Ignacy Morawski

Bayesian methods in forecasting – short introduction

Zarządzanie Zmianami : zeszyty naukowe nr 2, 45-56

2010

Ignacy Morawski*

Bayesian methods in forecasting — short introduction

Metody Bayesowskie w prognozowaniu — wprowadzenie

Streszczenie

Bayesowskie metody statystyczne są często używane przez ekonomistów w bankach centralnych do estymacji modeli prognostycznych. Podstawowa różnica między podejściem bayesowskim a podejściem statystyki klasycznej jest taka, że pierwsze z nich traktuje parametry modelu jako stochastyczne, drugie zaś przyjmuje wektor parametrów jako stałą. W zależności od wykorzystanej metody, różne są źródła ryzyka. W niniejszej pracy zaprezentowano podstawowe procedury związane z wykorzystaniem metod bayesowskich w prognozowaniu. Pokazano ich zastosowanie do modelu autoregresji pierwszego rzędu, jak też bardziej zaawansowane zastosowania do modeli VAR. Celem pracy jest stworzenie krótkiego przewodnika, który może być traktowany jako wstęp do bardziej zaawansowanych studiów.

Słowa kluczowe: statystyka bayesowska, metody bayesowskie, prognozowanie

Introduction

In every science researcher's beliefs play an important role: beliefs about what kind of scientific investigation should be undertaken, beliefs about tools to be used, variables to include etc. In Bayesian statistics these beliefs play exceptional role, because they directly influence the results of the research. Beliefs expressed in the form of prior distribution of parameter, about which we want make an inference, impact the final result of this inference. According to Bayes' rule:

$$P(\theta|X) \propto P(\vartheta)L(X|\vartheta)$$

The probability distribution of the parameter of interest is proportional to beliefs (prior) and the likelihood function of the model, computed at given data.

Forecasters also face the problem of expressing beliefs: about which forecasting model to choose, which variables to include, what relations between them to assume. These beliefs influence the results of the forecast, especially uncertainty, which is always related to forecasting. Researcher can never choose as many variables as possible and investigate relations between them, because there are no tools which would allow to reconcile precision with the size of the model. Some restrictions always

* Ignacy Morawski — student of the Bocconi University, graduate of the Warsaw University (Political Science), economist of WestLB Polska, columnist of „Rzeczpospolita", e-mail: ignacy.morawski@gmail.com.

have to be imposed. If they are mistaken, the forecast will be imprecise.

The Bayesian statistics supply tools which allow to impose gentle restrictions on forecasting models. On the one hand, using these tools a researcher can express his more or less strong beliefs about the model and limit the scope of randomness and uncertainty, on the other hand she can let the data speak more than under traditional ways of restriction.

In this short essay I discuss the following issues: 1. Simple example of AR(1) model and its forecasting application under Bayesian approach; 2. Vector autoregression and Minnesota prior; 3. Kalman filter and time varying parameters; 4. Structural models as a source of prior distribution.

1. Simple linear model

I can begin discussion of Bayesian methods in forecasting by introducing a simple linear autoregressive model. Its applications are rather limited, but it allows to present the Bayesian approach to time series forecasting, basic methods and results.

It is useful to start with the simplest possible statistical model: stationary AR(1), in which the present value of the variable x_t is determined by its past value x_{t-1} and a random forecast error e_t, which is distributed normally. Stationarity is not a necessary assumption, it is even an advantage of Bayesian methods that they do not require necessarily to investigate stationarity first, before doing an inference. However, this point will be discussed later. Here, for the simplicity purposes, we limit to the stationary version of the following model:

$$x_t = \alpha x_{t-1} + e_t \tag{1}$$

with

$$e_t \sim N\left(\mu, \sigma^2\right)$$

$$E\left[e_t e_{t-1}\right] = 0$$

$$x_t \left| x_{t-1} \right. \sim N\left(\alpha x_{t-1}, \sigma^2\right)$$

It is worth pointing out that all procedures related to AR(1) model can be easily extended to the autoregressive model of higher order AR(p), because as we can see:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} \ldots + \alpha_p x_{t-p} + e_t \tag{2}$$

can be represented as:

