Can Forecasting Performance of the Bayesian Factor-Augmented VAR be Improved by Considering the Steady-State?

- An application to Swedish inflation

Andreas Lindvall♣

First year Master's thesis
Lund University

May 24, 2017

Abstract

This paper investigates whether the forecasting performance of Bayesian factor-augmented VAR (BFAVAR) models can be improved by incorporating an informative prior on the steady-state of the time series in the system. The BFAVAR model is compared to the extended steady-state BFAVAR in an application to forecasting Swedish inflation, making use of data from 1996 to 2016. Results show that the out-of-sample forecasting performance of incorporating an informative prior into the BFAVAR models increase compared to an autoregressive model. When comparing BFAVAR models with and without an informative prior on the steady-state, the BFAVAR model with an informative prior marginally outperform the BFAVAR model without the informative prior. The results of this paper indicate that most of the gains in forecasting performance by incorporating an informative prior on the steady-state are associated with longer forecasting horizons.

Keywords: Bayesian factor-augmented VAR, Steady-state, Inflation, Out-of-sample forecasting precision

Supervisor: Joakim Westerlund, Lund University

♣ I would like to extend my gratitude towards my supervisor Joakim Westerlund for his guidance and patience during the development of this thesis.

Contents

1. Introduction

Forecasting future economic development is a major factor for policy-makers, consumers and investors. Many long-term private sector commitments are based on forecasts of the development of the general price level, for example, labor contracts and debts, such as mortgages. Central banks often publicly stress that, since monetary policy affects the economy with a lag due to the transmission mechanism, central banks has to be forward-looking in order to conduct good monetary policy. Being forward-looking relies on forecasts and projections about the future development of key variables, such as GDP growth and inflation. Svensson (2005) argue that the quality of the forecasts will affect the effectiveness of the central bank's monetary policy because the New Keynesian model makes explicit that optimal monetary policy depends on optimal forecasts.

Since the seminal work of Sims (1980), vector autoregressive (VAR) models have been the workhorse in forecasting and modeling macroeconomics. However, with VAR models the number of parameters increases rapidly with the inclusion of variables, which can lead to degree-of-freedom problems. On the other hand, including too few variables may not capture enough information, leading to omitted variation. This paper is particularly interested in two common ways of overcoming the issue of over-parametrization and omitted variation: Bayesian methods and dimensional reduction techniques.

As pointed out by Clements & Hendry (1998), a common source of poor forecasting accuracy on long horizons is a poorly estimated mean of the process. Since long horizon forecasts from stationary VAR models converge to the steady-state (unconditional mean) of the process, an informative prior on the steady-state might increase forecasting accuracy. In inflation forecasting, an informative prior on the steady-state is particularly interesting and readily available since inflation-targeting central banks actively work to reach an explicitly stated inflation target. Bayesian methods are commonly used in this setting due to its ability to incorporate prior information on the behavior of the time series in the model to shrink unnecessary parameters towards zero, thus making it able to handle more information and increase precision.

In a standard VAR setting, it is usually difficult to incorporate prior information on the steady-state. Therefore Villani (2005, 2009) suggest a slightly unconventional specification of the VAR model, resulting in a steady-state Bayesian VAR (BVAR) model, which allow for incorporating prior beliefs on the system in a more convenient way. Villani (2009) argues that a possible explanation for why the steady-state is handled rather casually is that it is expected to be fairly precise even without an informative prior on the steady-state. By comparing the forecasts of various macroeconomic variables from BVAR models with different priors and standard VARs estimated by maximum likelihood, Villani (2009) show that this is not always the case.

Österholm (2008) make use of Villani's BVAR model to investigate if forecasts of Swedish inflation and interest rates from Bayesian AR and VAR models can be improved by incorporating prior information on the steady state of the process. Österholm (2008) finds that there seem to be payoffs associated with using the steady-state prior with regards to forecasting accuracy when applied to interest rates; however, when applied to inflation the gains are modest. Villani's BVAR model has also been used by Adolfsson et al. (2007) and Mossfeldt & Stockhammar (2016), who find the BVAR to perform as good as, and in some cases, better than the models of the Riksbank and NIER[1] when forecasting different measures of inflation.

In the literature on forecasting in data-rich environments, Stock & Watson (1999, 2002) introduce the use of dimensional reduction techniques in order to extract more information from large data sets using primarily principal component analysis. Stock & Watson (2002) find that using a small number of factors from a large data set produce more accurate forecasts than a variety of benchmarks without factors. The results of Stock & Watson (2002) is confirmed and extended by Bernanke & Boivin (2003) who use the same methodology on different data. Bernanke & Boivin (2003) also find that the inclusion of factors eliminates omitted variation in VAR models, e.g. Sims' (1992) "price puzzle."

Bernanke et al. (2005) use principal components to extract information from a data set consisting of 120 macroeconomic time series and use these artificial variables in a VAR model in order to incorporate more information in a more parsimonious way. They argue that

[1] National Institute of Economic Research

their factor-augmented VAR (FAVAR) could reduce the over-parametrization problem with VAR models at the same time as overcoming the issue with omitted variation common in small VAR models. Laganá & Mountford (2005) use a similar approach in an application on determinants of the UK interest rate. As Bernanke et al. (2005), they find that the inclusion of factors eliminates omitted variation associated with small VAR models, but also that the use of their FAVAR model in forecasting is superior to a number of benchmark models without factors.

This paper extends the framework of Bernanke et al. (2005) by incorporating an informative prior on the steady-state of the system in a Bayesian setting. By making use of information from a large data set of disaggregated information related to Swedish inflation, this paper investigates if an informative prior on the steady-state of the process can improve the forecasting performance of the Bayesian factor-augmented VAR (BFAVAR). In a forecasting exercise similar to Mossfeldt & Stockhammar (2016), out of sample forecasts of Swedish inflation from steady-state BFAVAR models, BFAVAR models and the AR(1) are compared.

The forecasting exercise shows that an informative prior on the steady-state increase forecasting accuracy in an inflation forecasting setting. When adding an informative prior, the steady-state BFAVAR see an increase in forecasting precision compared to an AR(1) model and a standard BFAVAR model. The largest gain in forecasting performance by using the BFAVAR model with an informative prior on the steady-state compared to the AR(1) is 9 per cent at the longest horizon. The largest gain in forecasting precision by using the BFAVAR model *with* an informative prior on the steady-state compared to the BFAVAR model *without* is about 5 per cent on the eight quarter horizon. With respect to the magnitude of these results, the 9 per cent improvement has previously been considered as a non-negligible effect, while the 5 per cent improvement has previously been considered as a negligible-to-modest effect; see for example Stockhammar & Österholm (2016) and Österholm (2010).

The literature has in recent years been interested in Bayesian methods related to VAR models to a greater extent than FAVAR models. The contribution of this paper is thus twofold: Firstly, the framework of Bernanke et al. (2005) is extended by considering if incorporating an informative prior on the steady-state of the process can improve forecasting accuracy from BFAVAR models. Secondly, the broader contribution to the literature is in the estimation and forecasting evaluation of BFAVAR models applied to inflation in Sweden.

2. The Steady-State Bayesian FAVAR model

A BFAVAR can intuitively be thought of as an extension of the BVAR which makes use of a large amount of information in a more parsimonious way. This paper follows the two-step principal component approach used by Bernanke et al. (2005) and Stock & Watson (1998, 2002). The two-step approach begins by using static principal components to estimate the factors that summarize the most relevant information in the large data set.[2] By using principal components, a number of factors are obtained that are considerably smaller than the original 114 variables. As a result, the amount of information that can be handled by the model increases and hence, the chance of under-specifying the model decreases.

Following Bernanke et al (2005), let $\boldsymbol{X_t}$ be a $(n \times 1)$ vector of informational time series relating to the state of the Swedish economy; $\boldsymbol{Y_t}$ a $(m \times 1)$ vector of endogenous observed economic variables that is related to inflation and is a subset of $\boldsymbol{X_t}$; $\boldsymbol{F_t}$ a $(k \times 1)$ vector of unobservable factors that summarize most of the information in $\boldsymbol{X_t}$, where k is considerably smaller than n. The unobservable factors, $\boldsymbol{F_t}$, are thought of as diffuse concepts such as "economic activity" or "credit conditions." Since the standard Bayesian FAVAR model has no informative prior on the steady-state of the process, the estimated factors are incorporated into the framework of Villani (2009). This approach implies that the Bayesian FAVAR *without* an informative prior on the steady-state is a nested model of the Bayesian FAVAR *with* an informative prior on the steady-state.[3] The data-generating process can then be seen as the system of equations in (1). $\boldsymbol{X_t}$ is related to the unobservable factors $\boldsymbol{F_t}$ by the equation in (2).

$$\begin{bmatrix} \boldsymbol{F_t} \\ \boldsymbol{Y_t} \end{bmatrix} = \mathbf{B}(L) \begin{bmatrix} \boldsymbol{F_{t-1}} - \boldsymbol{\mu_1} \\ \boldsymbol{Y_{t-1}} - \boldsymbol{\mu_2} \end{bmatrix} + \boldsymbol{v_t} \tag{1}$$

$$\boldsymbol{X_t} = \boldsymbol{\Lambda}^f \boldsymbol{F_t} + \boldsymbol{\Lambda}^y \boldsymbol{Y_t} + \boldsymbol{e_t}. \tag{2}$$

[2] For a brief introduction to principal component analysis, see Appendix A.3. For a more detailed text, see Johnson & Wichern (2009).

[3] Note, the steady-state BVAR of Villani (2009) is also a nested model of the framework in (1) since, without the factors, the system reduces to the framework of Villani (2009). This model is however not of interest in this particular thesis.

