Central European University

Department of Economics

Bayesian model for forecasting Hungarian inflation:

A comparison of the performance of BVAR, VAR and ARIMA models

By
Miklós Lukács

Submitted to
Central European University
Department of Economics

In partial fulfillment of the requirements for the degree of Master of Economics

Supervisor: Professor Gábor Kőrösi

Budapest, Hungary
2009

Contents

Abstract

This thesis aims to conduct forecasts of the Hungarian inflation for different time horizons based on observations from January 1998 to March 2007. The forecasting precision of three different models are introduced: ARIMA, VAR and Bayesian VAR. The prior assumptions of the Bayesian model are specified with the Minnesota prior, which proves to be a proper setup for forecasting consumer prices, since the calculated RMSEs and MAEs shows that the BVAR framework is a more powerful tool to forecast inflation than the other two simple models lacking theoretical background. Moreover overfitting problem of the VAR model does not cause any problem in this framework.

Keywords: Bayesian VAR, forecast, inflation, Minnesota prior

List of tables and figures

1. Introduction

Inflation is one of the most often mentioned macroeconomic terms in every day life, because we perceive it directly, while for example the changes in GDP, export, import etc. are mostly known from the news. But this is not the reason why it is important. The expectation for the future level of this variable has very serious effects on the behavior of agents in the economy. The question is not how fast the prices increase, but whether we can predict it or not. If we could predict the inflation accurately even higher levels would not cause any problem but, as empirics shows, the predictability of inflation is negatively correlated with the magnitude. The costs that rise because of this uncertainty about the future level of inflation are severe. Mostly it is the responsibility of Central Banks to reduce this loss through their substantial and theoretically well funded forecasts.

Recently in the developed economies, inflation targeting has become the most popular and useful tool to stabilize prices. The major view is that price stability is essential to sustain the long run stable growth of the economy. In Hungary, the first and major objective of the national bank is to reach and sustain price stability. Its role, on the one hand, is to keep the prices as stable as possible that reduces ambiguity in the economy; on the other hand forecast the future level of macroeconomic variables. These forecasts are not only informative about the future, but they also shepherd the market, or expectations, towards stability. To some extent, if everybody believed that the inflation would materialize around the forecasted level that could conduct the economy towards the aimed state. This latter can only work if the forecasts are accurate and reliable, otherwise the central bank can not have an effect on the expectations. To be as credible as possible, many models are used to predict the path of each macroeconomic variable. Unfortunately – or fortunately – the complexity of the world can not be described by any model, as all of them have shortcomings; still they are useful to capture some of the underlying processes that keep the whole system in motion.

This entanglement can be partially handled if the prior beliefs of the analyst are incorporated in the model. Bayesian econometrics creates a very useful framework that can fill our knowledge about the behavior of certain variables in the model. In the literature, results are promising; Bayesian models outperform the standard ones that are usually applied as benchmark models to measure the goodness of fit of the forecasted values. In this work, I will show the forecasting performance of three different models that are widely used. In the Bayesian model the so called Minnesota prior will be applied, and its results will be compared, through loss functions, to the accuracy of an ARIMA and VAR model in terms of forecasting the Hungarian inflation for different time horizons. The structure of this thesis is the following. Chapter 2 contains the historical background, the development of concept of the Bayesian econometrics, the Minnesota prior and related works are described. In chapter 3 the Bayesian VAR model and the prior specification are presented. The results, the comparison of the three models, and alternative prior specifications are reported in chapter 4.

2. Historical background and literature overview

The so called Bayesian approach, developed in parallel with the classical statistics, was named after the prominent mathematician Thomas Bayes (1702-1761), though he was given credit for it only posthumously. It was Pierre-Simon Laplace who applied the general version of the theorem in celestial mechanics, reliability, jurisprudence and medical sciences (Stigler, 1986). This point of view of probability theory was oppressed by the more popular frequentist concept, statisticians simply neglected it until the 20th century. However outside of this area the approach had some supporters: Harold Jeffrey (1891-1989), a physicist, and Arthur Bowley (1869-1957), an econometrician, both applied the doctrine in their field (Bradley & Luis, 1996). The turning point came only in 1939, when Harold Jeffrey released his book *Theory of probability*; thereafter more and more publications appeared and the theory became very popular in many sciences. It became popular but there used to be shortcomings also, the calculation of heavy integrals set back the application of the concept. The real breakthrough was at the beginning of the 1980s due to the development of computational methods, the appearance of computers and the growing need for sophisticated methods. In the recent decades this approach became widely used in areas such as physics, social sciences, cybernetics, evolutionary biology, since it is very useful to understand cognitive systems.

Bayesian econometrics was first propagated by Arnold Zellner in the beginning of the 1970s (Zellner, 1971). After this, more and more publications and books appeared in this topic like that of Dale Priorier, which explained the Bayesian and frequentist approach in detail. Bauwens, Lubrano and Richard released an influential book in 1999 that dealt only with some particular areas of econometrics. All this was incorporated in Gary Koop's book in 2003 and also he gave a wide range of models for applied work with Bayesian method (Koop, 2003).

2.1 The difference between the frequentist an the Bayesian approach[1]

Bayesian methods have been applied in wide range of empirical or theoretical problems in numerous fields, and performed as well as any other available methods. All the properties of non-Bayesian finite and asymptotic properties hold for Bayesian method also. Common in both approaches is that they assume a model structure that depends on unknown parameters and a given data set, and also that the *y*, the data, is one realization of the data generating process. The main points of the classical concept are: (i) there exists a true parameter vector θ_0 that generated the observed data; (ii) the estimation, the test statistics, and the confidence intervals all depend on *y*, the data; and their properties are derived from a hypothetical repeated sampling; (iii) with a given model a hypothetic data set can be generated and from these the distribution of the estimates can be obtained; and (iv) the inference is based on the comparison of the original and the hypothetic data. In the Bayesian case (i) the parameter vector is a random variable with a given distribution; (ii) prior information is incorporated in the estimation; (iii) the data is used to revise the prior information; and (iv) inference is based upon one single realization of the data.