$$\begin{bmatrix} x_t \\ x_{t-1} \\ \cdots \\ x_{t-p} \end{bmatrix} = \begin{bmatrix} \alpha_1 \alpha_2 \ldots \alpha_p \\ 1 \;\; 0 \ldots 0 \\ \ldots \;\; \ldots \ldots \ldots \\ 0 \;\; 0 \;\; 1 \;\; 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \ldots \\ x_{t-p} \\ 0 \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \\ \ldots \\ 0 \end{bmatrix}$$

which is the vector autoregressive model of the form

$$x_t = \alpha_1 x_{t-1} + e_t$$

where almost all calculations from the AR model are applied. Vector autoregressive model will be discussed in details in the next chapter.

In classical statistics the forecast of the future values of $x = [x_{t+1}, x_{t+2}, \ldots, x_{t+j}]$ are computed as follows:

$$x_{t+j} = \alpha^j x_t + \sum_{k=1}^{j-1} \alpha^k e_{t+j-k} \tag{3}$$

Therefore the j-step ahead forecast is distributed as follows:

$$P\left(x_{t+j}\middle|x_t,\alpha,\gamma\right)=$$

$$=\frac{1}{\sqrt{2\prod\gamma}}\exp\left\{\frac{\left(x_{t+j}-\alpha^j x_t\right)^2}{2\gamma^2}\right\}$$

where γ^2 is a j-step ahead forecast error variance.

$$\gamma^2=\left(1+\alpha^2+\alpha^4+...+\alpha^{2(j-1)}\right)\sigma^2$$

And the density of the forecasted values from t+1 to t+j is a multivariate normal with autocorrelated and heteroskedastic errors (for detailed computations of multivariate normal distribution see: Hamilton [1994]).

$$P\left(x_{t+1},x_{t+2},...,x_{t+j}\middle|x_1...x_t,\alpha,\sigma^2\right)=$$

$$=\left(2\prod\right)^{-\frac{J}{2}}\left|\sum{}^{-1}\right|^{\frac{1}{2}}$$

$$\exp\left\{\left(x-x_t\right)'\sum{}^{-1}\left(x-x_t\right)\right\}$$

with:

$$x=\begin{bmatrix}x_{t+1}\\x_{t+2}\\\cdots\\x_{t+j}\end{bmatrix},\quad x_t=\begin{bmatrix}x_t\\x_t\\\cdots\\x_t\end{bmatrix}$$

and Σ is the variance-covariance matrix of the forecasted vector.

In Bayesian analysis the procedure is different, taking first into account that not only x_t is a random variable, but also the parameter α is a random variable. Therefore we have to proceed according to the following steps:

1. First, we assign a prior distribution to the parameters P(α, σ²).
2. Second, we compute a likelihood function of a vector of observations, expressed in terms of estimated OLS parameters.
3. Third, we compute a posterior distribution, which is the product of prior and the likelihood function, scaled by the parameter which guarantees integrity of the density.
4. Fourth, we compute the predictive density, which is unconditional on a given parameterization of the model:

(4)

$$P\left(x_{t+1},x_{t+2},...,x_{t+j}\middle|x_1...x_t\right)=$$

$$=\int_\Theta P\left(x_{t+1},x_{t+2},...,x_{t+j}\middle|x_1...x_t,\alpha,\sigma^2\right)$$

$$P\left(\alpha,\sigma^2\middle|x_1...x_t\right)d\Theta$$

Or we can write a forecast in a simpler way, as in equation (2), with the exception that the estimated parameter $\hat{\alpha}$ is a mean of posterior distribution.

Let see how it can work in practice. We can imagine that we have T observations of the variable x_t. They form a vector:

$$X=\begin{bmatrix}x_1\\x_2\\\cdots\\x_T\end{bmatrix}$$

The parameter α is a scalar. The vector of explanatory variables is X, the dependent variables is just the same vector, but

moved one step ahead. Moreover, we have a vector of errors:

$$E = \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_T \end{bmatrix}$$

with

$$E(e_t) = 0 \ and \ E(e_t e_t) = \sigma^2 I_T$$

Therefore we can write:

(5) $$X_t = \alpha X_{t-1} + E_t$$

The OLS estimation of the parameter α is given by $\hat{\alpha}$. The likelihood function is given by multivariate normal:

(6) $$P_X(X = x|\Theta) = (2\prod)^{-\frac{T}{2}} \left|\Omega^{-1}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} x \Omega^{-1} x'\right\}$$

where Ω is a variance — covariance matrix of the vector x, which is a function of parameters α and σ^2.