$\mathbf{B}(L)$ are polynomials of the lag operator; $\boldsymbol{v_t}$ is a vector of error terms; $\boldsymbol{\Lambda}^f$ is a $(n \times k)$ matrix of pattern loadings; $\boldsymbol{\Lambda}^y$ is a $(n \times m)$ matrix of loadings belonging to the endogenous variables and $\boldsymbol{v_t}$ is a vector of *iid* error terms assuming $\mathrm{E}(\boldsymbol{v_t}) = \mathbf{0}$ and $\mathrm{E}(\boldsymbol{v_t}\boldsymbol{v_t}') = \boldsymbol{\Sigma}$. $\boldsymbol{e_t}$ is a $(n \times 1)$ vector of error terms, allowing for weakly correlated errors.

Equation (2) represents the common forces that drive the dynamics of $\boldsymbol{X_t}$, which enables us to estimate the latent factors using principal component analysis, such that $\boldsymbol{F_t}$ are weighted combinations of all variables in $\boldsymbol{X_t}$. More precisely, by estimating the factors, we estimate the space spanned by the factors. The factors used in this model are the scores estimated from the principal components.

The main property of equation (1), $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, are both $(n \times 1)$ vectors describing the steady-state values of the variables in the system. That is, for each variable in the system, a prior interval is specified around the mean of each variable. The argument is that since long horizon forecasts from stationary VAR models converge to the steady-state of the process, an informative prior on the steady-state might increase forecasting accuracy. As the Riksbank has an explicitly stated inflation-target, it is particularly interesting to incorporate this information into the model. The specification in (1) implies that we can model the unconditional mean of the process explicitly even though the specification is nonlinear in its parameters. The restriction is thus in $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, meaning that when the BFAVAR is modeled without an informative prior on the steady-state, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$, the model reduces to the one used by Bernanke et al. (2005), which would in that case model the mean without any informative priors. The approach of modeling the unconditional mean without an informative prior on the steady-state is expected to be relatively precise; however, Villani (2009) show that this is not always the case and that an informative prior can in some cases increase precision.

As in Villani (2009), Österholm (2008) and Mossfeldt & Stockhammar (2016), the prior on the covariance matrix $\boldsymbol{\Sigma}$ follow the literature and is given by the non-informative standard Diffuse prior[4]

[4] The standard Diffuse prior is non-informative. As opposed to using the Minnesota prior on the covariance matrix, which replaces the covariance by an approximation (not necessarily a good one), the Diffuse prior is handled like a random variable.

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}]^{-(n+1)/2}.$$

Let $\boldsymbol{B} = (\boldsymbol{B_1}, \dots, \boldsymbol{B_p})'$. The prior on $vec(\mathbf{B})$ is a general multivariate normal distribution[5] given by

$$vec(\boldsymbol{B}) \sim N_{mn^2}(\boldsymbol{\theta_B}, \boldsymbol{\Omega_B}).$$

The prior on $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ is given by $\boldsymbol{\mu_1} \sim N(\boldsymbol{\theta}_{\mu_1}, \boldsymbol{\Omega}_{\mu_1})$ and $\boldsymbol{\mu_2} \sim N(\boldsymbol{\theta}_{\mu_2}, \boldsymbol{\Omega}_{\mu_2})$, which means that these priors refer to a 95% confidence interval around the mean of the series. These priors are specified in detail in Table 1.

Table 1: Steady State Prior Intervals for the Mean of the Variables in the Steady-State BFAVAR models

Variables	**Prior Interval**
CPI[6]	(0.19; 0.4)
GDP	(0.5; 0.75)
UNEMP	(5.0; 9.0)
FACTOR 1	(-0.21; 0.21)
FACTOR 2	(-0.21; 0.21)
FACTOR 3	(-0.21; 0.21)
FACTOR 4	(-0.21; 0.21)
FACTOR 5	(-0.21; 0.21)

Note: CPI inflation and GDP are measured in quarter-on-quarter percentage change. The unemployment rate is measured in per cent. The estimated factors are principal components. Prior intervals refer to a 95%-confidence interval around the mean of the series.

The prior intervals on the mean of inflation (CPI), gross domestic product (GDP) and unemployment (UNEMP) follow the literature; see for example Österholm (2010) and Stockhammar (2012). However, the prior interval on the mean of inflation is converted from a prior on the yearly change in inflation to a quarterly change. We do not have any information on the prior interval for the mean of the estimated factors a priori; however, they are estimated in the first step and we know that they are normalized and stationary. A prior interval on the mean of the factors is therefore constructed from the normal distribution and thus set to (-0.21; 0.21).

[5] The *vec* operator converts a $(m \times n)$ matrix A to a $(mn \times 1)$ column vector by stacking the columns of the matrix A on top of one another.

[6] The usual prior interval for the yearly inflation rate is (1; 3). Since the quarter-on-quarter percentage change is modeled and forecasted, this is transformed to be a quarterly prior. For the upper interval limit, 3, : $(1+3)^{\left(\frac{1}{4}\right)} - 1$, and for the lower interval limit, 1, : $(1+2)^{\left(\frac{1}{4}\right)} - 1$.

The priors on the dynamics, $\mathbf{B}$, are a slightly modified Minnesota prior and follow the literature. As opposed to using the traditional specification of a prior on the first lag equal to 1 and zero on all other lags, the prior mean on the first lag is set to 0.9 and zero on all other lags for the inflation and unemployment rate. This reflects the belief of some persistence in the inflation and the unemployment rate. Both the quarter-on-quarter percentage change in inflation and the unemployment rate (measured in per cent) were considered as stationary by the augmented Dickey-Fuller test over the period 1996Q3 – 2016Q4. The GDP is transformed to induce weak stationarity; therefore the prior mean on the first lag is set to 0.[7] Setting the prior mean on the first lag to 1 and zero on the subsequent lags takes its starting point in a univariate random walk and is thus not consistent with the steady-state model since it would not have a well-defined unconditional mean (non-stationary). Finally the hyperparameters, which describe how tight the priors on the dynamic coefficients in $\mathbf{B}$ are, follow the literature; the overall tightness is set to 0.2, the cross-variable tightness to 0.5 and the lag decay parameter to 1, see for example Österholm (2008), Villani (2009) and Stockhammar & Österholm (2016)

Since this paper uses the reduced form BFAVAR, there is no reason to subset $\boldsymbol{X_t}$ into fast-and-slow moving factors as Bernanke et al (2005) do, which is typically done when the objective is to estimate a structural model. Since the purpose of this paper is to forecast, we want to keep as much information in the factors as possible. Note that if the system in (2) is taken as true, estimating a VAR in $\boldsymbol{Y_t}$ would be miss-specified since the factors, $\boldsymbol{F_t}$, are omitted variables. This can lead to biased parameters and worse forecasting accuracy.

To create a base model similar to Bernanke et al. (2005) and Laganá & Mountford (2005), we let three variables: CPI, GDP, UNEMP, and the first factor enter the model from the start. The starting point is thus to estimate the first nested model in order to find the number of factors and lag lengths that give the smallest RMSFE. This is done through a step-wise selection process where the factors are then added and different lag lengths tested to find the model with the smallest RMSFE. In order to determine if the informative prior on the steady-state can improve forecasting accuracy, the lag length and number of factors have to be fixed for the models that are compared.

[7] This follows the recommendation by Carriero et al. (2013)

3. Data

The data set used to estimate the factors cover the time range from 1996Q3 to 2016Q4 and consist of 82 observations of 114 different time series. The time series that are originally on monthly frequency is transformed to quarterly frequency by averaging over the months making a quarter, covering the full time series. The 114 time series are on 11 different categories chosen to resemble those in Stock & Watson (2002), Bernanke et al. (2005) and Laganá & Mountford (2005). The variables in the data set are measures on variables related to Swedish inflation, covering the following 11 categories: real output and income, employment and hours, consumption, expectations, housing quantity and prices, stock and commodity prices, exchange rates, interest rates, money and credits, price indices and foreign variables. The category of foreign variables is included, similar to Gustafsson (2015), since Sweden is a comparatively small economy and reasonably affected by the state of larger economies. Some examples of such variables are US inflation, US unemployment, EU inflation, and measures of economic policy uncertainty in the EU and the US.

All time series are adjusted for holidays and seasonality, either directly from the database or by using the ARIMA-X13-Seats of the Census Bureau if any seasonality is detected.

After any seasonality is removed, the series are tested for any unit roots using the augmented Dickey-Fuller test (ADF). Since the framework outlined in section 2 assumes the time series in $\boldsymbol{X_t}$ to be stationary, series that were found to have any unit roots are suitably transformed to induce weak stationarity. This, together with normalizing the series to be mean zero with unit variance, is necessary in order extract principal components from the data. All transformations and a brief description of each time series are given in Appendix B, Table 2.

4. Empirical results

4.1 Model selection

The Diebold-Mariano (1995) test is often used to test whether differences in forecasting performance between two models is statistically significant. However, as pointed out by Clark & McCracken (2005), no formal tests currently exist for settings with recursively generated forecasts from nested Bayesian models and with a forecast horizon greater than one. Also, Armstrong (2007) argues that forecast significance tests are of little value in addition to the RMSFE criterion and focus should, therefore, be on practical relevance i.e. effect size, which in this case would be how much forecasting accuracy is improved by using the competing models. Therefore, the focus is on the RMSFE statistic as the criterion for choosing models, which is in line with the philosophy put forward by Armstrong (2007). The argument is that when choosing between competing models considered equally good a priori in a pure forecasting setting, the forecaster will choose the model with the smallest RMSFE, regardless if the difference is statistically significant.