[1] Based on Várpalotai (2008)

2.2 *The main idea of the Bayesian econometrics*

The basic idea relies on the learning process that starts from the knowledge that we have already learned about the world and all the incoming new information, which updates and becomes part of our *ex ante* beliefs. Throughout these loops we sharpen our comprehension of the mechanisms that we are interested in the most. The closer we are to the true nature of the underlying processes the better inference we can make. The main concept (Green, 2008) of the theory comes from the finding that the probability of event A given B occurred is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

For econometric purposes, this can be written in a more meaningful way

$$P(parameters|data) = \frac{P(data|parameters)P(parameters)}{P(data)}.$$

The expression in the denominator is only a constant, since it does not depend on the parameters that we are interested in, so we can simplify the above expression to

$$P(parameters|data) \propto P(data|parameters)P(parameters),$$

where $\propto$ stands for "is proportional to". As we dropped the denominator the expression on the left hand side is not a proper density, but it can be scaled up by the appropriate constant. On the right, we have the product of the joint distribution function of the observed variables given the parameters and marginal distribution of the parameters. Or we can rewrite right hand side as

$$(1) \qquad P\left(parameters \middle| data\right) \propto Likelihood\ \ function \times \Pr ior\ density\,,$$

where we have the joint distribution function given the parameters and the prior beliefs about the data generating process. This helps us to revise the information included in the data according to our knowledge that may come from intuition or theory.

2.3 Multivariate models

The above concept can be easily applied to single or multivariable cases. In this work the latter will be considered as a baseline model, since the forecast of consumer prices arises from a multivariate dynamic system. Vector Autoregression models are widely used in decision making, even if they have several shortcomings, but still it is a very handy framework to capture the dynamics of a multiple equation systems. These problems have already been profoundly discussed in the literature, and many solutions were suggested to overcome these (Ciccarelli and Rebucci, 2003). First drawback of this model is the overfitting; too many parameters have to be estimated that leads to huge loss of degrees of freedom, which causes inefficiency in the estimation. Second problem comes from making inference form the reduced form of VAR models, which is not more than simple description of the data.

The literature of this model is vast: many authors have contributed to develop the framework and raise it to a more theory based tool. With a priori knowledge about the data generating, we can restrict the original ad hoc structure to a more theory based one. The discovery of the relationship between stochastic trends of economic variables opened new dimensions in VAR analysis; Granger (1981), Engle and Granger (1987) and Johansen (1995) contributed a lot to the development of the cointegration analysis, so the Vector Error Correction (VECM) models became not only popular, but also useful application in econometrics.

Along with the above achievements, the Bayesian approach gained more space in publications. This method does not place any restrictions on the model coefficients; the uncertainty is taken into account when we formulate the prior distributions of the parameters of the model. This specification seems to be *ad hoc*, since there is no rule how to set the prior distribution; Leamer (1978) pointed out that "Two economists can legitimately make different inferences from the same data set." He suggested that alternative priors should be taken into account to have a broader view of the behavior of the model for different setups. The Bayesian method can be applied universally from binary choice problems through panel data, but in forecasting time series the multivariate VARs became more popular than univariate alternatives. Albeit Litterman (1980) was the first who recommended the Bayesian approach to VAR models to get around the overfitting problem, but the contribution of Thomas Doan, and Christopher Sims (1984) was important also. Their approach pushed the VAR model towards random walk, while in the most recent literature the dynamic stochastic general equilibrium models became more favored. [2]

[2] Ingram and Whiteman, 1994, Del Negro and Schorfheide, 2004, Canova, 2007

2.4 *The Minnesota prior*

In the 1980s, Robert Litterman was a teacher of University of Minnesota and at the same time he worked for the Federal Reserve Bank of Minneapolis, so the prior that he advocated is called Minnesota prior (Litterman, 1984 and 1986). His idea was: through a VAR estimation it can not be taken as granted that some coefficients are zero and also they may vary around a certain value. So he ordered a probability distribution to the parameter vector. The uncertainty is incorporated in the model, when the prior assumption is formulated. According to his point of view the optimal prior should be specified in such a way that the following – typical – time series regularities are taken into account: (i) trending of time series; (ii) the more recent series of a macroeconomic variable contains more information about the present state than past values; (iii) the own lagged values of a variable contains more relevant information than that of other variables (Ciccarelli and Rebucci, 2003). With all these properties, we have arrived at the random walk process.

The above properties of macroeconomic series can be captured if we assign probability distributions to the parameter vector. In Litterman's (1980) specification this was defined as (i) the coefficients of all lags except for the first one should be zero; (ii) the number of lags is negatively related to the variance of the coefficients; and (iii) the coefficient in an equation of the variables own lag is greater than that of the other explanatory variables. These properties and our prior knowledge can be compressed into set of hyperparameters $\Lambda = (\lambda_1 .... \lambda_H)$, where the Λ contains all our prior assumptions about the time series. These can control for the tightness of the overall model, the lagged and the other explanatory variables in each equation; they encompass the strength of our beliefs about the priors; and they can be responsible for the lag decay. This simplifies the problem to the estimation of H parameters, instead of all the parameters in the VAR equations.

2.5 *Forecasting inflation in Hungary*

In Hungary, the most precise inflation forecasts are prepared by the Central Bank; this is the one of the few institutions, which has predictions for longer horizons. Updating the forecasting method of consumer prices comes up from time to time. It is necessary because the assumptions change over time, and also there is a need to apply more and more sophisticated models for capturing the real nature of this variable.

Among numerous models and papers, one can track the evolution of modeling, and precision. From simple SARIMA models, very complicated DSGE models can be found as well. Models became more precise, more theoretically funded, but this does not mean that the simple model are worse than their counterparts. It was shown by Lieli (1999) that simple SARIMA model only extrapolates the past to the future, nevertheless it works well in the short run, even if it is unable to tell the long run course of the inflation. These simple models can serve as a reference point to evaluate more complicated models. In Hungary, inflation targeting regime was introduced in 2001, so the predictability of inflation became even more important. Fortunately this problem can be approached from many ways, Várpalotai (2003) came out with a paper, in which the inflation is estimated through cost factors. In his disaggregated level model, the inflation evolves from the underlying costs; in this structure both long-run and short run equations are estimated; that results a cointegration type framework, the consumer prices supposed to materialize around the level that is explained by the cost of factors.

