Simonato 1993), most of the preceding literature on forecasting the macroeconomic variables with BVARs has focused on using seasonally adjusted series. Although this is a very standard practice in economics, in the context of nowcasting it is not perfectly innocuous. Often when nowcasting, the interest lies on the actual changes on the levels of the economic variables, and not only on the trend that might be more informative about where the economy is headed in a slightly longer horizon. Pre-adjusting the series might also result in a loss of information about useful dynamics. However, incorporating the seasonal variation into the model often turns out to be a challenge that restricts and complicates the analysis.

Main contribution of this paper is to develop a seasonal Bayesian vector autoregressive model with Google search data for nowcasting the unemployment level in all 28 EU-countries, and to show that making a handful of simplifying modeling choices results in a model that is both competitive and easy to implement. Those simplifying modeling choices include a simplistic way of dealing with the so called *ragged edge* of the data and an efficient, theoretically well grounded, and easy to implement method for choosing the hyperparameters by exploiting the marginal likelihood function. That method is based on Giannone et al. (2015).

On top of using the official data releases of economic variables from Eurostat, the model utilizes Google search data, and in this paper various different methods for incorporating the Google data are considered. The most successful

method used in this paper for incorporating the Google data into the model is a so called *Google index*, that was originally proposed by Choi & Varian (2009). It was later included in a model similar to the one developed in this paper by Tuhkuri (2016*b*), and has ever since been used in the Etlanow forecasting project[1].

Although the Google search data has been previously studied in the context of macroeconomic forecasting, to the knowledge of the author's, the predictive power of Google data with BVAR models has not been tested before. BVAR models are widely considered to be the number one tool for short term macroeconomic forecasting, and by assessing the performance of Google search data in this context, this paper pursues to answer the question if the Google search data truly contains relevant information for macroeconomic forecasting, that is not already available in more orthodox sources of economic data. The findings of this paper are similar to those of the earlier literature. Google search data seems to provide significant, but very limited amount of additional information for forecasting of macroeconomic variables.

The seasonality in the model is incorporated through monthly dummy variables. One could argue in favor of a more sophisticated approach when dealing with seasonality, and alternatively a steady state prior could be used (Villani 2009). In one of the few recent studies concerning forecasting with non-seasonally adjusted series and BVAR models, Stelmasiak & Szafranski (2016) show that both dummy variables with Minnesota-type prior and a steady state prior work well, when predicting the non-seasonally adjusted inflation in Poland. They also report the steady state prior to have an edge over the dummy variable setup in forecasting performance, but that the differences in performance are not massive, and that they come with computational costs, more complex structure and need for stronger prior assumptions on seasonal factors. Using dummy variables allows us to preserve the flexible Minnesota-type prior of the model, which in turn allows for efficient sampling, minimum prior assumptions on seasonal factors and easier exploitation of marginal likelihood function for choosing the hyperparameters.

Typically, the information set available for nowcasting has several features that complicate the modeling design when dealing with vector autoregressive models. Often there is a great number of variables available to choose from, and a not-very-parsimonious lag structure easily causes the number of estimated coefficients to explode. The frequency of the variables might also differ and the differences in publication lags thus lead to a so called *ragged edge* of the data. The former issue is known as the *curse of dimensionality*. Typically, the dimensionality related issues are handled by using techniques that extract common factors of a large number of variables (Dynamic factor models) or by Bayesian shrinkage (e.g BVARs), whereas the ragged edge and the frequency related issues are tackled by exploiting the state space representation of the model and the Kalman filter.

The model in this paper is restricted to use only a few variables, which alleviates the issues discussed above. On top of Google data and the unemployment rate itself, the model only uses consumer price index and confidence indicators as additional explanatory variables. This is partly due to a lack of easily accessible and timely data that would be useful in predicting the unemployment,

[1] https://www.etla.fi/etlanow/

and partly because of the non informative prior on seasonal components of the model, which further restricts the capability of the model to handle a large number of variables.

As all the series are of monthly frequency, we have a very reasonable number of variables and because there are always observations available for every explanatory variable for the month in question when the official unemployment numbers are released, we are not required to exploit the state space representation of the model nor the Kalman filter. A simplistic method for dealing with the ragged edge, without compromising the model performance, by lagging the explanatory variables when needed prior to estimating the model, is proposed. This allows for exploiting the information on explanatory variables also from the periods where the official unemployment rates are yet to be released.

After country specific forecasts have been made, they can be used to create an aggregate forecast for the unemployment rate in the EU as a whole. The countries are given weights according to their share of the working population in the EU, and the point estimates of the EU forecast are essentially weighted averages of the country specific forecasts. However, in order to construct credible prediction intervals for these forecasts, normal approximation and sampling methods are needed. A not entirely innocuous assumption of the independence of the country specific forecasts is made, which allows for sampling from the joint distribution of normally approximated forecasting densities of the country specific forecasts. This results in approximated forecasting densities for unemployment rate in the EU, which are found to perform remarkably well.

The next section focuses on the structure of the model itself and discusses more elaborately on the modeling choices mentioned above. Also, the aggregation of the country specific forecasts to the EU level is discussed in more depth. The third section covers data related topics such as the variables used by the model, the Google search data and the ragged edge of the data. In the fourth section the forecasting ability of the model is assessed against a few benchmark models. For clarity, the assessment is done mostly from the point of view of point estimates and *root mean squared forecasting error* of the model, with different specifications. The results of this assessment support the modeling choices made in this paper. The inclusion of Google data into the model is found to yield modest improvements in terms of forecasting errors, and the proposed method for choosing the hyperparameters of the model is found to perform much better than using the so called rule-of-thumb values. Section five concludes.

2 Model

2.1 Seasonal Bayesian vector autoregressive model

The seasonal vector autoregressive (VAR) model can be represented as:

$$\boldsymbol{y}_t = \boldsymbol{D}\boldsymbol{s}_t + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + ... + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma}), \tag{1}$$

where $\boldsymbol{y}_t$ is an n-dimensional vector of observed variables at time t, $\boldsymbol{D}$ is a matrix of seasonal parameters, $\boldsymbol{s}_t$ is a 12-dimensional vector of seasonal dummy variables, $\boldsymbol{A}_1, ..., \boldsymbol{A}_p$ are the coefficient matrices and $\boldsymbol{\varepsilon}_t$ is a normally distributed vector of residuals with a covariance matrix $\boldsymbol{\Sigma}$.

With a large number of variables and lags, vector autoregressive models are good at capturing the complex dynamics of the economy, which leads to a good in-sample fit. However, when adding variables and lags to the model, the number of parameters grows quickly, which might lead to overfitting and poor out-of-sample forecasting performance. To mitigate this issue, priors on the parameters to be estimated can be imposed. By imposing prior believes, the model is shrunk towards a more parsimonious model and the overfitting problem can be resolved if the degree of shrinkage is chosen with care.