The first step in the model selection is derived from Bernanke et al. (2005) and Laganá & Mountford (2005). In choosing the Y_t-vector, the variables that enter the model from the start, one would like an, economically speaking, sensible model. The Y_t-vector is therefore chosen to be three variables: inflation (CPI), gross domestic product (GDP), unemployment (UNEMP) and the first factor. Since output (measured as GDP) and unemployment are common in macroeconomic models, this should form a reasonable base to evaluate the inclusion of the steady-state prior from. In order to specify the number of factors and number of lags, this paper conducts a step-wise out-of-sample forecasting exercise, similar to Mossfeldt & Stockhammar (2016). The methodology is thus similar to other studies using forecast precision to assess Granger causality of various variables for inflation (see for example Bachmeier et al. 2007, Gavin and Kliesen 2008, Berger and Österholm 2009, 2011 and Scheufele 2011). The forecasting exercise is of a recursive nature in the sense that the model is estimated for a training period of 1996Q3 – 2006Q3 and a 12 quarter forecast is made. The sample is then extended by one period, and the model re-estimated on the estimation window 1996Q3 – 2006Q4 and a new 12 quarter forecast is made. This process continues until the end of the sample, where the model is re-estimated for the period 1996Q3 – 2016Q3 with the last forecast made for the last quarter, 2016Q4. The step-wise exercise is conducted as follows:

1. The starting model with CPI, GDP, UNEMP and the first factor are estimated for the training period 1996Q3 – 2006Q3 and a recursive forecast is made according to the process above to evaluate the lag length that gives the smallest RMSFE. The forecasts are evaluated at the 1, 2 and 3-year horizon. The model with the lowest RMSFE advances to step 2.

2. The second of the estimated factors is added to the best performing model from the previous step. This process continues until there are no further gains in forecasting accuracy by adding another factor.

3. Finally, information on the steady-state is incorporated.

The model from step 2 and step 3 then advances to the forecasting comparison, i.e. the best BFAVAR model *without* an informative prior on the steady-state and the same BFAVAR model *with* an informative prior on the steady-state included. It is also tested if forecasting accuracy could be improved upon by adding an informative prior on the steady-state in step 2 before the estimated factors are added in the same step-wise manner.

Firstly, this process starts out by estimating and forecasting from the nested model, the BFAVAR without an informative prior on the steady-state. Secondly, the unrestricted model is estimated and forecasted i.e., the BFAVAR with an informative prior on the steady-state. The BFAVAR models with the same variables, the same number of factors and the same number of lags is then compared with the only difference being that one has an informative prior on the steady-state.

Since the first estimated factor accounts for about 96 % of the variation in $\boldsymbol{X_t}$, and the first five accounts for about 99 % of the variation in $\boldsymbol{X_t}$, including more factors than five might be unnecessary. In general, the forecasting exercise shows that including more than one factor actually decrease forecasting accuracy. Since the issue of number of factors to include is rather data-driven, there are numerous models in the literature using different number of factors. However, in an application to Sweden, Gustafsson (2015) also find that using one lag is the best.[8] Also, including more than six lags in general decrease the overall forecasting precision. The best performing models only include one lag. In general, adding more than one lag but less than six, only marginally increases the forecasting accuracy on the longest horizons at the cost of sharper decreasing accuracy on shorter horizons.

The notation of the models is as follows: $BFAVAR(n, f, p, \gamma)$, where n indicates the number of variables that enter the model from the start (which is always 3, following the base model); f indicates the number of factors; p indicates the number of lags and γ indicates if an

[8] Note, Gustafsson (2015) forecasts monthly inflation.

informative prior on steady-state is included. Note, in this notation γ only takes the value 1 to indicate if an informative prior on the steady-state is included in the BFAVAR model i.e., if μ_1 and μ_2 are included in equation (1). If $\gamma = 1$, μ_1 and μ_2 are included in equation (1) and the prior intervals refer to a 95% confidence interval around the mean of the series and are specified in Section 2, Table 1.

4.2 Forecast comparisons

By making use of the BFAVAR model *with* an informative prior on the steady-state and the BFAVAR model *without* an informative prior on the steady-state, the out-of-sample forecasting precision of the models applied to inflation is analyzed over the time span 1996Q3 – 2016Q4. The forecasting procedure is a recursive forecasting exercise in the sense that the estimation window is expanded for each forecast. More specifically, the model is estimated for a training period 1996Q3-2006Q3 and a forecast of twelve quarters is made for the period 2006Q4 – 2009Q4. In the next sequence, the estimation window is expanded by one period by including one more observation, the model is estimated for the period 1996Q3-2006Q4 and a new twelve-quarter forecast is made for the period 2007Q1 – 2010Q1. This process continues until the end of the sample, 2016Q3, with the last forecast being for 2016Q4. In the last step, the model is thus estimated for the period 1996Q3 – 2016Q3.

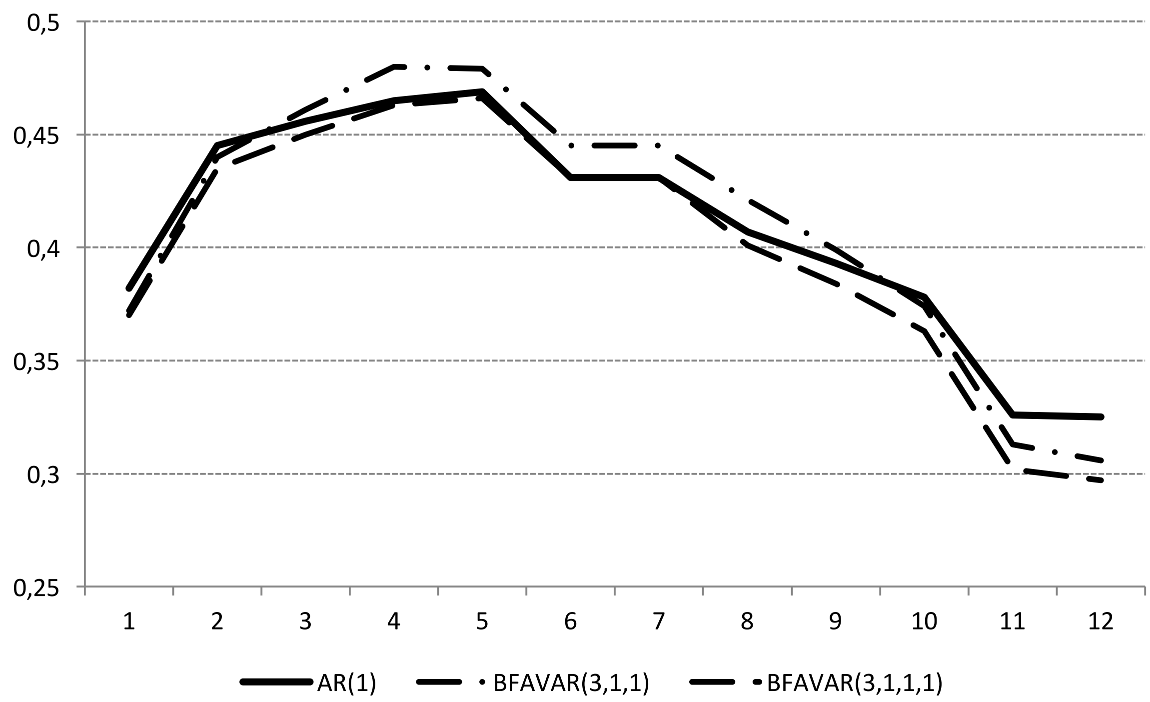
The numerical evaluations of the posterior distribution of the Bayesian models are conducted by a Gibbs sampler.[9] The Gibbs sampler makes use of a burn-in sample of 1000 draws which are discarded and the analysis is performed on the subsequent 20 000 draws. Discarding the first 1000 draws ensures that the draws being analyzed are stationary. Each of these simulations is used to forecast possible paths of the inflation, which are used to calculate the predictive density. All of these dynamic simulations are then used to approximate the point forecast by using the median of these simulations. For each model, the RMSFE is recorded and evaluated for quarter 1-12. I report the RMSFE statistics for each model as well as the relative RMSFE, defined as the RMSFE of the alternative model over the RMSFE of the benchmark, over the horizon 1-12 quarters. The AR(1) model is used as a benchmark and is the usual autoregressive model, however, including an intercept. The reduction in RMSFE is expressed as the reduction in percentage points compared to two different benchmarks; in the first comparison, the AR(1) is used as a benchmark and in the second comparison, the

[9] See Gelman et al. (2009) for further explanation.

BFAVAR *without* information on the steady-state. These results are reported in the following figures.

Figure 1 presents the RMSFEs for the steady-state BFAVAR model, BFAVAR(3,1,1,1), and the BFAVAR model, BFAVAR(3,1,1), together with the AR(1) model. We can see that the BFAVAR(3,1,1) model perform slightly better than the AR(1) on the first two-quarters and worse on quarters 3-9. The BFAVAR(3,1,1) seem to perform better on the longest forecasting horizons compared to the AR(1). If we consider the BFAVAR(3,1,1,1), with a steady-state prior included, we see that it performs slightly better than the AR(1) on the first five quarters and similar on horizons 5-7. For the horizon 7-12, the BFAVAR(3,1,1,1) outperforms the AR(1) model, with its largest precision gain being for the longest horizon, 12 quarters. As expected, information on the steady-state seems to increase precision on longer horizons, where the BFAVAR model with a steady-state prior outperform both the AR(1) model and the BFAVAR(3,1,1) model. By adding an informative prior to the BFAVAR(3,1,1) there seem to be precision gains over most horizons, but most prominent on the longest horizons, even though the effect does not seem to be especially large.