Though not the latest, but the most applicable model of the central bank is called Hungarian Quarterly Projection model (N.E.M.), which combines neo-Keynesian and neoclassical assumptions; the first is in effect in the short run, while the latter is applied for long run effects.[3] In terms of inflation the assumed price rigidities has important role in this model, while the overall equilibrium is connected through error correction mechanism. The other recently developed model is based on the dynamic stochastic general equilibrium framework; this is a much more sophisticated that accounts for real and nominal rigidities, and incorporates different type of frictions. The agents in the economy are assumed to perceive the macroeconomic variables in an adaptive way depending on their experience in the past. The model is estimated by Bayesian method, where posterior distribution of the parameters is derived – though random-walk Metropolis-Hastings algorithm – from the prior assumptions and rational expectations.[4]

[3] Benk, Jakab, Kovács, Párkányi, Reppa, and Vadas, 2006

[4] Precise description of this model and the results from forecasting can be found in: An estimated DSGE model of the Hungarian economy, Jakab, and Világi, 2008

3. BVAR model

3.1 Bayesian estimation

The general form of structural VAR models is

(2) $$A(L)y(t) = \varepsilon(t),$$

where *A(L)*, *m x m* matrix, contains all the coefficients and lag operators, $y(t)$ and $\varepsilon(t)$ are *m x 1* vectors with the endogenous variables and the error terms respectively at time *t*. m is the number of equations in the model. Let us denote the contemporaneous coefficient matrix with *A(0)*. The disturbance term is assumed to be normally distributed with zero expected value and identity covariance matrix:

$$E\left[\varepsilon(t)\varepsilon(t)' \middle| y(t-s),\ s>0\right] = I \text{ and } E\left[\varepsilon(t) \middle| y(t-s),\ s>0\right] = 0, \text{ for all } t.$$[5]

If we multiply both sides of (2) by $A(0)^{-1}$ we get

(3) $$y(t) = A(0)^{-1}A(L)y(t) + A(0)^{-1}\varepsilon(t),$$

so the distribution of the variables in the model will be

[5] Sims and Zha, 1998.

(4) $$y(t) \sim \mathrm{N}\left[A(0)^{-1}A(L)y(t), A^{-1}(0)A^{-1}(0)'\right].$$

From (4) we can derive the probability density function of the variables given the parameters and the data:

(5) $$p(y(t)|y(t-s),\ s>0; A(L)) = \sqrt{\frac{1}{2\pi}}|\mathrm{A}(0)|\exp\left[-\frac{1}{2}(A(L)y(t))'(A(L)y(t))\right], \text{ for all } t.$$

Those parts of this equation that do not contain relevant information about the parameters of the model can be omitted, so rewriting (5) gives

$$p(y(t)|y(t-s),\ s>0; A(L)) \propto |\mathrm{A}(0)|\exp\left[-\frac{1}{2}(A(L)y(t))'(A(L)y(t))\right],$$

where $\propto$ "stands for proportional to". The joint density function of all observations for all *t=1…T*:

(6) $$p(y(t)|y(t-s),\ s>0, t=1...T; A(L)) \propto |\mathrm{A}(0)|^{T}\exp\left[-\frac{1}{2}\sum_{t=1}^{T}(A(L)y(t))'(A(L)y(t))\right],$$

which is obviously gives us the likelihood function of the variables. The Bayesian inference from this expression is not obvious at all, but some simplifications can be introduced. Assuming that the correlation between the contemporaneous residuals is zero, allows us to break the system into *m* equations that can be estimated consistently by OLS (Zellner, 1962). To take advantage of this division redefine (4) for every equation

(7) $$Y_i = X_i \beta_i + u_i \text{ , i = 1 … m,}$$

where Y and u are $T\ x\ m$, X is $T\ x\ p$, β is $p\ x\ m$. So (6) can be further simplified[6] to

$$L(y(t),\, t = 0...T \mid A(L)) \propto \prod_i^m \left|\mathrm{A_{ii}}(0)\right|^T \exp\left[-\frac{1}{2} trace\big((Y_i - X_i\beta_i)'(Y_i - X_i\beta_i)A_{ii}(0)'A_{ii}(0)\big)\right].$$

Maximum likelihood is equivalent to the OLS, since the system can be estimated equation by equation, which simplifies the estimation.

(8) $$L(y(t),\, t = 0...T \mid A(L)) \propto \prod_i^m \left|\mathrm{A_{ii}}(0)\right|^T \exp\left[-\frac{1}{2} trace\Big((Y_i - X_i\hat{\beta}_i)'(Y_i - X_i\hat{\beta}_i)A_{ii}(0)'A_{ii}(0)\Big)\right].$$

To get the posterior distribution we need to define the prior distribution of the parameters of the model

(9) $$p(A_{ii}(0), \beta) = p(A_{ii}(0))\varphi\Big(vec(\hat{\beta}); \big(A_{ii}(0)'A_{ii}(0)\big)^{-1} \otimes \big(X_i^{'}X_i\big)^{-1}\Big),$$

where $\varphi\big(\beta; H(A_{ii}(0)\big)$[7] is the conditional distribution of β given the contemporaneous parameters. Combining (8) and (9), plugging them back to (1) we get the posterior distribution of the parameters given the data

[6] Zha, 1998.
[7] After Zha (1998) I assumed that the parameters in this framework are normally distributed.

$$(10) \qquad p(A_{ii}(0))\left|H\left(A_{ii}(0)\right)\right|^{-1/2}\left|\mathrm{A}_{\mathrm{ii}}(0)\right|^{T}\exp\left[-\frac{1}{2}trace\left(\hat{u}_i^{'}\hat{u}_i A_{ii}(0)' A_{ii}(0) - \hat{\beta}' H\left(A_{ii}(0)\right)^{-1}\hat{\beta}\right)\right],$$

where $\hat{u}_i$ is the residual from (7). The maximum of (10) gives the Maximum Likelihood estimates.

