ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Bachelor Thesis

Econometrie & Operationele Research

Demographic Structure and Macro-economic Trends: A VAR and BVAR forecasting comparison

Name student: Kiki Hajer

Student ID number: 476072

Supervisor: Wendun Wang

Second assesor: Andrea Naghi

Date final version: July 5th, 2020

Abstract

In this research, the effects of age structure on macroeconomic trends, estimated with a vector autoregression model, are investigated for 21 OECD economies for the period 1970 to 2014. Furthermore, we make conditional predictions of economic growth, incorporating demographic change, for each country. Those are compared with forecasts made with a Bayesian vector autoregression model, which is implemented with a diffuse prior and the Minnesota prior. To compare those forecasts with the conditional predictions and with each other, the performance measures RMSE and MAE are calculated for a total of five steps ahead. Our main results showed that the conditional predictions made with the VAR model performed best on the long-run. The Minnesota prior performed better than the diffuse prior, except for five steps ahead.

Contents

1 Introduction

A slowdown in macroeconomic performance has been observed, ever since the Great Recession. Concerns about the long-run prospects for developed economies, as a result of slow recovery and reducing productivity growth, have stirred up the secular stagnation debate. With decreased fertility and increased longevity, the shares of different age groups in the population shifted towards an increased share of retirees. It is common knowledge that all age groups differ in their savings behavior, investment opportunities, labor input and productivity levels. As a consequence, changes in the age structure of the population might have a large impact on macroeconomic performance. We therefore are interested in the impact age structure has on several key macroeconomic variables, which we will base on an already existing study. Additionally, we are interested in the long-run forecast for these economic variables. Multiple studies have shown that the Bayesian VAR model has better forecasting performance than the VAR model. We would like to investigate whether this result also applies to our data concerning the effect of age structure on economic growth. Hence, the central research question is formulated as follows:

How does age structure affect macroeconomic trends and can we obtain a better forecast for economic growth using Bayesian VAR compared to the VAR model?

This research is based on a study concerning the effect of age structure on macroeconomic trends. They made a conditional prediction for economic growth, incorporating demographics. Such a prediction of how demographic change will affect economic growth in the future can be of great relevance to the OECD countries treated in this research. That way, they can prepare for possible further ageing of the population and the consequences for economic growth that come with that. To build on this, we will try to improve the forecasts for economic growth by implementing the Bayesian VAR model. Earlier studies that compared VAR and BVAR forecasting concluded that the Bayesian approach has better forecasting performance than the VAR model. This made us question whether we could improve this prediction for economic growth by making Bayesian forecasts.

The dataset that is used in this research concerns annual data on 21 OECD countries for the period 1970 to 2014. This includes data on several economic variables which are used to calculate our key macroeconomic variables. There are data on real and nominal GDP, GFCF, National Savings, CPI, policy rates, and data on hours worked. Also, data on the change in real GDP for the period 2015 to 2019 to calculate the performance measures, and data on population is obtained for the period 1970 to 2019. To obtain the demographic effects on our six macroeconomic variables per age group, a VARX(1) model is implemented, with the data on population as exogenous element. The resulting demographic impact matrix is then used to make conditional predictions of economic growth for each country. For forecasting with the Bayesian VAR model, the introduction of a prior is necessary. In this research, two different priors will be implemented, namely the diffuse prior and Minnesota prior. The diffuse prior is a 'non-informative' prior which will be used as our benchmark. The Minnesota prior is a more informative one, which is extensively used in macroeconomic applications. By applying the Bayes Theorem on those priors, one can obtain the posterior distributions, which capture all useful information concerning the unobserved data. To compare the predicted and forecasted growth rates of real GDP of those three methods, the RMSE and MAE are calculated as performance measures, for a total of five steps ahead.

The results show, first of all, that the three different age groups all have different impact on the macroeconomic variables. The calculated performance measures on the forecasted growth rates tell us that the conditional predictions made with the demographic impact matrix are more accurate than both BVAR models, except for the first step ahead. Also, regarding the BVAR models, the implementation of the Minnesota prior performs better than the diffuse prior, except for five steps ahead. It seems that, even though BVAR models have better forecasting performance than VAR models, a conditional prediction made based on a long-run demographic impact matrix, in this case, performed better regarding forecasting. It should be noted that this does not mean that this conditional prediction will always be better at forecasting economic growth. It may be that with implementing different priors in the Bayesian VAR approach, those will perform better.

The remainder of this paper is structured as follows. In the next section, the problem field and used methods of this paper will be discussed more thoroughly, using earlier studies. Then, the data that is used in this paper will be specified. In Section 4, all used methods will be introduced and explained. After that, the results will be displayed and then reviewed in the discussion. Finally, the conclusion will summarise the main results and some last statements will be made, which include practical implications and suggestions for further research.

2 Literature review

It is well known that elderly, young dependents and working people all behave differently regarding savings, investments, labor input, and so on. Therefore, changes in the age structure of the population can have a large effect on the macroeconomy. Kuznets (1930) identified the so-called Kuznets cycles, which connected the changes in demographic structure to macroeconomic aspects. In addition, according to Gordon (2012), who asked how much further the frontier growth rate could decline, demography is one of the six "headwinds" of the observed reduction of economic growth. Multiple other studies have also shown that age structure affects economic growth and other macroeconomic aspects. For example, Andersson (2001), who studied the age-effects in Scandinavian countries, and Lindh and Malmberg (2009), who used a sample of EU countries. Both studies also found a positive influence on growth from the middle-aged group.