Minnesota type priors for vector autoregressions date back to Litterman (1979, 1980). Litterman's original prior formulation was based on stylized facts about macroeconomic data from the United States. He argued that most of the economic variables could be characterized by unit root processes and thus proposed that each variable should be shrunk towards a univariate random walk process. Although the Minnesota prior can be traced as far back as eighties, it still has many desirable properties and it has proven to be a very efficient approach for Bayesian shrinkage.

Later the Minnesota prior has been revised by for example Doan et al. (1984), Sims (1993), Litterman (1986), Kadiyala & Karlsson (1997) and Sims & Zha (1998). The prior used in this paper is very standard and the exposition follows closely the notation in Banbura et al. (2010) and Itkonen & Juvonen (2017) with a few minor adjustments. The prior mean for the coefficient matrix of the model is set as:

$$\mathbb{E}[(\boldsymbol{A}_l)_{ij} \mid \boldsymbol{\Sigma}, \boldsymbol{\delta}] = \begin{cases} \gamma_i, & \text{if } j = i, l = 1 \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

where $\gamma = 1$ for non stationary variables, $\gamma = 0$ for stationary variables and coefficient matrices$\boldsymbol{A}_1, ..., \boldsymbol{A}_p$ are assumed to be normally distributed. In other words, the prior suggests the non stationary variables to follow a first order unit root process, whereas stationary variables are suggested to be white noise. The prior for the covariance matrix is then set as:

$$cov[(\boldsymbol{A}_l)_{ij}(\boldsymbol{A}_r)_{hm} \mid \boldsymbol{\Sigma}, \boldsymbol{\delta}] = \begin{cases} \left(\frac{\lambda_1}{l^{\lambda_3}}\right)^2 \frac{\boldsymbol{\Sigma}_{ih}}{\boldsymbol{\Psi}_{jj}}, & \text{if } j = m, l = r \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where $l, r \in \{1, ..., p\}, i, j \in \{1, ..., n\}$, λ_1 accounts for overall tightness of the prior and λ_3 controls the lag decay rate. In other words, as hyperparameters λ_1 and λ_3 are given smaller values, the coefficients are shrank harder towards the cautious prior, thus controlling for overfitting. On the other hand, if $\lambda_1 = \infty$ the posterior coefficients coincide with the OLS estimates.

The term $\frac{\boldsymbol{\Sigma}_{ih}}{\boldsymbol{\Psi}_{jj}}$ accounts for different variances of the dependent and explanatory variables and thus scales the prior properly. Ψ is the prior mean of the covariance matrix of the residuals and it is set to $\mathbb{E}[\boldsymbol{\Sigma}] = \boldsymbol{\Psi} = diag(\sigma_1^2, ..., \sigma_n^2)$, where the diagonal elements are chosen to equal the estimated residual variance from the univariate ar-process for the corresponding variables. The elements on the diagonal could also be treated as hyperparameters to be optimized as in Giannone et al. (2015), and it must be noted that using the ar-residuals from the data as a part of the prior is not completely innocuous. However this practical approach has established itself in the literature and should not cause any practical concerns. To keep the complexity of the model at minimum, the

ar-residuals are thus used in the prior. However, in contrary to Giannone et al. (2015) the hyperparameter λ_3 is treated as endogenous and to be optimized, as opposed to just setting $\lambda_3 = 2$.

Next, a non informative but proper prior with zero mean is imposed on the seasonal parameters of the model, that also work as intercepts in the model. This prior is imposed by setting an arbitrary large prior variance λ_4 for the seasonal parameters. Unfortunately the Normal-Wishart prior does not allow for using any sensible informal prior on the parameter estimates of the seasonal components and this greatly reduces the number of variables able to be estimated efficiently with this choice of prior. Another option for vector autoregressive model with seasonal data would be to use a so called steady state prior (Villani 2009). Steady state prior would allow for sensible informal priors on the seasonal parameters of the model, but the number of variables could in that case also be restricted, because of the computationally more demanding nature of that particular prior. Also, choosing a sensible prior on the seasonal components of every EU country in the model would not be an arbitrary task either.

In addition to the priors described above, two additional priors are imposed to enhance the forecasting performance of the model. In the exposition of these priors the notation in Giannone et al. (2015) is closely followed.

As the Minnesota prior only supports first order unit root processes, Doan et al. (1984) suggested imposing a so called *sum-of-coefficients* (SOC) prior. SOC can be imposed by constructing a set of n artificial observations as follows:

$$\boldsymbol{y}^{+} = diag\left(\frac{\bar{\boldsymbol{y}}_0}{\lambda_5}\right) \tag{4}$$

$$\boldsymbol{x}^{+} = \left[0, \boldsymbol{y}^{+}, ..., \boldsymbol{y}^{+}\right], \tag{5}$$

where $\bar{\boldsymbol{y}}_0$ is an $n_u \times 1$ vector that contains the average of the p first observations for each variable, $\boldsymbol{y}^{+}$ is an $n_u \times n_u$ dimensional matrix, $\boldsymbol{x}^{+}$ is an $n_u \times (c + n_u p)$ dimensional matrix, $c = 12$ is the number of seasonal parameters by variable, and λ_5 is the hyperparameter controlling the strength of this prior. The parameter n_u stands for the number of non stationary variables in the model. This prior thus puts weight on the unit root processes of higher order, in addition to those of order one implied by the Minnesota prior. As mentioned, the original Minnesota prior reflected the stylized facts of the US macro economy in a simplest way possible. However, the possibility of variables to a priori follow a higher order unit root processes than of order one, and thus to have correlation among their own lags, better fits our perception of reality. Doan et al. (1984) show that imposing these believes through *sum-of-coefficient* prior can significantly improve the forecasting accuracy of a BVAR model. It is also important to notice that it makes no sense to impose this prior on stationary variables, if in the prior it is assumed that the stationary variables follow a white noise process.

The sum-of-coefficents prior is however not consistent with cointegration of macroeconomic variables, which motivated Sims (1993) to introduce a *dummy-initial-observation* (DIO) prior. It can be implemented by adding the following artifical observation to the dataset:

$$\boldsymbol{y}^{++} = \frac{\bar{\boldsymbol{y}}_0'}{\lambda_6} \tag{6}$$

$$\boldsymbol{x}^{++} \quad = \quad \left[\frac{1}{c\lambda_6}, ..., \frac{1}{c\lambda_6}, \boldsymbol{y}^{++}, ..., \boldsymbol{y}^{++}\right], \tag{7}$$

where $\bar{\boldsymbol{y}}_0'$ equals $\bar{\boldsymbol{y}}_0$, $\boldsymbol{x}^{++}$ is a $1 \times (c + np)$ vector, and λ_6 controls the strength of the prior. Thus, setting hyperparameters λ_5 and λ_6 to infinity would equal to ignoring the dummy observation priors altogether, whereas setting the hyperparameters to zero would put all the weight on these priors, therefore ignoring the data entirely.

