Charles University in Prague
Faculty of Social Sciences

RIGOROUS THESIS

Josef Stráský

Can Bayesian econometric methods outperform traditional econometrics in inflation forecasting?

Institute of Economic Studies

Supervisor: PhDr. Jaromír Baxa

Study Program: Economic Theory

2011

Acknowledgement

I would like to thank my supervisor Jaromír Baxa for the idea to write diploma thesis on the topic of Bayesian econometrics applied on inflation forecasting. I would also like to thank him for sharing his experience, leading me throughout the entire work and for giving me enough freedom to complete my thesis at my own pace. I would like to thank especially for his infinite patience during the completion of the thesis.

I would also like to thank Prof. Miloslav Vošvrda for letting me present the results of the diploma thesis on the doctoral seminar at Institute of Information Theory and Automation (Academy of Sciences of the Czech Republic).

The author gratefully acknowledges Grant Agency of Charles University (GAUK) for supporting the research on Bayesian econometrics and Monetary policy through the grant no. 91110.

Prohlášení

Prohlašuji, že jsem svou rigorózní práci vypracoval samostatně a použil jsem pouze podklady uvedené v přiloženém seznamu.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 14.2.2011 .....

Introduction to rigorous thesis

This rigorous thesis is based on my diploma thesis that has been defended in June 2010 and has been awarded by the dean of the Faculty of Social Sciences. The opponent's report on this master thesis is attached. The report is overall very positive and includes only two reproofs. The first is poor final form of the manuscript, mainly above-average amount of typos and misspellings. The other suggestion is to cite more references concerning inflation forecasting in Czech republic.

The manuscript has been carefully read through and plenty of typos have been corrected. Moreover some parts, mainly in theoretical chapters, have been reformulated to ensure their better readability.

Six related references have been added to Literature review section in order to meet opponent's objection.

Abstract

Forecasting of inflation has become crucial for both policy makers and private agents who try to understand and react to Central Bank decisions because many Central Banks implemented inflation targeting rules instead of control of monetary aggregates. Inflation forecasting is considered to be very complicated issue because univariate regression models and structural macroeconomic models are usually outperformed by naive random walk model. This work is intended for forecasting inflation in the Czech Republic by employing Bayesian econometric method (namely Bayesian vector autoregression - BVAR). Bayesian methods proved to be useful in inflation forecasting in developed countries (Fabio Canova: G-7 Inflation Forecasts: Random Walk, Phillips Curve or What Else?, 2007 [1]).

Bayesian econometrics is one of the fast developing fields of econometrics for past two decades. In the centre of the approach is Bayesian probabilistic theory based on conditional probabilities. This probabilistic approach is, however, computationally demanding. Fast computer evolution enables wide applications of Bayesian models. Model estimations are based on combining information from some prior beliefs and from the data. Many different sorts of models have their Bayesian variants (e.g. OLS) but the emphasis in this work is on Bayesian Vector autoregression (BVAR). One of the aims of the thesis is to become familiar with principles of Bayesian econometric and be able to use Bayesian approach in various models.

In this thesis, I compared the forecasting performance of various models by applying the Theil U-statistics. Since VAR models were able to outperform Random Walk in pseudo out-of-sample forecasts, I undertook an experiment with the aim to identify the best inflation predictors, that should be included within the VAR model. For this purpose I employed a set of almost 80 time series covering various economic indicators including forward looking variables extracted from surveys.

I have found that unemployment is never in the set of best predictors (rejection of Phillips curve as useful relationship), GDP measure appears only in the long term forecast, whereas forward looking indicators are important for shorter forecast horizons. Employing of BVAR models instead of VAR have brought mixed results. Out of sample predictions for years 2010 and 2011 are also provided. Variants of future research are briefly discussed.

Abstrakt

Vzhledem k tomu, že mnoho centrálních bank opustilo režim cílování peněžní zásoby a přešlo k režimu cílování inflace, stalo se předpovídání inflace zásadním, jak pro politické rozhodování, tak pro soukromé aktéry, kteří se snaží rozpoznat rozhodnutí centrální banky a reagovat na něj. Předpovídání inflace není jednoduché, neboť modely zahrnující jednu proměnnou, stejně jako strukturální makroekonomické modely, jsou, pokud se týká schopnosti předpovědi, překonávány naivní předpovědí, která je výsledkem modelu náhodné procházky. Cílem této práce je předpovídat inflaci v České republice použitím bayesovského ekonometrického modelu (konkrétně bayesovské vektorové autoregrese - BVAR). Bayesovské modely se osvědčily při předpovídání inflace ve vyspělých zemích (Fabio Canova: G-7 Inflation Forecasts: Random Walk, Phillips Curve or What Else?, 2007 [1]).

Bayesovská ekonometrie je jednou z nejrychleji se rozvíjejících oblastí ekonometrie za poslední dvě desetiletí. Ve centru bayesovského přístupu leží bayesovská pravděpodobnostní teorie, která je založena na konceptu podmíněné pravděpodobnosti. Tento přístup je ovšem výpočetně poměrně náročný a jeho praktické využití bylo umožněno teprve díky rychlému rozvoji výpočetní techniky. Odhady modelů jsou založené na kombinaci určitých předpokladů (priors) spolu s informací pocházející z naměřených dat. Většina ekonometrických modelů má dnes i své bayesovské varianty (včetně metody nejmenších čtverců), ovšem v této práci je využívána především bayesovská vektorová autoregrese (BVAR). Jedním z cílů této diplomové práce je seznámit se s principy bayesovské ekonometrie a osvojit si používání bayesovského přístupu v různých modelech.

V diplomové práci jsem porovnal předpovědi různých modelů pomocí Theilovy U-statistiky. Vzhledem k tomu, že modely typu VAR předčily model založený na náhodné procházce, jsem provedl experiment, jehož cílem bylo určit nejlepší prediktory pro předpověď inflace, které by měly být zahrnuty ve VAR modelech. Použil jsem proto rozsáhlý soubor čítající téměř 80 časových řad, které zachycují důležité ekonomické indikátory, včetně vpřed hledících indikátorů získaných na základě průzkumů.

Výsledkem hledání vhodných prediktorů je, že nezaměstnanost mezi ně nikdy nepatří (Phillipsova křivka se tudíž nepotvrdila jako využitelný vztah), statistiky zachycující HDP jsou použity pouze pro dlouhodobé předpovídání inflace, zatímco vpřed hledící indikátory jsou důležité pro krátkodobé předpovědi. Použití modelů typu BVAR místo VAR modelů přineslo smíšené výsledky. Uvedeny jsou též předpovědi inflace pro roky 2010 a 2011 a krátce jsou diskutovány možnosti dalšího výzkumu.

Contents

1 Introduction

Forecasting of inflation has become crucial for both policy makers and private agents who try to understand and react to Central Bank decisions because many Central Banks implemented inflation targeting rules instead of the control of monetary aggregates. Inflation forecasting is considered to be very complicated issue because univariate regression models and structural macroeconomic models are usually outperformed in inflation forecasting by naive Random walk model. This work is intended for forecasting inflation in the Czech Republic by employing Bayesian econometric method (namely Bayesian Vector autoregression - BVAR). Bayesian methods proved to be useful in inflation forecasting [1].

Bayesian econometrics is the developing field of econometrics for past two decades. Bayesian approach can be expressed by saying that anything that is unknown can be expressed by theory of probability. Variables in econometric models are considered as a realization of random variable in Bayesian econometrics.

Prior restrictions are imposed to the parameters of econometric model. So-called actualization principle is subsequently used (i.e. updating of the likelihood of a parameter from prior belief to posterior belief given the data). Many different classes of models have their Bayesian variants (e.g. OLS) but the emphasis in this work is on Bayesian vector autoregression (BVAR). Basic theoretical concepts of Bayesian econometrics and BVAR are described in the following chapter.

Bayesian methods are nowadays often used for inflation forecasting since other models (e.g. structural models or ARMA) are usually outperformed by naive Random walk forecast. Brief survey of literature can be found in the section 3.6.

In this thesis, I compared the forecasting performance of various models by applying the Theil U-statistics. Since VAR models were able to outperform Random walk in pseudo out-of-sample forecasts, I undertook an experiment with the aim to identify the best inflation predictors, that should be included within the VAR model. For this purpose I employed a set of almost 80 time series covering various economic indicators including forward looking variables extracted from surveys. Anticipating the results, I have found that unemployment is never in the set of best predictors (rejection of Phillips curve as useful relationship), GDP measure appears only in the long term forecast, whereas forward looking indicators are important for shorter forecast horizons. Employing of BVAR models instead of VAR brings mixed results.

Estimation procedure is based on Matlab environment including Econometric Toolbox by James P. LeSage [2]. The best inflation predictors of the time series are sorted out and VAR and BVAR forecast performances are compared. The contribution of BVAR models is discussed and the prospects for future work are outlined.

This thesis is organized as follows. Firstly Bayesian econometrics and BVAR concept are introduced. Introduction to inflation forecasting and literature review is provided in the next chapter. The estimation procedure is explained in detail in the chapter concerning the methodology. Chapter Results provides all the results including graphical output. Las two chapters discuss the results, raise the issues of future work and conclude.

2 Bayesian Econometrics

The most controversial issue about Bayesian Econometrics is that parameters of models are treated as random variables. We should deal with this controversy right in the beginning. The chief competitor to Bayesian econometrics is often called *frequentist econometrics*.[1] Frequentist econometricians say that parameters are not random but real and econometric estimation means trying to approach this real value of parameter. Here frequentists silently (or loudly) assume that there exist some 'true' model with real parameters, probably given by nature or God. I argue that there is only one precise model of reality and this is reality itself. But in any model we necessarily omit some relationships. And these omitted relationships can and do influence our model relations. Thus we have 'only' some estimate of parameter that is not real in any meaning. On the other hand Bayesian econometrics does not assume anything like 'true' parameters. It is based on a subjective view of probability, which argues that our uncertainty about anything unknown can be expressed using the rules of probability. On the other hand frequentists claim that probability can be defined only for repeating realizations of random process. Probability is thus frequency of given outcome. I consider this view being too restrictive and I would like to support the usefulness of Bayesian approach.

Following sections serves as brief but comprehensive introduction to Bayesian econometrics. This part follows Gary Koop's Bayesian Econometrics [3] and should serve as an overview of important concepts of Bayesian econometrics for interested readers. For understanding methodology and results of the thesis it might be particularly useful to read introductory section and section concerning Bayesian vector autoregression (BVAR). Readers familiar with the concept might skip whole chapter.

2.1 Introduction to Bayesian Theory

Main advantage of Bayesian approach is that Bayesian econometrics is based on a few simple rules of probability. Literally all the things we wish to do (estimate the parameters of the model, obtain prediction from the model or compare different models) involve the same, universal rules of probability.

Let us consider two random variables A and B. From the rules of

[1] This distinction comes from how probability can be regarded as. Frequentists consider probability to be the frequency of occurrence of some outcome of random event. On the other hand, form Bayesians' point of view, probability of some outcome captures all the information we have about the event.

probability:

$$p(A, B) = p(A|B)p(B) \tag{2.1}$$

where $p(A, B)$ is the *joint probability* of A and B occurring, $p(A|B)$ is the *conditional probability* of A given B, and $p(B)$ is the *marginal probability* of B. Now we can reverse the roles and rewrite the equation:

$$p(A, B) = p(B|A)p(A) \tag{2.2}$$

Putting these two equations together we get well-known *Bayes' rule*:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \tag{2.3}$$

In econometrics we typically work with models which depend upon parameters and we are usually interested in estimating these parameters (e.g. coefficients in regression model). Now, we move on in a little bit abstract manner. Let y be a vector or matrix of data and θ be a vector of matrix which contains the parameters of a model that tries to explain y. We have the data y and we are interested in estimating θ. We now use *Bayes' rule* and we replace B by θ and A by y in equation (2.3):

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{2.4}$$

We treat $p(\theta|y)$ as being of fundamental interest. It directly addresses the question what do we know about the parameters if we have got the data. Here we imply the treatment of θ as a random variable. Under this treatment the conditional probability of the unknown given the known is the best way of summarizing what we have learned. We are thus only interested in learning about θ so we reduce preceding equation since term $p(y)$ does not involve θ[2]:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{2.5}$$

The term $p(\theta|y)$ is referred to as the *posterior density*; the probability density function for the data given the parameters $p(y|\theta)$ as the *likelihood function* and $p(\theta)$ as the *prior density*, symbol $\propto$ means 'is proportional to' . Let us put it like lemma: **posterior is proportional to likelihood times prior**.

The prior $p(\theta)$ does not depend on data, it thus contains any non-data information. In other words it contains all we know about θ before we see

[2]Term $p(y)$ is of no particular importance and serves only as the normalization constant to preserve $p(\theta|y)$ being probability measure.

the data. Prior is another controversial aspect of Bayesian methods. Some technical econometricians would argue that employing some information that does not come of the data is just cheating. But even designing any model reflects some knowledge or presumptions that do not origin from the data. Thus we should regard setting the priors as a part of the model design (see the discussion about SVAR later on). In fact, compared to other methods, priors are useful for incorporating some prior idea about value of some parameters or structure of the model in an exact and transparent way.

Priors can be divided into informative, non-informative and empirical. Those informative are the 'rigorous' priors that employ some prior information. Non-informative priors deal with the controversy about preliminary information and do not use any prior information. Finally empirical priors employ into priors some information from data. These priors violate the basic premise of Bayesian methods (that prior $p(\theta)$ is independent on the data), but it works surprisingly well in practice.

The likelihood function $p(y|\theta)$ is the density of the data conditional on the parameters of the model. This can be seen as the data generating process (compare to common maximum likelihood estimation). For example, in the linear regression model we usually assume that errors are normally distributed. Thus, under Bayesian approach, it implies that $p(y|\theta)$ is normally distributed and depends upon regression coefficients and the error variance.

The posterior $p(\theta|y)$ summarizes all we know about parameters θ after seeing the data. We can thought about equation (2.5) as of an updating rule, where the data allows us to update our prior information about θ. The result is the posterior which combines both data and non-data information in a transparent way.

2.2 Simple regression model - an illustration

The aim of this section is to show how *natural conjugate prior* can be built, how posterior is estimated and how the Bayesian econometrics combines the information from prior and data. Let us consider the simplest regression model (single parameter - single explanatory variable, no constant term):

$$y_i = \beta x_i + \varepsilon_i \tag{2.6}$$

for $i = 1, ..., N$, where ε_i is an error term. Inclusion of error term can reflect measurement error, or it can reflect the fact that the model relationship is only an approximation of the true (real) relationship. Since no model can

completely describe reality (no model can be fitted through all data points - except of trivial cases), error term is inevitable.

2.2.1 The likelihood function

Assumptions about ε_i and x_i determine the form of the likelihood function. The standard assumptions are:

1. ε_i is i.i.d. $N(0, \sigma^2)$
2. x_i are either fixed (not random - this is conventional situation since we usually think that we have 'measured' data x_i) or, if they are random variables, they are independent of ε_i. In any case, these x_i are said to be dependent on some vector of parameters λ.[3] Potentially random x_i is thus given by probability density $p(x_i|\lambda)$. Note that parameters λ must be independent of β and σ^2.

Under these assumptions we allow x_i to be realization of random variable. Likelihood function is joint probability density function for the data conditional on the parameters (see (2.5)):

$$p(y_i, x_i|\beta, \sigma^2, \lambda) = p(y_i|x_i, \beta, \sigma^2)p(x_i|\lambda) \tag{2.7}$$

The distribution of x_i is not of interest and we will not explicitly include x and λ in our conditioning. We are thus interested in the likelihood function $p(y_i|\beta, \sigma^2)$. From the assumptions made about the errors we can build likelihood function. Firstly from assumption of the normality of errors ε_i:

$$p(y_i|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(y_i - \beta x_i)^2}{2\sigma^2}\right] \tag{2.8}$$

which is normal distribution with mean $E(y_i|\beta, \sigma^2) = \beta x_i$ and variance $var(y_i, \beta, \sigma^2) = \sigma^2$. Now we use assumption of ε_i being independent of ε_j for $i \neq j$. Thus $p(y|\beta, \sigma^2) = \prod_{i=1}^{N} p(y_i|\beta, \sigma^2)$, hence:

$$p(y|\beta, \sigma^2) = \frac{1}{(2\pi)^{\frac{N}{2}}\sigma^N} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta x_i)^2\right] \tag{2.9}$$

To see the similarity between Bayesian and common OLS estimation we can write this explicit likelihood function in a following form [3]. Derivation is

[3] We can create artificially the random variable x_i dependent on some λ from 'measured' values x_i.

straightforward though sizeable and messy:

$$p(y|\beta,\sigma^2) = \frac{1}{(2\pi)^{\frac{N}{2}}}\left\{exp\left[-\frac{h}{2}(\beta-\hat{\beta})^2\sum_{i=1}^{N}x_i^2\right]\right\}\left\{h^{\frac{\nu}{2}}exp\left[-\frac{h\nu}{2s^{-2}}\right]\right\} \tag{2.10}$$

where $\hat{\beta}$, s^2 and ν are the OLS estimator for β, standard error and degrees of freedom, respectively. h is the error precision $h = \frac{1}{\sigma^2}$. Let us note that equation (2.10) is a product of normal distribution for β and so called gamma distribution for h.

2.2.2 The prior

In this simple model we must elicit prior for β and h which we denote $p(\beta, h)$ and which will be determined in the form $p(\beta, h) = p(\beta|h)p(h)$. We search for *natural conjugate prior* that is defined by having the same form as the likelihood function. The importance of *natural conjugate prior* is that once both prior and likelihood function have the same form (the same probability distribution), the posterior is also of this distribution. [4] Thus we need $p(\beta|h)$ to be normal distribution and $p(h)$ to be gamma distribution. Priors can then be written in a form:

$$\beta|h \sim N(\underline{\beta}, h^{-1}, \underline{V}) \tag{2.11}$$

$$h \sim G(\underline{s^{-2}}, \underline{\nu}) \tag{2.12}$$

Prior hyperparameters $\underline{\beta}$, $\underline{V}$, $\underline{s^{-2}}$, $\underline{\nu}$ reflect researchers prior information. These new symbols - hyperparameters represent just numbers and their interpretation becomes clear from following section about *the posterior*. Note that bars under parameters (e.g. $\underline{\beta}$) denote parameters of prior density whereas over-bars (e.g. $\bar{\beta}$) denote parameters of posterior density.