Considering that multiple studies have shown that age structure affects economic growth, the question remains whether age structure can be used to make long-run forecasts of economic growth. Relying on known birth and death rates, demographics are quite predictable. Unfortunately, this does not particularly hold for economic variables, which are a lot harder to predict. However, Bloom et al. (2007) studied the implementation of age structure in growth models and found that this significantly improves its forecasting performance. Aksoy et al. (2019) did exactly this when they used a VAR model to make a conditional prediction of several macroeconomic variables, by employing likely demographic change as exogenous variable. The VAR model, originally launched by Sims (1980), possesses great flexibility and ability to fit the data, due to its rich parameterization. This comes with the risk of being overparameterized, which could result in unreliable coefficients, possibly being different from zero by accident. This would result in inconsistent model predictions and uncertainty about the future tracks projected by the model (Karlsson, 2013; Nicholson et al., 2014; Sevinç & Ergün, 2009; Villani, 2009). Litterman (1980) proposed the Bayesian VAR approach as a solution to this overfitting problem. This BVAR model has been put to the test multiple times and has proven its superior forecasting performance (Belloni, 2017; Litterman, 1986; McNees, 1986). The BVAR approach deals with the possible overparameterization of the VAR model by introducing a so-called prior, which contains information about the long-run properties of the data, independent from the observed data, to form the so-called posterior. A posterior expresses everything a researcher knows about the model parameters when looking at the data. There exists a variety of priors that can be used to implement the BVAR model. Our focus will lay on the implementation of two different priors. The first one is a 'diffuse' (or 'non-informative' or 'flat') prior, which is used in Bayesian inference. Those 'non-informative' priors are used to express the probability density function of the observed data as a function of the parameters (Miranda-Agrippino & Ricco, 2018). For this, the so-called diffuse prior or Jeffreys' (Jeffreys, 1961) prior will implemented. The diffuse prior was first introduced by Geisser (1965), Tiao and Zellner (1964), and is mostly used as benchmark, as will we. Alongside this prior, a more informative prior will be adopted. Informative priors are widely used in macroeconomic applications, since they summarise the researcher's personal beliefs in the form of a prior (Miranda-Agrippino & Ricco, 2018). For this, the Minnesota prior, will be implemented. The Minnesota prior, sometimes referred to as Litterman prior, was introduced by Doan et al. (1984) as a shrinkage prior. Litterman (1986) then reviewed this prior and multiple variations have been considered since. The idea behind the Minnesota prior is that each variable follows a random walk, with possible drift, which is a legitimate approximation of the behaviour of economic variables (Litterman et al., 1979).

Another interesting result came from Sevinç and Ergün (2009), who discussed a couple of priors, among which the Minnesota and diffuse prior. They compared their forecasting performances using the RMSE and found that the Minnesota prior performed as one of the best, while the diffuse prior performed worst. It should be noted that in this paper, since we replicate the implementation of the VAR model from Aksoy et al. (2019), this model is not used to make a forecast, but a conditional prediction. Conditional predictions are projections of the variables of interest on the future paths of different variables. Unconditional predictions, on the other hand, assume no knowledge of those future paths (Bańbura et al., 2015). This might provide important information, which could lead to interesting results in the forecasting comparison.

3 Data

The dataset used concerns annual data on 21 OECD countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States) over the period 1970-2014 (945 observations).

The data on population we use come from the World Population Prospects, the 2015 Revision (Nations, 2015). The age structures are calculated using the share of the so-called *de facto* population in the age group in the total population in each year. Note that UN population data measure those living in a certain country, not just its citizens, thus considering immigration.

Additionally, we use annual data on several macroeconomic variables, which are described in Section 4.1. Firstly, data on real and nominal Gross Domestic Product (GDP) were extracted from National Accounts, OECD. Secondly, Gross Fixed Capital Formation (GFCF) and National Savings, which are used to determine the investment and savings rate, respectively, were also drawn from National Accounts, OECD. Then, data on hours worked were extracted from Productivity Statistics, OECD. Finally, data on Consumer Price Index (CPI) and policy rates were extracted from International Financial Statistics/IMF. Data on the change in real GDP from 2015 to 2019, which are used to measure forecasting performance, were also drawn from National Accounts, OECD.

Some countries do not have complete datasets over the whole period, which makes the panel unbalanced. Data on hours worked are not available until 1983 for Greece and 1995 for Austria. Savings rates are not available until 1978 for France and Norway and not until 1990 for Switzerland. All other countries consist of complete datasets. MATLAB handles those missing values by removing the entire observation that contains at least one missing value.

4 Methodology

Since the main goal of this paper is to test the forecasting performance of the method introduced by Aksoy et al. (2019) against Bayesian VAR forecasting, we will first replicate their methods. Therefore, the notation, introduced in Section 4.1, and the methodology employed for the VAR model, described in Section 4.2, are adopted from their article. After that, we introduce the BVAR model, implementing two different priors.