In SUR type models the parameters can be estimated equation by equation, so $\hat{\beta}_i s$ can be obtained from restricted OLS estimates as

$$(11) \qquad \hat{\beta}_i = \left(X_i^{'}X_i + iR'R\right)^{-1}\left(X_i^{'}Y_i + iR'r\right),$$

in this way our prior beliefs and the information from the data can be incorporated. This is true if R does not include cross-equation constraints, so it can be partitioned to R_is. The restrictions in (11) have the following form:

$$R_i\beta_i = r_i + v_i, \qquad v_i \sim N(0, \lambda^2 I).$$

In R and λ we can specify our initial beliefs or assumptions about the data generating process. These restrictions are going to be the same in case of every equation, so i can be omitted.

3.2 Specifying priors

Litterman (1986) pointed out that the macroeconomic variables carries several properties of the random walk process, so in his framework he set the priors as the variables could be described as

$$y_{it} = y_{it-1} + \eta_{it}$$

for every endogenous variables. The mean of the parameter of the first lag is set to one, while that of the further lags to zero, for all *i*s. Intuitively the variance of the parameters of the lagged variables is decreasing with the distance from *t*. Litterman introduced restrictions on the reduced form coefficients of the covariance matrix. The prior standard deviation of β_i for parameters of own lags is

(12) $$\frac{\lambda_1}{\sigma_j},$$

for parameters of other variables is

(13) $$\frac{\lambda_1 \lambda_2}{\sigma_j p^{\lambda_3}}.$$

where λ_1 is responsible for the overall tightness of beliefs on A(0); λ_2 is the relative tightness of the model; λ_3 is responsible for decreasing the variance if we increase the lag length. In (12) the σ s are scale factors in each equation, and *p* is the lag length. These are obtained from univariate autoregression on each variable. Litterman (1986), instead of this scale factors, used $\frac{\sigma_i}{\sigma_j}$ but in this way the disturbances in the structural equation are normalized to one (Sims and Zha, 1998).

The true values of the scale factors are not known, so we have to estimate them. Since we assumed that we can estimate the equations equation by equation it is easy to calculate the residuals from OLS models.

4. Forecasting

As it was mentioned before, the application of VAR models in forecasting simultaneous systems has a vast literature; at the same time it is criticized because of the lack of theoretical background. Along the contribution of numerous authors the BVAR models overcame this problem since they incorporate our prior knowledge in the estimation. This may come from intuition or theory; it helps us to outperform the unrestricted VAR model in terms of forecasting.

4.1 The scheme

To get two year-ahead forecast for the inflation I will estimate the fitted values recursively: if the last observation in the sample is at time t and the forecast horizon is h, then the first estimation for period $t+1$, then for $t+2$ and so on until $t+h$. This means that all together for each model there will be 24 forecasts for each model. The next step after the forecast is to evaluate the results, in applied econometrics the most used tools for this are the Root Mean Squared Error (RMSE) and Mean absolute error (MAE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\ ,$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|.$$

Both loss functions measure the average magnitude of the error, but the RMSE punishes more the larger deviations from the true value, while the in the MAE each difference has equal weight, but the direction does not matter in either case. It is worth using them together even the concept is similar, but the RMSE is always greater equal to the MAE. The difference between them has also implications: greater discrepancy between them means higher variance in the forecast around the true value.

All this will be applied to compare the results of the projections of an ARIMA, a multivariate VAR, and a BVAR model for five different horizons, I will calculate these loss functions for one month, one quarter, six months, one year, and two years.

4.2 The variables

The future level of inflation does matter for every agent in the economy from the individuals to the firms; the uncertainty about the price changes causes uncertainty in the future payouts, the profits. To diminish this source of global ambiguity Central Banks have a major role in forecasting inflation among many macroeconomic variables. They are not like fortune tellers, but they can make inference from the past. On the one hand there are, as in any other science, laws in the economy that take effect under certain circumstances; on the other hand this can be supported by measured variables in the recent past. The first gives the theory, the second the evidence and the combination of the two can help to make inferences about the future. Since in macroeconomics mainly time series are used the past carries relevant information about the future. The whole idea is about capturing the key features, the underlying mechanisms that play major role of the examined process.

Inflation is probably one of the most endogenous variables, so it is not hard to find variables which explains its variation. There are several groups of variables that can be good indicators of the change of inflation. It can be grasped through prices (domestic, external), demand and financial variables (Kenny, Meyler and Quinn, 1998).

In this work I will use five variables to forecast inflation, four of them are price related and one comes from domestic demand. I chose variables in the analysis such that only the lagged values of them affect consumer prices; this is essential because, on the one hand, I assumed before that there is no contemporaneous effect present between the variables; on the other hand there is not too much theory based to explain the inflation with its components. Although the picture is not that clear at all, it would be also interesting to see how sectoral prices influence the inflation in a dynamic system. Let us consider that fuel prices affects every single sector of the economy, inflation in manufacturing will soon or later appear in the service sector, the wage level affects the demand for goods and it is clearly not independent from the food and service prices. But all these do not happen at the same time; for example an exchange rate shock appears earlier in the tradeable goods sector than in the non-tradeables. But at the end of the day, with these variables we could not really capture any underlying process that would determine the whole system.

I wanted to choose variables that capture the substantial directions of changes in the variables. I included the consumption and the wage as demand side variables; the unit labor cost and the fuel price as supply side variables; and the exchange rate that affects both the consumer and the producer prices simultaneously. These variables are probably able to capture better the data generating process, and provide a more accurate projection.

4.3 The data

For the price variables data was available for monthly frequency, but the nominal consumption is only reported for quarters. To overcome this mismatch there are two options: one is to make the monthly data quarterly, and the other is the opposite. I approximated the monthly values of the consumption with an algorithm that minimized the squared sum of the change from month to month in a way that the overall consumption did not chang in any quarter.[8] The used sample ranges from January 1998 to March 2009, i.e. an eleven year period with 134 observations. The price series are represented as levels, the exchange rate is the average monthly price of one Euro expressed in Hungarian Forint. I chose this period because I did not have observations for the consumption, and the *cpi* was volatile in the beginning and the middle of the 90s; I assumed that the last eleven years are more relevant than the previous times.