2.2.3 The posterior

Since we used natural conjugate prior we have a posterior of the same form:

$$\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s^{-2}}, \bar{\nu}) \tag{2.13}$$

where

$$\bar{V} = \frac{1}{\underline{V}^{-1} + \sum x_i^2} \tag{2.14}$$

[4] The importance lies in the fact that natural conjugate prior ensures that posterior is of some well-defined probability distributions whose important properties (mean, variance etc.) can be (mostly) calculated analytically. Hence no posterior simulation (see below) is needed.

$$\bar{\beta} = \bar{V}\left(\underline{V}^{-1}\underline{\beta} + \hat{\beta}\sum x_i^2\right) \tag{2.15}$$

$$\bar{\nu} = \underline{\nu} + N \tag{2.16}$$

and $\bar{s}^{-2}$ is defined through

$$\bar{\nu}\bar{s}^2 = \underline{\nu}\underline{s}^2 + \nu s^2 + \frac{(\hat{\beta} - \underline{\beta})^2}{\underline{V} + \left(\frac{1}{\sum x_i^2}\right)} \tag{2.17}$$

From equations (2.14)-(2.17) we can see how Bayesian methods combine prior and data information in a very simple model. Common Bayesian estimate for β is $\bar{\beta}$, which is a weighted average of the OLS estimate $\hat{\beta}$ and the prior mean $\underline{\beta}$. The weights are proportional to $\sum x_i^2$ and $\underline{V}^{-1}$, respectively. Where $\underline{V}^{-1}$ is the confidence in the prior. The higher $\underline{V}$ the less certain we are about the values of β before seeing the data and consequently the smaller weight is given to the prior information in the posterior. On the other hand term $\sum x_i^2$ is proportional to the variance of the data and reflects the confidence of data's best guess for β, OLS estimator $\hat{\beta}$.

Similarly, the posterior sum of squares ($\bar{\nu}\bar{s}^2$) can be interpreted as the sum of prior sum of squares ($\underline{\nu}\underline{s}^2$), OLS sum of squares ($\nu s^2$) and a term which measures the conflict between prior and data information.[5] Equation (2.14) can be interpreted by saying that posterior precision is an average of prior precision ($\underline{V}^{-1}$) and data precision ($\sum x_i^2$). Thus posterior variance of β incorporates transparently both prior and data information. Now using some algebra (namely integrating out the dependency on h from the posterior distribution), we can derive some important properties of posterior that allow for continuing comparison to common OLS model:

$$E(\beta|y) = \bar{\beta} \tag{2.18}$$

$$var(\beta|y) = \frac{\bar{\nu}\bar{s}^2}{\bar{\nu} - 2}\bar{V} \tag{2.19}$$

$$E(h|y) = \bar{s}^{-2} \tag{2.20}$$

$$var(h|y) = \frac{2\bar{s}^{-2}}{\bar{\nu}} \tag{2.21}$$

Note that in this simple example we do not need neither numerical integration nor posterior simulation. These equations also illustrate how the Bayesian approach combines data and non-data information. We emphasized

[5]Note that natural conjugate prior can be interpreted as arising from fictitious data set, thus we can interpret $\underline{\nu}$ as a prior sample size since it plays the same role as N.

similarities between Bayesian approach and OLS estimation. But two differences are remarkable. Firstly, Bayesian approach incorporates non-data information (in the form of prior). And secondly Bayesians treat β as a random variable, whereas frequentists interpret $\hat{\beta}$ as a random variable.

2.2.4 Non-informative prior

Natural conjugate prior, used in last section, ensures that prior information enters in the same manner as data information and this helps with posterior elicitation. In many cases researchers may be able to agree on what a sensible prior might be (e.g. from economic theory). However, sometimes very different priors can be tenable. Two Bayesian strategies addressing these problem can be carried out.

1. *Prior sensitivity analysis* can reveal the dependency of result on different priors. Special case is *extreme bound analysis* that analyses the bound of results with any sensible prior taken.

2. *Noninformative prior* means setting $\underline{\nu} = 0$ and $\underline{V}^{-1} = 0$ in our simple case.

   Results of Bayesian estimation are then equivalent to the OLS estimation. This approach has undesirable property that such prior is *improper* since its density is not valid because does not integrate to one. Unfortunately non-informative priors are improper in most models. Loosely speaking, such non-informative prior takes the form of uniform density and such density integrates to infinity over $(-\infty, \infty)$. Thus such non-informative prior is improper.

2.2.5 Extension 1 - Many explanatory variables

It is straightforward (and for skilful econometricians even easy) to extend previous one parameter model to many parameters model. The only magic is matrix notation. If we use again natural conjugate prior we get qualitatively identical results. Particularly important is that our posterior is again well known distribution - in this case it is multivariate t distribution. Concerning computational issues, note that it is easy to take random draws from this distribution.

2.2.6 Extension 2 - Other priors

Until now we used natural conjugate prior with two *conditionals* $p(\beta, h) = p(\beta|h).p(h)$ with normal and gamma distributions, respectively. Now, let us assume prior $p(\beta, h) = p(\beta).p(h)$ with $p(\beta)$ being normal and $p(h)$ gamma distribution. Likelihood remains the same as before and thus posterior can be evaluated. Algebra is not so different and the exact derivation is not provided since qualitative considerations are in the center of the debate.

If we create the posterior we can find out that the *joint* posterior density ($p(\beta, h|y)$) does not take the form of any well-known density and hence, cannot be directly used in a simple way for posterior inference. The *conditionals* of the posterior are, however, simple. $p(\beta|y, h)$ and $p(h|\beta, y)$ remain normal and gamma distributions, respectively. It must be stressed that $p(\beta, h|y) \neq p(\beta|y, h).p(h|\beta, y)$ and thus even from 'nice' conditionals we do not have information we need ('nice' posterior). In fact posterior simulator must be used (see Bayesian Computation section).

2.2.7 Extension 3 - Inequality constraints

Imposing inequality constraints on the parameters (e.g. coefficients in the linear regression model) is something that the researcher often wishes to do. Imposing stationarity into model with autocorrelated errors might be of particular importance. These constraints can all be written in the form $\beta \in A$, where A is relevant region. We can impose such constraints simply through the prior under Bayesian framework. When constructing the prior, statement $\beta \in A$ is equivalent to statement that all other regions of parameter space are *a priori* excluded, thus they receive a prior probability weight of 0. Such prior information can be combined with any other prior information (e.g. with natural conjugate prior as we did before). As a result we usually get posterior that is truncated to the region where $\beta \in A$. In linear regression we get posterior $p(\beta|y)$ as a multivariate t distribution truncated to the region $\beta \in A$. As a result of inequality restriction we might get posterior that is not of any analytical form. It is common that posterior simulation called *importance sampling* must be employed (see below).

2.3 Bayesian Computation

Historically, the computation issues are the reason for the inferior status of Bayesian econometrics. The computing revolution has led to a sharp growth of using of Bayesian methods in many fields. In the previous very simple

example we could calculate all integrals analytically. This is a very rare case. In fact we often need to estimate some integrals numerically. In this section we will discuss in detail methods based on *Monte Carlo integration*, namely the *Gibbs sampler* and the *Metropolis-Hastings* algorithm.

Equation (2.5) defines the posterior and, importantly enough, does not contain any integral. Moreover posterior density $p(\theta|y)$ summarizes all information about parameters. However, it is rarely possible to present all the information about $p(\theta|y)$ in a compact form. Moreover numerical summaries that we might wish to present usually involves integration. This numerical summary can be for instance some point estimate (e.g. mean and median). The *posterior mean* of any element is calculated as:

$$E(\theta_i|y) = \int \theta_i p(\theta|y) d\theta \tag{2.22}$$

This integral can be analytically evaluated only in few simple cases (one of these was presented in the illustrative OLS example), otherwise it must be evaluated by some numerical method.

We are also interested in the degree of uncertainty associated with the point estimate. The most common such measure is the *posterior standard deviation*, which is the square root of the *posterior variance*, that is calculated:

$$var((\theta_i|y)) = E(\theta_i^2|y) - [E(\theta_i|y)]^2 \tag{2.23}$$

This formula requires to integrate (2.22) and moreover:

$$E(\theta_i^2|y) = \int \theta_i^2 p(\theta|y) d\theta \tag{2.24}$$

Similarly we can be interested whether some parameter is positive:

$$p(\theta_i \leq 0|y) = \int_0^\infty p(\theta|y) d\theta^i \tag{2.25}$$

Generally, all these calculations have the form:

$$E(g(\theta)|y) = \int g(\theta) p(\theta|y) d\theta \tag{2.26}$$

where $g(\theta)$ is a function of interest. Even the predictive density falls in this framework if we set $g(\theta) = p(y^*, y, \theta)$. The main exception are the marginal likelihood and quantiles of the posterior density. These are not evaluated using this form and are usually easier to obtain. Serious problem arises when $E(g(\theta)|y)$ does not exist (e.g. mean of the Cauchy distribution).

2.3.1 Monte Carlo integration

There are several methods for evaluating (2.26), but the dominant approach is *posterior simulation*. This approach allows to approximate the value of $E(g(\theta)|y)$ by sequence of some computationally manageable steps.

There exist many posterior simulators and these are all applications of *laws of large numbers* and *central limit theorem*.

Monte Carlo integration:

Let $\theta^{(s)}$ for $s = 1, ..., S$ be a random sample from $p(\theta|y)$, and define

$$\hat{g}_S = \frac{1}{S} \sum_{s=1}^{S} g\left(\theta^{(s)}\right) \tag{2.27}$$

then $\hat{g}_S$ converges to $E(g(\theta)|y)$ as S goes to infinity.

This approach allows us to approximate $E[g(\theta)|y]$ by taking random draws from the posterior. $\theta^{(s)}$ is referred to as a *draw* or *replication*. $\hat{g}_S$ is a good approximation of $E(g(\theta)|y)$ only if S is sufficiently large. Using central limit theorem and properties of normal distribution we can get the approximate result

$$Pr\left[-1.96\frac{\sigma_g}{\sqrt{S}} \leq \hat{g}_S - E(g(\theta)|y) \leq 1.96\frac{\sigma_g}{\sqrt{S}}\right] = 0.95 \tag{2.28}$$

where $\sigma_g = var[g(\theta)|y]$, which is in practice unknown but can be approximated by Monte Carlo integration procedure. The term $\frac{\sigma_g}{\sqrt{S}}$ is known as the *numerical standard error*. Unfortunately, it is not always possible to do Monte Carlo integration since algorithms for taking random draws do not exist for all probability densities.

2.3.2 Gibbs Sampler

Let us remind that in the extension 2 of our illustrative model (section 2.2.6) we have posterior ($p(\beta, h|y)$) that does not take form of any well-known density whereas its conditionals ($p(\beta|y, h)$ and $p(h|y, \beta)$) are of defined distributions. Thus we are interested to simulate random draws from posterior (this is what we want but cannot directly do) using random draws from conditionals (these are often perfectly available).

We have vector of parameters θ (in illustrative model $\theta = (\beta, h)$[6]). Let us divide θ into various *blocks* as: $\theta = (\theta_1, \theta_2, ..., \theta_B)$, where each θ_j is scalar

[6]More precisely θ equals transpose of (β, h), since θ is vector.

or vector of parameters. (In illustrative model it is convenient to set $\theta_1 = \beta$ and $\theta_2 = h$.)

Following the motivation of our example of having 'nice' conditionals we now define generally *full conditional posterior distributions*:

$$p(\theta_1|y, \theta_2, ..., \theta_B), p(\theta_2|y, \theta_1, \theta_3, ..., \theta_B), ...p(\theta_B|y, \theta_1, \theta_2, ..., \theta_{B-1}). \qquad (2.29)$$

(In regression model with independent priors these conditionals are normal distribution $p(\beta|y, h)$ and gamma distribution $p(h|y, \beta)$.)

Let us consider now case $B = 2$. Suppose that we have initial random draw θ_2^0 from $p(\theta_2|y)$.[7] Now we make a draw $\theta_1^1$from (well known) conditional $p(\theta_1|y, \theta_2)$.[8] From properties of marginal and joint densities it follows that this θ_1^1 is a valid draw of θ_1 from whole posterior $p(\theta|y)$. This we got employing $p(\theta|y) = p(\theta_1|y, \theta_2)p(\theta_2|y)$. Analogously we can write: $p(\theta|y) = p(\theta_2|y, \theta_1)p(\theta_1|y)$. Thus having draw θ_1^1 we can draw new θ_2^1 from conditional $p(\theta_2|y, \theta_1)$. Hence we have first complete valid draw $\theta^1 = (\theta_1^1, \theta_2^1)$ from full posterior $p(\theta|y)$. Moreover we can use θ_2^1 to make a new draw θ_1^2 and continue indefinitely. This procedure of sequentially drawing from conditional posterior distributions is called *Gibbs sampling*.

The problem arises that sometimes it is hard to find initial draw θ_2^0. Usually we choose θ_2^0 just 'somehow' and then hope (or test or prove) that this particular choice of initial draw does not contaminate our draws from $p(\theta|y)$. Moreover it is common, after having our set of S draws, to discard first S_0 draws to eliminate the effect of the initial draw. Remaining S_1 draws can be used to create estimates of posterior features of interest:

$$\hat{g}_{S_1} = \frac{1}{S_1} \sum_{s=S_0+1}^{S} g\left(\theta^{(s)}\right) \qquad (2.30)$$

where $\hat{g}_{S_1}$ converges to $E[g(\theta)|y]$ for S_1 going to infinity.

We employ here Monte Carlo integration but we need to note that its assumption of random draws is not fulfilled. Particularly our draws θ^s is not independent from θ^{s-1}. This dependency on last value means that sequence is a *Markov Chain*. There are many posterior simulators with this property. Such posterior simulators have the general name of *Markov Chain Monte Carlo* algorithms (MCMC).

[7] Lower index represents ordinal number of parameter - second parameter (would be h in the example). Upper index represents ordinal number of the draw (zero draw - initial draw). This number, in principal, runs to very high values through the procedure.

[8] Thus we have first draw of first parameter.

2.3.3 MCMC diagnostics

We use various diagnostics to see whether estimated results of *MCMC* algorithms are reliable. Common MCMC diagnostic is numerical standard error. We can employ central limit theorem (similarly to 2.28) and get numerical standard error being equal to $\frac{\sigma_g}{\sqrt{S_1}}$. But our new σ_g is higher than in the case of original Monte Carlo integration because it has to compensate for the fact that θ^s is a correlated sequence.

Another diagnostic is based on the intuition that if the sufficiently large number of draws have been taken, the estimate of $g(\theta)$ based on the first half of draws should be the same as the estimate based on the second half. More precisely we can divide our S_1 of draws into first set S_A, middle set S_B and last set S_C. After discarding middle set we can construct so called *convergence diagnostic* (CD) that compares estimates based on sets of draws S_A and S_C.

$$CD = \frac{\hat{g}_{S_A} - \hat{g}_{S_C}}{\frac{\sigma_A}{\sqrt{S_A}} + \frac{\sigma_C}{\sqrt{S_B}}} \to N(0,1) \tag{2.31}$$

In words, convergence diagnostic is the difference between the estimates divided by the sum of numerical standard errors. Moreover this converges to normal distribution thus critical values of normal distribution can be used. Large values of this diagnostic suggests that estimates are quite different form each other and hence we have not taken enough replications (S_1 is too small).

MCMC diagnostics are usually quite informative though two exceptions must be expressed.

1. When posterior is bimodal (e.g. comprises of two distinct normal distributions that are located far from each other in parameter space) then Gibbs sampler may not switch from one to the other. All the result would be missing one of the two normals that comprise the posterior. Unfortunately, MCMC diagnostics will not explore this.

2. When initial draw θ^0 is too far from the region of the parameter space where most of the probability lies then this region might not be found and MCMC diagnostics might even not realize this. Both these cases occur because of effect of the initial draw has not worn off. Common practise is to run Gibbs Sampler several times using a different values for θ^0.

To be more precise, let us take m different initial values from very different regions of the parameter space $\theta^{0,i}$, in the jargon *overdispersed*

starting values. Now we can run Gibbs sampler m times and get m sequences $g(\theta^{s,i})$. From these we can count *within sequence variance* that represents the variance of Gibbs sampler sequence with using one particular initial value. The average of these is *average within sequence variance*:

$$W = \frac{1}{m}\frac{1}{S_1 - 1}\sum_{i=1}^{m}\sum_{s=S_0+1}^{S}\left[g(\theta^{s,i}) - \hat{g}^i_{S_1}\right]^2 \tag{2.32}$$

We can also compute *between sequence variance*:

$$B = \frac{S_1}{m-1}\sum_{i=1}^{m}\left(\hat{g}^i_{S_1} - \sum_{i=1}^{m}\hat{g}^i_{S_1}\right)^2 \tag{2.33}$$

Note that W is an estimate of $var[g(\theta)|y]$ but heavily underestimates this variance in two considered problematic cases (because some trial sequences might not find some part of the posterior probability in parameter space). On the other hand, B is based on overdispersed starting values and efficiently cures problematic cases. B overestimates $var(g(\theta)|y)$. It can be shown that

$$\hat{var(g(\theta)|y)} = \frac{S_1 - 1}{S_1}W + \frac{1}{S_1}B \tag{2.34}$$

is a good estimate of $var(g(\theta)|y)$. Finally we get commonly presented MCMC convergence diagnostic:

$$\hat{R} = \frac{\hat{var(g(\theta)|y)}}{W} \tag{2.35}$$

Values of $\hat{R}$ near one indicates that the Gibbs sampler has successfully converged. Role of thumb suggests that values greater than 1.2 indicate poor convergence. $\sqrt{\hat{R}}$ is called the *estimated potential scale reduction*.

2.3.4 Importance Sampling

When we employed Gibbs Sampler we did not need analytic form of whole posterior but we needed 'nice' conditionals. Unfortunately we often get such posterior that even its conditionals are not of any analytical form. Extension 3 of our illustrative linear regression model (section 2.2.7) might serve as a motivation example.

Firstly let us introduce the distribution function of parameters called *importance function* $q(\theta)$. We can construct this distribution in principle anyhow, but we seek for two main features. First is that it must be easy to

take random draws from this distribution and second $q(\theta)$ must approximate $p(\theta|y)$ reasonably well (particularly $q(\theta) > p(\theta|y)$ for any θ). After having importance function we employ this posterior simulator:

Importance sampling

Let θ^s for $s = 1, .., S$ be a random sample from $q(\theta)$ and define

$$\hat{g}_S = \frac{\sum_{s=1}^{S} w(\theta^s) g(\theta^s)}{\sum_{s=1}^{S} w(\theta^s)} \tag{2.36}$$

where

$$w(\theta^s) = \frac{p(\theta = \theta^s|y)}{q(\theta = \theta^s)} \tag{2.37}$$

then $\hat{g}_S$ converges to $E[g(\theta)|y]$ as S goes to infinity (under weak assumptions).

In words, we use weighted averaging instead of simple averaging, where weight always describes how well the importance function approximates posterior in a particular point of randomly chosen parameter. We then do not need to take random draws form posterior since we only take random draws from importance functions and compare value of this importance function with the computed value of posterior (we just put into the posterior function actual random draw from the importance function).