4.1 General notation

First, some general notation will be introduced. In the panel VAR we have a vector of macroeconomic variables, $\mathbf{Y}_{it}$, and a vector of population shares, $\mathbf{W}_{it}$, for countries $i = 1, 2, ..., n$.

In the benchmark specification, we have the following six endogenous variables: growth rate of the real GDP, g_{it}, share of investment in GDP, I_{it}, share of personal savings in GDP, S_{it}, the logarithm of hours worked per capita, H_{it}, real short-term interest rate, rr_{it}, rate of inflation, π_{it}. These six macroeconomic variables can be written in a vector as

$$\mathbf{Y}_{it} = (g_{it}, I_{it}, S_{it}, H_{it}, rr_{it}, \pi_{it})',$$

where g_{it} is calculated by the change in real GDP between time t to $t+1$. I_{it} and S_{it} are computed by dividing the national GFCF and national savings by national GDP, respectively. H_{it} is calculated by taking the logarithm of the hours worked divided by the total population. rr_{it} is calculated by subtracting inflation from the Central Bank (CB) rate and π_{it} is the change in CPI between time t and $t+1$.

The age structure is represented by the population shares of three different groups. Firstly, the young dependents, aged 0-19. Secondly, the working-age population, aged 20-59. Thirdly, the old dependents, aged 60 and older. The share of age group $k = 1, 2, 3 \quad (0-19, 20-59, 60+)$ in the total population is denoted by w_{kit}. Since $\sum_{k=1}^{3} w_{kit} = 1$, if all age structure shares are included, there is exact collinearity. We can deal with this by restricting the coefficients to sum to zero by using the twp element vector $(w_{kit} - w_{3it})$ as $\mathbf{W}_{it}$. The coefficient of the oldest age group is then recovered from $\beta_3 = -\sum_{k=1}^{2} \beta_k$.

4.2 VAR model

We hope to find the long-run impact of changes in the age structure on our key macroeconomic variables. Since the dynamic relation between those variables can be complicated, as well as identifying and estimating such a (linearized) system, we implement a reduced form under the assumption that it can be written as a vector autoregressive model. We estimate the following augmented panel VARX(1) model, conditional on the exogenous variables represented by $\mathbf{W}_{it}$:

$$\mathbf{Y}_{it} = \mathbf{a}_i + \mathbf{A}\mathbf{Y}_{i,t-1} + \mathbf{D}\mathbf{W}_{it} + \mathbf{u}_{it} \tag{1}$$

Then, knowing $\mathbf{a}_i$, $\mathbf{A}$, and $\mathbf{D}$, the long-run equilibrium for the system is given by:

$$\mathbf{Y}^*_{it} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{a}_i + (\mathbf{I} - \mathbf{A})^{-1}\mathbf{D}\mathbf{W}_{it} \tag{2}$$

in which $\mathbf{D}_{LR} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{D}$ symbolizes the effect of demographic variables. The long-run contribution of demography to each variable in each country can be identified by obtaining the demographic attractor for the economic variables, for every time t,

$$\mathbf{Y}^D_{it} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{D}\mathbf{W}_{it} = \mathbf{D}_{LR}\mathbf{W}_{it} \tag{3}$$

4.2.1 Conditional predictions

Not only are we interested in the demographic effects on the macroeconomic variables, we would also like to know the effects of the changes in age structure per country. To determine these, the demographic predictions for each country ($\mathbf{W}_{i,t+h}$) are used, which are also obtained from the UN World Population Prospects, 2015 Revision. To make a conditional prediction of the effect of expected demographic changes, the long-run demographic impact matrix $\mathbf{D}_{LR}$ is used. This gives:

$$\mathbf{Y}_{i,t+h} = \mathbf{D}_{LR}(\mathbf{W}_{i,t+h} - \mathbf{W}_{i,t}) + \mathbf{Y}_{i,t} \tag{4}$$

This model will be implemented to make predictions for the period of 2015 to 2025. The predictions made for 2015 to 2019 will be used to measure the performance of the different forecasting methods.

4.3 Bayesian VAR model

In this section, we will introduce two different priors for the Bayesian VAR model, which will be used for forecasting economic growth to see whether we can beat the conditional predictions made by model (4). The first one is the diffuse prior, mainly used as a default prior, which will operate as our benchmark. The second one that will be applied is the Minnesota prior, which is a more informative prior. First, the basis of the methodology of the Bayesian approach will be explained, using Belloni (2017), Documentation (2020), Karlsson (2013), Miranda-Agrippino and Ricco (2018), Rao Kadiyala and Karlsson (1993).

A BVAR model treats the parameters of a VAR model as random variables. Writing our VARX(1) model (1) in matrix form gives:

$$\mathbf{Y}_{it} = \mathbf{z}_{it}\mathbf{\Lambda}_i + \mathbf{u}_{it} \tag{5}$$

in which $\mathbf{\Lambda}_i = [\mathbf{a}_i \quad \mathbf{A} \quad \boldsymbol{\delta} \quad \mathbf{D}]'$, $\mathbf{z}_{it} = \begin{bmatrix}1 & \mathbf{Y}'_{i,t-1} & 0 & \mathbf{W}_{it}\end{bmatrix}$, and $\mathbf{u}_{it}$ is a 6×1 vector of random, serially uncorrelated, multivariate normal innovations. Under the typical assumption in macroeconomic literature, $\mathbf{u}_{it}$ has mean $\mathbf{0}$ and 6×6 covariance matrix $\mathbf{\Psi}_i$.