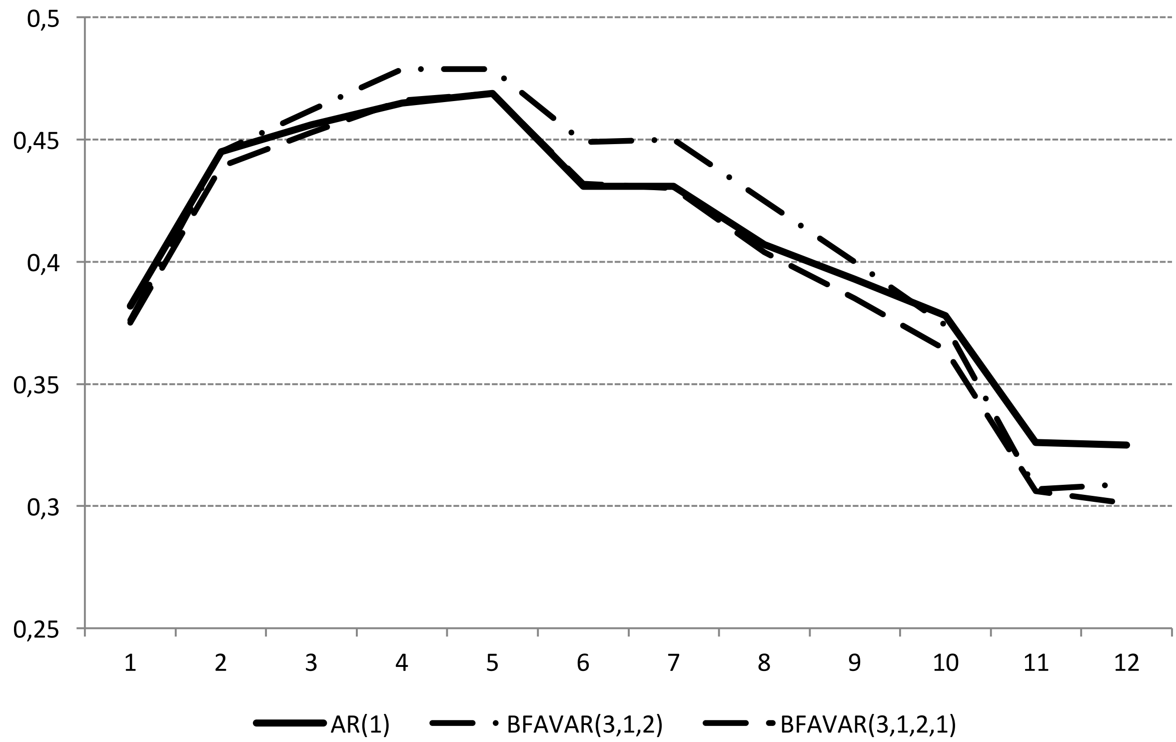
Figure 1: RSMFE for the best BFAVAR models and the AR(1) model, 2006Q3 – 2016Q3



Note: The RMSFEs are given in percentage points on the vertical axis over the horizon 1-12 quarters on the horizontal axis. See section 4.1 for a description of the notation.

Figure 2 presents the RMSFEs for the second best models: the steady-state BFAVAR model, BFAVAR(3,1,2,1), and the BFAVAR model, BFAVAR(3,1,2), together with the AR(1) model. In Figure 2, a similar pattern emerges even though the models are slightly different. As in the previous figure, the BFAVAR(3,1,2) perform slightly better compared to the AR(1) model on the first quarter and seem to have a hard time outperforming the AR(1) for quarters 3-9. The BFAVAR(3,1,2) seem to gain most forecasting precision in the longer horizons, quarter 10-12. By including an informative prior on the steady-state, the performance compared to the AR(1) seems to change. When the informative prior on the steady-state is included, the BFAVAR(3,1,2,1) model outperform the AR(1) model on the three first quarters as opposed to only the first. On the fourth to the seventh quarter, the AR(1) and the BFAVAR(3,1,2,1) are similar in performance. The largest gains by using the BFAVAR(3,1,2,1) model take place on the eighth quarter and forward, with the maximum gain in forecasting precision for the model with an informative prior on the steady-state occurring on the two last quarters, quarter 11 to 12.

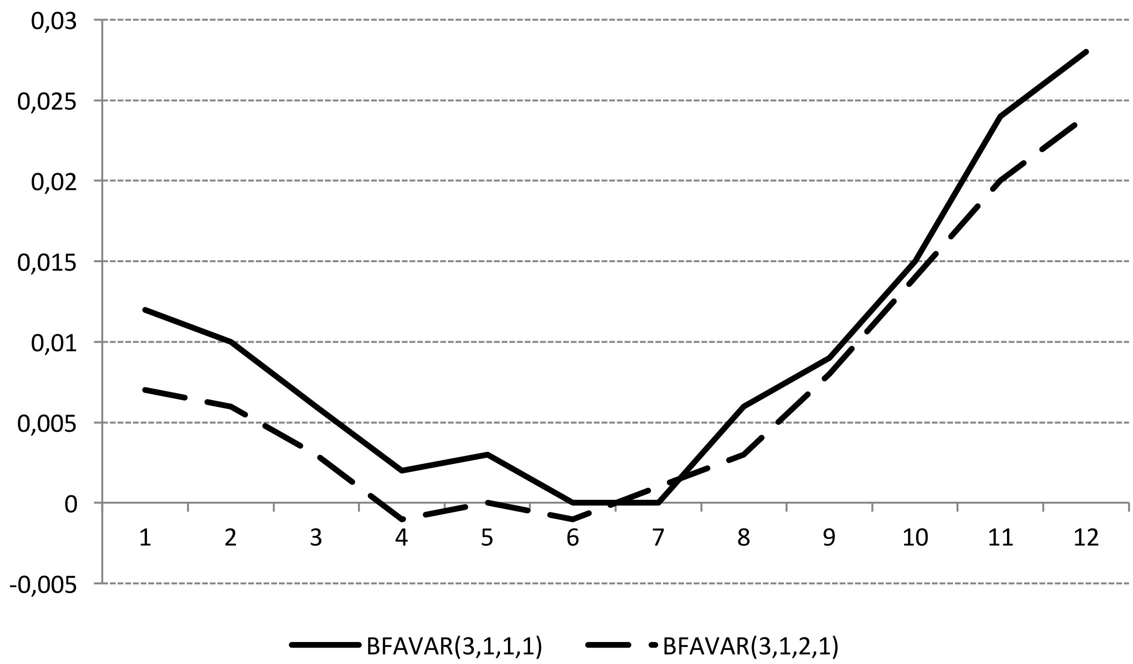
Figure 2: RMSFE for the second best BFAVAR models and the AR(1) model, 2006Q3 - 2016Q3



Note: The RMSFEs are given in percentage points on the vertical axis over the horizon 1-12 quarters on the horizontal axis. See section 4.1 for a description of the notation.

In Figure 3 we see the reduction in RMSFE for the two best BFAVAR models considered above with an informative prior on the steady state incorporated, compared to the AR(1) model. The reduction is the difference between the RMSFE of the AR(1) and the RMSFE of each competing model. This means that a positive number indicates that the BFAVAR models perform better than the AR(1) model. As we see in Figure 3, the BFAVAR(3,1,1,1) outperform the AR(1) model on all horizons except quarter six and seven, whereas the BFAVAR(3,1,2,1) perform worse than the AR(1) model on the horizon 4-6 quarters. We see that, as the forecasting horizon increase, the both BFAVAR models seem to continuously improve their performance relative to the AR(1) model, each reaching their maximum reduction in RMSFE compared to the AR(1) model at the last forecasting horizon, quarter 12. The maximum reduction in RMSFE of using the BFAVAR(3,1,1,1) model compared to the AR(1) is at most 0.028 percentage points, which translates into a reduction in RMSFE of about 9 per cent at the last quarter.[10]

Figure 3: Reduction in RMSFE for the BFAVAR models with an informative prior on the steady-state compared to the AR(1), 2006Q3 – 2016Q3



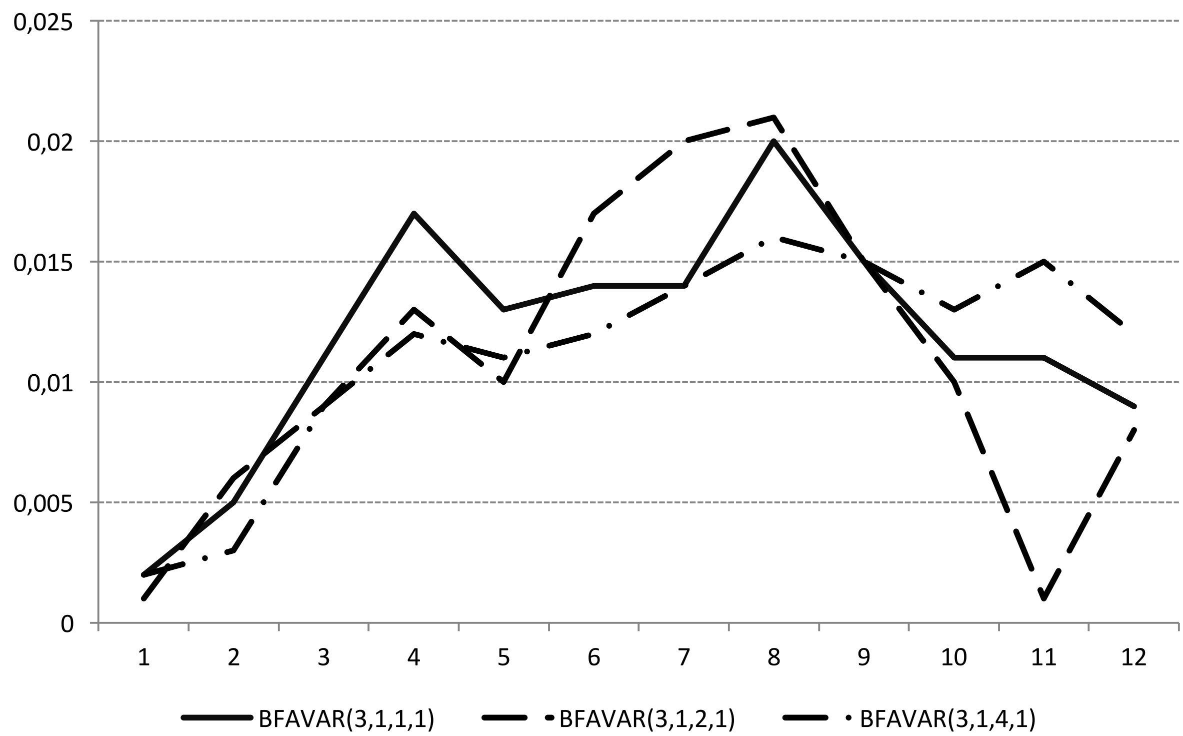
Note: Reduction in RMSFE is given in percentage points on the vertical axis. Forecasting quarters is on the horizontal axis. A positive number indicates that the models have a higher forecasting accuracy than the AR(1) model.