Main advantage of the importance sampling are very weak assumptions on posterior (loosely saying we need only $p(\beta|y)$ and $E[g(\theta)|y]$ existing). Its biggest disadvantage is a need of careful choice of $q(\theta)$. Unless $q(\theta)$ does not approximate the posterior well enough, then weight $w(\theta^s)$ is virtually zero for almost every draw, thus S may need to be enormous. As a result importance sampling is far less popular than e.g. Gibbs sampler, because it usually involves hunting for, justifying and fine tuning of convenient importance functions.

Fortunately, in some cases, choosing importance function is quite natural. In Extension 3 of our illustrative OLS model (section 2.2.7) we imposed inequality restrictions on parameters ($\beta \in A$). As a result we got truncated posterior that have lost its analytical form. Under framework of importance sampling it is natural (and successful) to define importance function $q(\theta)$ as unrestricted posterior with weights being equal to 1 if particular draw from parameters fulfils imposed restrictions, zero otherwise: $w(\theta^s) = 1 \quad (\beta^s \in A)$. This strategy simply involves drawing from the unrestricted posterior and discarding draws which violate the inequality restrictions.

2.3.5 Metropolis-Hastings algorithm

Metropolis-Hastings algorithm can be be viewed as a whole class of algorithms. It combines importance sampling with its generality and Gibbs sampler including its cook-book fashion.

What we called importance function in section *Importance sampling* will now be called *candidate generating density*, $q(\theta^{s-1};\theta)$. This notation indicates that *candidate draw* θ^* is taken of the random variable θ whose density depends on 'previous' density θ^{s-1}. As with the Gibbs sampler, the current draw depends on the previous draw (unlike importance sampling). Thus Metropolis-Hastings algorithm is an MCMC algorithm and the drawn values are often referred to as a chain.

With Metropolis-Hastings algorithm, we weight all draws equally (on the contrary from importance sampling), but not all candidate draws are accepted. New steps in this cookbook algorithm are that we count *acceptance probability* $\alpha(\theta^{s-1},\theta^*)$ after taking the candidate draw. And now we set $\theta^s = \theta^*$ with probability $\alpha(\theta^{s-1},\theta^*)$ and we set $\theta^s = \theta^{s-1}$ (i. e. refuse new draw) with probability $1-\alpha(\theta^{s-1},\theta^*)$.

Acceptance probability is chosen to be highest in areas where posterior probability is highest. This ensures that if current θ^{s-1} is in an area of high posterior probability it will tend to stay here whereas if θ^{s-1} is in an area with low posterior probability the algorithm will tend to move quickly away. Under this intuition we want a candidate draw (θ^*) to be accepted with high probability if it is in a region of higher posterior probability than θ^{s-1}. In other words we want the algorithm to stay longer in high probability regions but still it must be able to visit the areas of low probability as well. Following choice has these desirable properties:

$$\alpha(\theta^{s-1},\theta^*) = min\left[\frac{p(\theta=\theta^*|y)q(\theta^*;\theta=\theta^{s-1})}{p(\theta=\theta^{s-1}|y)q(\theta^{s-1};\theta=\theta^*)},1\right] \tag{2.38}$$

We must stress that equally as this algorithm combines advantages of importance sampling and Gibbs sampler it also contains disadvantages of both. From Gibbs sampler arises the need for finding a good initial draw θ^0 since each draw is dependent on the previous one. And, more importantly, from importance sampling arises the problem of searching for convenient generating density $q(\theta^{s-1};\theta)$. There is a myriad of possible strategies to choose candidate generating density. I will comment two of the common.

The Independence Chain Metropolis Hastings Algorithm

The name suggests that we use candidate generating density that is independent on the previous draw $q(\theta^{s-1};\theta) = q^*(\theta)$. The algorithm is then

very similar to the importance sampling. We also need good approximation to the posterior. The only difference is that instead of different weighting under importance sampling, we now use acceptance probability whether to stay (thus similar to higher weight) or to quickly move on (similar to low weight). Under notation from importance sampling we can rewrite acceptance probability in simple and understandable form:

$$\alpha(\theta^{s-1}, \theta^*) = min\left[\frac{w(\theta = \theta^*)}{w(\theta = \theta^{s-1})}, 1\right] \tag{2.39}$$

In words, the acceptance probability is the ratio of importance sampling weights evaluated at the old and candidate draws.

The most important issue is still choosing candidate density. For a regression (even non-linear), we know that maximum likelihood estimator is asymptotically normal thus we can state that, if sample size is reasonably large the posterior might be approximately normal with mean $\hat{\theta}_{ML}$ and variance $var\left(\hat{\theta}_{ML}\right)$. However, it is more common to use t-distribution because it has been found that the candidate generating density should have tails which are at least as fat as those of the posterior.

The Random Walk Chain Metropolis-Hastings Algorithm

With Random Walk Chain Metropolis-Hastings Algorithm we make no attempt (we do not need) to approximate the posterior. Candidate generating density is chosen to wander widely. Formally we generate candidate draws according to random walk:

$$\theta^* = \theta^{s-1} + z \tag{2.40}$$

where z is *increment random variable*. The acceptance probability ensures that the 'random' walk chain moves in the appropriate direction:

$$\alpha(\theta^{s-1}, \theta^*) = min\left[\frac{p(\theta = \theta^*|y)}{p(\theta = \theta^{s-1}|y)}, 1\right] \tag{2.41}$$

The choice of density for z determines the precise form of the candidate generating density. Convenient choice is the multivariate normal distribution. Thus $q(\theta^{s-1}; \theta)$ is normal distribution with mean θ^{s-1} and variance-covariance matrix Σ. All we then need is to choose Σ in order to get optimal acceptance rate. Rule of the thumb says that optimal acceptance rate for univariate problem is 0.45 and with higher dimensions decreases asymptotically to 0.23.

Metropolis-within-Gibbs

We showed Gibbs Sampler as a simple method how to sequentially combine random draws from posterior conditionals to get random draws form

posterior. Moreover we showed that Metropolis-Hastings algorithm allows us to take draws from posterior efficiently without having posterior of any analytic form. Metropolis-within-Gibbs algorithm uses Metropolis-Hastings algorithm to provide draws from some (or one or all) posterior conditionals and then involves all Gibbs Sampler procedure. This method is very common since many models have posterior where most of the conditionals are easy to draw from, but few of them do not have convenient form.

2.4 Model Comparison - principles

Suppose we have m different models, M_i for $i = 1, ..., m$, which all seek to explain y. The posterior for the parameters is then written as in (2.4):

$$p(\theta^i|y, M_i) = \frac{p(y|\theta^i, M_i)p(\theta^i, M_i)}{p(y, M_i)} \tag{2.42}$$

We thus have posterior, likelihood and prior for each model M_i.

We use the Bayes' rule to derive a probability statement about what we do not know (whether model M_i is good or not) conditional on what we know (data y). We thus count *posterior model probability* $p(M_i|y)$ using (2.3):

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)} \tag{2.43}$$

$p(M_i)$ is *prior model probability*, it does not contain information from the data so it measures how likely we believe that M_i is correct before we see the data. $p(y|M_i)$ is the *marginal likelihood*. We can integrate (2.42) over θ^i and use $\int p(\theta^i|y, M_i)d\theta^i = 1$ so we get $p(y|M_i)$:

$$p(y|M_i) = \int p(y|\theta^i, M_i)p(\theta^i, M_i)d\theta^i \tag{2.44}$$

For comparing two models directly we usually use *posterior odds ration*:

$$PO_{ij} = \frac{p(M_i|y)}{p(M_j|y)} = \frac{p(y|M_i)p(M_i)}{p(y|M_j)p(M_j)} \tag{2.45}$$

We thus get rid of $p(y)$ that is usually hard to compute directly. The non-informative choice $p(M_i) = p(M_j) = \frac{1}{m}$ is commonly made (i.e. all models have the same prior probability of being the true one). In this case we speak about *Bayes factor*, defined as:

$$PO_{ij} = \frac{p(y|M_i)}{p(y|M_j)} \tag{2.46}$$

These principles are developed in many different ways for being applicable to different types of models. Details concerning *highest posterior density intervals, Savage-Dickey density ratio, Gelfland-Dey method, posterior predictive p-value* or *Bayesian model averaging* can be found in [3].

2.5 Bayesian prediction

We might be interested in predicting some future, unobserved data y^* given the observed data y. As usual Bayesian reasoning suggests to use probability of what we do not know (y^*) conditional on what we know (y). That is *predictive density* $p(y^*|y)$. It is known from probabilistic and statistical rules that we can obtain marginal density from joint density through integration:

$$p(y^*|y) = \int p(y^*, \theta|y)d\theta \tag{2.47}$$

We can now rewrite term term inside the integral using another known rule about conditional probability and get:

$$p(y^*|y) = \int p(y^*|y, \theta)p(\theta|y)d\theta \tag{2.48}$$

This form is particularly convenient since it involves posterior. Let us assume the case of simple regression model, we then need to assume that some new values of explanatory variable (x^*) are known. We rewrite last equation:

$$p(y^*|y) = \int\int p(y^*|y, \beta, h)p(\beta, h|y)d\beta dh \tag{2.49}$$

Since ε^* is independent of any ε_i, y and y^* are independent, it holds $p(y^*|y, \beta, h) = p(y^*|\beta, h)$. This is likelihood function (see 2.8). Remaining term is the posterior. Both these terms are of well known analytical form in this OLS case. Thus we can also solve equation (2.49) analytically and we find that $p(y^*|y)$ is t distribution with mean βx^*, variance $\frac{\bar{\nu}\bar{s}^2}{\bar{\nu}}\left(1 + \bar{V}x^{*2}\right)$, and degrees of freedom $\bar{\nu}$. This result provides point predictions and measures of uncertainty associated with the point prediction.

New issues arise when $p(y^*|y)$ is not of analytical form. We then cannot get complete $p(y^*|y)$ similarly as we cannot get precise form of posterior. But in the same manner as we extract information from posterior by simulation (see eq. 2.26) we can extract virtually any predictive feature of interest in the form $E[g(y^*)|y]$. For instance, calculating the predictive mean of y^* implies

setting $g(y^*) = y^*$, calculating the predictive variance $g(y^*) = y^{*2}$. We are thus interested in calculating:

$$E[g(y^*)|y] = \int g(y^*)p(y^*|y)dy^* \tag{2.50}$$

From its similarity to posterior simulation (2.26) we can repeat Monte Carlo integration theorem:

Let $y^{*(s)}$ for $s = 1, ..., S$ be a random sample from $p(y^*|y)$, and define

$$\hat{g}_Y = \frac{1}{S}\sum_{s=1}^{S} g\left(y^{*(s)}\right) \tag{2.51}$$

then $\hat{g}_Y$ converges to $E(g(y^*)|y)$ as S goes to infinity.

Following strategy can be employed using this formula,. We take a draw θ^s (or in our particular regression β^s and h^s) from posterior using posterior simulator (e.g. Gibbs sampler, Importance sampling, Metropolois-Hastings algorithm). After that we can finally compute y^* from $p(y * |\beta^s h^s)$, which is counted as a likelihood function and is often of analytical form. Thus small generalization of posterior simulator can provide also information about any predictive feature of interest.

2.6 VAR and BVAR

Before we analyse BVAR approach we shortly revise the vector autoregression (VAR) to address its main advantages and problems.

2.6.1 VAR

Vector autoregression is an n-equation, n-variable linear model in which each variable is explained by its own lagged values, current values of the remaining $n - 1$ variables and their past values. Under this framework rich dynamics in multiple time series can be captured. VAR is coherent and credible approach to data description, forecasting, structural inference and policy analysis. However, structural inference and policy analysis are more difficult due to identification problem between causation and correlation.

Reduced form VAR expresses each variable as a linear function of its own past values and the past values of all other variables plus serially uncorrelated error term. Each equation is estimated by OLS and the number of included

lags can be determined by different methods. Unfortunately, if the different variables are contemporaneously correlated with the others - as they typically are in macroeconomic applications - then the error term in the reduced form model will also be correlated across equations, which is in contradiction with the assumptions of *reduced form VAR*.

Recursive VAR includes some contemporaneous values as regressors. This is usually systematized by some ordering of equations. As a result, error term in each regression equation is uncorrelated with the error in the preceding equations. Problem is that all the results depend on the ordering of equations and variables and there are $n!$ possible recursive VARs representing all possible orderings.

Structural VAR uses economic theory to sort out the contemporaneous links among the variables, thus ordering of equations. Structural VAR (SVAR) requires identifying assumptions under which correlations can be interpreted causally. The number of SVARs is limited only to the fantasy of the researcher. SVAR is also common name for any VAR in that contemporaneous relationships occur.

Estimates of the VAR's coefficients or R^2 are often unreported. More important are other analyses. *Granger-causality statistics* examine whether lagged values of one variable help to predict the dependent one. *Impulse response analysis* exhibits the response of current and future values of each of the variables to a one-unit increase in the current value of one of the VAR errors, under assumptions that it is only one-period shock. This thought experiment makes more sense when the errors are uncorrelated across equations, so impulse responses are typically calculated for recursive and structural VARs.

Ultimate test of a forecasting model is its out-of-sample performance. However, it is common to test so-called *pseudo out-of-sample* performance. We thus estimate model using some old data and make prediction that can be immediately compared to the newer but already known data. We can predict in different forecast horizons that are interesting in the practical application. It is common to compare prediction performance to simpler models like univariate AR models or even to its its naive variant: Random walk models.

Standard VARs miss non-linearities, conditional heteroscedasticity and breaks in parameters. VAR model (mainly small VAR models) are also often unstable. However, adding variables to VAR creates the biggest complication. Nine-variable, four-lag VAR has 333 unknown parameters including intercepts. Macroeconomic series data cannot provide reliable estimates of all these coefficients. In the case when there are more parameters

than observations, then such model cannot be estimated. However even when there are less parameters than observations, but still quite an amount, then we speak about *overparametrization* or *overfitting*, since model estimates and forecasts are not reliable any more. Generally, this problem is reffered to as 'degrees of freedom' problem.

Common solution is to impose *ad hoc* restrictions to parameters acording to more or less plausible assumptions from economic theory. Some relationships are then suppressed (this effectively means setting some groups of parameters to zero) and thus number of parameters is reduced (so called 'parsimonious VAR approach'). Such models then go from being *overparametrized* to being *overidentified* [4].

Another option, prefferd in this work, is to impose common structure on the coefficients using Bayesian methods.

2.6.2 BVAR

Bayesian VAR approach was introduced by Litterman (1979), expanded by Doan, Litterman and Sims (1984) and somewhat summarized in Litterman (1985) with 'five years of experience' proving the forecasting ability of BVAR models [5]. BVAR approach proved to be a flexible and effective forecasting method [6].

BVAR model with general prior is atheoretical or statistical time-series model as well as VAR without any structural restriction to parameters. This advantage over macroeconomic structural models is already pointed in Litterman, 1985 [5]. Short quotation right from the second paragraph of this forty-pages article is more than illustrative: '(BVAR) does not require judgemental adjustment. *Thus it is a scientific method* which can be evaluated on its own, without reference to the forecaster running the model' (Litterman, 1985). However, complete specification requires, among others, to specify the model variables. Variables are often selected according to their economic plausibility (and thus dependent on the forecaster's judgement). Nevertheless, in this work, variables are selected according to their pseudo out-of-sample forecast ability. Litterman's call for scientific method is then satisfied.

Under Bayesian framework, *a priori* information can be incorporated into models through priors. Priors impose general restrictions on parameters avoiding the problem of overfitting. Many different priors were defined and used, though the majority of works still keep in with originally proposed prior by Litterman: Minnesota prior.

2.6.3 Minnesota Prior

Minnesota prior was originally proposed by Litterman, Doan and Sims (1984). Priors restrict parameters of the VAR model and they are in the form of normal distribution. Normal distribution is defined by mean and variance thus prior for each parameter is defined by its mean and variance.

Concerning the means, Minnesota prior is based on the belief that random walk is a good proxy for the behaviour of economic variables through the time.[9] This means that mean of the prior for the parameter is equal to one for the first lag of the dependent variable in each equation. All other parameters (connected to all other variables and also to all higher lags of the dependent variable) are endowed with the prior with zero mean. Minnesota prior thus take the form [2]:

$$\underline{\beta_i} \sim N\left(1, \sigma^2_{\beta_i}\right) \tag{2.52}$$

$$\underline{\beta_j} \sim N\left(0, \sigma^2_{\beta_j}\right) \tag{2.53}$$

- $\underline{\beta_i}$ - parameters associated with first lag of dependent variable in each equation
- $\underline{\beta_j}$ - all other parameters

This implies that each variable included in the model is assumed to be dependent mainly on its own first lag. Higher-order lags and lags of other model variables are thus viewed as less important.

The prior variances $\sigma^2_{\beta_i}$ specify uncertainty about the prior means $\underline{\beta_i} = 1$, and $\sigma^2_{\beta_j}$ indicates uncertainty regarding the means $\underline{\beta_j} = 0$. Since we are interested in model without judgemental adjustment and generally model can contain a large number of parameters, it is useful to employ certain formula[10] to generate the standard deviations as a function of small number of hyperparameters. This approach allows to specify individual prior variances for a large number of coefficients in the model using only few hyperparameters:

$$\sigma_{ijk} = \theta w(i,j) k^{-\phi} \left(\frac{\hat{\sigma}_{uj}}{\hat{\sigma}_{ui}}\right) \tag{2.54}$$

[9] It should be preferred to use 'random walk with drift' when dealing with non-stationary series. This was originally proposed by Litterman [5]. In this work we use priors without drift because we predominantly use series expressed in growth rates and we are interested in forecasting of inflation which is stationary series.

[10] This was suggested already by Doan, Litterman and Sims, 1984

- $\hat{\sigma}_{ui}$- estimated standard error from a univariate autoregression involving variable i[11]
- k - number of lag
- θ - overall tightness parameter
- ϕ - decay parameter
- w - weighting parameter or matrix element of weight matrix (see below)

The variances of coefficients on lags of other than the dependent variable are not scale invariant. Thus term $\left(\frac{\hat{\sigma}_{uj}}{\hat{\sigma}_{ui}}\right)$ is a necessary scaling factor. Hyperparameter θ indicates the tightness of the 'random walk' restriction, or the relative weight of the distribution of all priors (the higher θ the lower weight of the prior). Another restriction is based on the assumption that, in general, second-order lags contain more information than third-order lags and so on. Using harmonic lag decay function, the information content decay with $k^{-\phi}$ and so the prior tightens to its zero mean. The higher is the decay parameter ϕ the faster is the decay.