After stacking the row vectors $\mathbf{Y}_{it}$, $\mathbf{z}_{it}$, and $\mathbf{u}_{it}$ in the usual way, for $t = 1, ..., T$, we can write model (5) as follows:

$$\mathbf{Y}_i = \mathbf{Z}_i\mathbf{\Lambda}_i + \mathbf{U}_i \tag{6}$$

Now, assigning j to be the j-th column vector, the above equation can be expressed for each variable as

$$\mathbf{Y}_{ij} = \mathbf{Z}_i\boldsymbol{\lambda}_{ij} + \mathbf{u}_{ij} \tag{7}$$

In Bayesian VAR, all parameters are treated as random variables which have their own probability distribution. The researcher's beliefs are collected in the prior distribution of the parameter, $p(\mathbf{\Lambda}_i)$ which is independent of the observed data. Another important aspect of this method is the likelihood function, with the observed data conditional on the parameters:

$$L_i(\mathbf{Y}_{i,1}, ...\mathbf{Y}_{i,T}, |\mathbf{\Lambda}_i, \mathbf{W}_{it}) = \prod_{t=1}^{T} f_i(\mathbf{Y}_{it}|\mathbf{Y}_{i,t-1}, \mathbf{Y}_{i,t-2}, ...; \mathbf{\Lambda}_i; \mathbf{W}_{it}) \tag{8}$$

By applying the Bayes Theorem, we can now derive the posterior distribution, which contains information about the observed data, and the prior beliefs of the parameters, which looks like:

$$p(\mathbf{\Lambda}_i|\mathbf{Y}_{i,1}, ...\mathbf{Y}_{i,T}; \mathbf{W}_{it}) = \frac{p(\mathbf{\Lambda}_i)L_i(\mathbf{Y}_{i,1}, ...\mathbf{Y}_{i,T}|\mathbf{\Lambda}_i, \mathbf{W}_{it})}{p(\mathbf{Y}_{i,1}, ...\mathbf{Y}_{i,T})} \tag{9}$$

In Bayesian forecasting, the predictive distribution, conditional on the observed data, can be written as $p(\mathbf{Y}_{i,T+1:T+h}|\mathbf{Y}_{i,1}, ..., \mathbf{Y}_{i,T}; \mathbf{W}_{it})$, in which the future datapoints $\mathbf{Y}_{i,T+1:T+h} = (\mathbf{Y}'_{i,T+1}, ..., \mathbf{Y}'_{i,T+h})'$ are conditional on the observed data $\mathbf{Y}_{i,1}, ..., \mathbf{Y}_{i,T}, \mathbf{W}_{it}$. To determine this, the distribution of future observations, conditional on the parameters and observed data is needed:

$$f_i(\mathbf{Y}_{i,T+1:T+h}|\mathbf{Y}_{i,T}, \mathbf{Y}_{i,T-1}, ...; \mathbf{\Lambda}_i; \mathbf{W}_{it}) = \prod_{t=T+1}^{T+h} f_i(\mathbf{Y}_{it}|\mathbf{Y}_{i,t-1}, \mathbf{Y}_{i,t-2}, ...; \mathbf{\Lambda}_i; \mathbf{W}_{it}) \tag{10}$$

After using the Bayes Theorem, equation (8), equation (10), and rewriting using equation (9) the predictive distribution can now be written as:

$$\begin{aligned} &p(\mathbf{Y}_{i,T+1:T+h}|\mathbf{Y}_{i,1}, ..., \mathbf{Y}_{i,T}; \mathbf{W}_{it}) = \\ &\int f_i(\mathbf{Y}_{i,T+1:T+h}|\mathbf{Y}_{i,T}, \mathbf{Y}_{i,T-1},...; \mathbf{\Lambda}_i; \mathbf{W}_{it})p(\mathbf{\Lambda}_i|\mathbf{Y}_{i,1}, ...\mathbf{Y}_{i,T}; \mathbf{W}_{it})d\mathbf{\Lambda} \end{aligned} \tag{11}$$

This distribution captures all useful information concerning the unobserved data. In this paper, the forecasts made with the Bayesian VAR models will be for the period 2015 to 2019 and will be based on our complete dataset (1970-2014).

The next two sections describe the methodology for the diffuse and Minnesota prior, which were obtained using Documentation (2020), Kotzé (n.d.), Miranda-Agrippino and Ricco (2018), Rao Kadiyala and Karlsson (1993), Sun and Ni (2004).