The Dickey-Fuller test showed that all the variables – except for the *cpi* and the consumption – are first order integrated. Surprisingly the inflation with Dickey-Fuller[9] and Philips-Perron tests proved to be stationary, but if one glances at the series it is obvious that it can not be. After taking the first difference it much more resembled a zero mean stationary process. The consumption was second order integrated.

Though it is not relevant in this paper, the first difference of the consumption and the consumer price index proved to be cointegrated for the examined period. In every series – except for the exchange rate, the unit labor cost and the fuel price – seasonality is present, so it was necessary to treat this; otherwise we would not be able to make good inferences and the Minnesota prior can not be applied to seasonal data either. I used Tramo/Seats seasonal

[8] Datasource: http://mnb.hu/engine.aspx?page=mnbhu_statisztikai_idosorok
[9] See in appendix: Table A.

adjustment to get the series with the desired properties. Finally, to get economically meaningful and interpretable results, I took the logarithm of the series.

To conduct an out of sample forecast we need to split up the data to sample period and forecast horizon. In this case the latter runs from January 1998 to March 2007, and the forecast interval begins in April 2007 and finishes in March 2009, which is all together 24 months. As a base of the comparison I compared the precision of the four models for five different horizons: one month, one quarter, six months, one year, and two years.

4.4 The models

4.4.1 ARIMA

As a benchmark, a simple ARIMA will be used since every zero mean covariance stationary stochastic process can be represented by an ARMA model; it can be expressed as the combination of a deterministic autoregressive and a stochastic moving average process (Greene, 2002). This seems to be ad hoc specification, but this model is favored in forecasting, because it captures very well the dynamics of the macroeconomic time series, at least in the short run.

As it turned out in the data analysis inflation is not stationary, but first order integrated, it is necessary to work with the first differences. The most appropriate specification is

$$\Delta cpi_t = 0.35\Delta cpi_{t-1} + 0.16\Delta cpi_{t-3} + 0.22\Delta cpi_{t-4} - 0.13\Delta cpi_{t-7} + 0.22\Delta cpi_{t-11}$$ [10],

with these variables the residuals are not serially correlated, the sample contains 99 observations and the $R^2 = 0.24$.

4.4.2 VAR

Applying the same notation as in (3) results in

$$y(t) = B(L)y(t) + u(t),$$

which is the reduced form of (2), where $B(L) = A^{-1}(0)A(L)$ and $u(t) = A^{-1}(0)\varepsilon(t)$. According to the Akaike, Schwarz, and Hannan-Quinn information criteria the optimal lag length is 12.[11] All together 84 parameters have to be estimated that is way too many considering the low number of observations. Since we assumed no contemporaneous effects in the model we could estimate this equation by equation in the SUR framework.

4.4.3 Bayesian VAR

To be able to compare the forecast precision of this model with that of the VAR, I used the same lag length in this framework. But here I have to first compress all prior belief into the hyperparameters. As Litterman (1986) applied several priors in his work, first I will

[10] See in appendix: Table B.
[11] See in appendix: Table C.

show the outcome from some of these. Obviously we can not exclude the data when we form our prior belief from the data, since in (12) and (13) we directly use the residuals from models that we estimated using the data. Litterman (1986) showed if we use the Minnesota prior with $\lambda_1 = \lambda_2 = 0.5$ and $\lambda_3 = 1$, we can outperform the unrestricted was. Actually this setup was the loosest in his analysis, so in the first step I will apply these in my model to compare them with the alternative models.

4.5 *The results*

After all estimation that are described above I forecasted the future level of inflation in two years horizon. The outcomes, the RMSEs and the MAEs, can be seen in Table 4-1 and in Table 4-2. In both cases the BVAR model provided better projection than the other two, which coincides with the results of Litterman (1986), but this is the less surprising in these tables. If one compares only the results from the AR and VAR model, the results are astonishing, because the VAR performs very poorly compared to the univariate autoregression. Considering that the AR process is nested in the VAR, intuition might say that it should perform worse, but as the tables show this not the case.

Table 4-1: Root mean squared errors of forecasts

	RMSE				
	one month	one quarter	six months	one year	two years
AR	0.003	0.009	0.015	0.034	0.054
VAR	0.015	0.017	0.018	0.061	0.129
BVAR	0.001	0.007	0.011	0.032	0.061

The results from the RMSE analysis show that in the short run, up to one year, the BVAR model gave the most accurate projection, but for longer horizon the ARIMA proved to be at least as good. At the same time the VAR's RMSEs are mostly twice as large as the other two's. This can be attributed to the very low number of degrees of freedom, since we had to, in this case, estimate 84 parameters from 134 observations.

Table 4-2: Mean absolute errors of forecasts

	MAE				
	one month	one quarter	six months	one year	two years
AR	0.003	0.008	0.013	0.029	0.049
VAR	0.015	0.016	0.016	0.037	0.095
BVAR	0.001	0.006	0.010	0.026	0.053

We can conclude the same from Table 4-2, but the difference is that the VAR does not seem as bad as before, implying that the variance of the forecasted values around the true series is larger than in the other two alternative methods.

4.6 Further priors

Kenny, Meyler, and Quinn (1998) forecasted Irish inflation with BVAR models. They examined several cases of the estimation and finally[12] for a model where the wage and some demand side variables are included they suggested the following priors: $\lambda_1 = \lambda_2 = 0.3$ and $\lambda_3 = 1$. Their reasoning was the following: between the wage and the inflation there is no

[12] For a simple three variable (domestic HICP, foreign HICP, and exchange rate) model they suggested $\lambda_1 = 0.3$, $\lambda_2 = 0.8$ and $\lambda_3 = 1$.

strong structural relationship, but it still can contain important inference about the future realization of the consumer prices. Not surprisingly the results[13] are very much the same as before with the difference that the RMSEs became larger in case of the one year and two years horizon.