The weight parameter needs more explanation since it allows some useful manipulation. This restriction is based on the assumption used for each equation, that the variable's own lags contain more information than the lags of other variables. This might be represented by scalar weight parameter w. For example, if we want to restrict prior of other variables more strongly than the dependent variable prior, we set w being equal to e.g. 0.5, which means lowering such prior variances by one half.

By this general prior specification we fulfil the requirements for the model to be atheoretical, without judgement adjusting and also giving reliable results even for many variables models with literally any amount of included lags.[12]

2.6.4 Weight parameter under matrix notation

Last discussed prior restriction - weight parameter - might be extended to a matrix notation. Following weight matrix represents the same example that was given in previous paragraph:

[11]This means that for each variable an univariate autoregression model with some meaningful number of lags must be estimated in order to get standard error $\hat{\sigma}_{ui}$ of the residuals.

[12]Until it is computationally feasible or econometrically meaningful.

$$w = \begin{pmatrix} 1 & 0.5 & \cdots & 0.5 \\ 0.5 & 1 & & 0.5 \\ \vdots & & \ddots & \vdots \\ 0.5 & 0.5 & \cdots & 1 \end{pmatrix}$$

'Ones' on the diagonal represent that the prior variances of the coefficients corresponding to the lags of the dependent variables in each equation are not further tightened by this 'weight' restriction. However 'halves' in the rest of the matrix represent that there is imposed further shrinkage of prior to its zero mean value. These priors correspond to the coefficients of the lags of all variables except of the dependent one in each equation. Rows, obviously, correspond to the equations of vector autoregression. Columns then correspond to the different variables (diagonal terms thus refer to the dependent variable lags as it is clear from previous discussion). One matrix element refers to whole variable in the equation disregarding the number of lags, since lag dependence of the prior variance is already given by the decay function.

2.6.5 Structural restrictions in BVAR

Matrix notation of weight parameter allows to impose different restrictions to the priors' variances of different variables in each equation. Setting off-diagonal weight parameter close to zero[13] results in setting the coefficient value efficiently to zero. This is very similar to imposing restriction (setting the parameters to zero) in the VAR models. Through weight parameter can thus be implied any structural assumptions. This, of course, contradicts the original Litterman's idea of BVAR being atheoretical model. However this setting might be very useful in the models where some structural restrictions are undoubted or when some variable has obviously bigger influence on the dependent variable than others. This is of particular interest when e.g. some economic activity over different regions is simultaneously modelled. It is clear that neighbouring regions affect each other more than distant regions. This kind of information that does not origin from the measured data can be efficiently implied by weight parameter under matrix notation [2].

2.6.6 VAR as a special case of BVAR

Interestingly enough, VAR models can be regarded as special case of BVAR model. We can equalize the variances of Minnesota prior over all variables

[13]It is not possible to set the parameter equal to zero, since prior would be improper (non-integrable).

by setting weight parameter $w = 1$. We can also erase lag decay function by setting $\phi = 0$. Thus all the variances are now equal. We know that main difference between VAR and BVAR is that VAR does not use any prior infomation. In principle, this can be done by setting overal tightness of the prior parameter distribution to be infinite ($\theta = \infty$). However, this is not really possible. Nevertheless efficiently non-informative prior can be attained by setting the tightness to be sufficiently large (e.g. $\theta \doteq 1000$) [1]. BVAR estimated using this choice of the parameters then corresponds to the VAR model.

2.6.7 Other priors

Modified Litterman prior Though it is not obvious, Minnesota prior corresponds only to the situation of reduced VAR form [4], because it assumes that the covariance matrix of the residuals of the model is diagonal, fixed and known. However, this does not hold if we intend to use structural model rather than the reduced form, thus allow some variables to be mutually contemporaneously dependent [7]. Improved prior is required to allow such more flexible implementation of Minnesota prior. This improvement was captured by Sims and Zha (1998) [8]. They added a prior for covariance matrix of the residuals in the form of inverse-Wishart distribution. This generalized prior is often called inverse-Wishart prior even though priors for coefficients are still normally distributed.

It must be pointed that at the extent of this generalization, the weight parameter must be erased from the Minnesota prior. The reason is that once we allow contemporaneous relationships in VAR model, we get, in fact, simultaneous equations. In this case it is impossible to distinguish in each equation which variable is in fact dependent, because it depends on the chosen transformation from structural to reduced form. As a result we cannot distinguish between the lags of the dependent variable ('own lags') and lags of other variables ('others lags'). Therefore inclusion of weight parameter is of no meaning.

There are many different extensions to Minnesota prior that allow for different flexible specifications. These extensions are usually referred to as *modified Litterman prior* [9].

Random walk averaging prior There are myriads of different priors suitable for BVAR estimation. Random walk average prior, suggested by LeSage and Krivelyova (1999) [10], alters not only the variances of parameters but also their means. Basic idea is very similar to the idea of weight parameter under matrix notation. In the weight matrix we can express

higher impact of certain variable to the dependent one through widening the prior distribution that is spread around zero mean. The ratio of the weight parameters in one row of the matrix (one equation in the VAR) express the relative importance of the parameters in the equation. However, under Minnesota prior, this relative importance is represented only through the variance of the prior.

On the other hand, we are interested in setting the mean of the prior to be correspondent with the importance of the parameter in random walk averaging prior. This can be done in an analogy with weight matrix in a following manner. Firstly, let us normalize each row of the weight matrix (the row then sums to 'one'). Secondly, we use this new parameters as the mean of the priors for the first lag of corresponding variable (means of all subsequent lags remain being 'zero' - as in the Minnesota prior). Sum of the means of the priors in each equation is equal to 'one'. In fact, the means of the priors of the first lags of all variables sum to 'one' and the others are zero. It is similar to Minnesota prior where the mean of the first lag of the dependent variable is equal to 'one' and all the others are zero (thus also means sum to 'one'). Hence, both discussed prior are based on the assumption of random walk behaviour.

The advantage of random walk averaging prior is that it enlarges the distinction between important and unimportant variables in each equation. It can also suppress the autoregressive behaviour that is always heavily imposed by Minnesota prior. Serious disadvantage is that the time series for the variables in the model need to be scaled or transformed to have equal magnitudes. Otherwise, it would make little sense to say that the mean value of the inflation is equal to one half of US GDP measured in US dollars. This shortcoming may limit the usage to certain data set or call for some transformation of the data. More detailed discussion, computational solution and illustrative example can be found in [2].

2.6.8 Bayesian computer software

The myriad of possible priors and likelihoods, we can choose, make it difficult to construct a universal Bayesian computer package. Several software packages that are useful for doing Bayesian analysis in certain classes of model exist though many Bayesian econometricians create their own programs in programming languages such as e.g. matrix programming environment Matlab.[14] Some useful Bayesian packages such as BUGS

[14]Though implementation in quite traditional programming language Fortran can be also found [11].

(Bayesian Inference Using Gibbs Sampling), BACC (Bayesian Analysis, Computation and Communication) or LeSage's Econometric Toolbox [2], which was extensively used in this work, can be linked to Matlab.

3 Inflation Forecasting

3.1 Phillips curve

The traditional Phillips curve is heavily discussed trade-off relationship between inflation and unemployment It is the most common econometric basis for prediction of inflation, however the usefulness of the Phillips curve has been questioned by several authors. Critiques follow two main directions. The first is that there exist economic variables (e.g. confidence indices) that allow for more accurate inflation forecasts [12], moreover parameters of models based on Phillips curve change over time. The second is that any forecast based on Phillips curve is worse than naive forecasts or simple univariate autoregression [13]. This stream of critique makes Phillips curve certainly inappropriate for efficient inflation forecasts.

3.2 Other approaches

There were attempts to forecast inflation using large number of different models, each with a single predictor. However, most of the models do not forecast inflation more accurately than naive steady state forecasts. Moreover, even when a model has relatively higher predictive power, it tends to be unstable over time. Thus even if model has good predictive power in one subperiod has little or no propensity to have good predictive power in another subperiod.

Two approaches recently proved improved forecasting accuracy. First is based on using large datasets - large number of predictive variables. It is thus necessary to impose restriction on the huge amount of parameters e.g. by employing BVAR estimation. Second approach consists of averaging the forecasts of different models. This approach dates back to the work of Granger and Bates (1969) [14]. It is widely discussed that the best predictive performance is obtained by equal weight averaging of forecasts from many models. It is unlikely that the 'true' optimal weights of many different models are exactly equal, but the error, introduced by estimating these weights, may more than offset any benefits [15]. However, Bayesian model averaging (BMA) is successfully used in current literature. This method averages the models according to their posterior model probability (see section 2.4) [11].

3.3 Inflation targeting

Forecasting of inflation rates has become crucial for both policy makers and private agents who try to understand and react to Central Bank decisions, since many Central Banks implemented inflation targeting rules instead of controlling monetary aggregates. The transmission of monetary policy to inflation and other real economic variables is not perfectly understood. Nevertheless policy interventions are assumed to affect the economy with a considerable lag. Successful policy thus depends upon accurate forecasts of relevant variables, mainly target variable - inflation. Apart of exploring the monetary transmission, it is forecasting issue that needs to be dwelled on for successful monetary policy conducting.

Since this work is concerned with forecasting inflation in Czech republic, it must be said that Czech national bank started to be pure inflation targeter by January 1998.

3.4 Pseudo out-of-sample forecasts

Literally indefinite amount of models can be produced attempting to forecast inflation. However, an economist would have to keep all her models in mind (computer) and run time consuming real-time experiments to validate the models. Another option is to run pseudo out-of-sample experiment. Once the model is specified, we allow it to use only some old data and produce forecasts for the horizon for which the real data are also already known. It allows immediate comparison between model pseudo forecast and the real value. This is by far the most common procedure in the literature and it is also used in this work.

However, Litterman [5] warns against overestimating forecast ability of pseudo out-of-sample forecasts, because it is very difficult to know what kind of after-the-fact information was used for generating the specifications of the model. Regarding VAR and BVAR modelling it is mainly the choice of the variables and the tuning of the model. This doubt calls for testing the robustness of the model (mainly the stability of the parameters over different periods). Revisions of the data must be also considered. Only data that were readily available might be used.

3.5 Forecasting and vector autoregression

Estimated values provided by *reduced form* of VAR are in fact one-step-ahead forecast. We are actually more interested in multi step forecasting. Common solution (also used in this work) is to use the chain rule of forecasting. This means that estimated one-step-ahead forecasts are taken as the basis for two-step-ahead forecasts and so on.

Another option is estimating the model whose dependent variables are directly the (pseudo) future value and explanatory variables are lagged by intended forecasting horizon (let's say one year). Thus only the information that was available one year before is used for estimating the dependent variable. Similarly, using current values we can directly estimate the value of the dependent variable one year ahead. This approach can also be used in simple regression model [11]

Additional variables that are not estimated by the model can be included on the right hand side of the VAR equations. These variables then can influence dependent variables. This variables are thus exogenous (exogenous external block of variables) [16]. Introducing this exogenous block is usually connected with assumption of 'small open economy'. This block then represents influence of foreign sector onto domestic economy (typically oil prices or GDP growth of some important trading partner). However, inducing this exogenous block effectively disallow using multi-step chain forecasting, since it would be then necessary to employ separate models for each exogenous variable. In this context direct estimation described in the last paragraph appears to be very beneficial.

3.6 Literature review

This section covers the most important results that influenced this work. It is in no meaning comprehensive summary of literature concerning inflation targeting.

Litterman [5], apart of other already discussed things, provides short discussion about fundamental scarcity of the data in economics. One of the reasons is the phenomenon of business cycles. In fact, business cycles need to be predicted, but even if we measure macroeconomic data with higher frequency (e.g. monthly) and we have the time series as long as possible, the number of the business cycles included in the data set is quite low, thus hard to predict by statistical methods. Moreover structure of economy and government policies are constantly changing. However, this uncertainty

about the structure of the economy can be overcome by employing Bayesian estimation techniques. Litterman provides results from five year real-time experiment and shows that BVAR estimation and forecasting show better performance than commercially available forecasts. He also partly defends common criticism that the time-series models never forecast turning points, nevertheless, he admits that time series modelling is suitable for short-term forecasts, whereas structural models are needed to capture the turning points.

One of the most influential articles for this work is by Fabio Canova [1]. This 'horse-race' article compares different approaches to forecast inflation in G7 countries. ARIMA models are used along with bivariate theory-based models, VAR and BVAR models. Important forecasting horizons are set to be one quarter, one year and two years (used also in this work). All the models are estimated recursively and pseudo out-of-sample forecasts are compared to the real values via Theil-U statistics (see in detail below).[1] AR models are generally better than naive steady-state forecast. Theory-based models improve over univariate specification only at long horizons. Anyway, theory-based models tend to be unstable over chosen subperiods and over different countries (Phillips curve specification is relatively stable, but not very useful). BVAR models are better then VAR, but the perception of turning points is unimpressive. Significant improvement over univariate models can only be attained by employing time-varying coefficients.

Bikker (1998) [6] provides estimations of BVAR models for EU-7 and EU-14 countries and compares forecasts of these models to forecasts by OECD. Author used 15 time series concerning the most important economic indicators, but each model consists of eight variables only. BVAR forecasts compare well to OECD forecasts at both one year and two years forecasting horizons.

Authors Ballabriga and Castillo (2003) [16] provide BVAR model for forecasting aggregate EMU inflation. They conclude that forecasting yields favorable results with respect to forecasts of other analysts. Important influence to inflation comes from external sector - GDP growth in outer states and commodity prices.

Bayesian model averaging method is succesfully used to forecast inflation by Wright (2009) [15] and Jacobson and Karlson (2004) [11]. Both articles conclude that Bayesian averaging outperform simple equal-weight averaging models. It can also be concluded that there is a dramatic difference in forecasting performance between one year a two years forecasting horizons.

The most related paper from Czech economic environment is Borys

[1]Theil-U statistics was successfully used already by Litterman as a Theil coefficient.

and Horváth (2007) [17]. Paper concerns the understanding the transmission mechanism of monetary policy to inflation and other real economic variables. Principle component analysis is employed onto large number of economic time series to overcome the problem of limited number of variables that can be included in VAR model. Factor augmented VAR (FAVAR) model is subsequently estimated. Provided discussion concerning the data is important for this work. Sample is restricted to the data from 1998 on, since inflation targeting has been adopted by Czech national bank by January 1998. While other studies often employ quarterly data, given the length of the sample authors decided to work at monthly frequency (and so it is in this work). Authors also discuss the drawback of VAR literature for its backward-looking dimension. On the other hand, inflation targeting monetary policy is typically forward looking. However, this fundamental drawback of VAR can be weaken by including forward looking variables.

Modelling of the inflation in the Czech republic has also been undertaken by Golinelli and Orsi (2001) [21] using multivariate cointegration empirical models. Other papers focus more on evaluation of inflation targeting in Czech republic (e.g. [22], [23], [24], [25], [26] etc.).

Joiner (2002) and Bloor and Matheson (2009) [7] used BVAR to describe monetary policy effects in Australia and New Zealand respectively. However, there must be used methodology which incorporates restrictions in both contemporaneous and lagged relationships in the model [8] to decompose particular effects in BVAR.

Finally, it must be stressed, that there are already some developed methods that allow for using Bayesian framework in identification of popular DSGE models (e.g. An and Schorfheide, 2007 [19]).

4 Data

Following Borys and Horváth [17] we use data from January 1998. This is the time of adoption of pure inflation targeting by Czech National Bank. Due to the reduced length of the sample we decided to use monthly data. When it is possible, we chose data in the form of percentage change on the same period of the previous year. Data were downloaded on the 8th of April 2010 and have not been furthermore updated in any way.

4.1 Data sources

4.1.1 ARAD database

The data were downloaded from two different sources. First source was the database ARAD. ARAD is a public database presenting time series of aggregated statistical data. Most of these data origins from Czech National Bank (CNB), however data from external sources (mainly Czech Statistical Office - CZSO) are also provided. Choice of the time series is heavily limited by their availability. Only data that covers whole period from January 1998 were used, altogether 30 time series:

- discount rate, lombard rate - by the end of each month
- PRIBOR (3 months, 6 months, 1 year)
- registered unemployment
- sources and the use of the monetary base and their components (levels)
- GDP (by the type of expenditure) and its components - percentage change on the same period of the previous year
- gross capital formation and its components - percentage change on the same period of the previous year
- export and import and their components - percentage change on the same period of the previous year

The last three groups of data origin from Czech statistical office and are available only with quarterly frequency. Monthly data were evaluated by employing linear interpolation between two neighbouring values. As a result these last three groups of data are generally available with one quarter delay.

4.1.2 OECD database

As a second source serves database OECD Stats that is available for subscribers. 49 time series were downloaded from this database, however only 27 of them are available immediately:

- consumer and producer prices and their components (percentage change on the same period of the previous year)
- share prices, industrial production (ratio to trend, smoothed), business confidence - percentage change on the same period of the previous year
- overnight interbank rate, exchange rate (USD, percentage change on the same period of the previous year of monthly averages)
- block of indicators from business tendency surveys (concerning manufacturing industry, construction industry, retail trade, etc.) - levels

Other time series are only available with one quarter lag (see discussion below):

- industrial production, retail trade, broad money, imports and exports - percentage change on the same period of the previous year
- harmonized unemployment
- leading indicator (amplitude adjusted) - OECD main cyclical indicator providing qualitative information on short-term economic movements
- GDP and its components, service exports and imports (balance of payments), total exports and imports and their difference - percentage change on the same period of the previous year
- current account as a percentage of GDP

Last two groups of time series are published only with quarterly frequency thus linear interpolation was again employed.

Preceding lists of used variables serve only as a summary. Complete list of data can be found on the attached CD and their detail description is publicly available.

4.2 Forecasting horizons

We use percentage change on the same period of previous year of consumer price index as inflation data. This choice is quite common (e.g. [11]). This time series origins from OECD data, thus this choice allows for direct comparison of forecasting performance between other European countries in future work. However, it obscures the comparison to CNB forecasts.

All the data were divided into two groups. In the first group, data were available until March 2010 - together 44 time series (27 by OECD, 17 from ARAD). In the other group, the data were available only up to December 2009 - 35 additional time series (22 by OECD, 13 from ARAD).

If we would like to provide reliable forecast one quarter ahead (out-of-sample forecast) and we use only the data from the first group, it is enough to compute the forecast three steps (months) ahead. However, if we allow for using data from second group we need to employ longer forecast (up to 6 steps) in order to get forecast one quarter to the future. It must be admitted that in estimating our model this has not be taken into account and the data from both groups were mixed.

Forecasting horizons were set to 3, 12 and 24 steps (months) for one quarter, one year and two years predictions, respectively.

Revisions of the data were not analysed. Model must be adjusted for both availability and revisions of the data in the future work.