Finally, the artificial observations above are stacked together with the observed data, thus constructing the $\boldsymbol{X}$ and $\boldsymbol{Y}$ data matrices to be used in the estimation of the model. The Normal-Wishart prior used here is a natural conjugate prior for normal multivariate regressions (Karlsson 2012). It assumes a normally distributed coefficient matrix $\boldsymbol{A}$ conditional on the covariance matrix $\boldsymbol{\Sigma}$, that follows an Inverse-Wishart distribution. Due to convenient properties of the Normal-Wishart prior, the estimates can be obtained through direct sampling, which makes the estimation procedure computationally very efficient. The computational efficiency combined with the benefits of Bayesian shrinkage, has lead to a development of very large Bayesian vector autoregressive models with even hundred or more variables (see e.g Banbura et al. 2010, Koop 2013). However, these large BVAR models have always used seasonally pre-adjusted data and thus the non informativeness of the prior on the intercepts has not become an issue. With monthly seasonal dummies, we must restrict our number of variables to be reasonably small to avoid overfitting through seasonal components, but in the other hand we can make use of the almost arbitrarily long lag lengths made possible by Bayesian shrinkage.

2.2 Hyperparameter choice

Above, the prior of the model has been specified as a function of five separate hyperparameters $(\lambda_1, \lambda_3, \lambda_4, \lambda_5, \lambda_6,)$. These hyperparameters do not include λ_2, as in the original Minnesota prior it would portray the difference in prior variance between coefficients on own lags and coefficients on lags of other variables. To impose the Normal-Wishart prior, λ_2 must however be normalized to unity.

The hyperparameters controlling the lag decay and the prior variance of the intercepts are often set to two and to some arbitrarily large number, respectively. For choosing the other three, $\lambda_1, \lambda_5, \lambda_6$, there have been various different approaches in the literature. The *rule of thumb* values of 0.2, 1 and 1, respectively, originally proposed by Sims & Zha (1998), have proved to perform well in many cases, and consequently they are a popular choice in the literature.

As variables or lags are added to the model, the hyperparameters should also be revised in order to account for the increased risk of overfitting. Therefore, to compare models of different size the choice of hyperparameters should reflect the number of variables in the model. Two popular choices have been to minimize the out-of-sample forecasting error of some subjectively chosen time interval, or to choose the hyperparameters in a way that the in-sample fit stays the same for all the models of different size (see eg. Banbura et al. 2010). These two approaches however lack a solid theoretical foundation and neither of the approaches have proved to consistently outperform the rule-of-thumb values provided by Sims & Zha (1998). This has further increased the popularity of the rule of thumb values in the literature.

Giannone et al.(2015) however emphasize the fact that "the distinction between parameters and hyperparameters is mostly fictious and made only for convenience". This means that the hyperparameter choice could be treated similarly to estimating the other parameters of the model. They propose imposing a hierarchical structure in the case of a Normal-Wishart prior. The prior structure thus becomes:

$$\pi\left(\boldsymbol{A} \mid \boldsymbol{\Sigma}, \boldsymbol{\delta}\right) \pi\left(\boldsymbol{\Sigma} \mid \boldsymbol{\delta}\right) \pi\left(\boldsymbol{\delta}\right), \tag{8}$$

where $\boldsymbol{\delta}$ stands for the vector of hyperparameters and π is a probability density function. As without the hierarchical structure it could be written as:

$$\pi\left(\boldsymbol{A} \mid \boldsymbol{\Sigma}\right) \pi\left(\boldsymbol{\Sigma}\right) \tag{9}$$

In addition to a theoretically well grounded approach for choosing the hyperparameters, the hierarchical structure allows for taking account for the estimation uncertainty of the hyperparameters as well. This would however require the implementation of a Metropolis algorithm for sampling, and we would have to part ways with the computationally efficient direct sampling.

An approach that simplifies the methodology formalized in Giannone et al. (2015), and takes an advantage of the theoretically well grounded approach for choosing the hyperparameters, while retaining the original Normal-Wishart structure that allows for computionally efficient direct sampling, is proposed here. It is achieved by using the numerical mode of the hyperposterior distribution as the hyperparameter vector in the non-hierarchical setup. This comes with a price of only not accounting for the estimation uncertainty in the hyperparameters.

Giannone et al. (2015) show that applying Bayes' law implies that the hyperposterior distribution is proportional to the marginal likelihood (ML) times the hyperprior distribution.

$$p\left(\boldsymbol{\delta} \mid \boldsymbol{y}\right) \propto p\left(\boldsymbol{y} \mid \boldsymbol{\delta}\right) p\left(\boldsymbol{\delta}\right) \tag{10}$$

Conveniently, the use of Normal-Wishart natural conjugate prior guarantees that the marginal likelihood can be written in a closed form, since it follows a matrixvariate t-distribution (Karlsson 2012). The hyperprior can then be defined on every hyperparameter separately, as long as we assume the hyperparameter values to be a priori independent of each other. Here, similar gamma denitites as in Giannone et al. (2015) are chosen, with the exeption of λ_3 of course, which was treated as exogenous in their study. The prior gamma densities for $\lambda_1, \lambda_3, \lambda_4, \lambda_5, \lambda_6$ are thus chosen so that the modes are equal to 0.2, 2, 50, 1, 1, and the standard deviations are 0.4, 0.4, 30, 1 and 1, respectively.

With the marginal likelihood and hyperprior distribution defined, the mode of the hyperposterior distribution can easily be found by using the numerical optimization methods. In the model of this paper, an 'L-BFGS-B' algorithm, originally formalized in Byrd et al. (1995), is used. The mode acquired this way depends on the data, since it essentially maximizes the marginal likelihood function. The mode, and hence the hyperparameters used, therefore adjust automatically to the changes in the data. If for example a variable is added to the model, the hyperparameters adjust accordingly.

The approach described above for finding the optimal hyperparameters has also very convenient properties in the context of nowcasting. Giannone et al.

(2015) show that using the mode of the hyperposterior distribution is optimal in the sense that it minimizes the one-step-ahead out-of-sample forecasting error, which serves as a great starting point for building a nowcasting model. The model in this paper also strives to nowcast the unemployment in each of the 28 EU-countries, and it is convenient to have the hyperparameters to adjust automatically for each country in every period, without the need for subjective considerations.