[10] At quarter 12: the reduction in RMSFE for the BFAVAR(3,1,1,1) model compared to the AR(1) model is 0.028 percentage points and the RMSFE of the AR(1) model is 0.325, which gives (0.028/0.325)*100 = 9 per cent. The reduction in RMSFE is expressed as per cent of the RMSFE of the univariate models.

From Figure 1-3 we see that the BFAVAR model without an informative prior on the steady-state struggle to outperform the AR(1) model on the medium-term horizons. However, when an informative prior on the steady state is included, the BFAVAR model increase its performance relative to the AR(1) model over most horizons.

Figure 4 presents the reduction in RMSFE by making use of an informative prior on the steady-state compared to having no information on the steady-state. Here, a comparison of the BFAVAR(3,1,1,1), BFAVAR(3,1,2,1) and the BFAVAR(3,1,4,1) is made. The last model is the third best performing model and is included for comparison. As in Figure 3, the reduction in RMSFE of the models is expressed as the difference between the BFAVAR model *without* an informative prior on the steady state and the corresponding BFAVAR model *with* an informative prior, i.e. the restricted and the unrestricted models. We see that all BFAVAR models with an informative prior reach their maximum gain in performance compared to the BFAVAR models without an informative steady-state prior in the eighth quarter.

Figure 4: Reduction in RMSFE by using the steady-state BFAVAR compared to the BFAVAR, 2006Q3 - 2016Q3



Note: Reduction in RMSFE is given in percentage points on the vertical axis. Forecasting quarters is on the horizontal axis. A positive number indicates that the models have a higher forecasting accuracy than the BFAVAR model without information on the steady-state.

At most, the gain of making use of the informative prior on the steady-state is 0.02 percentage points for the BFAVAR(3,1,1,1) model, which translates to a reduction in RMSFE of about 5 per cent compared to the model without steady-state.[11] The performance gain of the BFAVAR models using an informative prior on the steady-state compared to the BFAVAR models without the informative prior is expressed as per cent of RMSFE of the corresponding BFAVAR model without an informative prior.

The reduction in RMSFE of the steady-state BFAVAR as expressed in per cent of the RMSFE of the AR(1) is, at best, 9 per cent. This reduction is considerably less than the reductions of 25 per cent reduction found in the comprehensive study of a wide range of forecasting models by Faust and Wright (2013). However, these results are not completely comparable since Faust & Wright (2013) use a slightly different benchmark model. The 9 per cent reduction in RMSFE compared to the benchmark is more in line with the results found by Stockhammar & Österholm (2016) and Beechey & Österholm (2010), which is about 6-9 per cent reductions in RMSFE. Both above-mentioned papers argue that a reduction of 9 per cent in RMSFE is non-negligible with respect to economic significance.

The resulting reduction in RMSFE by incorporating an informative prior on the steady-state from this forecasting exercise of 5 per cent is a rather modest gain and on the borderline of being of economic significance. For example, Österholm (2010) finds a BVAR gaining about 7 per cent to a benchmark as a modest gain. As a comparison, while investigating if forecasting performance of Bayesian AR and VAR models can improve by incorporating an informative prior on the steady-state, Österholm (2008) also find that gains in precision are modest when forecasting inflation. Österholm (2008) find more prominent gains by incorporating prior information when forecasting interest rates. Note that Österholm (2008) use a different benchmark model, the Naïve model, which he argue are a reasonable benchmark given the high persistence in inflation. In this study, the commonly used AR(1) is used as the main benchmark and the Naïve model is simply included for reference. As we can see in Appendix C, Table 3, all three BFAVAR models with an informative prior on the steady-state and the ones without an informative prior clearly outperform the Naïve model on all horizons.

[11] At quarter 8 the reduction in RMSFE for the BFAVAR(3,1,1,1) model is 0.002 and the RMSFE of the BFAVAR (3,1,1) model is 0.421. Thus (0.02/0.421)*100 gives about 5 per cent.

In research related to incorporating informative priors on the steady-state related to VAR models, reductions in RMSFEs compared to various benchmarks is commonly found. For example, Mossfeldt & Stockhammar (2016) find reductions in RMSFE when forecasting goods and services inflation of about 37 per cent while making use of an informative prior on the steady-state in their BVAR model. The important difference in this setting is that the BFAVAR models in this study have included factors, which might on its own contribute to more precise forecasts in general and a more precise estimation of the steady-state in particular.

The base model in this application is basically taken at face value in order to be sensible economically speaking and to focus on the effect of including an informative prior on the steady-state. The only thing that varies with model choice is the inclusion of more than one estimated factor and the lag length. In this particular case, since the first factor accounts for about 96% of the variation in the large data set and the first five factors accounts for about 99% of the variation it is rather reasonable that forecasting performance is not increased significantly when additional factors are added to the model. In general, including more than one factor does not seem to increase forecasting performance for any model in this setting.

Similar results are also found in Gustafsson (2015). When forecasting monthly changes in inflation using a BFAVAR model without an informative prior on the steady-state, Gustafsson (2015) find that including more factors than one typically do not increase the overall forecasting performance for the best models. On the other hand, Laganá & Mountford (2005) find that adding as much as five factors to a benchmark VAR consisting of three variables reduces the RMSFE compared to a benchmark VAR with one included factor. Laganá & Mountford (2005) find that the five first factors, on average, explain a third of the variation in their data set. It is therefore not surprising that they find that adding more factors seem to contribute to the forecasting performance. They find that they would need 12 factors to explain about half of the variation in their data, which they argue is not practical to include given the short length of their data.

5. Conclusion

This paper has investigated whether out-of-sample forecasting precision can be increased by incorporating an informative prior on the steady-state in BFAVAR models. This issue was investigated by an out-of-sample forecasting exercise of inflation. The forecasting exercise shows that an informative prior on the steady-state incorporated in BFAVAR models marginally increase forecasting accuracy compared to BFAVAR models without the informative prior. The largest gain by using the BFAVAR model with an informative prior on the steady-state compared to the non-informative BFAVAR model is about 5 per cent, which can reasonably be considered a negligible-to-modest effect.

When comparing the forecasting accuracy of the steady-state BFAVAR with the AR(1), the RMSFE is reduced by approximately 9 per cent. This result can, in comparison to similar studies, be considered as more of a non-negligible effect.

This study has provided limited evidence for including a steady-state prior in BFAVAR models. There do, however, seem to exist some results in favor of this methodology compared to the AR(1) model, at least on longer forecasting horizons. However, there is one caveat related to the BFAVAR models. The factors estimated by principal components are estimated over the whole sample, meaning that there might be some overfitting bias.

Further research into distinguishing between low-inflation and high-inflation regimes might be able to specify different priors on inflation depending on which regime are current, leading to better forecasting accuracy in these different regimes. In a setting when the inflation target is incorporated at face value, the model using an informative prior on the steady-state tends to overestimate the future path of inflation and the time to reach the prior interval when inflation is far below the inflation target.

6. References

Adolfson, M, Andersson, M., Lindé, J, Villani, M, & Vredin, A. (2007), "Modern forecasting models in action: Improving macroeconomic analyses at central banks." *Riksbank Research Paper Series* 23.

Bachmeier, L. J., Leelahanon, S., & Li, Q. (2007), "Money Growth and Inflation in the United States", *Macroeconomic Dynamics*, 11, 113−127.

Beechey, M. & Österholm, P. (2010), "Forecasting Inflation in an inflation-targeting regime: A Role for informative steady-state priors". *International Journal of Forecasting*, 26, 248-264.

Berger, H. & Österholm, P. (2011), "Does Money Granger Cause Inflation in the Euro Area? Evidence from Out-of-Sample Forecasts Using Bayesian VARs", *Economic Record*, 87, 45−60.

Bernanke, B. & Boivin, J. (2003), "Monetary policy in a data-rich environment." *Journal of Monetary Economics* 50.3: 525-546.

Bernanke, B., Boivin, J. & Eliasz, P. (2005) "Measuring the Effects of Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) approach." *Quarterly Journal of Economics*, Feb 387: 422.

Carriero, A., Clark, T. & Marcellino, M. (2013), "Bayesian VARs: specification choices and forecast accuracy." *Journal of Applied Econometrics*.

Clements, MP. & Hendry, DF. (1999), "Forecasting Economic Time Series". Cambridge University Press: Cambridge.