5 Methodology

For all the computations, except of some basic manipulations in MS Excel, was used mathematic environment Matlab version 2007b[1]. Econometric toolbox for Matlab by James P. Le Sage [2] was extensively used for estimating VAR and BVAR models.

5.1 RW and AR model

Ljung-Box test and Augmented Dickey Fuller test were used for initial tests whether inflation data are random or follow random walk or can be generally further modelled. Both Ljung-Box test and Augmented Dickey Fuller test are preprogrammed in GARCH toolbox as `lbqtest` and `dfARTest`, respectively. Lag structure of Ljung-Box is chosen to be 24, which also serves as degrees of freedom for chi-square distribution (asymptotic distribution of Q-statistice that is calculated employing Ljung-Box test).

Random walk (or naive or steady state) forecasts are easily performed only by shifting the vector of data forward by required forecasting horizon (forecast horizon - steps of forecast - were in all functions denoted as parameter k, thus we will reffer to forecast horizon also in text as k). These RW forecasts then serve as a benchmark for forecasts provided by all other models.

ARMA modelling is based on function `armax` from System Identification toolobx. This function estimates ARMA model, once lags of both AR and MA processes are stated. However, it is straightforward to program short script (using other functions from the toolbox such as `selstruc` and `predict`), which automatically chooses optimal lag for AR process (according to Akaike criterion), estimates the ARMA model and provides vector of forecasts in the required horizon. Except of data and forecast horizon, we need to input also lag of the MA process, since this cannot be computed by the procedure. Thus best AR lag is set according to the optimal estimation of the model (Akaike criterion), whereas optimal MA lag is set manually according to pseudo out-of-sample forecast performance (from comparison of forecast and the data).

When computing the forecast at time t, model uses only data upto time $t - k$. Model computes forecast k-steps forward employing the chain rule. Then time t is moved one step forward (thus for estimation we use one further value of the data). Model is re-estimated and new k-steps ahead forecast is

[1] Version 7.5.0.342, released on 15 August 2007

computed. The result of the procedure is then the vector of forecasts.

It is clear that pseudo out-of-sample forecast close to the beginning of the data cannot be computed (up to forecast horizon plus number of lags) and few more values are not reliable, since model estimation is there based on a very short period of the data. It makes sense to test forecast accuracy only several steps from the beginning of the data (we usually use whole first half of the data for model learning and only the second half is compared to the actual forecast; exceptions are explicitly stated in results chapter). Short script, which was written for estimating AR model, employs functions that were programmed to compute following statistics concerning the accuracy of the forecast.

5.2 Forecast comparison

Forecast accuracy is calculated from comparing pseudo out-of-sample forecasts vector and the vector of the real data. Forecast is absolutely accurate when both vectors are equal. There are several statistics used for comparing forecast to the data.

5.2.1 Mean square error

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(f_t - y_t)^2 \tag{5.1}$$

- f - vector of forecast
- y - vector of data
- T - the length of the time period that was used for comparing the forecast and the data

Mean square error is common statistics used for estimation of differences between two vectors (e.g. also between data and regression line - errors). The lowest bound is obviously zero (absolutely exact forecast), there is no upper bound. The lower value the better forecast. Compared to Mean absolute error (MAE), MSE places a greater penalty on large forecast errors. However, both these statistics fail to provide information on forecasting accuracy relative to the scale of the series examined.

5.2.2 Theil statistics

This subsection follows [20]. We use two different Theil statistics in this work. First of these provides scaling to MSE.

Theil1

$$TheilU_1 = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(f_t - y_t\right)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} y_t^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T} f_t^2}} \tag{5.2}$$

Meaning of the symbols is the same as in the definition of MSE.

This statistics takes into account scale of the time series and the forecast. It is bounded between 0 and 1, with values closer to 0 indicating greater forecasting accuracy. However, this statistics proved to be misleading in some cases. Our data do not have zero mean (in fact, except of very few values, inflation is positive). Thus forecast can systematically underestimate the data (produce significant errors) but the scaling term $\sum_{t=1}^{T} f_t^2$ remains very low. On the other hand, if the forecast systematically overestimates the data (such that the MSE can be equal in both cases - and so is the nominator of Theil1 statistics), the term $\sum_{t=1}^{T} f_t^2$ is now much higher and as a result Theil1 statistics is lower and signalizes improved forecast. Moreover, it may easily happen that even obviously worse forecast (by MSE and from the charts comparing forecasts with the data) shows improved forecast accuracy according to Theil1 statistics. To sum up, this statistics can be used for forecast comparison only combined with non-scaled statistics like MSE and, preferably, with graphical representation.

Theil2 The notation follows [20]. However, in the literature is often used Theil-U statistics [1], Theil statistics or Theil coefficient [5]:

$$TheilU_2 = \frac{\sqrt{\sum_{t=1}^{T-k}\left(f_{t+k} - y_{t+k}\right)^2}}{\sqrt{\sum_{t=1}^{T-k}\left(y_t - y_{t+k}\right)^2}} \tag{5.3}$$

- f - vector of forecast
- y - vector of data
- T - time period that was used for comparing the forecast and the data (see discussion above in previous subsection)

Theil2 statistics is the most extensively used statistics for comparing pseudo out-of-sample forecast to the real data throughout the work. Detailed

inspection shows that in the nominator we find again square root of MSE. In the denominator is square root of errors of random walk (steady state) forecast. Theil2 statistics thus compares given forecast to potential random walk forecast.

The lower is the value of the Theil2 statistics, the better is the forecast. '0' represents exact forecast. '1' describes forecast that is the same accurate as random walk forecast. Values lower than '1' signalize that forecast is better than steady state forecast, on the other hand, values higher than '1' signalize that the forecast is worse than the random walks forecast, thus quite useless.

Theil2 statistics allows for comparison between the forecasts until the potential random walk forecast remains unchanged (e.g. for one forecasting horizon). However, comparison between the forecasts is problematic when we also change random walk forecast (change forecasting horizon, truncate the data,...). It cannot then be distinguished, whether given forecast improved or random walk forecast became less accurate.

5.3 VAR and BVAR - pseudo out-of-sample forecast

Production of pseudo out-of-sample forecast is one of the main issue of estimation procedure in this work. Generating these forecasts is almost the same for both VAR and BVAR, thus it will be described at once.

Procedure is based on two important functions from James P. LeSage's Econometric Toolbox [2]. These are called `varf` and `bvarf`. They compute multi-step ahead forecasts of VAR and BVAR model, respectively. The input of the functions is matrix of the data, number of lags of the model, the number of required forecasting steps and the beginning of the forecasting (usually the end of the data that are available to the model). Function `bvarf` also uses parameters specifying the prior of BVAR model.

The functions estimate VAR and BVAR models. Multi-step forecasts are computed for all variables of the estimated model. Forecasts are produced by chain procedure up to given horizon. Each step of the forecast is written in the output (the 12th row of this forecast is the one year ahead forecast from the end of the data, when employing monthly data). Thus if we want e.g. 12 steps ahead forecast we need to compute all the forecasts up to 12 and then we use only the last one.

I programmed two new functions `varpred` and `bvarpred` based on the two preprogrammed. These functions are intended to compute vector of pseudo out-of-sample forecast for given forecasting horizon. Procedure

is multi-step. The function `varpred` takes some part of the data (e.g. first half) and let the function `varf` compute forecasts up to the given forecasting horizon. `varpred` function then saves the last value (row) of the forecast, because this is the forecast for given horizon. Then one more value (row) of the data is added (procedure moves one step forward), model is again re-estimated (!), and next forecast is computed (by `varf` function). Procedure moves up to the end of the data (in fact only up to k steps before the data end, so even the last forecast can be compared to the last value of the data). The result is vector of the forecasts, computed for given forecast horizon.

Finally, this forecast vector is compared to the data and values of MSE and both Theil statistics are computed. Vector of forecasts is also used in graphical representation.

5.4 VAR and BVAR - finding good predictors

Procedure that is described in previous section can be used after the predictors of inflation are chosen. Following procedure describes how to choose predictors according to their forecast accuracy.

Procedure is a straightforward generalization of previous functions `varpred` and `bvarpred`. We use these functions whose input is given matrix of data (inflation plus chosen predictors) and output is, among others, Theil2 statistics. Newly programmed functions `varpredn` and `bvarpredn` use parameter n that defines how many variables should be included in the VAR or BVAR model (including inflation). Functions also use all the data that are available[2]. Functions generate all the combinations of predictors (according to the parameter n and number of used time-series) and produce the matrix of corresponding data for each combination. Functions `varpred` and `bvarpred` are consequently used for each generated data matrix and the forecast accuracy statistics are saved along with the correspondent choice of the predictors. Finally, predictors combinations are sorted with respect to Theil2 statistics. As a result we get MSE and Theil statistics and the correspondent predictors for the best models according to Theil2 statistics.

[2]The more time-series we use, the longer is the computational time. This might be substantial limit to the amount of used time-series. In our case we needed to divide the data to halves to lower the computational burden.

5.5 Turning points

I programmed two additional function to analyse further the quality of the forecast. Function `tp` analyse whether the forecast is able to find turning points in the data.

Let us have vector of data y. We define upturn of y at t if true:

$$y_{t-2} > y_{t-1} > y_t < y_{t+1} < y_{t+2} \tag{5.4}$$

downturn of series y_t is then defined by:

$$y_{t-2} < y_{t-1} < y_t > y_{t+1} > y_{t+2}. \tag{5.5}$$

The definition allows that the trend of the data can be changed (if the change is not direct) without measuring the turning point. On the other hand, we can measure the turning point and the data do not show any long-term change in the trend. Turning points analysis must always be accompanied by graphical representation.

Function `tp` is designed to find all the turning points in the data y and then to find out whether turning point (of the same kind) is present also in the forecast. However, forecast usually hardly predicts the turning point exactly at the same date (especially when employing monthly data). That is why we add parameter q that can be manually set. If $q = 0$ then turning point of the forecast must be found exactly at the same date as the turning point in the data to be evaluated as the successfully found turning point. However, if $q = 2$[3] then turning point is evaluated to be successfully found even if turning point in the forecast is found up to 2 positions earlier or 2 position later.

Analysing the turning points can be used as an additional statistics to compare quality of the forecast. However, if there are not enough turning points in the data, the results of this statistics are less useful.

5.6 Direction of forecast

This additional function analyses, whether forecast follow the same direction (upward or downward) as the data, over the forecasting horizon (let us say one year). Function analyses whether the value of the data one year ahead is lower or higher than the current value. The same is evaluated for the forecast. If the direction of the data disagrees with the forecast function

[3]This choice we use throughout the work.

adds '1' to 'wrong direction variable'. Function scans through the data. Finally, function divide its 'wrong directions' by the number of points that it has searched through. Output e.g. 0.40 means that at the 40% of the data points the one year ahead forecast is higher than is the current forecast (that have been established one year ago) but the future data value will be lower than is the current value (or vice-versa). The function always analyses only difference between two points that can be quite far from each other, thus we can hardly speak about 'slopes' of the forecast and of the data.

The best forecasts should receive the value close to zero, on the other hand, completely wrong forecast receives '1'.

5.7 Out-of-sample forecasts

Out-of-sample forecasts need only tiny modification of `varpred` and `bvarpred` functions. Once the model is chosen (predictors and number of lags), we can use these functions to provide the vector of the forecast. The only difference is that now (functions `varpredout` and `bvarpredout`) we use the final values of the data to compute forecast by chain rule (the end of the data were originally not used in this respect). No MSE or Theil statistics can be calculated, since we do not have any real data to be compared. The only output is then the vector of the forecasts that is as long as is the forecast horizon.

It must be noted that only the last value of the out-of-sample forecast need the last value of the data, since the forecasts are still computed on the basis of given forecast horizon.

All the programmed functions with comments about its inputs and outputs are given in the Appendix A.

6 Results

6.1 Random walk and AR model

Inflation data time series consists of 144 values, however only last 80 observations are shown in the chart 6.3, since only these last 80 values were compared with different forecasts from AR model. Previous data are left for model learning. Three cycles can be seen in the figure 6.3, the first two quite indistinctive and, on the contrary, the last one longer in period and much more apparent due to outstanding amplitude. Since we have just three cycles in the data to be forecasted, we can recall the discussion of the scarcity of economic data [5].

Data were tested by Ljung-Box test whether they are random. This hypothesis was rejected with p-value below computer discrimination ability.[1] The rejection of random data hypothesis can be also observed form autocorrelation functions (ACF) chart 6.1. Data were also tested by augmented Dickey-Fuller test for a unit root. Hypothesis was also strongly rejected.[2]

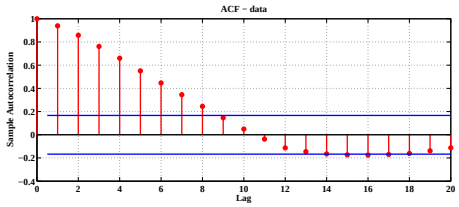

Figure 6.1: ACF - data

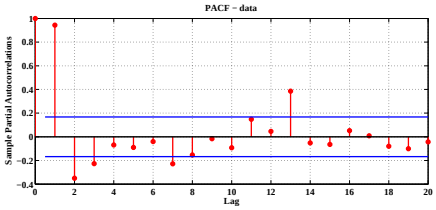

Figure 6.2: PACF - data

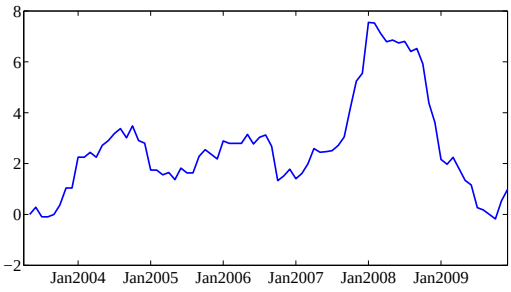

Figure 6.3: Data

Figure 6.4: AR - estimation

[1] pValue = 0; Qstat = 534.1447; CriticalValue = 36.4150
[2] pValue = 0; TestStat = -3.5524; CriticalValue = -1.9428

Data were modeled by univariate ARMA model. Since adding MA lags did not lead to improvement in forecast accuracy, AR(3) model was finally chosen (autoregression lag according to Akaike criterion, choice is confirmed by the chart of partial autocorrelation function - figure 6.2). Values, which can be estimated by the model using step-by-step calculation, are shown in figure 6.4 along with the data (blue line - data, green line - estimate, for more detailed legend - see below).[3]

Residuals of this model were again tested by Ljung-Box test. Hypothesis that the residuals are random was again rejected[4], however autocorrelation functions charts for both residuals (figure 6.5) and squared residuals (figure 6.6) do not show any autocorrelation effects remaining, except of some seasonality in residuals at one year period.

Due to no clear correlations remaining in the squared residuals, GARCH analysis was not undertaken.

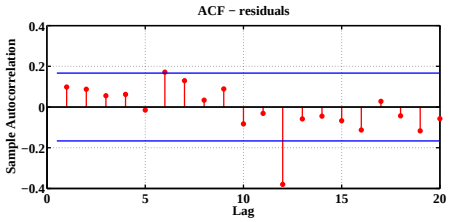

Figure 6.5: ACF - r

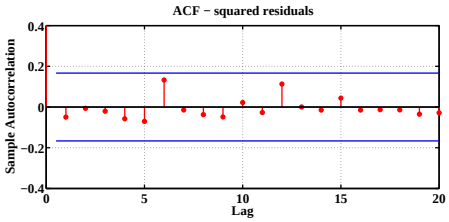

Figure 6.6: ACF - r^2

Vector of forecasts was estimated from AR(3) model by chain rule. Forecast was compared to the data in means of MSE and both Theil statistics. Values of two additional functions (concerning turning points and correct direction of forecast) were also calculated. Results are compared to steady state naive forecast (Random Walk - RW) and summarized in table 6.1. Numbers 3, 12 and 24 refer to forecasting horizon of one quarter, one year a two years, respectively.

Direct comparison between Random Walk and Autoregressive model is given by figures 6.7 to 6.12. In these figures (and all other similar figures) we skip legend since it would worsen the readability, so the legend is given now and for all:

- blue line - data (always the same values)

[3]This calculated values from the AR(3) model can be regarded to as one step forecast ($k = 1$).

[4]pValue = 0.0048; Qstat = 45.6736; CriticalValue = 36.4150

Table 6.1: RW and AR models results

Model	MSE	Theil1	Theil2	Turns	Directions
RW-3	1.3658	0.1770	1	0	0.39
RW-12	9.1068	0.4584	1	2	0.85
RW-24	7.1524	0.4404	1	0	0.68
AR-3	1.1734	0.1722	0.9263	0	0.42
AR-12	7.1944	0.5029	0.8881	2	0.85
AR-24	7.9378	0.6597	1.2767	1	0.64

- green line - forecast FOR given time (no matter at what time it had been generated)
- vertical axis - percentage change on the same period of the previous year of consumer price index (CPI) - provided by OECD Stats
- horizontal axis - time in months (period under consideration slightly changes among different models and horizons)

It is clear that RW forecasts only shift the data curve by given number of steps forward (this is the steady state forecast - the inflation after 12 months will be the same as it is today, figures 6.7, 6.9 and 6.11). AR model estimates the data quite successfully (fig. 6.4), on the other hand, it is not particularly suitable for forecasting, since it improves RW forecast just slightly (figs. 6.8, 6.10), and sometimes not even that (fig. 6.12).

Using the figures we may provide some discussion to the results in table 6.1.[5] Concerning Theil2 statistics, RW obviously returns value of '1'. AR-3 and AR-12 models are better than the RW, whereas AR-24 is significantly worse. More over two counter-intuitive results can be identified.

Firstly, Theil1 statistics is lower (better) for RW-12 than for AR-12, but MSE is much higher (worse). The reason can be explored by looking at appropriate charts (6.9, 6.10). RW-12 is worse forecast (especially in the end of the observed period), that is why MSE is high, however, RW-12 forecast is generally higher (containing higher values) comparatively to AR-12, which usually underestimates the real data. As a result scaling factor in Theil1

[5]We remind that the lower MSE the better the forecast. The same holds for both Theil statistics. Theil1 statistics is bounded between '0' and '1'. Theil2 compares forecast to steady state forecast (this benchmark is represented by '1'). The higher value of turning points statistics (maximum is 5) the better turning point prediction. The lower direction statistics the better the forecast follows the slope of the data (bounded between '0' and '1').

statistics is higher in the case of R-12 and resulting statistic value is lower, signalizing better forecast. This supports discussion in section 5.2.2 that Theil1 statistic must be taken into account with a great care.