2.3 Forecasting

In the model developed in this paper, monthly forecasting horizons up to six periods ahead are considered. Since most EU countries publish their official unemployment rates few weeks after the end of the respective month, in the nowcasting context forecasting periods from current month to five months ahead are considered. The forecasts are obtained by simulating the predictive posterior distribution. As mentioned above, the Normal-Wishart structure allows for direct sampling from the posterior distribution. Since draws are also independent, much less draws are needed than for example in the case of a Metropolis algorithm.

Next the process of drawing from the posterior predictive distribution is briefly described. First, a draw from the posterior Inverse-Wishart distribution of the covariance matrix $\boldsymbol{\Sigma}$ is drawn. After this, a draw from the normally distributed posterior distribution of the coefficient matrix $\boldsymbol{A}$, conditional on the previously drawn $\boldsymbol{\Sigma}$, is drawn. This accounts as one draw from the posterior distribution of the coefficient matrix. Thus, repeating this 10 000 times results in the same number of independent draws from the posterior distribution of the coefficient matrix.

The predictive posterior distribution follows a t-distribution and can be approximated as follows (see eg. Karlsson 2012). After every draw from the posterior distribution of the coefficient matrix, a single draw from the predictive distribution of the one-period-ahead forecast, is produced by adding a normally distributed error term, conditional on the covariance matrix $\boldsymbol{\Sigma}$ drawn before, to the forecast. After the one-period-ahead forecast, two- to six-periods-ahead forecasts can be obtained similarly, conditioning on the already obtained forecasts of the shorter horizon. The credible prediction intervals can then be obtained from these predictive distributions. Finally, the equation 11 illustrates this recursive approach for approximating the predictive distribution, with a notation following closely to that of Karlsson (2012).

$$\tilde{\boldsymbol{y}}_{T+h} = \sum_{i=1}^{h-1} \boldsymbol{A}_i \, \tilde{\boldsymbol{y}}_{T+h-i} + \sum_{i=h}^{p} \boldsymbol{A}_i \, \tilde{\boldsymbol{y}}_{T+h-i} + \boldsymbol{D} \, \boldsymbol{s}_{T+h} + \boldsymbol{u}_{T+h} \, , \tag{11}$$

where $\tilde{\boldsymbol{y}}_{T+h}$ is the draw from the predictive distribution, h is the forecast horizon, p is the lag length of the model, $\boldsymbol{A}$, $\boldsymbol{D}$ and $\boldsymbol{s}_{T+h}$ are defined as before, and $\boldsymbol{u}_{T+h}$ is an error term drawn from a multinormal distribution with a zero mean and a covariance matrix $\boldsymbol{\Sigma}$.

2.4 EU aggregation

In order to obtain nowcasts for the unemployment level in the whole EU, the country specific forecasts need to be aggregated. The aggregation is essentially done in a very straight forward fashion, by giving each country a weight relative to it's share of the labor force in the EU. However, with this approach two complications emerge.

First, the official unemployment rates are reported only up to one decimal place, which when aggregating 28 separate forecasts leads to a systematic additional measurement error. Optimally, the non rounded figures for the unemployment rate should thus be computed and then used in the model. However, the official unemployment rate is a survey based statistic, which by construction has some magnitude of measurement error in it. If the structural error caused by rounding, is then estimated to be small enough, it can be assumed not to have a significant effect on the point estimates, and it could then only be taken into account when constructing the credible prediction intervals. The systematic error caused by rounding in this study is estimated to be approximately 0.05 percentage points on average. The error is not negligible, but it is small enough to justify the usage of the rounded official figures, as long as the systematic error is taken into account in the credible prediction intervals.

Second, different EU countries have different publication lags for the unemployment figures and this must be taken into account when producing timely nowcasts based on the latest data releases. Because of this, the model in this paper produces the EU forecasts from the past month to three months ahead, as opposed to the current month to five months ahead interval used in the country level forecasts. For example, if in April one country has yet to publish it's official figures for February, the forecast can only be aggregated up to July, since the model produces forecast on a country level only five periods ahead.

Taken into account the issues discussed above, the aggregated unemployment nowcast for the EU is produced as follows. The first forecast is done for the month before the reference month. With every country the predictive distribution of the forecast with the shortest horizon is chosen. If the official figure is available, the standard deviation of the predictive distribution for that country is considered to be zero. The predictive distributions are then given a weight corresponding to the labor force of a country. Next, the structural error discussed above is incorporated to the model by constructing an additional normally distributed predictive distribution with mean zero and standard deviation equal to the estimated structural error. When aggregating, the country specific forecasts are assumed independent. It is acknowledged that this assumption might be a little unrealistic, but in the short term the correlation between the labor markets of individual countries should not pose too big of a problem. The empirical assessment of the accuracy of these forecasts gives support to this assumption.

Using the independence assumption, the approximated EU-level predictive distribution is obtained by taking a million draws from the joint distribution of the 28 independent and normally approximated predictive distributions, with each distribution given a weight proportional to the country specific labor force.

3 Data

The lack of informative prior on seasonal components of the model leads to a problem of overfitting intercepts discussed earlier. Because of this issue, the number of variables in the otherwise very flexible model has to be restricted. After specifying several models with different number of variables and inspecting their out-of-sample forecasting performance, using only four or five variables looks to be minimizing the out-of-sample forecasting error. Earlier reasonably scarce literature on the seasonal BVAR models with non informative priors on seasonal components supports this finding (eg. Stelmasiak & Szafranski 2016). The fact that BVAR models with seasonally adjusted data are known to have in principle no upper limit on the number of variables highlights the magnitude of this problem (see eg. Banbura et al. 2010, Koop 2013).

On the other hand, there might not be that many useful, easily accessible and timely variables, that would also be available for every EU country. With the exception of confidence indicators and consumer price index data, most of the monthly available data provided by Eurostat has impractically long publication lags from the point of view of nowcasting problem discussed in this paper. Thus, a choice of adding only a country specific economic sentiment indicator and consumer price index to the model is made.

The economic sentiment indicator is composed of a survey based consumer confidence and business sentiment indicators. The biggest advantage of the survey based *soft data*, as opposed to the so called *hard data*, is the nonexistent publication lag coupled with it. From the perspective of nowcasting it is very practical for explanatory variables for the reference month to be available before the official figures of the variable to be nowcasted.

The consumer price index as well is usually available only days or weeks after the end of the reference month. And while there is no clear consensus on the common dynamics of the inflation and unemployment, the consumer price index was observed to have a modest but significant amount of predictive power on the level of unemployment in the model developed in this paper, and it is therefore included in the model.