4.3.1 Diffuse prior

One of the most popular non-informative priors is the combination of the 'diffuse' prior and the Jeffreys' prior, which are, respectively, a constant prior for $\boldsymbol{\lambda}_i$ and a prior for $\boldsymbol{\Psi}_i$, which can be combined as a joint prior distribution:

$$p(\boldsymbol{\lambda}_i, \boldsymbol{\Psi}_i) \propto |\boldsymbol{\Psi}_i|^{-(m+1)/2} \tag{12}$$

with m being the number of variables included in $\mathbf{Y}_{it}$. The conditional posterior distributions can then be easily derived from the prior distribution by applying Bayes' Theorem and look like this:

$$\boldsymbol{\lambda}_i|\boldsymbol{\Psi}_i, \mathbf{Y}_{it}, \mathbf{W}_{it} \sim \mathcal{N}_{(mp+r+1_c+1_\delta)\times m}(\overline{\mathbf{M}}_i, \overline{\mathbf{V}}_i, \boldsymbol{\Psi}_i) \tag{13}$$

$$\boldsymbol{\Psi}_i|\mathbf{Y}_{it}, \mathbf{W}_{it} \sim \text{ Inverse Wishart}(\overline{\boldsymbol{\Omega}}_i, \overline{\mathbf{v}}_i) \tag{14}$$

where p equals the number of lags, r the number of exogenous variables, and 1_c and 1_δ are indicators which equal 1 when, respectively, a constant and trend are included in the model and equal 0 if not. Furthermore, $\overline{\mathbf{M}}_i = \left(\sum_{t=1}^{T} \mathbf{z}'_{it}\mathbf{z}_{it}\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}'_{it}\mathbf{Y}'_{it}\right)$ is a matrix of means, $\overline{\mathbf{V}}_i = \left(\sum_{t=1}^{T} \mathbf{z}'_{it}\mathbf{z}_{it}\right)^{-1}$ is the among-coefficient scale matrix, $\overline{\boldsymbol{\Omega}}_i = \sum_{t=1}^{T} \left(\mathbf{Y}_{it} - \overline{\mathbf{M}}'_i\mathbf{z}'_{it}\right)\left(\mathbf{Y}_{it} - \overline{\mathbf{M}}'_i\mathbf{z}'_{it}\right)'$ is the scale matrix, and $\overline{\mathbf{v}}_i = T + k$ is the degrees of freedom.

4.3.2 Minnesota prior

This prior distribution was originally proposed by Litterman (1986). Since all equations are treated separately, writing the prior beliefs of the i-th equation looks like:

$$p(\boldsymbol{\lambda}_{ij}) \sim \mathcal{N}(\tilde{\boldsymbol{\lambda}}_{ij}, \tilde{\boldsymbol{\Sigma}}_{ij}) \tag{15}$$

Then, again applying Bayes' Theorem, the posterior distribution is given by

$$\boldsymbol{\lambda}_{ij}|\mathbf{Y}_i, \mathbf{W}_i \sim \mathcal{N}(\overline{\boldsymbol{\lambda}}_{ij}, \overline{\boldsymbol{\Sigma}}_{ij}) \tag{16}$$

where $\overline{\boldsymbol{\Sigma}}_{ij} = (\tilde{\boldsymbol{\Sigma}}_{ij}^{-1} + \boldsymbol{\psi}_{jj}^{-1}\mathbf{Z}_i'\mathbf{Z}_i)^{-1}$ and $\overline{\boldsymbol{\lambda}}_{ij} = \overline{\boldsymbol{\Sigma}}_{ij}(\tilde{\boldsymbol{\Sigma}}_{ij}^{-1}\tilde{\boldsymbol{\lambda}}_{ij} + \boldsymbol{\psi}_{jj}^{-1}\mathbf{Z}_i'\mathbf{Y}_{ij})^{-1}$. $\boldsymbol{\psi}_{jj}$ are the diagonal elements of $\boldsymbol{\Psi}_i$.

The Minnesota prior can be interpreted as a hyperparameter structure for the joint prior distribution of $(\boldsymbol{\lambda}_i, \boldsymbol{\Psi}_i)$, which is used to obtain a parsimonious model by arranging the coefficient matrix of model (5). This prior was introduced as so-called *shrinkage* prior and thus it considers tuning parameters for the center of shrinkage and tightness of shrinkage. This center of shrinkage can be identified as the prior mean of the coefficients. Furthermore, the Minnesota method sets all non-diagonal elements of coefficient matrix $\mathbf{A}$ equal to zero. The diagonal elements typically take values in the interval $[0, 1]$. The tightness of shrinkage can be identified as the prior variance of the coefficients:

$$\text{Var}(\mathbf{A}_{q,ls}|\boldsymbol{\Psi}) = \begin{cases} \frac{v_0}{q^d}, & l = s \\ \frac{v_\times}{q^d}\frac{\sigma_l^2}{\sigma_s^2}, & l \neq s \end{cases} \tag{17}$$

Here, v_0 is specified as the tightness on the prior means of all self lags of $\mathbf{A}$, d is the speed of tightness decay, $v_\times$ equals the tightness on the prior means of all cross-variable lag coefficients of $\mathbf{A}$, and σ_l^2 is the prior response variance.

4.4 Performance measures

The forecasting performance of the models explained above will be measured, using the forecasted growth rates of real GDP, and compared using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for a total of 5 steps ahead on the period 2015 to 2019. The RMSE and MAE are calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(g_{it} - \hat{g}_{it})^2} \tag{18}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|g_{it} - \hat{g}_{it}| \tag{19}$$

With 21 countries, n equals 21. Real values g_{it} were extracted from the data. The forecasted (or predicted) growth rates are defined as $\hat{g}_{it}$.