Among other macro variables Litterman (1986) specified priors ranging from 0.1 to 0.5 and showed that for some variables the tighter the prior the more accurate the forecast. This was not true either for the inflation; in his work tightening the variance of the prior variance of the inflation did not improve significantly the results.

4.6.1 Sensitivity analysis

The initial setup of the model depends on how we judge the behavior of the variables in the future. We may allow for larger variance in the long run or *vice versa:* it depends on the perception of the underlying processes of the analyst. The forecast horizon of interest may differ also, for example we might use different hyperparameters for one month ahead and one year ahead forecasts. As we can see in Table 4-3, there can be significant differences among priors.[14]

[13] See in appendix: Table D-E.
[14] For the MAEs see appendix: Table F.

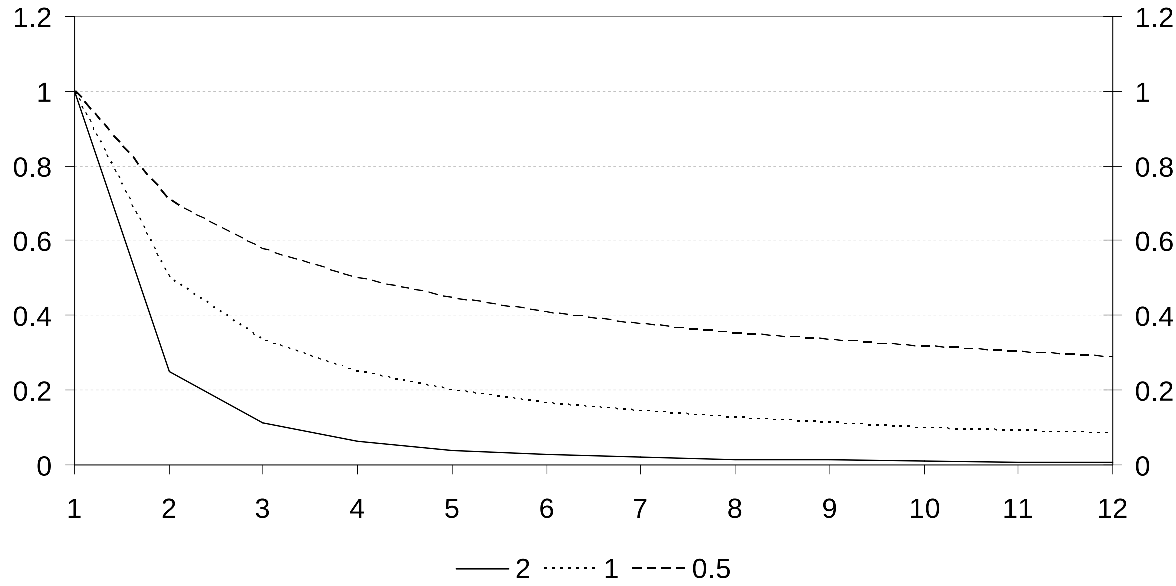
Table 4-3: RMSEs for different priors

Priors	RMSE					
	one month	one quarter	six months	one year	two years	average
1	0.003	0.006	0.021	0.029	0.033	0.018
0.75	0.001	0.002	0.004	0.010	0.036	0.011
0.5	0.001	0.007	0.011	0.032	0.061	0.022
0.4	0.001	0.007	0.014	0.038	0.067	0.026
0.3	0.001	0.007	0.015	0.039	0.067	0.026
0.2	0.001	0.005	0.012	0.033	0.057	0.022
0.1	0.000	0.002	0.005	0.017	0.027	0.010

It is clear that 0.75 and 0.1 priors fit the best for the forecast horizon; they are equally good if every period has equal weight in the objective function, but this is not necessarily the case. For short run horizon, (up to one year) it is more beneficial to use 0.75, but for one month or two years 0.1 gives more accurate projection. This may be true for this specific sample, but it is uncertain how the results would change if the sample would be shorter or longer. This work could be extended with an optimization that finds the optimal priors that fits the best for all possible sample sizes.

The third hyperparameter that has not been changed yet, λ_3, is responsible for the speed at which the variance of the lagged parameters are diminishing as they get further from time *t*. If we assume that the past less relevant we can set this parameter greater or otherwise smaller. This feature of the time series may differ over time, for example the exchange rate pass through depends on the volatility or the magnitude of the change. Figure 4-1 shows how the rate of decay change as λ_3 varies.

Figure 4-1 The rate of decay for lagged variables



If we recall (13), this figure tells us that our assumption about the variances can be crucial also. To demonstrate this I recalculated the loss functions for all these hyperparameters. The results[15] changed the optimal prior in both cases. First I used as a lag decay $\lambda_3 = 2$, in this case the most precise forecast materialize if $\lambda_1 = \lambda_2 = 0.1$ and $\lambda_1 = \lambda_2 = 0.2$. While when $\lambda_3 = 0.5$, so we do attribute a greater role for lagged values of the explanatory variables, $\lambda_1 = \lambda_2 = 0.5$ hyperparameters outperformed all the others. Here also true that this is not a general solution, this can be the case for only this sample it may alternate for shorter or longer ones.

As we have seen, we get less and more precise forecast for each period depending on how we set the hyperparameters in the beginning. If we focus on only one period, then it is simple: we set the priors such that they give the most accurate prediction. If more periods are taken into consideration then we have to assign weights to the loss function of each horizon. The optimal hyperparameters can be found through an iterative procedure. One has to take the

[15] See appendix: Table G-J.

shortest but still reasonable sample and extend it gradually up to the t (if we forecast for h horizon, and we have $t+h$ observations); and forecast for h period ahead. Applying an appropriate algorithm (for example gradient method) the optimal hyperparameters can be estimated. But this strategy might raise some questions: in this way the specification is getting more *ad hoc*, while the main advantage of the Bayesian econometrics is that it combines our prior knowledge or intuition about the interaction between variables and not choosing the best fitting model. The analyst should stick to the theory more, since the hyperparameters that arise through this optimization may work well on the forecasting horizon but there is no guarantee that the behavior of the variables do not change over time. And it is hard to explain why exactly those parameters were chosen; if the forecasting procedure is not connected to theory it is condemned to fail soon or later.