3.1 Google data

The fourth variable in the model is extracted from the Google search data, provided by Google Trends[2]. The earlier literature on using the Google search data for forecasting of macroeconomic variables suggests that Google searches might provide useful information for predicting the unemployment rates (see eg. Tuhkuri 2016*b*, 2015, Koop & Onorante 2016).

As mentioned in the first section, to the knowledge of the author's, the predictive power of Google data in macroeconomic forecasting has not been tested before in the context of Bayesian vector autoregressive models. The results in this study regarding the predictive power of the Google search data are however very similar to those of the earlier literature. Google search data seems to provide a significant but very limited amount of useful information for forecasting of macroeconomic variables. These results are assessed more closely in the fourth section.

[2]https://trends.google.com/trends/

The Google search terms used in this study are almost the same ones as in the Etlanow project documented in Tuhkuri (2016*a*). The search terms for different countries are provided by 29 European research institutions and they can be accessed from the Etlanow project website[3].

The Google data consists of multiple time series for each country, a single time series containing the search intensity of a specific unemployment related search term in a given geographic area. Two to fifteen search terms per country were used. The Google data is available starting from the year 2004, which greatly reduces the length of the other data used in the model as well.

As mentioned earlier, the seasonal BVAR model seems to perform optimally with only a handful of variables, thus dimensionality reduction techniques must be considered in order to incorporate the search intensity data to the model in an adequate manner.

In the Etlanow project, Tuhkuri (2016*a*) uses a so called Google Index, originally proposed by Choi & Varian (2009), for dimensionality reduction. The Google index is a fairly simple method for compressing the data from multiple series into a one single series by summing the search intensities of every search term at a given point in time. More formally the Google index can be defined as follows:

$$I_t = \left(\frac{\frac{K_t}{G_t}}{max\left(\frac{K_t}{G_t}\right)} \right) \times 100 \ , \tag{12}$$

where I_t is the value of the index at time t, K_t is the amount of searches with a given set of keywords and G_t is the amount of all Google searches in a given geographic area at time t.

In this study, various other methods for incorporating the Google data were considered as well. In total, 47 different dimensionality reduction methods were considered to extract the information from the Google search data. First, using each of those 47 methods, the out-of-sample forecasting errors when forecasting the unemployment rate in Finland were computed. Then those errors were compared against the errors of the benchmark model containing no Google data at all and the most promising methods for dimensionality reduction were chosen. Among the most promising methods were the *Google Index*, traditional *principal component analysis* (PCA) and *robust principal component analysis* (RPCA). Next, the out-of-sample forecasting errors for every EU country, using each of these methods, were computed and compared against the benchmark model.

Addition of Google data did not generate great improvements in forecasting performance of the model. For horizons from one to three steps ahead the models with and without Google data showed no significant differences in performance. For forecasting horizons from four to six periods ahead however, the model with *Google Index* seemed to perform a little, but significantly, better than the rest of the models and thus it was chosen to be the method for dimensionality reduction in the model. More thorough assessment of these results can be found in the fourth section as well.

[3]https://www.etla.fi/en/etlanow/

3.2 Ragged edge

The *ragged edge* of the data means that the last observation of every variable is not from the same period. For example, if one variable is published at the end of the reference month, while the other is published with a delay of several weeks, simply dismissing the published data of the other variable until the other gets published as well, would allow us to use vector autoregressive models. This however, would lead to a loss of available information and thus to a suboptimal forecast.

In the literature there have been different approaches for tackling this issue, and most of them have taken the advantage of the Kalman filter (eg. Itkonen & Juvonen 2017, Schorfheide & Song 2015, McCracken et al. 2015). In the case of the model in this paper, a far simpler approach is however deemed to be sufficient and Kalman filter is not required.

As the model uses only *soft data* as additional explanatory variables, the variable to be forecasted (ie. the unemployment rate) is always the last variable to be published for a given period. The consumer price index and the confidence indicators are always released at latest a couple weeks after the end of the reference month, while daily Google search data to be aggregated to a monthly level is available almost in real time. Official unemployment figures for most EU countries are usually published by the end of the month following the reference month.

Therefore, as new data becomes available the explanatory variables can simply be lagged, so that the last observation of every variable is from the same period as the last official unemployment figure. The lagging causes no conceptual problems when only the forecasting densities are of interest, as is the case of the model developed in this paper.

4 Forecasting performance

The assessment of the forecasting performance of the model is executed by producing pseudo out-of-sample forecasts.[4] It means that for every period to be examined, the forecast is made using only the information set that would have been available at the time. It is of course impractical to use the exact publication dates of different variables for these pseudo out-of-sample forecasts, and thus some assumptions must be made. Forecasts in this section are produced at the end of every month, and for most countries it is reasonable to assume that at the end of every month there is Google search data available for the current month, while the last observation of every other variable is from the month before that. In practice this means that for every forecast the Google data is lagged by one period, and after that the data is cut so that the last observation of every variable is from the month before the reference month.

In the literature assessing the forecasting performance of the BVAR models the most commonly used error measure is the *Root Mean Squared Forecasting Error* (RMSFE), and it is used as the primary error measure in this study as well. RMSFE measures the squared forecasting error of every observation and then takes the square root of the mean of those observations for improved in-

[4] All the analysis and computations presented in this chapter were carried out using R software package (R Core Team 2018).

	Random Walk	Seasonal RW	Seasonal AR	Seasonal VAR	Seasonal BVAR
Austria	0.37	**0.33**	0.37	0.39	0.34
Belgium	0.23	0.21	0.42	**0.15**	**0.15**
Bulgaria	0.30	0.34	0.37	**0.17**	0.18
Croatia	0.75	0.44	0.85	0.38	**0.25**
Cyprus	0.92	0.68	1.06	**0.51**	0.56
Czech Republic	0.33	0.39	0.35	0.26	**0.24**
Denmark	0.20	0.16	0.24	0.29	**0.15**
Estonia	**0.45**	0.52	**0.45**	0.55	0.52
Finland	0.89	0.53	0.56	0.50	**0.39**
France	0.32	0.17	0.31	**0.16**	**0.16**
Germany	**0.26**	0.34	**0.26**	0.41	**0.26**
Greece	0.98	1.05	1.02	0.76	**0.69**
Hungary	0.19	0.34	**0.17**	**0.17**	0.19
Ireland	0.30	0.32	0.57	0.15	**0.14**
Italy	0.81	0.55	**0.53**	0.54	0.59
Latvia	0.25	0.35	0.40	0.26	**0.24**
Lithuania	0.49	0.47	0.74	0.43	**0.42**
Luxembourg	0.22	0.15	0.27	**0.14**	0.15
Malta	0.14	0.22	0.34	0.19	**0.16**
Netherlands	0.28	0.17	0.21	0.14	**0.13**
Poland	0.27	0.54	0.35	0.14	**0.13**
Portugal	0.30	0.33	0.34	**0.18**	0.19
Romania	0.29	0.32	0.42	0.29	**0.28**
Slovakia	0.21	0.45	0.35	0.15	**0.14**
Slovenia	0.32	0.23	0.43	**0.13**	**0.13**
Spain	0.37	0.39	0.33	**0.16**	**0.16**
Sweden	0.66	0.41	0.45	0.36	**0.32**
United Kingdom	0.15	0.16	0.16	0.11	**0.10**
EU28	0.22	0.17	0.16	0.11	**0.09**
Median	0.30	0.34	0.37	0.22	**0.19**