5 Results

5.1 Demographic effects

After estimating the VARX(1) model from equation (1) and determining the long-run equilibrium from equation (2), the long-run demographic impact matrix is obtained, shown in Table 1. All those actions were performed using MATLAB. One thing that can be observed is that the youngest age

Table 1: Long-run demographic impact matrix, $\mathbf{D}_{LR}$

	β_1	β_2	β_3
g	0.0666	0.0981	-0.1647
I	0.1363	-0.0534	-0.0829
S	0.1763	0.0971	-0.2734
H	-0.1393	0.2201	-0.0808
rr	0.1851	0.6799	-0.8650
π	0.6614	-0.9134	0.2520

group has a negative impact on hours worked, which seems legit, since children do not contribute in this part of the process.

5.1.1 Predicted growth rates per country

Using the demographic impact matrix $\mathbf{D}_{LR}$ and the data drawn from World Population Prospects, concerning data from 2010 to 2024, predictions for the growth rate of real GDP were made using equation (4). The results per country are shown in Table 2 for 2015 and 2025 in the second and third column, respectively. The first column shows the sample average taken from the period 1970 to 2010. The fourth column shows the change between the predicted growth rates of 2015 and those of 2025. Overall, the results suggest that for all countries in our sample, the changing demographics reduce long-term GDP growth. The intensity of the drop varies between countries. For example, for Austria it is 0.82 percentage point and for Sweden 0.37. Since the results in Table 2 are ‘only’ conditional predictions, which were made using model (4), we are now looking at forecasting with a Bayesian VAR model instead.

Table 2: Average predicted impact on GDP growth per country (in percent)

Country	Sample average (1970-2010)	Projected at 2015	Projected at 2025	Change (2015-2025)
Australia	3.16	3.01	2.42	-0.59
Austria	2.51	2.22	1.40	-0.82
Belgium	2.32	2.09	1.45	-0.64
Canada	2.87	2.45	1.56	-0.89
Denmark	1.94	1.72	1.18	-0.54
Finland	2.77	2.39	1.83	-0.56
France	2.36	1.97	1.32	-0.65
Greece	2.42	2.18	1.35	-0.83
Iceland	3.62	3.24	2.58	-0.66
Ireland	4.67	4.11	3.48	-0.63
Italy	2.06	1.72	1.13	-0.59
Japan	2.79	2.16	1.63	-0.53
Netherlands	2.50	2.19	1.58	-0.61
New Zealand	2.43	2.02	1.39	-0.63
Norway	3.14	2.99	2.52	-0.47
Portugal	2.94	2.52	1.77	-0.75
Spain	2.91	2.31	1.44	-0.87
Sweden	2.12	2.02	1.65	-0.37
Switzerland	1.66	1.43	0.92	-0.51
United Kingdom	2.28	2.09	1.60	-0.49
United States	2.92	2.58	1.91	-0.67

5.2 Forecasting with Bayesian VAR

When computing the forecasts of the growth rates with the Bayesian VAR model, the function `simsmooth` was implemented into MATLAB. Since this is a function that returns 1000 random draws each time it is run, we ran this function several times for each country and used the one that seemed most appropriate. To create a Bayesian VAR model with diffuse priors, the Matlab function `diffusebvarm` was used. This function specifies the joint prior distribution as given in equation (12). In order to implement the Minnesota prior, the Matlab function `bayesvarm` was used, since this function automatically employs these prior settings according to the original hyperparameter structure of Litterman (1986). Further discussion of the Bayesian VAR forecasts can be found in the next section.

5.2.1 Graphical representation

For comparison, the graphs below in Figure 1 show the observed growth rates, the growth rates predicted using the VAR model, and the growth rates that were forecasted using the Bayesian VAR models, for each country. From the graphs, one can easily observe that the differences between the forecasts made with the diffuse and Minnesota prior do not seem to be extremely big. Also, sometimes the conditional predictions made with model (4) seem a lot more accurate, for example for Denmark, than the Bayesian VAR forecasts, and sometimes it is the other way around, for example for Norway. We will get a better idea of all of their forecasting performances in Section 5.3.