If we have strong beliefs according to which we set the hyperparameters, but it turns out that there exist alternative priors that can perform much better in any circumstances, then we should revise our prior beliefs. In the Bayesian framework, the outcome can be considered as a weighted average of the results from the data and our prior beliefs, but still if we do not specify our priors carefully alternative methods can perform better than this complicated model.

5. Conclusions

It is no wonder that the Bayesian statistic based modeling is gaining more space in applied econometrics; applying theory, intuition or both throughout the analysis besides the data makes the concept very appealing. As I showed in this thesis, the Bayesian VAR model outperforms the ARIMA and VAR models; the *root mean squared errors* and the *mean absolute errors* clearly show that these models perform poorly compared to the BVAR. But this is only one reason why this framework should be given credit; it also solves the very often cited shortcoming of the Vector Autoregression models: the number of the parameters to be estimated is only a small fraction of VAR model's. In this particular example, when the sample period included only 135 observations and 84 parameters, not surprisingly the variance of the estimated parameters was large, so this model could not provide useful information for the future value of the inflation. While in some cases even an ARIMA model captured very well the dynamics of the time series and performed as good as the Bayesian, the VAR model was always lagging behind. The problem with this *ad hoc* model is that we can not really implant our knowledge in this framework, so for more sophisticated forecasts it might be hard to back up our findings by theory in a model where only the lagged variables examined time series are included.

I applied the Minnesota prior, after Litterman (1980), which introduces very simple assumptions about the behavior of macroeconomic variables: they follow random walk and the parameters of lagged variables have zero expected value and their variance die out with higher lags. Even this setup was good enough to forecast the future better than the ARIMA and the VAR. This framework has already been extended to a more sophisticated method, with the DSGE framework; Del Negro and Schorfheide (2004) showed that there are alternative models that beats the Minnesota prior in terms of forecasting.

Bayesian econometrics is becoming a competitive alternative framework. In my case Bayesian VAR was estimated that served as a solution for the mentioned overfitting problem of the unrestricted VAR model, moreover I could incorporate some prior beliefs about the behavior of the analyzed variables.

Bayesian method provides a very flexible framework, which can easily be made stricter or looser depending how much the priors are credible. It is not always easy but necessary to incorporate as much prior knowledge as possible in the model. This thesis shows that even parsimonious assumptions about the behavior of macroeconomic variables can bring us closer to the true nature of the underlying processes.

Reference

Benk, Szilárd, Jakab, Zoltán, Kovács M. András, Párkányi Balázs, Reppa, Zoltán, and Vadas, Gábor. 2006. "The Hungarian Quarterly Projection Model (NEM)" MNB Occasional Papers 2006/60, Magyar Nemzeti Bank (The Central Bank of Hungary)

Bradley, P. & Louis, T. 1996. "Bayes and Empirical Bayes Methods for Data Analysis", London: Chapman & Hall.

Canova, Fabio.2007. *Methods for applied macroeconomic research*. Princeton University Press. Chapters 9-11.

Ciccarelli Matteo and Rebucci, Alessandro. 2003. "Bayesian VARs: A Survey of the Recent Literature with an Application to the European Monetary System", IMF Working Papers 03/102, International Monetary Fund.

Del Negro, Mand and Schorfheide Frank. 2004. „Priors from equilibrium models for VARs", International Economic Review, Vol. 45, pp 643-673.

Doan, Thomas, Robert Litterman, and Christopher Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions", Econometric Reviews, Vol. 3, 1-100.

Geweke, John and Whiteman, Charles. 2006. *Bayesian Forecasting*, Handbook of Economic Forecasting Volume 1, Pages 3-80.

Greene, H. William. 2002. *Econometric Analysis*, 5th edition

Ingram, Beth F. and Whiteman, Charles. 1994. "Supplanting the 'Minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors", Journal of Monetary Economics, Elsevier, vol. 34(3), pages 497-510, December

Jakab, M. Zoltán, and Világi, Balázs. 2008. "An estimated DSGE model of the Hungarian economy", MNB working papers, 2008/9 Magyar Nemzeti Bank (The Central Bank of Hungary)

Kenny, Geoff, Meyler, Aidan and Quinn, Terry. 1998. "Bayesian VAR Models for Forecasting Irish Inflation", Central Bank of Ireland and Financial Services Authority of Ireland

Koop, Gary. 2003. *Bayesian Econometrics*, John Wiley and Sons, New York.

Leamer, Edward, E. 1978. "Specification searches: ad hoc inference with nonexperimental data", Wiley

Lieli, Róbert. 1999. "Az idősormodelleken alapuló inflációs előrejelző modellek: Egyváltozós módszerek" MNB Füzetek, 1999/4

Litterman, B. Robert. 1980. “A Bayesian Procedure for Forecasting with Vector Autoregressions”, Federal Reserve Bank of Minneapolis

Litterman, B. Robert, 1986. “Forecasting With Bayesian Vector Autoregressions -- Five Years of Experience”, Journal of Business & Economic Statistics 4, 25-38.

Stigler, M. Stephen.1986. *The history of statistics* Harvard University press. Chapter 3.

Várpalotai, Viktor. 2008: “Modern Bayes-i ökonometrián alapuló idősormodellek és empirikus elemzések (Simasági priorok alkalmazása az üzleti ciklusok szinkronizációjának mérésére és az infláció előrejelzésére)”, Phd thesis, Corvinus University Budapest.

Zellner, Arnold. 1962. “An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias”, Journal of the American Statistical Association 57: 348–368.

Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics*, J. Wiley and Sons, Inc., New York

Zha, Tao. 1998. “Block Recursion and Structural Vector Autoregressions”, Federal Reserve Bank of Atlanta, GA 30303

Appendix

Table A: Result of the unit root test of the cpi

Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(CPI)				
Method: Least Squares				
Sample (adjusted): 1998M03 2007M03				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
CPI(-1)	-0.005718	0.002073	-2.758235	0.0068
D(CPI(-1))	0.322202	0.090981	3.541405	0.0006
C	0.038999	0.012986	3.003228	0.0033
R-squared	0.233202			

Table B: ARIMA model for the differenced inflation

Dependent Variable: D_CPI				
	Coefficient	Std. Error	t-Statistic	Prob.
D_CPI(-1)	0.351990	0.092997	3.784955	0.0003
D_CPI(-3)	0.157959	0.095435	1.655150	0.1012
D_CPI(-4)	0.218556	0.097016	2.252784	0.0266
D_CPI(-7)	-0.127176	0.116512	-1.091529	0.2778
D_CPI(-11)	0.223440	0.088012	2.538746	0.0128
R-squared	0.238025			

Table C: Lag order selection

VAR Lag Order Selection Criteria						
Endogenous variables: CONS CPI EXCH FUEL ULC_M ULC_S WAGE						
Included observations: 99						
Lag	LogL	LR	FPE	AIC	SC	HQ
0	1387.245	NA	1.83e-21	-27.88375	-27.70025	-27.80950
1	2484.474	2017.128	1.17e-30	-49.06009	-47.59214	-48.46615
2	2659.595	297.1738	9.26e-32	-51.60797	-48.85557	-50.49435
3	2917.683	401.4707	1.41e-33	-55.83198	-51.79513	-54.19866
4	3038.279	170.5394	3.55e-34	-57.27836	-51.95705	-55.12535
5	3084.343	58.62797	4.25e-34	-57.21906	-50.61330	-54.54636
6	3229.312	164.0047	7.39e-35	-59.15782	-51.26760	-55.96542
7	3359.639	129.0109	1.89e-35	-60.80079	-51.62613	-57.08871
8	3435.848	64.66223	1.65e-35	-61.35047	-50.89135	-57.11870
9	3619.210	129.6494	1.98e-36	-64.06484	-52.32127	-59.31337
10	3838.833	124.2316	1.50e-37	-67.51179	-54.48376	-62.24062
11	4068.876	97.59368	1.43e-38	-71.16920	-56.85673	-65.37835
12	4365.950	84.02115*	7.91e-40*	-76.18082*	-60.58388*	-69.87027*

* indicates lag order selected by the criterion

Table D: RMSEs for $\lambda_1 = \lambda_2 = 0.3$ and $\lambda_3 = 1$ priors

	RMSE				
	one month	one quarter	six months	one year	two years
AR	0.003	0.009	0.015	0.034	0.054
VAR	0.015	0.017	0.018	0.061	0.129
BVAR	0.001	0.007	0.015	0.039	0.067

Table E. MAEs for $\lambda_1 = \lambda_2 = 0.3$ and $\lambda_3 = 1$ priors

	MAE				
	one month	one quarter	six months	one year	two years
AR	0.003	0.008	0.013	0.029	0.049
VAR	0.015	0.016	0.016	0.037	0.095
BVAR	0.001	0.006	0.010	0.026	0.053

Table F: MAEs for $\lambda_3 = 1$

Priors	MAE					
	one month	one quarter	six months	one year	two years	average
1	0.003	0.005	0.017	0.026	0.031	0.016
0.75	0.001	0.001	0.003	0.007	0.028	0.008
0.5	0.001	0.006	0.010	0.026	0.053	0.019
0.4	0.001	0.006	0.012	0.031	0.059	0.022
0.3	0.001	0.007	0.015	0.039	0.067	0.026
0.2	0.001	0.004	0.010	0.027	0.050	0.019
0.1	0.000	0.002	0.004	0.013	0.024	0.009

Table G: RMSEs for $\lambda_3 = 2$

Priors	RMSE					
	one month	one quarter	six months	one year	two years	average
1	0.003	0.010	0.017	0.036	0.062	0.026
0.75	0.002	0.009	0.016	0.036	0.064	0.025
0.5	0.001	0.006	0.012	0.033	0.060	0.022
0.4	0.001	0.005	0.010	0.029	0.055	0.020
0.3	0.000	0.003	0.007	0.024	0.046	0.016
0.2	0.000	0.002	0.004	0.017	0.032	0.011
0.1	0.000	0.002	0.004	0.014	0.020	0.008

Table H: MAEs for $\lambda_3 = 2$

Priors	MAE					
	one month	one quarter	six months	one year	two years	average
1	0.003	0.009	0.015	0.031	0.055	0.023
0.75	0.002	0.008	0.014	0.031	0.057	0.022
0.5	0.001	0.005	0.010	0.027	0.052	0.019
0.4	0.001	0.004	0.008	0.024	0.047	0.017
0.3	0.000	0.003	0.006	0.019	0.040	0.014
0.2	0.000	0.001	0.003	0.013	0.027	0.009
0.1	0.000	0.001	0.003	0.011	0.017	0.007

Table I: RMSEs for $\lambda_3 = 0.5$

Priors	RMSE					
	one month	one quarter	six months	one year	two years	average
1	0.005	0.020	0.055	0.105	0.341	0.105
0.75	0.002	0.009	0.027	0.037	0.057	0.026
0.5	0.000	0.002	0.003	0.014	0.037	0.011
0.4	0.000	0.004	0.006	0.026	0.049	0.017
0.3	0.000	0.005	0.010	0.033	0.057	0.021
0.2	0.000	0.005	0.012	0.033	0.055	0.021
0.1	0.000	0.003	0.008	0.020	0.032	0.013

Table J: MAEs for $\lambda_3 = 0.5$

Priors	MAE					
	one month	one quarter	six months	one year	two years	average
1	0.005	0.016	0.045	0.090	0.265	0.084
0.75	0.002	0.007	0.022	0.033	0.051	0.023
0.5	0.000	0.001	0.002	0.010	0.029	0.009
0.4	0.000	0.003	0.005	0.019	0.042	0.014
0.3	0.000	0.004	0.008	0.027	0.050	0.018
0.2	0.000	0.004	0.009	0.027	0.049	0.018
0.1	0.000	0.003	0.006	0.017	0.029	0.011