Table 1: *Root mean squared one-step-ahead forecasting errors of out-of-sample forecasts from first month of 2014 to the third of 2018. With seasonal RW, VAR and BVAR models the seasonality is accounted for by monthly seasonal dummy variables.*

terpretation. Other error measure that is often considered is the *Mean Absolute Forecasting Error* (MAFE), which does not punish for extreme errors as much as the RMSFE does. On top of these measures, it is important for a forecasting model to be unbiased. This can be inspected by computing the *Mean Forecasting Error* (MFE) of the forecasts. For the unbiased model this should be close to zero. In this study all the forecasts were observed to be reasonably unbiased, as is often the case with BVAR models, and reporting other measures on top of RMSFE would not add much informational value. Reporting more measures would also lead to a decreased interpretability of the tables, and therefore only RMSFE figures are reported in this section. Using MAFE instead of the RMSFE would not change any qualitative results reported.

As mentioned in the third section, the data used in this model starts as late as 2004. In addition to that, the seasonal BVAR is also fairly complex and for efficient estimation of the 256 parameters[5] in the model, using at least 10 years of data is desirable. This shortens our out-of-sample forecasting time interval to a little bit over four years of data, from the first month of 2014 to the third of 2018. While the data is thus reasonably *short*, it is however quite *wide*, since we have data from 28 separate countries, which alleviates the issue.

One further aspect to pay attention to, when executing this out-of-sample forecasting study, is how the dimensionality reduction of the Google search data is to be executed. If the Google index is used for dimensionality reduction, no further attention is required when executing the study, but if principal component analysis is used, the principal component must be computed separately for each period prior to every forecast from the raw search term intensities. It is not enough to cut the already computed series of principal components to achieve *true* out of sample forecast, since earlier observations of this series may contain information from yet to be observed observations.

Another aspect worth mentioning is that since the unemployment rate is a survey based measure, the revisions to the data do not play as big of a role as in the case of GDP. Therefore, there is no compulsive need to worry about data revisions.

4.1 Benchmarks

In order to conduct any meaningful analysis of the forecasting performance of the model, we need to have some benchmarks to compare the model against. First naive benchmark model to be used is the *Random Walk* (RW). It accounts to forecasting each period, that there is to be no change at all from the last observation in the variable to be forecasted. Since the model is seasonal and the data is not seasonally adjusted, the other naive benchmark model that could be considered is the *Seasonal Random Walk* (SRW). SRW could be constructed in a lot of ways. For example, it could be defined as a no-change-forecast from the observation one year ago. However, SRW defined this way was found to perform very poorly for most countries, especially on short horizons. For one-period-ahead forecasts it performed just a little better than RW for only three countries, where the seasonal variation in the unemployment rate seemed to be the strongest[6]. Due to these issues, the SRW is defined in this paper as an

[5] Hyperparameters not included

[6] These countries were Finland, Italy and Sweden

ordinary RW adjusted for seasonal variation by estimating seasonal constants for every month. The constants are estimated by *ordinary least squares* after which the SRW forecast is defined as a sum of no-change forecast and the difference in monthly constants estimated. This results in a naive benchmark model, that should more closely resemble the forecasting performance of an ordinary RW observed in the case of seasonally pre-adjusted data.

Third benchmark model to be used is the seasonal autoregressive (AR) model of order one from Tuhkuri (2016*a*). It uses the same Google search data and Google index as the model developed in this paper. The model can be written as follows:

$$log(y_t) = \beta_0 + \beta_1 log(y_{t-1}) + \beta_2 log(y_{t-12}) + \beta_3 google + e_t, \tag{13}$$

where y_t is the unemployment rate at time t and *google* is the value of the Google index at time t, or at the latest time available. As exactly the same model as in Tuhkuri (2016*a*) is used, the model only produces forecasts from h = 0 to h = 3.

The last benchmark model to be used is an ordinary *vector autoregressive model* (VAR) equipped with monthly dummy variables to account for seasonal variation. As the model developed in this paper does not have more than four variables, the Bayesian VAR (BVAR) might not have too big of an edge over the ordinary VAR and the difference in performance might turn out to be not too massive. The lag length of the VAR-model prior to every forecast is selected by choosing the model specification with the smallest *Akaike information criterion* (AIC). Otherwise the VAR is defined as the BVAR earlier, only without the prior.

Median RMSFE	h = 0	h = 1	h = 2	h = 3	h = 4	h = 5
Random Walk	0.30	0.49	0.66	0.80	0.88	0.94
Seasonal RW	0.34	0.42	0.51	0.58	0.65	0.68
Seasonal AR	0.30	0.47	0.57	0.65	-	-
Seasonal VAR	0.22	0.32	0.42	0.48	0.52	0.56
Seasonal BVAR	**0.19**	**0.30**	**0.40**	**0.45**	**0.47**	**0.50**

EU aggregate RMSFE	h = 0	h = 1	h = 2	h = 3
Random Walk	0.22	0.34	0.44	0.52
Seasonal RW	0.34	0.42	0.51	0.58
Seasonal AR	0.16	0.30	0.38	0.43
Seasonal VAR	0.11	0.13	0.15	0.17
Seasonal BVAR	**0.09**	**0.12**	**0.14**	**0.15**

Table 2: *Root mean squared forecasting errors for all the horizons of out-of-sample forecasts from first month of 2014 to the third of 2018.*

4.2 Out-of-sample performance

In order to compare the performance of the models on a country level, the Table 1 presents the one-step-ahead root mean squared forecasting errors of the models for every country. In practice the one-step-ahead forecast is the nowcast for the

current month that uses the information of the preceding months. The seasonal BVAR seemed to perform better than the benchmark models in most of the countries. In only eight out of twenty-eight countries did the seasonal BVAR not provide the best performance. In most cases however, the BVAR produced much more accurate forecasts, while at worst it was only barely less accurate than the most accurate model. In other words, in none of the 28 countries did any benchmark model yield significantly more accurate forecasts than the BVAR. Those few times the BVAR performed worse than a benchmark model, could also be owing to the short length of the sample used for computing the errors.