Secondly, Theil2 statistics is better for AR-12 model than for AR-3. The reason lies in fact in RW models. RW-3 is still quite succesful, thus the basis of comparison is better (the denominator of Theil2 statistics is low)[6]. On the other hand, RW-12 forecast is poor. As a result AR-12 shows bigger improvement with respect to RW-12 than AR-3 to RW-3. Hence, it must be always taken into account that value of Theil2 statistics depends heavily on how successful is the random walk forecast. The accuracy of RW forecast must be considered also in the discussion of results of VAR and BVAR models represented by Theil2 statistics.

Concerning the turning points, there are only 5 turning points in the data, two downturns and three upturns. Two turning points were identified by both RW-12 and AR-12 (one upturn in the beginning of the data and one downturn in the middle). However, these findings were just fortunate. Identified turning points in the forecast represent only local change of the slope instead of long-lasting change in trend. The only conclusion is that this statistics (having only 5 turning points in the data) does not say much about forecast accuracy in the mean of prediction of turning points.

We can non-surprisingly conclude from the second additional statistics, that one quarter forecast follows the direction of the data best. However, interesting is very poor performance of one year ahead forecast. High value of this statistics alerts that forecast moves often in the opposite direction than the data. This can be confirmed from the charts 6.9 and 6.10. Forecast and the data seem to be in anti-phase for both RW-12 and AR-12 models. This is no longer true for 24 steps ahead forecasts and thus the statistics 'directions' improves.

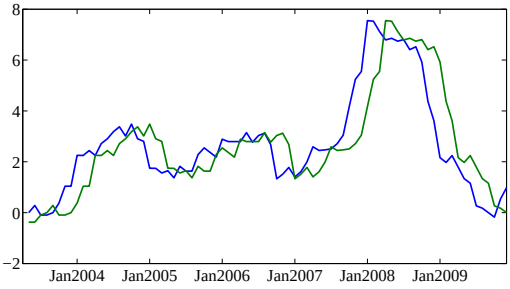

Figure 6.7: RW-3

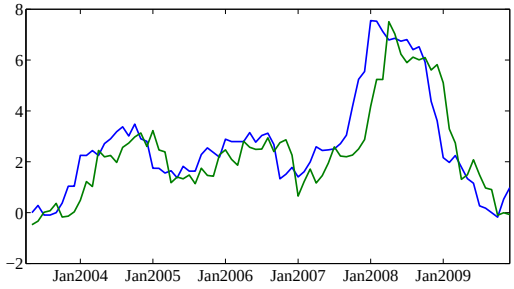

Figure 6.8: AR-3

[6]The square root of MSE of RW models is in fact the denominator in Theil2 statistics

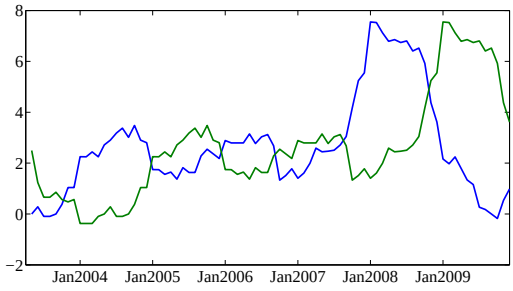

Figure 6.9: RW-12

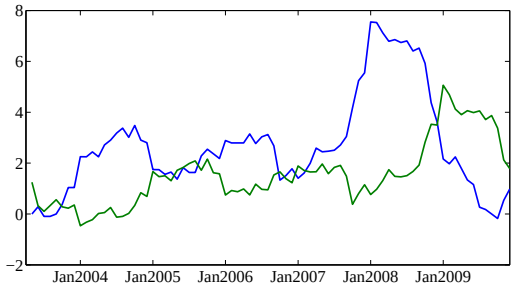

Figure 6.10: AR-12

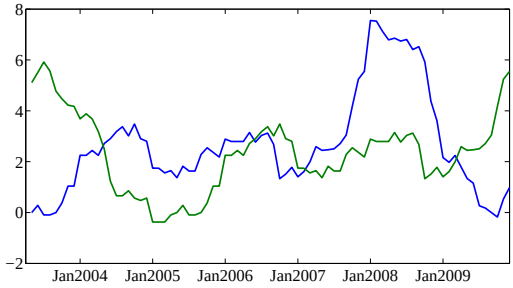

Figure 6.11: RW-24

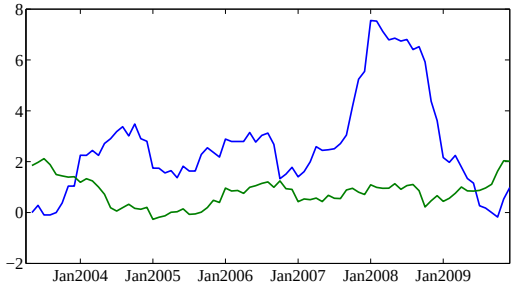

Figure 6.12: AR-24

To conclude, AR model does not provide accurate forecasts and, moreover, it does not markedly improve Random walk forecasts (in fact AR forecast is even worse under 24 steps horizon). This section also covered the principles of presentation of the results that will be kept throughout the chapter.

6.2 VAR and BVAR

This section covers the most important results of this work - results from both VAR and BVAR modelling. Results are divided according to forecasting horizon. However, first we describe the estimation procedure, discuss the optimization of BVAR models and comment out-of-sample forecasts.

6.2.1 Estimation procedure

First we use functions described in methodology to find the best predictors according to the Theil2 statistics. We use only second half of the forecast and the data for the comparison.[7] First half is devoted only to learning of

[7] p = 0.5

the model. (When we compute the Theil2 statistics, the truncated sample is furthermore shortened by the number of forecasting steps. Data that are used for comparison in the case of two years forecast horizon are thus too short. That is why we use 70% of the data for the comparison in this case.)

Unfortunately, due to computation burden, it is impossible to analyse all 79 time series at once[8]. Data set was thus divided into two groups[9]. In the first half we chose ten best predictors, according to the procedure described in the next paragraph. These time series were then added into second half and the search for best predictors was repeated.

This time consuming procedure was undertaken simultaneously for both VAR and BVAR models (the sets of best predictors were very similar for both types of models). We considered models with up to three dependent time series (equations in VAR)[10]. Bigger models would have allowed much more combinations of the predictors and thus computational burden would have risen substantially. The procedure was, of course, repeated absolutely separately for each forecasting horizon. Number of lags was also varied during the estimation procedure (best forecast accuracy has been usually found with 3 to 5 lags). Finally, top ten predictors were identified.

Detailed analysis was performed among this reduced dataset, employing bigger models (arbitrarily big, e.g. including all ten time series) and concerning the stability of the respective models. Reduced data were produced for this reason. Data were shortened from the beginning by 12 and 24 values, respectively. Thus the learning sample differed significantly. It might have been more accurate to delete some data from the period where forecast is already compared to data, however, too short data would have been then left for the forecast comparisons. Anyway, some models proved to be unstable even after this slight change, whereas others remained in top rated models. These stable models were then chosen (with optimized number of lags) as the best VAR and BVAR models for each forecasting horizon. The results of these models are shown separately for each horizon and are denoted by numbers (VAR2, BVAR2).

6.2.2 Optimization of BVAR models

BVAR models were further optimized in order to get improved forecast by adjusting the parameters defining the variances of the priors. Parameters were optimized 'manually' one by one and forecast was always immediately

[8] Will be significantly improved by using more powerful computer in future work

[9] Division follows the data availability, as discussed in the chapter Data

[10] $n = 3$

computed. Optimization continued until change of single parameter did not provide improved forecast (according to Theil2 statistics). This procedure, of course, does not necessarily find global extreme (the most accurate forecast). However, forecast accuracy does not vary dramatically when we fine tune the final parameters. It is also checked, whether it is not advantageous to use different number of lags with new parameters (if so, then parameter optimization must be undertaken again with new number of lags). Optimized BVAR models are denoted by letter 'b' (e.g. BVAR2b) and optimized parameters of prior variance are shown along with used lags. Used predictors are the same as in the correspondent non-optimized model.

6.2.3 Out-of-sample forecasts

The out-of-sample forecast is computed in the same manner as pseudo out-of-sample forecast. Only such data are used that were available 3 (or 12 or 24) months before the declared forecast. Thus we need all the data (up to December 2009) only for the last value of each forecast.

Forecasts are also compared to inflation forecasts of Czech National Bank (CNB). CNB provides its forecast that are based on the percentage change on the same period of the previous year of the consumer price index. This time series is measured by Czech Statistical Office. However, the methodology of the Czech Statistical Office is obviously (see table 6.2.4) different from OECD methodology.[11] Hence, CNB forecast is only partially comparable to provided forecasts.

One and two years forecasts can be compared to real data in the future. Such real-time experiment will allow for decision, which models are stable over time. However, all models seem to underestimate inflation in years 2010 and 2011, at least when compared to CNB forecasts.[12]

Following three sections concerning three forecasting horizon are organized as follows. First we revise the predictors that are used by all the models concerning given forecast horizon. Predictors for each model and the number of lags are then presented. Models are enumerated (e.g. VAR2 and BVAR2), the same number for both VAR and BVAR means that the same predictors were used for both. Optimized BVAR parameters are also given. Subsequently the table with forecasting comparison results is shown. Finally, out-of-sample forecasts are provided. Some comments on the results are provided continuously.

[11] Notes on the methodology can be found on the web pages of both institutions www.czso.cz; www.oecd.org.

[12] Some more comments are provided in the chapter Discussion.

6.2.4 Forecasting horizon - one quarter

For this horizon, three different combinations of parameters were chosen. First combination consists of three time series (except of inflation series) and other two consist of four predictors (in fact three from the first model and one more added predictor in each of these subsequent models). Altogether we used 5 different predictors[13].

- Share price - provided by OECD, belongs to financial indicators, national all-share price index, based on arithmetic average of the closing daily values
- Leading indicator - provided by OECD, belongs to cyclical indicators, composite indicator formed by aggregating a variety of component indicators, designed to provide qualitative information on short-term economic movements
- Volume of sterilisation - provided by CNB, belongs to monetary policy factors of sources of the monetary base, end of month position
- Exports of goods and services - provided by CZSO[14], belongs to national accounts
- Consumer prices - food - provided by OECD, consumer price index comprising only food as a product group

It is of particular interest that no GDP measure appeared in these top five indicators, nor in the top ten (from national accounts only export have appeared). Moreover no measure of interest rate is present (e.g. PRIBOR), not speaking about unemployment at all. Broad money measure appeared in the top ten as well as business confidence. Appearance of cyclical indicator is not accidental at all - all the successful models used some cyclical indicator. This is also proved by forecasting with one year horizon.

Following three combinations of parameters led to successful model in both VAR and BVAR variants, BVAR models were subsequently optimized.

VAR1 (lags - 4), BVAR1 (lags - 6), BVAR1b (lags - 5)

[13]Details on each predictor can be found on the web pages of OECD and CNB, depending on the origin of particular time series.

[14]Data are collected (calculated) by Czech statistical office (CZSO) and published also through time series database ARAD that is run by CNB.

- predictors: Share price, Leading indicator, Volume of sterilisation

  VAR2 (lags - 4), BVAR2 (lags - 7), BVAR2b (lags - 5)

- predictors: Share price, Leading indicator, Volume of sterilisation, Exports of goods and services

  VAR3 (lags - 4), BVAR3 (lags - 7), BVAR3b (lags - 5)

- predictors: Share price, Consumer prices - food, Leading indicator, Volume of sterilisation

- BVAR1b - parameters: $\theta = 0.5$, $w = 1$, $\phi = 1$[15]
- BVAR2b - parameters: $\theta = 0.5$, $w = 1$, $\phi = 1$
- BVAR3b - parameters: $\theta = 0.5$, $w = 1$, $\phi = 1$

We must note that optimized parameters define bigger variance than the default and this means loosening the prior. Improved forecast is thus reached when the BVAR is less restricted, this may signalize that the assumptions of the prior are not perfectly fulfilled by the time series under consideration (as it is partly suggested by Augmented Dickey-Fuller test).

Table 6.2: VAR and BVAR results - one quarter

Model	MSE	Theil1	Theil2	Turns	Directions
VAR1	0.8917	0.1361	0.7845	0	0.31
BVAR1	1.1220	0.1537	0.8828	1	0.34
BVAR1b	0.8941	0.1377	0.7857	0	0.31
VAR2	0.9181	0.1355	0.7904	3	0.31
BVAR2	1.1178	0.1486	0.8872	0	0.33
BVAR2b	0.9262	0.1365	0.8055	3	0.34
VAR3	1.0368	0.1481	0.8466	1	0.39
BVAR3	1.1303	0.1537	0.8881	1	0.36
BVAR3b	0.9472	0.1422	0.8164	3	0.40

Few conclusions can be made from the results in the table 6.2. Firstly, BVAR in its default specification is always worse than the VAR. Secondly,

[15]It might be worth to repeat the default values: $\theta = 0.1$, $w = 0.5$, $\phi = 1$.

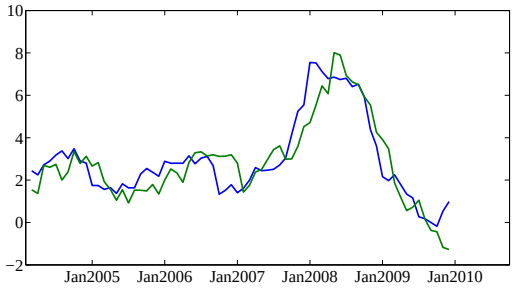

Figure 6.13: VAR1-3

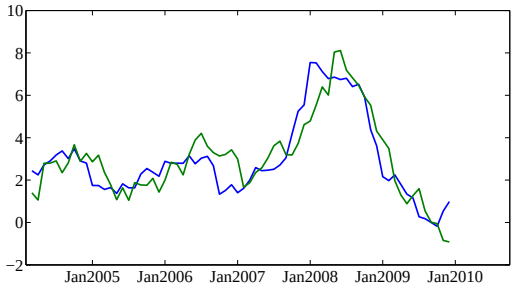

Figure 6.14: VAR2-3

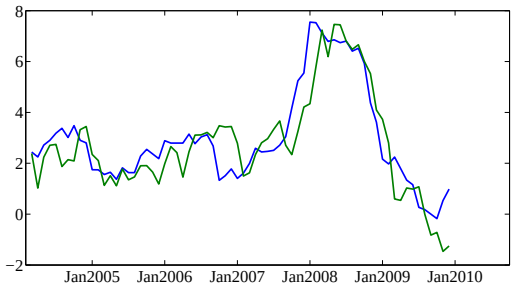

Figure 6.15: VAR3-3

once BVAR parameters are adjusted, BVARs reach the VAR models in forecast accuracy. Further we can state that turning points are found with more often by models that show generally better prediction ability (does not hold for VAR1 model). No clear consideration can be made about the accuracy of directions of the forecast (the third group of models falls out slightly worse).

It must also be noted that bigger models did not outperform these (quite) small in the forecast accuracy (not even in the BVAR form, which is quite surprising if not disappointing). **However, the most important result is that VAR and BVAR models (under improved specification) clearly outperformed autoregressive and random walk models at one quarter forecasting horizon.** This can be observed from both the table 6.2 and the figures 6.13 to 6.21.

Out-of-sample forecasts of particular models can be compared from the table 6.3. Only data up to December 2009 are used. Data from the first quarter 2010 that are already provided by OECD and CNB (in fact by CZSO) are added in table 6.3. OECD inflation measurement is probably heavily influenced by rise of the excise duty by January 2010, which affected mainly fuel prices. Data by CNB does not show any inflation jump in the

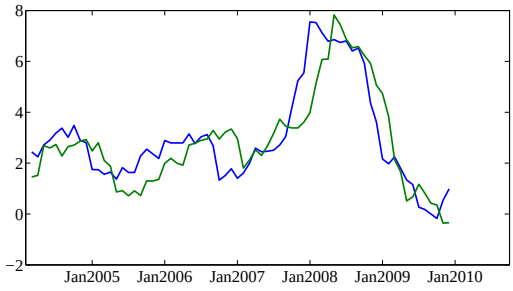

Figure 6.16: BVAR1-3

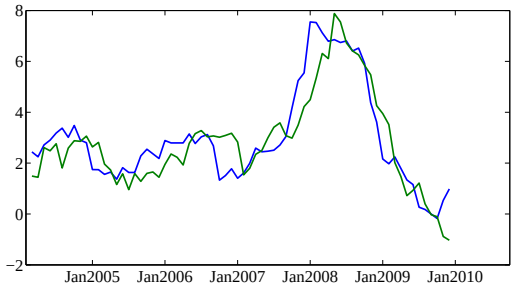

Figure 6.17: BVAR1b-3

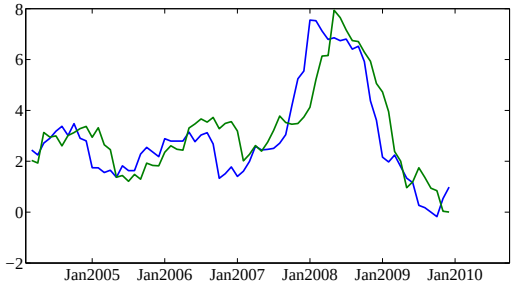

Figure 6.18: BVAR2-3

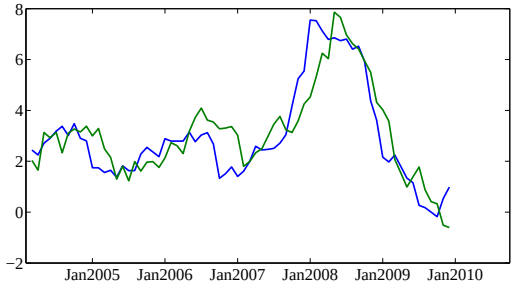

Figure 6.19: BVAR2b-3

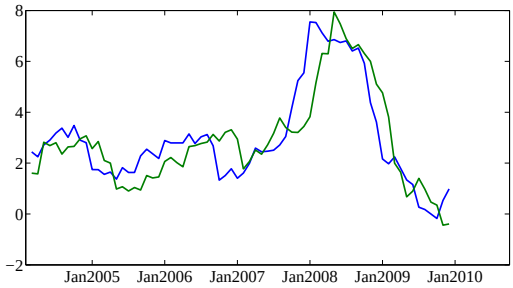

Figure 6.20: BVAR3-3

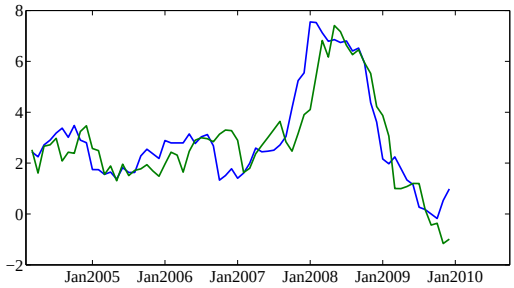

Figure 6.21: BVAR3b-3

beginning of the year.