Faust, J. & Wright, J.H. (2013), "Inflation Forecasting" in *Handbook of Forecasting*, Elliott, G. & Timmermann, A. eds., 2−56.

Gavin, W. T. & Kliesen, K. L. (2008), "Forecasting Inflation and Output: Comparing Data-Rich Models with Simple Rules", *Federal Reserve Bank of St Louis Review*, 90, 175−192.

Gelman, A, Ca, J., Stern, H. & Rubin, D. (2009), "Bayesian data analysis". Vol. 2. Chapman & Hall/CRC.

Gustafsson, O. (2015), "Forecasting Swedish CPI inflation – A Bayesian factor augmented VAR approach", MSc thesis, Department of Economics, Stockholm University.

Johnson, A. R. & Wichern, W. D. (2009), "Applied Multivariate Statistical Analysis." 6th edition. Pearson Prentice Hall NJ

Lagana, G. & Mountford, A. (2005), "Measuring Monetary Policy in the UK: a Factor Augmented Vector Autoregression Model Approach". *The Manchester School* 73.s1: 77-98.

Mossfeldt, M. & Stockhammar, P. (2016), "Forecasting Goods and Services Inflation in Sweden". *Working Paper* No. 146, National Institute of Economic Research.

Scheufele, R. (2011), “Are Qualitative Inflation Expectations Useful to Predict Inflation?”. *Journal of Business Cycle Measurement and Analysis*. 2011/1, 29−53.

Sims, C. (1980), “Macroeconomics and reality.” *Econometrica*: Journal of the Econometric Society: 1-48.

Stock, J. & Watson, M. (1999) “Forecasting inflation.” *Journal of Monetary Economics* 44.2: 293-335.

Stock, J. & Watson, M. (2002), “Forecasting using principal components from a large number of predictors.” *Journal of the American statistical association* 97.460: 1167-1179.

Stockhammar, P. & Österholm, P. (2016), “Do Inflation Expectations Granger Cause Inflation?”, *Working Paper No. 145*, National Institute of Economic Research.

Svensson, L.E.O. (2005), “Monetary Policy with Judgment: Forecast targeting”, *International Journal of Central Banking*, 1, 1−54.

Villani M. 2005. “Inference in vector autoregressive models with an informative prior on the steady state.” *Working paper No. 181*, Sveriges Riksbank.

Villani, M. (2009), “Steady-State Priors for Vector Autoregressions”, *Journal of Applied Econometrics*, 24, 630-650.

Österholm, P. (2008) “Can Forecasting Performance Be Improved by Considering the Steady State? An Application to Swedish Inflation and Interest Rate.” *Journal of Forecasting* 27, 41–51 (2008)

Appendix A – Bayesian Statistics and Principal Components

A.1 Bayesian inference in VAR models

Introducing shrinkage using Bayesian methods commonly starts with the researcher having prior beliefs or a priori knowledge about the parameters. The aim is to shrink unnecessary parameters towards zero to conserve degrees-of freedom. The clearest distinction between the Bayesian and the frequentist approach is that in the Bayesian approach, we condition on the observed data (see Gelman 2009 for a good introduction on Bayesian statistics).

The first step in Bayesian data analysis is to set up a full probability model. We therefore need a model providing the joint probability distribution for θ and y, in order to make probability statements on θ given y. Following the notation of Gelman (2009), the joint probability density function can be written as a product of two densities: the prior distribution $p(\theta)$, specified by the researcher, and the sampling distribution, $p(y|\theta)$, coming from the data:

$$p(\theta, y) = p(\theta)p(y|\theta). \quad (3)$$

The second step is to condition on the known value of the data, y. Using Bayes' rule[12], we get the posterior distribution we want to estimate:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \quad (4)$$

where y is the actual observations and θ are the model parameters. $p(\theta|y)$, the posterior distribution, is the conditional probability distribution of the unobserved values of interest. $p(y)$ is for scaling and does not depend on θ. With fixed y, $p(y)$ can be considered a constant and thus omitted. Therefore, equation (4) can also be written as the unnormalized posterior distribution:

$$p(\theta|y) \propto p(\theta)p(y|\theta), \quad (5)$$

[12] $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, where A and B are events and $P(B) \neq 0$

where $\propto$ means "proportional to". The resulting parameter is now a combination of the prior distribution and the data distribution and there are some intuitive results that follow from the relationship in (5):

(i) If the prior is uninformative, the posterior distribution is mainly determined by the data.

(ii) If the prior is informative, the posterior distribution is a combination of the prior and the data.

(iii) The more informative the prior is, the more data is needed in order to change our beliefs since the posterior will then be mainly driven by the prior information.

(iv) If there is a lot of data, the data will dominate the posterior distribution in the sense that the prior gets less weight when combining the prior and the data, which will make the information from the data more important.

In many empirical applications, the integration of $p(\theta|y)$ can be impossible. Therefore, numerical integration based on Monte Carlo simulation methods are used. A commonly used method is the Gibbs sampler; see e.g. Gelman et al. (2009) for further explanation.

The choice of priors can be purely subjective or based on empirics. There is a variety of priors to choose among and the literature has become quite extensive on the issue. The prior on the dynamics considered here is the most commonly used Minnesota prior. The prior on the covariance matrix is the standard Diffuse prior.

A.2 The Minnesota Prior

The priors used by Doan, Litterman and Sims (1984) and Litterman (1986) became known as the Minnesota prior due to their connection to the University of Minnesota and the Federal Reserve Bank of Minneapolis. This prior is based on replacing Σ with an estimate $\hat{\Sigma}$, which simplifies the prior compared to other priors (e.g. Diffuse-and Normal Wishart, which treats the covariance matrix as a random variable). The original Minnesota prior assumed Σ to be diagonal, which simplifies it even further. When we assume no correlation between the errors in each equation, then each equation of the VAR can be estimated one at the time and the elements on the diagonal becomes the estimated variance of each separate equation $\hat{\sigma}_{ii} = s_i^2$. Even though this approach simplifies computation, replacing the Σ with the estimate $\hat{\Sigma}$ might be disadvantageous since we are replacing an unknown matrix by an estimate (and not necessarily a good one).

By replacing Σ with the estimate $\hat{\Sigma}$, it is now fixed and we only have to specify a prior mean and variance for the coefficient matrix α. Following the notation of Koop & Korobilis (2010) the Minnesota prior assumes:

$$\alpha \sim N(\underline{\alpha}_{Mn}, \underline{V}_{Mn}). \tag{6}$$

An advantage of the Minnesota prior is that it leads to a simple posterior distribution in the sense that it only involves the Normal distribution:

$$\alpha | y \sim N(\overline{\alpha}_{Mn}, \overline{V}_{Mn}), \tag{7}$$

where

$$\overline{V}_{Mn} = \left[\overline{V}_{Mn}^{-1} + \left(\hat{\Sigma}^{-1} \otimes (X'X)'\right)\right]^{-1},$$
$$\overline{\alpha}_{Mn} = \overline{V}_{Mn}\left[\underline{V}_{Mn}{}^{-1}\underline{\alpha}_{Mn} + \left(\hat{\Sigma}^{-1} \otimes X\right)'y\right].$$ [13]

As Koop & Korobilis (2010) points out, the disadvantage with replacing Σ with the estimate $\hat{\Sigma}$ leads to a non-Bayesian way of treating Σ by ignoring the uncertainty in this parameter. Also, the Minnesota prior typically use a prior mean that reflects a distinct random walk behavior ($\underline{\alpha}_{Mn}$ are set to 1 for the first own lag and 0 for the rest). In our case it is more relevant to set the prior mean for the coefficient on the own first lag to 0.9, reflecting a prior belief that our variables show a fair degree of persistence, but not unit root behavior. Otherwise, the Minnesota prior would be inconsistent with the BFAVAR since a random walk does not have a well-defined unconditional mean.

[13] ⊗ denotes the Kronecker product, which results in a block matrix. For example:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \otimes \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} a\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} & b\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\ c\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} & d\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \end{bmatrix}$$

Recall that the Minnesota prior assumes the prior covariance matrix, $\underline{V}_{Mn}$, to be diagonal. We can then let $\underline{V}_i$ be the block of $\underline{V}_{Mn}$ with the K coefficients in equation i and $\underline{V}_{ij}$ be its diagonal elements. A common implementation of the Minnesota prior is then to set the hyperparameters to:

$$\underline{V}_{ij} = \begin{cases} \frac{\underline{a}_1}{r^2}, \text{for coefficients on own lag r for r} = 1, \ldots, \text{p} \\ \frac{\underline{a}_2 \sigma_{ii}}{r^2 \sigma_{jj}}, \text{for coefficients on lag r of variable j} \neq \text{i for r} = 1, \ldots, \text{p}, \\ \underline{a}_3 \sigma_{ii}, \text{ for coefficients on exogenous variables} \end{cases} \quad (8)$$

This simplifies the prior by letting us choose three scalars $\underline{a}_1, \underline{a}_2$ *and* $\underline{a}_3$ rather than specifying all elements of $\underline{V}_{Mn}$. Also, this specification captures the property that, as lag length increases, the coefficients are increasingly shrunk towards zero and that by setting $\underline{a}_1 > \underline{a}_2$, own lags are more likely to be important predictors than lags of other variables. The exact choice of values for these hyperparameters depends on the empirical applications but what should be noted is that these hyperparameters determines how tightly to shrink the prior variance towards the random walk. If the prior variance is set to infinity, it will resemble estimating a standard reduced form VAR, while if the prior variance is set to 0 it is equivalent of modelling a random walk for each equation.