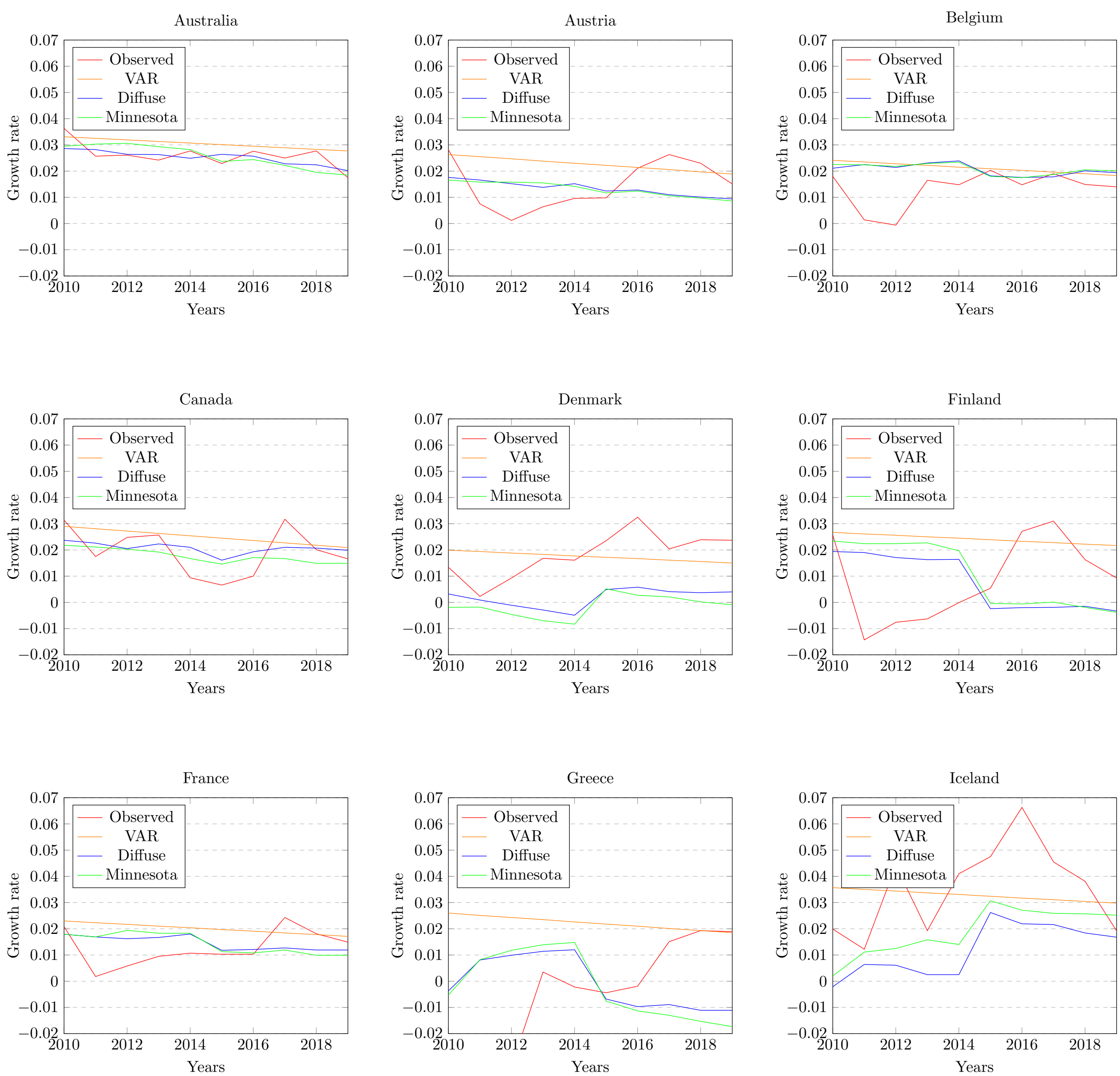

Figure 1: Observed and forecasted growth rates using the VAR model and BVAR models with diffuse and Minnesota prior for each country

5.3 Forecasting performance

The results of the forecasting performance measures are shown in Table 3. The values of RMSE and MAE tell us how well each model fits the data for all steps ahead. A lower value of each forecasting performance indicates a better fit. The first thing that comes to notice, is that the values of RMSE

Table 3: RMSE and MAE for VAR model and BVAR models with diffuse and Minnesota prior

	RMSE					**MAE**				
	2015	2016	2017	2018	2019	2015	2016	2017	2018	2019
VAR	0.0133	0.0130	0.0119	0.0124	0.0083	0.0112	0.0094	0.0078	0.0081	0.0065
Diffuse	0.0103	0.0167	0.0233	0.0224	0.0165	0.0083	0.0118	0.0167	0.0153	0.0109
Minnesota	0.0093	0.0159	0.0230	0.0219	0.0173	0.0076	0.0117	0.0161	0.0151	0.0124

and MAE get bigger for further steps ahead for both Bayesian VAR models, while those values for the VAR model become smaller. This can be caused by the fact that the conditional prediction was made using the demographic impact matrix, which was obtained from the long-run equilibrium in equation (3). It can also be observed that the prediction made with the VAR model performed better than both Bayesian VAR forecasts, except for the first step ahead. This can be due to the fact that with the VAR model a conditional prediction was made, instead of a forecast. Even though Bayesian VAR models in general have better forecasting performance than VAR models, it seems in this case they could not compete with a conditional prediction. Also, from earlier research, it was expected that the Minnesota prior would perform better than the diffuse prior. Even though the differences are not that high, the Minnesota prior does perform better, except for five steps ahead.

6 Discussion

Since the first part of this paper was replicated from Aksoy et al. (2019), we will compare the results and try to explain the differences. First of all, due to the reunification, they only included Germany in their predictions of the growth rates. Their predictions were made using only the data of 2010 until 2014, which made it possible to include Germany here. We decided not to do that, since our forecasts made with Bayesian VAR were based on our complete dataset, which we did not have for Germany. Also, we did not perform the Wald test on the elements of the demographic impact matrix, since we were not able to get reliable results from this.

Regarding Table 1, the results are not identical, but the general interpretation seems to be the same. The signs (negative or positive) are the same, except for a few, and the difference in magnitude between age groups also seems consistent with their results. The differences in values can be due to the fact that, even though the data were obtained from the same databases, there were differences between our database and theirs. The same goes for the results displayed in Table 2, in which the values do not exactly match. However, for most countries, the differences in magnitude between the countries seems to add up and the change in predicted growth rate between 2015 and 2025 also seems quite compatible with the results of Aksoy et al. (2019).