On the other hand, in many countries the ordinary VAR was approximately as accurate as the BVAR. There were however a few countries, such as Finland, Germany and Denmark, in which the VAR performed significantly worse, possibly due to overfitting. These results emphasize the fact that the seasonally non-adjusted data poses a major challenge for the BVAR models, and the gains of using Bayesian techniques remain very limited when working with vector autoregressive models and seasonally non-adjusted data.

The forecasting errors produced by the BVAR for EU aggregate forecast were almost half the errors of the seasonal AR-model and clearly smaller than those of the VAR as well. The same story holds for the median errors. The median RMSFE of the BVAR for a country level forecast was almost half the median RMSFE of the seasonal AR-model, and a little bit less than that of a VAR. Random walks produced slightly more accurate median forecasts than the AR-model, but the errors of the EU level forecasts were clearly smaller for the AR-model.[7]

Thus, at least the one-step-ahead forecasts of the seasonal BVAR seem to be most accurate of the models compared, the VAR claiming the second place by a narrow margin. In Table 2 however, is the median RMSFE of the country level forecasts and RMSFE for the EU aggregate forecast, with forecasting horizons up to six-periods-ahead[8]. The median forecasts of the BVAR seemed to be more accurate than those of the VAR for all forecasting periods, with both of these models outperforming the other benchmarks by a wide margin. The forecasting errors of the BVAR for EU aggregate forecasts were also smaller than those of the benchmark models, the VAR model coming as a close second.

The forecasting errors of the BVAR model thus seemed to be the smallest for every forecasting horizon and the difference in performance between the BVAR and the VAR stayed reasonably constant over all the horizons.

4.3 Performance of Google variables

As mentioned in the third chapter, the addition of the Google data leads only to minor improvements in the performance of the model. Similar findings regarding the limited but statistically significant role of Google data in forecasting of macroeconomic variables are presented for example in Tuhkuri (2015, 2016*b*) and Koop & Onorante (2016). In this study the seasonal nature of the model restricted greatly the number of variables in the BVAR model, and thus only the

[7]The median was chosen here instead of the mean, so that the extremely bad forecasts would not affect the metric more than the extremely good ones. All the qualitative results would however hold had the mean being chosen instead of the median.

[8]Here *six-periods-ahead* equals $h = 5$

Median RMSFE	h = 0	h = 1	h = 2	h = 3	h = 4	h = 5
No Google	**0.18**	0.30	0.39	0.47	0.51	0.55
Google Index	0.19	0.30	0.40	**0.45**	**0.47**	**0.50**
PCA	0.19	0.30	**0.38**	0.46	0.52	0.55
RPCA	0.20	**0.29**	0.39	0.47	0.51	0.58

EU aggregate RMSFE	h = 0	h = 1	h = 2	h = 3
No Google	**0.09**	**0.12**	0.15	0.18
Google Index	**0.09**	**0.12**	**0.14**	**0.15**
PCA	0.10	0.13	0.16	0.20
RPCA	**0.09**	0.13	0.16	0.19

Table 3: *Root mean squared forecasting errors of the seasonal BVAR model for all the horizons of out-of-sample forecasts from first month of 2014 to the third of 2018, with different methods for dimensionality reduction.*

one dimensional specification of Google data could be considered. That might have been a major reason for the lack of increased forecasting performance when testing the several dimensionality reduction techniques. As in the case of a seasonally adjusted BVAR model one would not have to worry about the number of variables, the inclusion of Google search data with more principal components, or even without any dimensionality reduction steps, could turn out to be more successful. This is something to consider in further studies on the matter but falls beyond the scope of this study.

The Table 3 presents the root mean squared forecasting errors for median and EU aggregate forecasts using no Google data at all and three different dimensionality reduction techniques. As can be seen from the table, the modest improvements in forecasting performance of the model with Google Index become evident, to a little surprise, only when forecasting three or more periods ahead. The principal component analysis or the robust principal component analysis did not seem to perform any better than the model without any Google data. This was probably because the first principal component alone might not contain much information of interest for the forecasting problem at hand. For example, the first component might only capture the long term trend of the series, while the Google index, despite of it's simplicity, might be better able to capture relevant information about the unemployment.

Another aspect that might have affected the performance of Google variables is how the Google search terms were chosen. More time and effort were probably used for choosing the search terms for some countries, and less for others. A rigorous testing and inspection of different sets of search terms were found to improve the out-of-sample forecasting performance of the model for Finland, but unfortunately the same testing could not be carried out in the same way for the other countries, as different languages set certain restrictions for the author.

Even though Google search data did not seem to provide major improvements in forecasting performance, it however seems to contain some useful information for economic forecasting. Addition of Google search data to the model should thus be considered especially when large set of explanatory variables is not a problem, or when the forecasting accuracy of the model is of great importance.

Median RMSFE	h = 0	h = 1	h = 2	h = 3	h = 4	h = 5
Sims-Zha	0.20	0.34	0.44	0.51	0.57	0.61
Hyperposterior Mode	**0.19**	**0.30**	**0.40**	**0.45**	**0.47**	**0.50**

EU aggregate RMSFE	h = 0	h = 1	h = 2	h = 3
Sims-Zha	0.10	0.14	0.18	0.21
Hyperposterior Mode	**0.09**	**0.12**	**0.14**	**0.15**

Table 4: *Root mean squared forecasting errors of the seasonal BVAR model for all the horizons of out-of-sample forecasts from the first month of 2014 to the third of 2018, using rule-of-thumb values of Sims & Zha (1998) and hyperposterior mode.*

4.4 Assessment of the hyperparameter choice

One interesting modeling choice that we can assess by inspecting the out-of-sample forecasting errors of the model is the hyperparameter choice. In the second section, a method for choosing the hyperparameters according to the numerical mode of the *hyperposterior distribution* was proposed. In order to see if that method truly improved the forecasting performance of the model, we can compare the out-of-sample errors attained by using the numerical mode to those that would have been attained using some other method for hyperparameter choice. A natural choice for a benchmark is a model with the rule-of-thumb values originally proposed by Sims & Zha (1998). These rule-of-thumb values are also the prior means for the hyperparameters when constructing the prior distribution for the hyperparameters.