With regards to optimal choices of hyperparameters, Carriero et al. (2013) find very small losses, and even gains, by adopting specification choices that make BVAR modelling fast and simple. Carriero et al. (2013) find that cross-variable tightness of 0.5 is better with respect to accuracy than a value of 0.2. They find that fixing the lag length and hyperparameters according to their idea of simple Bayesian estimation is hard to beat, meaning a lag length of a year and hyperparameters as before. In general, they find that fixing lag length on a bit longer than optimum (if optimum is not known) is a good idea.

A.3 Principal Component Analysis

Principal component analysis is primarily a tool for explaining the variance-covariance structure of a large set of variables through a number of linear combinations of these variables. It is possible to create as many principal components as there are variables in the data set, however, often much of the variability can be explained using the principal components such that p principal components contain as much, or almost as much, information as the original k variables, Johnson & Wichern (2009). For example, we are interested in p new variables which are linear combinations of the variables in the larger data set:

$$\begin{aligned} \delta_1 &= w_{11}x_1 + w_{12}x_2 + \cdots + w_{1p}x_p \\ \delta_2 &= w_{21}x_1 + w_{22}x_2 + \cdots + w_{2p}x_p \\ &\vdots \\ \delta_p &= w_{p1}x_1 + w_{p2}x_2 + \cdots + w_{pp}x_p \end{aligned}, \tag{9}$$

where $\delta_1, \delta_2, \dots, \delta_p$ are principal components, w_{ij} are weights and x_i are the original variables. Before estimating the principal components, the variables have to be weakly stationary and the data has to be normalized in order to get the variables in comparable units, otherwise variables of high scale would account for relatively too much of the total variation.

The principal components in (11) are calculated given three conditions: [14]

(i) The first variable accounts for as much of the total variation as possible in the data, the second accounts for as much as possible of the variation left and so on.
(ii) The squared weights sum to one. This condition ensures a proper scale of the variables, for example the variance of a new variable could otherwise be altered by scaling up or down the weights.
(iii) $w_{i1}w_{j1} + w_{i2}w_{j2} + \cdots + w_{ip}w_{jp} = 0 \quad \forall\, i,j = 1,2,\dots,p, i \neq j$ i.e., the principal components are orthogonal.

When choosing the numbers of principal components to create from the original data set, there is some rule of thumb. One suggestion is to keep adding components until there is a distinct "bend" in the scree plot and then include all components before the "bend" including the "bend" itself. Another popular suggestion is to keep adding components as long as the eigenvalue is greater than one, Johnson & Wichern (2009). However, different ways of deciding the optimal number of factors are not necessarily optimal when deciding the number to be included in the FAVAR model, Bernanke et al (2005). The principal components can then be used to calculate the scores, which are then used in the model.

[14] For the complete maximization problem, see Johnson & Wichern (2009).

Appendix B – Data Description and Transformation

Below follows a short description of the variables used for each model. The first column is the name of the variables, the second is how it is measured, the third is a code for how seasonality is treated, the fourth is how the series is transformed and the fifth is a short description of the variable. If the seasonal code is 0, no seasonal adjustment is made, if the code is 1 the ARIMA-X13 SEATS of the Census Bureau is used and if the seasonal code is 2 the series is adjusted from the data base. For the transformation code column: 0 means no transformation, 1 means the natural log, 2 means first difference, 3 means first difference of the natural log, 4 means two first differences and 5 means two first difference of the natural log. * Indicates that the monthly series is transformed to quarterly by averaging over the months making a quarter.

Table 2: Data Description and Transformations for BFAVAR model

Time Series	**Measure**	**Season code**	**Transformation code**	**Description**
Real Output and Income				
IPI_AGG	index 2010=100	2	2	Industrial production in manufacturing, mines and minerals (quarterly)
IPI_MAN	index 2010=100	2	3	Industrial production in manufacturing (quarterly)
IPI_FOOD	index 2010=100	2	3	Industrial production in food, tobacco and alcohol (quarterly)
IPI_LUMBER	index 2010=100	2	3	Industrial production in lumber and planning mill (quarterly)
IPI_PAPER	index 2010=100	2	3	Industrial production in paper industry (quarterly)
IPI_STEEL & METAL	index 2010=100	2	3	Industrial production in steele and metal mill (quarterly)
IPI_MOTOR	index 2010=100	2	3	Industrial production in Motor industry(quarterly)
HH_DISP	Million SEK	1	3	Household disposable income (quarterly)
GDP	Million SEK	2	3	Gross domestic product (quarterly)
GROSS_CAP	Million SEK	2	3	Gross capital formation (quarterly)
CHANGE_INVENT	Million SEK	2	3	Change in capital spent on inventories (quarterly)
IMPORTS	Million SEK	2	3	Total imports (quarterly)
EXPORTS	Million SEK	2	3	Total exports (quarterly)
Employment & Hours				
EMP_GOV	Number employed 16-64	0	3	Number employed in governmental sector (thousands) *
EMP_MUN	Number employed 16-64	0	3	Number employed in municipalities (thousands) *
EMP_PRIV	Number employed 16-64	2	3	Number employed in private sector (thousands) *
EMP_TOT	Number employed 16-64	0	3	Number employed in total (thousands) *
UNEMP_RATE	Percent of age 16-64	0	2	Unemployment rate as percentage of working age population
EMP_RATE	Percent of age 16-64	0	2	Employment rate as percentage of working age population *

UNEMP_LABOUR	Percent of age 16-64	0	2	Unemployment rate as percentage of labour force *
NOT_LABOUR	Number of inactive 16-64	0	2	Share of not participating in the labour force as percentage of working age population *
HRS_TOT	Hours in millions	2	3	Total numbers worked within and outside of Sweden (quarterly)
HRS_AGRICULTURE	Hours in millions	2	3	Total numbers worked in agriculture (quarterly)
HRS_MIN_MAN_QUAR	Hours in millions	2	3	Total numbers worked in mining, manufacturing and quarrying(quarterly)
HRS_MAN	Hours in millions	2	3	Total numbers worked in manufacturing industry (quarterly)
HRS_ENG_ENV	Hours in millions	2	3	Total numbers worked in energy and environmental industry (quarterly)
HRS_CONS	Hours in millions	2	3	Total numbers worked in construction industry (quarterly)
HRS_PROD_SERVICES	Hours in millions	2	3	Total numbers worked in producer of services sector (quarterly)
HRS_RETAIL	Hours in millions	2	3	Total numbers worked in retail industry (quarterly)
HRS_TRANSPORT	Hours in millions	2	3	Total numbers worked in transportation sector (quarterly)
HRS_HOT_REST	Hours in millions	2	3	Total numbers worked in hotel and restaurant sector (quarterly)
HRS_INFO	Hours in millions	2	3	Total numbers worked in information and communications sector (quarterly)
HRS_FIN	Hours in millions	2	3	Total numbers worked in financial and insurance industry (quarterly)
HRS_RESEARCH	Hours in millions	2	3	Total numbers worked in research and development sector (quarterly)
HRS_EDUC_HEALT_SOCIAL	Hours in millions	2	3	Total numbers worked in education, health and social sector (quarterly)
HRS_CULT	Hours in millions	2	3	Total numbers worked in culture (quarterly)
HRS_PUB_ADM	Hours in millions	2	3	Total numbers worked in public and administrative sector (quarterly)
VACANCIES	Amount	2	3	Vacant job positions *
Consumption				
IND_CONS	Million SEK	0	3	Household individual consumption (reported quarterly)
GOV_BUDGET_EPX	Million SEK	1	3	Government budget expenditure *
HH_CONS_EXP	Million SEK	2	3	Household consumption expenditure excluding non-profitable organizations (reported quarterly)
Expectations				
OWN_NOW	Scale 1:6	0	2	Question: How is your economy right now? (individuals) *
OWN_12	Scale 1:6	0	2	Question: How is your economy in 12 months? (individuals) *
SWE_NOW	Scale 1:6	0	2	Question: How is the Swedish economy now? (individuals) *
SWE_12	Scale 1:6	0	2	Question: How is the Swedish economy in 12 months? (individuals) *
SAVE_NOW	Scale 1:6	0	2	Question: Is it favorable to save now? (individuals) *
EXPP_RS	Index	0	2	Expected selling prices for firms in retail sales sector (quarterly)
EXPP_NDG	Index	0	2	Expected selling prices for firms in non-durable goods sector (quarterly)
EXPP_M	Index	0	2	Expected selling prices for firms in motor sector (quarterly)
EXP_CPI_12	Index	0	2	Expected inflation in 12 months for private sector (quarterly)
Housing				