7 Conclusion

In this paper we studied the impact from age structure on several macroeconomic variables and whether forecasting economic growth with Bayesian VAR models gives more accurate results than a conditional prediction using a VAR model. All of this was summarised in the following research question: *How does age structure affect macroeconomic trends and can we obtain a better forecast for economic growth using Bayesian VAR compared to the VAR model?*

In order to estimate the effect of age structure on those variables, we created three different age groups (young dependents, working people, old dependents). After obtaining the demographic impact matrix, which was computed using a VARX(1) model, we found to what extent each age group affects our six macroeconomic variables. This differs per variable and per age group. Next, a conditional prediction of economic growth was made, using this demographic impact matrix. In order to see whether forecasting using a BVAR model would obtain better results, we implemented this model with two different priors: the diffuse prior and the Minnesota prior. Their forecasting performances were compared using the RMSE and MAE for a total of five steps ahead. The results showed that the conditional prediction using the VAR model made more accurate forecasts than both BVAR models, except for one step ahead. This can be explained by the fact that this conditional prediction was made using the demographic impact matrix, which was obtained from the long-run equilibrium. Also, the Minnesota prior performed better than the diffuse prior, except for five steps ahead.

A practical implication of this research is that one should note the difference between a long-run steady state and a long-run estimate. The estimates obtained in this research issue a long-run forecast for economic growth, conditional on the age group shares. After that, the effect of demographics on economic growth is measured on the long run. However, it could be that the age structure progresses towards a steady-state demographic distribution. Since this process is not modelled in this research, we cannot provide a straightforward estimate of the effect of possible demographic convergence to a steady-state on economic growth. Nevertheless, the demographic changes predicted by the UN that are used in that part of this paper may have already accounted for this possible convergence. Finally, there exist a lot of different priors for Bayesian VAR aside from the two priors that were implemented in this paper. For further research, one could consider implementing different priors to find whether those priors are able to outperform the conditional prediction made with the VAR model.

References

Aksoy, Y., Basso, H. S., Smith, R. P., & Grasl, T. (2019). Demographic structure and macroeconomic trends. *American Economic Journal: Macroeconomics*, *11*(1), 193–222.

Andersson, B. (2001). Scandinavian evidence on growth and age structure. *Regional Studies*, *35*(5), 377–390.

Bańbura, M., Giannone, D., & Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of forecasting*, *31*(3), 739–756.

Belloni, A. (2017). Comparing the forecasting performance of var, bvar and u-midas.

Bloom, D. E., Canning, D., Fink, G., & Finlay, J. E. (2007). Does age structure forecast economic growth? *International Journal of Forecasting*, *23*(4), 569–585.

Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, *3*(1), 1–100.

Documentation, M. (2020). Matlab documentation. *Matlab*, R2020a.

Geisser, S. (1965). Bayesian estimation in multivariate analysis. *The Annals of Mathematical Statistics*, *36*(1), 150–159.

Gordon, R. J. (2012). *Is us economic growth over? faltering innovation confronts the six headwinds* (tech. rep.). National Bureau of Economic Research.

Jeffreys, H. (1961). Theory of probability (3rd ed.). Oxford: Clarendon Press.

Karlsson, S. (2013). Forecasting with bayesian vector autoregression, In *Handbook of economic forecasting*. Elsevier.

Kotzé, K. (n.d.). Bayesian vector autoregression models.

Kuznets, S. S. (1930). *Secular movement in production and prices: Their nature and their bearing upon cyclical fluctuations*. Houghton Mifflin; company, Boston.

Lindh, T., & Malmberg, B. (2009). European union economic growth and the age structure of the population. *Economic Change and Restructuring*, *42*(3), 159–187.

Litterman, R. B. (1980). *A bayesian procedure for forecasting with vector autoregressions*. MIT.

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

Litterman, R. B. Et al. (1979). *Techniques of forecasting using vector autoregressions* (tech. rep.).

McNees, S. K. (1986). Forecasting accuracy of alternative techniques: A comparison of us macroeconomic forecasts. *Journal of Business & Economic Statistics*, *4*(1), 5–15.

Miranda-Agrippino, S., & Ricco, G. (2018). Bayesian vector autoregressions.

Nations, U. (2015). World population prospects: The 2015 revision. *United Nations Econ Soc Aff*, *33*(2), 1–66.

Nicholson, W. B., Wilms, I., Bien, J., & Matteson, D. S. (2014). High dimensional forecasting via interpretable vector autoregression. *arXiv preprint arXiv:1412.5250*.

Rao Kadiyala, K., & Karlsson, S. (1993). Forecasting with generalized bayesian vector auto regressions. *Journal of Forecasting*, *12*(3-4), 365–378.

Sevinç, V., & Ergün, V. (2009). Usage of different prior distributions in bayesian vector autoregressive models. *Hacettepe Journal of Mathematics and Statistics*, *38*(1), 85–93.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Sun, D., & Ni, S. (2004). Bayesian analysis of vector-autoregressive models with noninformative priors. *Journal of Statistical Planning and Inference*, *121*(2), 291–309.

Tiao, G. C., & Zellner, A. (1964). On the bayesian estimation of multivariate regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 277–285.

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