It is clear that our forecasts follow decreasing trend from the end of the used for two months further. In fact the most restricted models (BVAR without adjustment) show the lowest decrease below zero inflation. However, forecast for the March 2010 was already very good.

Table 6.3: Out of sample - one quarter

Forecast-3	VAR1	VAR2	VAR3	BVAR1	BVAR1b
Jan2010	-1.3	-0.9	-1.0	-0.4	-1.0
Feb2010	-1.2	-1.0	-1.2	-0.3	-1.1
Mar2010	-0.1	0.2	0.1	0.6	0.1

Forecast-3	BVAR2	BVAR2b	BVAR3	BVAR3b	OECD	CNB
Jan2010	-0.1	-0.6	-0.3	-0.9	1.2	0.7
Feb2010	-0.1	-0.7	-0.2	-1.1	0.0	0.6
Mar2010	0.8	0.5	0.7	0.2	0.3	0.7

6.2.5 Forecasting horizon - one year

For one year horizon we estimated two different VAR models (VAR1 and VAR2) and two different BVAR models (BVAR2 and BVAR3) plus adjusted BVARs. Models' labels indicate the fact that two models (VAR2 and BVAR2) use the same predictors. BVAR3 model uses five predictors, whereas others only three. BVAR variant of the VAR1 is not shown since it does not follow high forecast accuracy of its counterpart. This holds vice-versa for BVAR3 model.

We need altogether six different predictors. The sets of the predictors used with one year and one quarter horizons are completely disjoint.

- Broad money - provided by OECD, belongs to financial indicators, M3 monetary aggregate, seasonally adjusted
- Order books indicator - provided by OECD, belongs to business tendency and consumer opinion surveys indicators
- PRIBOR1year - provided by CNB, one year PRIBOR
- Producer prices - manufacturing - provided by OECD, producer pice index that covers the branch of manufacturing
- Exports of goods - provided by OECD, belongs to international trade, seasonally adjusted
- Business confidence indicator - belongs to cyclical indicators, collected from Business tendency and consumer opinion surveys

Order books indicator deserves further comment since it proved to be very useful, since it appeared literally in all successful. Indicator represents

data that are collected from surveys from 1100 enterprises covering 65% of total turnover in the manufacturing sector. The enterprises answer every month to the question 'Do you consider your current overall order books to be... + more than sufficient (above normal), = sufficient (normal for the season), - not sufficient (below normal)'. This indicator can be considered as future looking and important for future behaviour of real economy, thus it is not surprising that it is useful in forecasting. Business confidence indicator plays similar role.

It must be again noted that no direct GDP measure neither unemployment are represented in the top ten predictors.

VAR1 (lags - 3)

- predictors: Broad money, Order books indicator, PRIBOR1year

VAR2 (lags - 3), BVAR2 (lags - 5), BVAR2b (lags - 5)

- predictors: Broad money, Producer prices - manufacturing, Order books indicator, PRIBOR1year

BVAR3 (lags - 3), BVAR3b (lags - 3)

- predictors: Broad money, Producer prices - manufacturing, Order books indicator, Exports of goods, Business confidence indicator
- BVAR2b - parameters: $\theta = 0.8$, $w = 1$, $\phi = 5$
- BVAR3b - parameters: $\theta = 0.8$, $w = 1$, $\phi = 10$

BVAR parameters optimization again calls for loosening the restrictions of the prior, however in this case only for the first lag. The effect of further lags is strongly suppressed by enormous decay parameter.

From the table 6.4 we can see that the best model is again of the VAR type. However BVAR3b outperformed all other VAR models in the means of Theil2 statistics. Poor performance of RW model under one year forecast horizon allows for eminent values of Theil2 statistics. However, contrary to some results from literature [1], there is no doubt that forecast performance of VAR models outperform the forecast ability of both Random walk and AR models (though some robustness and stability tests might be required to support this statement).

Table 6.4: VAR and BVAR results - one year

Model	MSE	Theil1	Theil2	Turns	Directions
VAR1	2.4639	0.2626	0.4782	0	0.35
VAR2	3.4734	0.2758	0.5305	3	0.39
BVAR2	3.7616	0.3000	0.5780	2	0.39
BVAR2b	3.3297	0.2760	0.5212	3	0.37
BVAR3	3.7224	0.2880	0.5434	2	0.39
BVAR3b	3.5454	0.2784	0.4983	3	0.37

It is also of interest that the best model (VAR1) did not address single turning point in the data since all the other models found 2 or 3 out of 5. Directions statistics proves again to be quite useless.

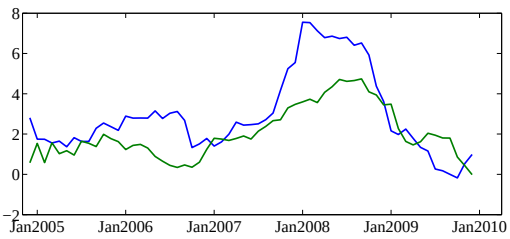

Figure 6.22: VAR1-12

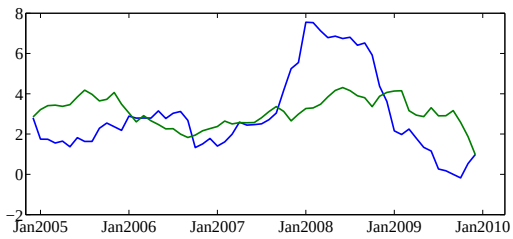

Figure 6.23: VAR2-12

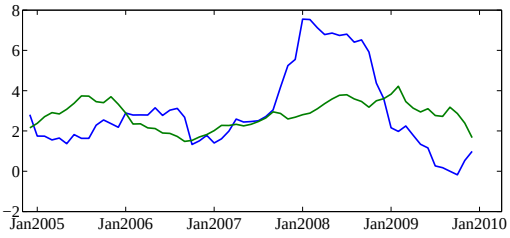

Figure 6.24: BVAR2-12

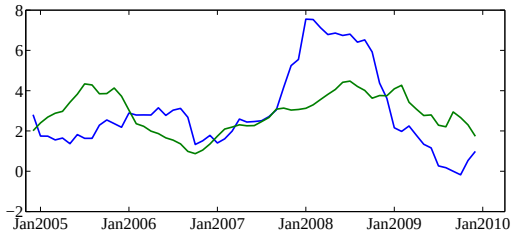

Figure 6.25: BVAR2b-12

From the figures 6.22 to 6.27, we can observe that the forecast does not look like moved and deformed copy of the data (it was the case of one-quarter horizon). Cycles in the data seem to be at least partially forecasted (mainly by VAR1 model), however the forecasts do not catch steep decrease in the end of the data and furthermore forecasts then swing down to negative values. It is also visible that the data that are used for forecast comparison are slightly shortened due to longer forecasting horizon and the need for calculation the Theil2 statistics.

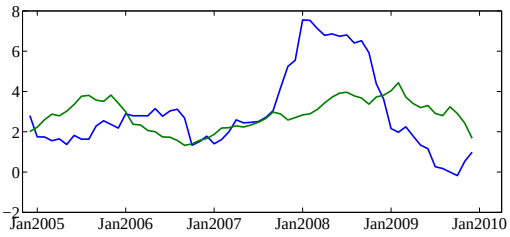

Figure 6.26: BVAR3-12

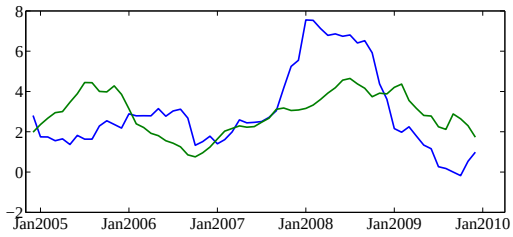

Figure 6.27: BVAR3b-12

Table 6.5: Out of sample - one quarter

Forecast-12	VAR1	VAR2	BVAR2	BVAR2b	BVAR3	BVAR3b	CNB
Jan2010	-0.8	0.4	1.1	1.1	-0.7	-0.6	0.7
Feb2010	-0.5	0.7	1.0	1.2	-1.7	-1.7	0.6
Mar2010	-0.6	1.3	1.5	1.8	-0.7	-0.7	0.7
Apr2010	-1.8	0.2	0.5	0.4	-0.3	-0.3	0.9
May2010	-1.6	-0.2	0.6	0.6	-0.2	-0.2	0.9
Jun2010	-1.7	-0.6	-0.1	-0.2	-0.6	-0.7	0.9
Jul2010	-1.3	-0.4	-0.1	-0.3	-0.4	-0.6	1.4
Aug2010	-1.5	-1.1	-0.7	-0.9	-0.7	-0.9	1.9
Sep2010	-1.4	-1.2	-0.9	-1.5	-0.7	-1.1	2.0
Oct2010	-0.1	-0.1	-0.3	-1.0	-0.7	-1.3	2.2
Nov2010	0.0	0.0	-0.3	-1.0	-0.5	-1.0	2.3
Dec2010	0.2	0.6	0.4	-0.4	0.2	-0.4	2.1

Table 6.5 shows out-of-sample forecasts computed at one year forecast horizon (thus all the available data are necessary only for the last value of the forecast). Forecasts are also depicted in figure 6.28.

CNB forecasts are added for the comparison. First three months are already known data, next three months (up to June 2010) are monthly forecasts, however then CNB provides only quarterly forecasts. To get monthly data, we put these forecasts to the middle month of the quarter and missing values are then computed by linear interpolation between two neighbouring known values. CNB forecasts also reflect its inflation target - 2% (see Discussion).

It is very probable that all the forecasts underestimate the future values of inflation. Models VAR2, BVAR2 and BVAR2b show the shortest swing to negative values and thus are expected to be the most successful.

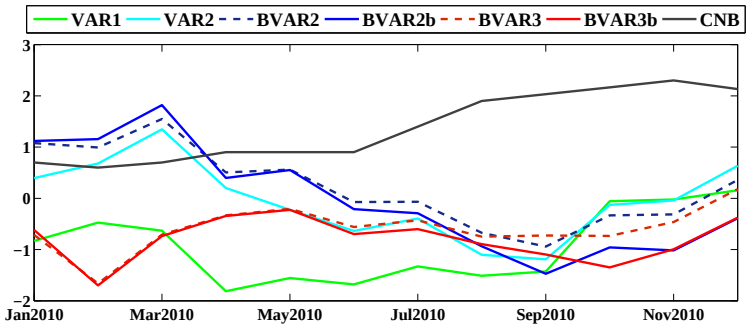

Figure 6.28: Forecast-12

6.2.6 Forecasting horizon - two years

In the line with the literature, two years ahead forecasting using statistical models revealed to be very complicated issue. The biggest problem arises with the stability of models and the best predictors. Already our easy robustness check (shortening the learning data) appeared to have an incredible effect on the top rated models. It was then very hard to decide which predictors allow for the most accurate forecasts. Strong dis-balance between VAR and BVAR models resulted in that no set of predictors allow for successful forecast employing both VAR and BVAR models. Finally, three different VAR models and only one BVAR model (plus its adjustment) were estimated. Each model uses four predictors. However, we have altogether eight predictors.

- Consumer prices - energy - provided by OECD, consumer price index comprising only energy as a product group
- Leading indicator - provided by OECD, belongs to cyclical indicators, composite indicator formed by aggregating a variety of component indicators, designed to provide qualitative information on short-term economic movements
- Overnight interbank rate - provided by OECD, belongs to financial indicators
- GDP - provided by CZSO, constant prices
- GDP - Gross fixed capital formation - provided by CZSO, constant prices

- GDP - provided by OECD, constant prices, seasonally adjusted
- GDP - Gross fixed capital formation - provided by OECD, constant prices, seasonally adjusted
- Service exports - provided by OECD, belongs to balance of payments, derived from both surveys and administrative records

This set of variables (no matter that doubled due to two different data sources) might be theoretically supported. This is in accordance with the hypothesis that theoretically based models are superior when dealing with long-term horizons. However, unemployment is again missing.

VAR1 (lags - 4)

- predictors: Consumer prices - energy, Leading indicator, Overnight interbank rate, GDP (CZSO)

VAR2 (lags - 4)

- predictors: GDP (OECD), Overnight interbank rate, GDP (CZSO), Gross fixed capital formation (CZSO)

VAR3 (lags - 1)

- predictors: Consumer prices - energy, GDP (OECD), Gross fixed capital formation (OECD), Gross fixed capital formation (CZSO)

BVAR4 (lags - 4), BVAR4b (lags - 3)

- predictors: Consumer prices - energy, GDP (OECD), Service exports, Leading indicator
- BVAR4b - parameters: $\theta = 0.06$, $w = 0.5$, $\phi = 1$

It must be noted that VAR3 is not really proper VAR model since using only one lag is not common choice. Including only one lag reminds in fact structural model, just instead of contemporaneous values the first lag is used and furthermore one autoregression parameter. Anyway this specification appeared to give the best forecast. On the contrary to other forecasting horizons, BVAR optimization now calls for tightening the prior, thus for

Table 6.6: VAR and BVAR results - two years

Model	MSE	Theil1	Theil2	Turns	Directions
VAR1	5.3775	0.3299	0.9574	1	0.33
VAR2	4.6570	0.3492	0.8544	0	0.35
VAR3	3.2757	0.2919	0.7751	0	0.24
BVAR4	2.8181	0.2474	0.5164	2	0.43
BVAR4b	2.7053	0.2418	0.4898	2	0.39

stronger restriction. This may be appointed to that unrestricted forecasts are very poor thus prior restriction comparatively improves the performance.

The superior performance of BVAR models is obvious from the table 6.6. However, this is partly due to poor performance of other models. Anyway, BVAR forecast accuracy is far better than performance of other used method (especially AR model). In fact, the forecast performance is better (see MSE and Theil1 statistics) than one year ahead forecast, which is suspicious. Thus robustness and stability of this model should be further tested, e.g. by real-time experiments.

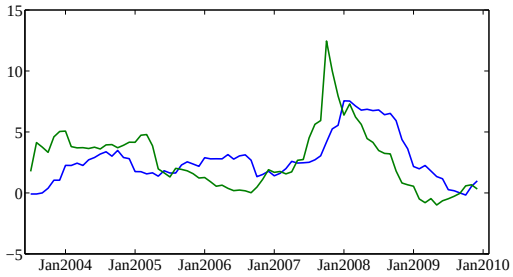

Figure 6.29: VAR1-24

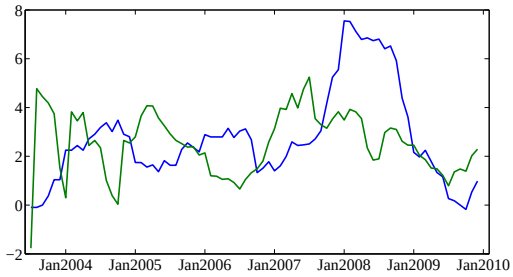

Figure 6.30: VAR2-24

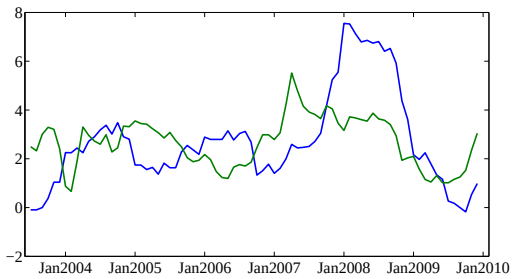

Figure 6.31: VAR3-24

It is clear from figures 6.29 to 6.33 that VAR forecasts are very very wild and thus almost useless. On the other hand BVAR forecasts seem to

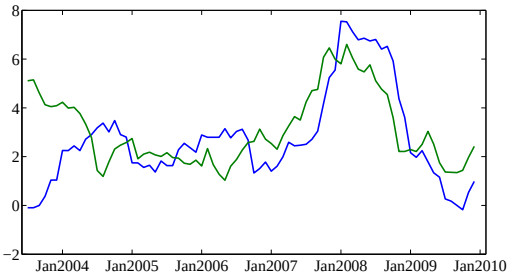

Figure 6.32: BVAR4-24

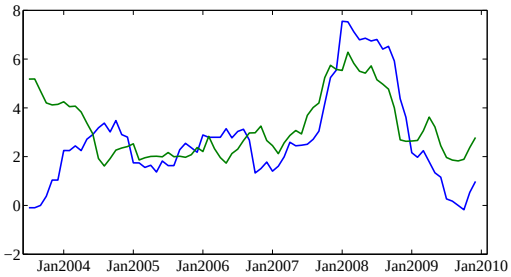

Figure 6.33: BVAR4b-24

be very successful. However, the completely wrong beginning of the BVAR forecasts suggest that fortune might have played important role.

Table 6.7: Out of sample - two years

Forecast-24	VAR1	VAR2	VAR3	BVAR4	BVAR4b	CNB
Jan2010	0.7	2.4	2.4	3.4	3.4	0.7
Feb2010	-0.6	2.6	2.8	4.5	4.1	0.6
Mar2010	-0.7	2.9	2.7	3.8	3.8	0.7
Apr2010	-0.9	3.1	2.8	3.5	3.3	0.9
May2010	0.1	3.6	3.1	3.6	3.5	0.9
Jun2010	0.1	4.1	3.1	4.4	3.8	0.9
Jul2010	0.5	4.3	3.3	4.7	4.0	1.4
Aug2010	0.5	3.7	2.8	4.0	3.6	1.9
Sep2010	1.0	3.9	2.9	3.8	3.2	2.0
Oct2010	0.9	4.7	2.7	3.3	2.9	2.2
Nov2010	0.3	3.4	2.5	2.9	2.5	2.3
Dec2010	0.4	2.9	2.4	1.9	1.7	2.1
Jan2011	-1.0	2.4	2.2	1.3	1.1	2.0
Feb2011	-0.4	1.0	2.0	0.7	-0.5	1.8
Mar2011	-0.9	0.1	2.3	0.1	-1.9	1.8
Apr2011	-2.0	-1.6	1.3	-1.1	-3.5	1.8
May2011	0.4	1.9	2.4	-1.2	-3.1	1.8
Jun2011	-0.2	0.2	2.1	-1.4	-3.4	1.8
Jul2011	0.0	-1.7	2.0	-1.7	-3.5	1.8
Aug2011	1.7	0.6	2.7	-1.4	-3.7	1.8
Sep2011	1.0	0.3	2.9	-0.7	-2.9	1.8
Oct2011	1.4	-0.3	2.6	-0.8	-2.7	1.8
Nov2011	2.2	2.2	3.2	-0.2	-1.7	1.9
Dec2011	2.5	2.0	3.7	1.8	-0.1	1.9

Out of sample forecasts that are summarized in table 6.7 and depicted in figure 6.34 show different behavior for VAR and BVAR models. VAR predictions are very wild, whereas those of BVAR models follow clear pattern. However, VAR predictions might be more accurate, because once they did not catch fast decline in the end of used data, they do not swing deep into negative values as the BVAR does. Time-series models seem to be inappropriate for providing inflation predictions two years ahead.