REAL_EST_PRICE_IND	Index	0	3	Index over real estate prices in Sweden (reported quarterly)
SMH_Q	Number/1000	1	3	Number of small houses sold (reported in thousands)
HOL_Q	Number/1000	1	3	Number of holiday houses sold (reported in thousands)
RENT_Q	Number/1000	1	3	Number of rental real estates sold (reported in thousands
INDU_Q	Number/1000	1	3	Number of industrial real estates sold (reported in thousands)
FARM_Q	Number/1000	1	3	Number of farming real estates sold (reported in thousands). Note observation for 1996-1998 is made quarterly by averaging.
SMH_P	Number/1000	0	3	Purchasing price small houses sold (mean)
HOL_P	Number/1000	0	3	Purchasing price holiday houses sold (mean)
RENT_P	Number/1000	0	3	Purchasing price rental real estates sold (mean)
INDU_P	Number/1000	0	3	Purchasing price industrial real estates sold (mean)
FARM_P	Number/1000	0	3	Purchasing price farming real estates houses sold (mean). Note observation for 1996-1998 is made quarterly by averaging.
SMH_BAS/TAX	Number/1000	0	3	Ratable value of small houses sold (mean)
HOL_BAS/TAX	Number/1000	0	3	Ratable value of holiday houses sold (mean)
RENT_BAS/TAX	Number/1000	0	3	Ratable value of rental real estates sold (mean)
INDU_BAS/TAX	Number/1000	0	3	Ratable value of industrial real estates sold (mean)
FARM_BAS/TAX	Number/1000	0	3	Ratable value of farming real estates sold (mean). Note observation for 1996-1998 is made quarterly by averaging.
SMH_P/TAX	Number/1000	0	3	Purchasing price over ratable value of small houses sold (mean)
HOL_P/TAX	Number/1000	0	3	Purchasing price over ratable value of holiday houses sold (mean)
RENT_P/TAX	Number/1000	0	3	Purchasing price over ratable value of rental real estates sold (mean)
INDU_P/TAX	Number/1000	0	3	Purchasing price over ratable value of industrial real estates sold (mean)
FARM_P/TAX	Number/1000	0	3	Purchasing price over ratable value of farming real estates sold (mean). Note observation for 1996-1998 is made quarterly by averaging.
Stock & Commodity prices				
OMXS30	Index 2010=100	0	3	Swedish stock index of the 30 largest companies *
BRENT	Dollar/Barrel	0	3	Price of Brent oil in US dollar (quarterly)
COPPER	Dollar/Metric ton	0	3	Price of copper in US dollar (quarterly)
IRON	Dollar/Metric ton	0	3	Price of iron ore in US dollar (quarterly)
ALUM	Dollar/Metric ton	0	3	Price of aluminum in US dollar (quarterly)
NICKEL	Dollar/Metric ton	0	3	Price of nickel in US dollar (quarterly)
Exchange rates				
YEN	YENSEK	0	2	YENSEK exchange rate *
EUR	EURSEK	0	2	EURSEK exchange rate *
GBP	GBPSEK	0	2	GBPSEK exchange rate *
USD	USDSEK	0	2	USDSEK exchange rate *
KIX	Index	0	2	Weighted exchange rate index by NIER *
TWC	Index 1992=100	0	2	Weighted exchange rate index by the Riksbank *
Interest rates				
STIBOR1M	Percentage	0	2	Stockholm Interbank Official Rate 1 months *

STIBOR3M	Percentage	0	2	Stockholm Interbank Official Rate 3 months *
STIBOR6M	Percentage	0	2	Stockholm Interbank Official Rate 6 months *
SSVX1M	Percentage	0	2	Treasury bill 1 month *
SSVX3M	Percentage	0	2	Treasury bill 3 months *
SSVX6M	Percentage	0	2	Treasury bill 6 months *
GVB2Y	Percentage	0	2	Government bond 2 year *
GVB5Y	Percentage	0	2	Government bond 5 year *
GVB10Y	Percentage	0	2	Government bond 10 year *
BOOBL2Y	Percentage	0	2	2 yeas housing bond *
BOOBL5Y	Percentage	0	2	5 year housing bond *
Money and Credits				
M0	Million SEK	0	3	Currency and digital money on the market (narrow money) *
M3	Million SEK	0	3	M1 + debt and deposits up to two years maturity *
GOV_DEBT	Million SEK	0	3	Government external debt *
Prices				
CPI	Percentage change	1	0	Change in Swedish consumer price index (quarterly)
CPIF	Percentage change	0	0	Change in Swedish consumer price index deducing effects from housing interest rates
EXPI	Percentage change	0	0	Change in export price index
HMPI	Percentage change	0	0	Change in home market prices
IMPI	Percentage change	0	0	Change in import price index
ITPI	Percentage change	0	0	Change in domestic resource price index
PPI	Percentage change	0	0	Change in producer price index
TRIM85	Yearly percentage change	0	2	The Riksbank's measure of underlying inflation (see http://www.riksbank.se/sv/Statistik/Makroindikatorer/Underliggande-inflation/) *
UND24	Yearly percentage change	0	2	The Riksbank's measure of underlying inflation (see http://www.riksbank.se/sv/Statistik/Makroindikatorer/Underliggande-inflation/) *
Foreign variables				
US3M	Percentage	0	2	US 3 month interest rate *
US6M	Percentage	0	2	US 6 month interest rate *
FED_INTEREST	Percent	0	2	Federal reserve interest rate *
EU_INFLATION	Percentage change	0	0	Change in EU inflation *
US_EPUI	Index	0	2	US economic policy uncertainty index. High value of index indicates high uncertainty etc. *
US_UNEMP	Percent	0	2	US unemployment rate *
US_INFLATION	Percentage change index: 2010=100	2	0	Change in US inflation rate (quarterly)
EU_EPUI	Index	0	2	EU economic policy uncertainty index. High value of index indicates high uncertainty etc. *
US_FSI	Index	0	2	St Louis FED Financial Stress Index. High value of index indicates high uncertainty etc. quarterly

Appendix C – RMSFEs

Table 3: RMSFEs for BFAVAR models

Horizon	1	2	3	4	5	6	7	8	9	10	11	12
BFAVAR without steady-state												
BFAVAR(3,1,1)	0.372	0.44	0.461	0.48	0.479	0.445	0.445	0.421	0.399	0.374	0.313	0.306
BFAVAR(3,1,2)	0.376	0.445	0.462	0.479	0.479	0.449	0.45	0.425	0.4	0.374	0.307	0.309
BFAVAR(3,1,4)	0.379	0.451	0.47	0.486	0.487	0.454	0.454	0.428	0.405	0.382	0.323	0.32
BFAVAR with steady-state												
BFAVAR(3,1,1,1)	0.37	0.435	0.45	0.463	0.466	0.431	0.431	0.401	0.384	0.363	0.302	0.297
BFAVAR(3,1,2,1)	0.375	0.439	0.453	0.466	0.469	0.432	0.43	0.404	0.385	0.364	0.306	0.301
BFAVAR(3,1,4,1)	0.377	0.448	0.461	0.474	0.476	0.442	0.44	0.412	0.39	0.369	0.308	0.308
Benchmark												
AR(1)	0.382	0.445	0.456	0.465	0.469	0.431	0.431	0.407	0.393	0.378	0.326	0.325
Naive	0.409	0.542	0.588	0.665	0.665	0.643	0.699	0.708	0.633	0.597	0.554	0.503

Note: the univariate model refers to the univariate BVAR model. The Naïve model is a random walk benchmark such that: $\hat{\boldsymbol{x}}_{\boldsymbol{t+h|t}} = \boldsymbol{x_t}, \qquad (\boldsymbol{h = 1, \ldots H})$.

Table 4: Relative RMSFE of BFAVAR models compared to AR(1), 2006Q3 - 2016Q3

Horizon	1	2	3	4	5	6	7	8	9	10	11	12
BFAVAR without steady-state												
BFAVAR(3,1,1)	0.97	0.99	1.01	1.03	1.02	1.03	1.03	1.03	1.02	0.99	0.96	0.94
BFAVAR(3,1,2)	0.98	1.00	1.01	1.03	1.02	1.04	1.04	1.04	1.02	0.99	0.94	0.95
BFAVAR(3,1,4)	0.99	1.01	1.03	1.05	1.04	1.05	1.05	1.05	1.03	1.01	0.99	0.98
BFAVAR with steady-state												
BFAVAR(3,1,1,1)	0.97	0.98	0.99	1.00	0.99	1.00	1.00	0.99	0.98	0.96	0.93	0.91
BFAVAR(3,1,2,1)	0.98	0.99	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.94	0.93
BFAVAR(3,1,4,1)	0.99	1.01	1.01	1.02	1.01	1.03	1.02	1.01	0.99	0.98	0.94	0.95

Note: the relative RMSFE is calculated as a ratio between the univariate model and the competing model. A number smaller than 1 indicates lower RMSFE for the competing model.

Table 5: Reduction in RMSFE for the best models compared to the AR(1) model, 2006Q3 - 2016Q3

Horizon	1	2	3	4	5	6	7	8	9	10	11	12
BFAVAR without steady-state												
BFAVAR(3,1,1)	0.01	0.01	-0.01	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	0.01	0.02
BFAVAR(3,1,2)	0.01	0.00	-0.01	-0.01	-0.01	-0.02	-0.02	-0.02	-0.01	0.00	0.02	0.02
BFAVAR(3,1,4)	0.00	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	0.00	0.00	0.01
BFAVAR with steady-state												
BFAVAR(3,1,1,1)	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.03
BFAVAR(3,1,2,1)	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02
BFAVAR(3,1,4,1)	0.01	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	0.01	0.02	0.02

Note: the reduction in RMSFE is calculated as the difference between the univariate model and the competing models. A negative number indicates that the RMSFE is smaller for the univariate model than the competing model.

Table 6: Reduction in RMSFE for BFAVAR with steady-state compared to BFAVAR without steady-state

Horizon	1	2	3	4	5	6	7	8	9	10	11	12
BFAVAR with steady-state												
BFAVAR(3,1,1,1)	0.002	0.005	0.011	0.017	0.013	0.014	0.014	0.020	0.015	0.011	0.011	0.009
BFAVAR(3,1,2,1)	0.001	0.006	0.009	0.013	0.010	0.017	0.020	0.021	0.015	0.010	0.001	0.008
BFAVAR(3,1,4,1)	0.002	0.003	0.009	0.012	0.011	0.012	0.014	0.016	0.015	0.013	0.015	0.012

Note: the reduction in RMSFE is calculated as the difference between the BFAVAR model *with* steady-state and the competing model, BFAVAR *without* steady-state. A negative number indicates that the RMSFE is smaller for the BFAVAR model without steady-state than the competing model.