The Table 4 presents the results of this assessment. The out-of-sample point estimates seem to be more accurate with all the forecasting horizons when using the hyperparameter values implied by the mode of the hyperposterior distribution. Using this theoretically well grounded and computationally efficient method seems to yield significantly more accurate forecasts, than using the common rule-of-thumb values would. It must however be noted that the model in this study was fairly parsimonious, as despite of it's 256 parameters it only had four variables. In Giannone et al. (2015) it is suggested, that the rule-of-thumb values might be closer to optimal with more variables in the model. Also, only the accuracy of the point estimates is assessed here, and to attain a full picture of the performance of the model, the whole predictive distributions should be assessed. This could lead to speculation in favor of using the whole hyperposterior distribution to account for hyperparameter uncertainty, instead of using only the mode of the hyperposterior distribution. Computational costs of using the whole hyperposterior distribution are however significant, and it might not always be advisable to part ways with the computationally efficient way of not imposing the full hierarchical layer to the model. Empirical evidence in this study suggests that using only the mode of the hyperposterior distribution, instead of the whole hyperposterior distribution, yields very competitive results and should be considered whenever optimal forecasting performance is pursued, and the computational efficiency of the model must be considered.

5 Conclusions

In this paper, a seasonal Bayesian vector autoregressive model for nowcasting the unemployment rate in the EU-countries, and in the EU as a whole, is developed. This paper adds to the reasonably scarce literature on economic forecasting of seasonally non-adjusted variables with Bayesian vector autoregressive models. It is shown, that even with reasonably simple modeling choices a seasonal model for unemployment with considerable forecasting accuracy can be build. Further research on the topic could include a comparison of seasonally adjusted forecasts attained from a typical BVAR model with seasonally adjusted data, and from a seasonal model like the one developed in this paper. This way the effect of addition of seasonal factors on the dynamics of the model and on it's forecasting accuracy could be assessed.

Another contribution of this paper was to test the predictive power of Google search data and various methods for incorporating it within an already accurate BVAR model. Dimensionality reduction methods were not found to lead to any improvements in forecasting accuracy of the model in this paper. This could however be due to the fact, that the first principal component alone did not contain enough information to be of use in forecasting. Unfortunately, adding more than one principal component to the model was found to decrease the accuracy of the model due to non-informative prior on the seasonal components. In further studies on the topic, the predictive power of Google search data could be tested in a context of a BVAR model using seasonally adjusted data, since there is found to be in practice no upper bound on the number of variables in those kind of models.

Finally, this paper contributes to the already existing empirical literature on the hyperparameter choice with Bayesian vector autoregressive models. A method exploiting the numerical mode of the hyperposterior distribution is proposed, and the out-of-sample forecasting study of 28 separate countries implies that using the marginal likelihood function for hyperparameter choice, instead of the rule-of-thumb values often used in the literature, can lead to a significant increase in forecasting accuracy.

References

Banbura, M., Giannone, D. & Reichlin, L. (2010), 'Large bayesian vector auto regressions', *Journal of Applied Econometrics* **25**, 71–92.

Barnett, W. A., Chauvet, M. & Leiva-Leond, D. (2016), 'Real-time nowcasting of nominal gdp with structural breaks', *Journal of Econometrics* **191**, 312–324.

Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. (1995), 'A limited memory algorithm for bound constrained optimization', *SIAM Journal on Scientific Computing* **16**(5), 1190–1208.

Choi, H. & Varian, H. (2009), Predicting the present with google trends, Working papers.

Doan, T., Litterman, R. B. & Sims, C. A. (1984), 'Forecasting and conditional projection using realistic prior distributions', *Econometric Reviews* **3**(1), 1–100.

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

Itkonen, J. & Juvonen, P. (2017), 'Nowcasting the finnish economy with a large bayesian vector autoregressive model', *BoF Economics Review* .

Kadiyala, K. R. & Karlsson, S. (1997), 'Numerical methods for estimation and inference in bayesian var-models', *Journal of Applied Econometrics* **12**(2), 99–132.

Karlsson, S. (2012), Forecasting with bayesian vector autoregressions, Working Papers 12, Örebro University, School of Business.

Koop, G. M. (2013), 'Forecasting with medium and large bayesian vars', *Journal of Applied Econometrics* **28**(2), 177–203.

Koop, G. M. & Onorante, L. (2016), Macroeconomic nowcasting using google probabilities, Working papers.

Litterman, R. (1979), Techniques of forecasting using vector autoregressions, Working Papers 115, Federal Reserve Bank of Minneapolis.

Litterman, R. (1980), A bayesian procedure for forecasting with vector autoregression, Working papers, Massachusetts Institute of Technology.

Litterman, R. (1986), 'Forecasting with bayesian vector autoregressions-five years of experience', *Journal of Business & Economic Statistics* **4**(1), 25–38.

McCracken, M. W., Owyang, M. T. & Sekhposyan, T. (2015), Real-time forecasting with a large, mixed frequency, bayesian var, Reserve bank of st. louis working paper series.

Modugno, M. (2011), Nowcasting inflation using high frequency data, Working Paper Series 1324, European Central Bank.

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Raynauld, J. & Simonato, J.-G. (1993), 'Seasonal bvar models: A search along some time domain priors', *Journal of Econometrics* **55**, 203–229.

Schorfheide, F. & Song, D. (2015), 'Real-time forecasting with a mixed-frequency var', *Journal of Business I& Economic Statistics* **33**, 366–380.

Sims, C. (1993), A nine-variable probabilistic macroeconomic forecasting model, *in* 'Business Cycles, Indicators and Forecasting', National Bureau of Economic Research, Inc, pp. 179–212.

Sims, C. & Zha, T. (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review* **39**(4), 949–68.

Stelmasiak, D. & Szafranski, G. (2016), 'Forecasting the polish inflation using bayesian var models with seasonality', *Central European Journal of Economic Modelling and Econometrics* **8**(1), 21–42.

Tuhkuri, J. (2015), Big data: Do google searches predict unemployment?, Master's thesis, University of Helsinki, Faculty of Social Sciences, Department of Political and Economic Studies.

Tuhkuri, J. (2016*a*), 'A Model for Forecasting with Big Data, Forecasting Unemployment with Google Searches in Europe'.

Tuhkuri, J. (2016*b*), Forecasting unemployment with google searches, Working Papers 35, ETLA.

Villani, M. (2009), 'Steady-state priors for vector autoregressions', *Journal of Applied Econometrics* **24**(4), 630–650.

Elinkeinoelämän tutkimuslaitos

The Research Institute of the Finnish Economy

ISSN-L 2323-2420
ISSN 2323-2420 (print)
ISSN 2323-2439 (pdf)

Tel. +358-9-609 900
www.etla.fi
firstname.lastname@etla.fi

Arkadiankatu 23 B
FIN-00100 Helsinki