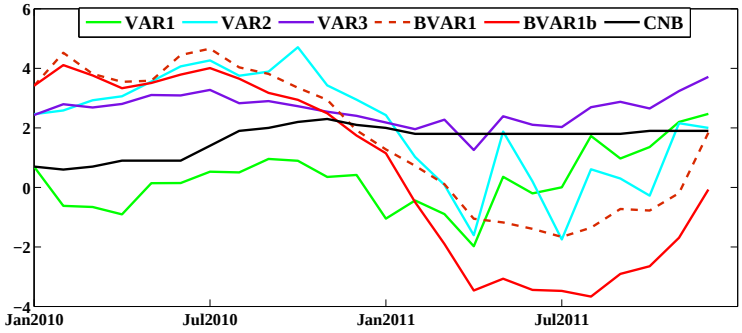

Figure 6.34: Forecast-24

7 Discussion

The chapter Discussion points out the most important results of the work and comments them shortly. However, it also reveals all the imperfections and thus opens lots of issues for future work.

7.1 Data

The choice of the data was as wide as possible, but from the two sources only - OECD Stats and the database ARAD by Czech National Bank. One of the reason is that these data are easily downloadable. However, the data from Czech Statistical Office should also be included into the used data, since ARAD database contains only an extract of these data. Including some more time series is considered to be beneficial. These might be some important commodity prices (oil, steel, gold...) and important economic indicators from other important countries (Germany, USA, etc.). However, these time series should be contained in the external exogenous block of the data and usage of this block is strongly limited because we use chain rule of forecasting. Once direct future estimation is used (see Future work), the exogenous data will be easily included into estimation and forecasting procedure.

The data sources commonly provide the methodology of the measurement and comments on data revisions. These comments must be taken into account when using the data to ensure that forecast is reliable, i.e. that we use in pseudo out-of-sample only such data that were available.

Time series should also be divided by their availability (they partly were) and if the availability of the time series is lagged behind the real time, the forecasting horizon must be appropriately prolonged once this series is used. It is quite possible that comparatively less powerful models will be used just because the data they use are immediately available.

7.2 RW and AR models

Random walk model served as a benchmark for all other models, since Theil2 statistics was used as the main tool for comparison. It is important to depict this benchmark estimation. We could see very poor performance of RW model at one year forecasting horizon which clearly helped VAR and BVAR models to reach excellent values of Theil2 statistics (incomparably better then in Canova [1]).

AR model is suitable for estimation of the data. Structure of its

residuals does not seem to allow for some further modelling (though this is not confirmed by Ljung-Box test). Autocorrelation functions also revealed some seasonality in inflation data. It might be useful to use some seasonal variable when modelling the data. AR prediction power is very poor and particularly in the case of two years horizon it is even worse than naive RW forecast. AR model also serves as the benchmark for subsequent VAR and BVAR modelling.

7.3 VAR and BVAR

Foremost, we must point that all the shown VAR and BVAR models (the best VAR and BVAR models) have higher forecasting accuracy than both AR and RW models.

On the other hand it must be pointed that generally believed idea that BVAR modelling ensures improved forecasting accuracy over VAR models cannot be confirmed. Results are mixed. The only case in that BVAR was superior over VAR is the two years horizon. Nevertheless, this might be attributed to poor performance of VAR models. Concerning other two forecasting horizons, BVAR showed roughly the same forecasting accuracy.

It is of interest that in the case of one quarter and one year horizons, the optimizing of BVAR parameters always lead to loosening of the prior. This might suggest that the prior is not appropriate in these cases and some other prior specification should be used. On the other hand, under two years horizon (BVAR was superior), the prior variances were tightened by optimizing the parameters, thus prior seems to be more accurate in this case.

One of the possible priors might be random walk averaging prior proposed by LeSage [10]. However, it is not particularly favourable choice since all the time series would have to be scaled. The prior assumes solely positive impact on the dependent variable and moreover there must be some mechanism that reveal which variables are important before seeing the data to elicit precise form of the prior. It might be also helpful to include contemporaneous variables into BVAR by employing approach by Waggoner and Zha [18].

The biggest concern about the reliability of forecast lies in the problems of the stability and the robustness of models, mainly the stability of the choice of the predictors. Real-time experiment and adding further data might reveal whether these models are successful. In fact some robustness test might have been already undertaken, but main problem is the scarcity of the data (it is common problem of the data from Central European states since the period

without any major structural changes is still very short). Such stability tests usually comprise of dividing the data into subperiods and checking the model stability. However, subperiods in our case would be so short that estimates in each of the subperiod would not be reliable anyway.

7.4 Future work

Many issues that might be covered in future work are already mentioned in previous section. It comprises mainly of checking the availability and the revisions of the data, employing different prior and testing stability of the predictors.

Interesting issue of direct future value estimations has also been already partly discussed. Such model is of this general form ([11], [1]):

$$y_{t+k} = \beta_0 x_t + \beta_1 x_{t-1} + ... \tag{7.1}$$

y_{t+k} may represent only inflation data or whole set of dependent variables. x_t is a set of explanatory variables (may include the lags of the dependent variables as well), k represents the forecasting horizon (e.g. 12 or 24). The notation should emphasize that required future value is directly estimated from currently available data. For pseudo out-of-sample forecast we use only data that were available 'k steps before or earlier' to estimate the 'current' value of the dependent variable. By this process we get rid of chain type step-by-step forecasting.

This approach seems to be quite advantageous. It allows to include external exogenous block of the data. Moreover data might be organized according to their availability, disregarding what time they in fact describe. (This might lead to easier carrying of statistical inference, however interpretation would then be quite obscured.) Finally, there is a promise of improved forecasting accuracy.

Depending on the type of the dependent variables (whether it is just one variable or whole set) and on the type of explanatory variables (whether there are lags of the dependent variables) we may use (Bayesian) OLS estimation or (B)VAR. These two types of estimation are not, in fact, so different. Other estimation techniques are also available (GLS, Ridge regression, SUR etc.).

Inflation forecasts should also take into account monetary policy rules of the Central Bank (so called conditional estimation), because monetary policy can seriously affects the inflation in the important forecasting horizons. (In fact the reasoning is vice-versa, important horizons are those at which

monetary policy can influence the inflation.) Assessing the monetary policy is also very important and interesting issue. Moreover Bayesian estimation techniques are suitable for analysing monetary policy effects [4], [7].

Finally, it might be of interest, and it is not so demanding (apart of searching for all useful data), to provide inflation forecasts also for other (Central European) countries. Inter alia, it is interesting analysis of stability of inflation predictors among various countries.

8 Conclusion

This work starts with double theoretical introduction. The first is comprehensive introduction to Bayesian econometrics and Bayesian estimation techniques with emphasis on Bayesian vector autoregression and Minnesota Prior. The second theoretical introduction comments the importance of inflation forecasting under inflation targeting governed by Central Bank. Generally, the issue of forecasting in economy and particularly forecasting by VAR are discussed. Short literature review is also provided. Important forecasting horizons are stated to be one quarter, one year and two years ahead.

The work focuses on Czech inflation forecasting. Monthly data (almost 80 time series) were downloaded from two sources: OECD Stats and ARAD, the latter is the time series database by Czech National Bank (CNB) and Czech Statistical Office (CZSO). The choice of the data and handling with them is not particularly elaborated and it is discussed throughout the work.

Time consuming and computationally demanding procedure of estimation is described in a detail. Main features are realization in Matlab supplemented by James P. LeSage toolbox [2], multi-step choice of the best predictors, optimization of BVAR parameters and calculation of pseudo out-of-sample and out-of-sample forecasts by chain rule. Forecast comparison statistics (particularly Theil statistics) are presented and extensively used.

Random walk and Autoregressive models were estimated as the benchmark for VAR and BVAR modelling. AR model does not show good forecast accuracy (in the case of two years horizon even worse than naive Random walk forecast). VAR and BVAR modelling outperform both RW and AR models for all three forecast horizons in forecasting accuracy. The set of the best predictors is completely different for each forecasting horizon. Interesting is that neither GDP measure nor unemployment are among the best predictors for one quarter and one year forecast horizon. On the other hand, GDP is one of the used predictors in models for two years ahead forecasts along with e.g. fixed capital formation (this agrees with the literature that argues that structural relationships are useful when forecasting horizon is long). Overall success of various OECD indicators based on business and consumers surveys is also surprising.

For the two shorter horizons (one quarter and one year) BVAR forecasts are hardly better than the VAR. Moreover, optimizing of the BVAR parameters comprises of loosening the prior. This suggests that Minnesota prior is not particularly appropriate. Different situation arises under two years horizon. BVAR models clearly outperform the VAR models and

optimized parameters of BVAR models tightened the prior heavily.

Out-of-sample forecasts are compared to forecasts of Czech National Bank. Almost all the forecasts showed improbable swing into negative values of inflation during year 2010 or 2011. All the results are discussed in detail and fairly enough issues are opened for future work.

The only reliable answer to the title question is: 'it depends.'

9 Bibliography

[1] Canova, Fabio; G-7 Inflation Forecasts: Random Walk, Phillips Curve or What Else?; Macroeconomic Dynamics, 2007

[2] LeSage, J. P.; Applied Econometrics using MATLAB; Department of Economics, University of Toledo, 1999

[3] Koop, Gary; Bayesian Econometrics, John Wiley &Sons Ltd, 2003

[4] Joiner, Alex; Monetary Policy Effects in an Australian Bayesian VAR Model; Australasian Macroeconomics Workshop, 2002

[5] Litterman, Robert B.; Forecasting with Bayesian Vector Autoregressions - Five Years of Experience; Federal Reserve Bank of Minneapolis, 1985

[6] Bikker, J. A.; Inflation Forecasting for Aggregates of the EU-7 and EU-14 with Bayesian VAR Models; Journal of Forecasting, 1998

[7] Bloor, Chris; Matheson, Troy; Analysing shock transmission in a data-rich environment: A large BVAR for New Zealand, Reserve Bank of New Zealand Discussion Paper Series, 2009

[8] Zha, T., Sims, C. A.; Bayesian Methods for Dynamic Multivariate Models; International Economic Review, 1998

[9] Robertson, J. C., Tallman, E. W.; Prior parameter uncertainty - Some implications for forecasting and policy analysis with VAR models; Federal Reserve Bank of Atlanta, 1999

[10] LeSage, J. P., Krivelyova, A.; A Spatial Prior for Bayesian Vector Autoregressive Models; Journal of Regional Science, 1999

[11] Jacobson, Tor; Karlsson Sune; Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach, Journal of Forecasting, 2004

[12] Stock, J. H., Watson M. W.; Forecasting Inflation; Journal of Monetary Economics, 1999

[13] Atkeson, A., Ohanian, L. E.; Are Philips curves useful for forecasting inflation?; Federal Reserve Bank of Minneapolis, 2001

[14] Granger, C. W. J., Bates, J.; The combination of forecasts; Operation Research Quarterly, 1969

[15] Wright, J. H.; Forecastin US Inflation by Bayesian Model Averaging; Journal of Forecasting, 2009

[16] Ballabriga, F. C.; Castillo, S.; BBVA-ARIES: A Forecasting and Simulation Model for EMU; Journal of Forecasting, 2003

[17] Borys, M. B.; Horváth, R.; The Effects of Monetary Policy in the Czech Republic: Evidence from Factor Augmented VAR, Czech National Bank Working Paper, 2008

[18] Waggoner, D.F., Zha T.; A Gibbs simulator for restricted VAR models; Federal Reserve Bank of Atlanta, 2000

[19] An, Sungbae; Schorfheide, Frank; Bayesian analysis of DSGE models, Econometric Reviews, 2007

[20] Cook, Steven; Understanding the construction and interpretation of forecast evaluation statistics using computer-based tutorial exercises; Swansea University, published by Economics Network, 2006

[21] Gollinelli, Roberto, Orsi, Renzo; Modelling inflation in EU accession countries: the case of the Czech Republic, Hungary and Poland; University of Bologna, presented at the seminar "East European Transition and EU Enlargement: A Quantitative Approach", 2001

[22] Kotlán, Viktor; Navrátil, David; Inflation Targeting as a Stabilization Tool: Its Design and Performance in the Czech Republic; Czech Journal of Economics, 2003

[23] Matoušek, Roman; Taci, Anita; Direct Inflation Targeting and Nominal Convergences: The Czech Case; Open economies review, 2003

[24] Babecký, Jan; Podpiera Jiří; Inflation Forecasts Errors in the Czech Republic: Evidence from a Panel of Institutions; EconPapers, 2008

[25] Kulhánek, Lumír; Indicators for Inflation Targeting in the Czech Republic; Monetary Policy in Transforming Economies, 1998

[26] Coats, Warren (editor); Inflation Targeting in Transition Economies: The Case of the Czech Republic, Czech National Bank and International Monetary Fund, 2000

10 Appendix A - List of used functions

This appendix contains the list of all used functions. We provide only names of functions that are contained in the Matlab toolboxes. Functions that I consequently programmed are accompanied with short comment on the inputs and outputs. All functions can be found on the attached CD, however without proper documentation.

- `lbqtest` - Ljung-Box Q test
- `dfARTest` - Augmented Dickey-Fuller test
- `autocorr` - Computes and plots autocorrelation functions
- `parcorr` - Computes and plots partial autocorrelation functions

`arma2` - programmed script

- inputs: `y` - data vector, `nc` - number of MA lags, `k` - forecast horizon
- output: `ypar` - vector of forecast

```
function createfigure80(y,f,s,name,sizex,sizey)
```

- inputs: `y` - data matrix, `f` - forecast matrix, `s` - the column number of the variable that we want to depict, `name` - name of the chart, `sizex` - horizontal size of the chart, `sizey` - vertical size of the chart
- output: automatically generated chart depicting the forecast and the data in the .eps format

```
function [MSE,th1] = theil1(y,f,s)
```

- inputs: `y` - data matrix, `f` - forecast matrix, `s` - the column number of the variable that we want to forecast
- outputs: `MSE` - Mean square error of the forecast, `th1` - Theil1 statistics

```
function [th2] = theil2(y,f,s,k)
```

- inputs: `y` - data matrix, `f` - forecast matrix, `s` - the column number of the variable that we want to forecast, `k` - forecast horizon
- output: `th2` - Theil2 statistics

`function[MSE,th1,th2,y3,f]=varpred(y,k,nlag,s,p)`

- inputs: `y` - data matrix, `k` - forecast horizon, `nlag` - number of lags included in the model, `s` - the column number of the variable that we want to forecast, `p` - fraction of the data used to compare to forecast with the data
- outputs: `MSE` - Mean square error of the forecast, `th1` - Theil1 statistics, `th2` - Theil2 statistics, `y3` - truncated data that are compared to the forecast, `f` - forecast matrix

`function[MSE,th1,th2,y3,f]=bvarpred(y,k,nlag,s,p,tight,weight,decay)`

- inputs: `y` - data matrix, `k` - forecast horizon, `nlag` - number of lags included in the model, `s` - the column number of the variable that we want to forecast, `p` - fraction of the data used to compare to forecast with the data, `tight` - tightness parameter θ, `weight` - weight parameter w (scalar or matrix), `decay` - decay parameter ϕ
- outputs: `MSE` - Mean square error of the forecast, `th1` - Theil1 statistics, `th2` - Theil2 statistics, `y3` - truncated data that are compared to the forecast, `f` - forecast matrix

`function[e2,e3]=varpredn(z,k,nlag,s,p,n)`

- inputs: `z` - complete data matrix, `k` - forecast horizon, `nlag` - number of lags included in the model, `s` - the column number of the variable that we want to forecast, `p` - fraction of the data used to compare to forecast with the data, n - number variables included in VAR
- outputs: `e2` - matrix that contains in each row in this order: MSE, Theil1, Theil2 and reference numbers of the used predictors (according to their column number in z matrix), matrix $e2$ is sorted by ascending order with respect to Theil2, `e3` - the first tenth of $e2$ matrix - just for better orientation if $e2$ is too big

`function[e2,e3]=bvarpredn(z,k,nlag,s,p,tight,weight,decay,n)`

- inputs: `z` - complete data matrix, `k` - forecast horizon, `nlag` - number of lags included in the model, `s` - the column number of the variable that we want to forecast, `p` - fraction of the data used to compare to forecast with the data, `tight` - tightness parameter θ, `weight` - weight parameter w (scalar or matrix), `decay` - decay parameter ϕ, n - number variables included in VAR

- outputs: `e2` - matrix that contains in each row in this order: MSE, Theil1, Theil2 and reference numbers of the used predictors (according to their column number in z matrix), matrix $e2$ is sorted by ascending order with respect to Theil2, `e3` - the first tenth of $e2$ matrix - just for better orientation if $e2$ is too big

  `function [otp,ftp,tpr,otpm,ftpm] = tp(y,f,q)`

- inputs: `y` - data vector, `forecast vector`, `q` - parameter that allows forecast not to hit the turning point in the data precisely

- outputs: `otp` - number of turning points in the data, `ftp` - number of correspondent turning points in the forecast, `tpr` - ftp/otp, `otpm` - positions of turning points in the data, `ftpm` - positions of correspondent turning points in the forecast,

  `function [dm,dh,dl,dt] = td(y,f,k)`

- inputs: `y` - data vector, `forecast vector`, `k` - forecast horizon

- outputs: `dm` - all positions marked by either '1' (forecast rises whereas data decrease), '-1! (vice-versa), or '0' (forecast follows the same direction as the data, `dm` - number of points from where the forecast rises and the data decrease, `dl` - number of points from where the forecast decreases and the data rise, `dt` - total number of points where do statistics was evaluated

  `function[f]=varpredout(y,k,nlag)`

- inputs: `y` - data matrix, `k` - forecast horizon, `nlag` - number of lags included in the model

- output: `f` - forecast matrix

  `function[f]=bvarpredout(y,k,nlag,tight,weight,decay)`

- inputs: `y` - data matrix, `k` - forecast horizon, `nlag` - number of lags included in the model, `tight` - tightness parameter θ, `weight` - weight parameter w (scalar or matrix), `decay` - decay parameter ϕ

- output: `f` - forecast matrix

11 Appendix B - Contents of the enclosed CD

Text - contains the diploma thesis in .pdf and also all the source texts (.tex) and all the used pictures (.eps)

Data - contains all the data, both originally downloaded and also after different manipulations, lists of the top variables from the first group of the data and lists of the top ten variables for each forecasting horizon are provided

Functions - contains all newly programmed Matlab functions, which are described in Appendix A, functions should be copied to the directory containing the VAR and BVAR functions by James P. LeSage to work properly

Reults - collection of MS Excel tables containing the results after each estimation step