Now we need a prior distribution for parameters α and σ^2. In this example we will use so called Jeffreys prior, which is given by:

$$P(\Theta) = \frac{1}{\sigma^2}$$

The use of this particular prior in this case is only for presentation purposes. There exist a vast literature on what kind of prior distributions should be used in certain applications. In this chapter we only want to show the basic procedures, therefore we use the simple prior.

Combining the likelihood function and the prior distribution we get a posterior distribution function of the parameters:

(7) $$P(\Theta|X) \ \alpha \ P_X(X = x|\Theta) P(\Theta)$$

Because the likelihood function can be expressed in terms of OLS parameters only, the posterior distribution is the function of parameters. Next we need a predictive density function.

Then the predictive density is computed:

(8) $$P(x_{t+1}, x_{t+2}, ..., x_{t+j}|x_1 ... x_t) = \int_\Theta P(x_{t+1}, x_{t+2}, ..., x_{t+j}|x_1 ... x_t, \alpha, \sigma^2) P(\alpha, \sigma^2|x_1 ... x_t) d\Theta$$

The numerical method is required to compute this density. Then the mean and standard errors are computed, together with confidence intervals.

Or, again, we can write a forecast as in equation (2):

(9) $$x_{t+1} = \alpha^{*j} x_t + \sum_{k=1}^{j-1} \alpha^{*j} e_{t+j-k}$$

where α^{*j} is the mean of posterior distribution of α.

2. Bayesian vector autoregression and Minnesota prior

In this chapter I want to extend the basic procedures presented above and add some discussion about priors. The simple AR(1) model can be rarely used in applied economics, because relations in the economy are too complicated to be represented by just one autoregressive relation. More commonly used are vector autoregressions, so called VAR, which are often combined with the Bayesian approach, yielding so called Bayesian VAR, or BVAR.

Economists often use VAR models for forecasting purposes. They select a group of variables and allow them to interact linearly with their past values — of their own and each other. If a researcher imposes no strong restriction on the model's parameters, we often talk about unrestricted VAR — UVAR. It is often said that in such models "data speak" — the estimation depends almost only on the data, any beliefs supplied by the researcher are irrelevant. Such models perform quite well if the number if variables is small. Why? Because they incorporate only few forecast errors, therefore mean squared forecast error is low. On the other hand, it seems obvious that very often researchers want to investigate relations between many variables, which incorporate many forecast errors. In such a case UVAR may perform poorly. First, the mean squared forecast error becomes higher. Second, the more data we have and more relations we want to estimate, the more probable it is that we pick up some accidental relations, so called random effects. As Todd [1984, p. 3] points out, the data may be estimated "too well". *The coefficients are so numerous that the statistical procedure can choose them to also fit many of the less important features of the historical data, features which often reflect merely accidental or random relationships that will not recur and are of no use in forecasting.*

One way of avoiding such a problem is to use structural approach — base the statistical model on the structural model, what is done by including only few variables, believed to represent structural interactions in the economy. In such models economic theory is used as a source of restriction on the number of variables and parameters. But some researchers point out that this restriction is too rigid. As Todd [1984, p. 3] writes: *Excluding variables from an equation amounts to certainty that their coefficients are zero. Certainty is an absolute belief, not subject to revision by any amount of historical evidence. So such exclusion restrictions also amount to assigning coefficients of zero to the variables regardless of historical evidence.*

Bayesian approach may be the other way of imposing restrictions on the statistical model's parameters. The beliefs of a researcher are not imposed in the form of rigid exclusion of some parameters or variables, but in the form of assigning probability distribution to every parameter, allowing some of them to be estimated more "freely" and others — to be restricted around some values. Doing so we can allow data to have larger influence on some parameters, and smaller — on the others.

The most commonly used way to impose prior distributions is to use so called Minnesota prior. The name comes from the fact that this way of proceeding was developed by economists coming from Minnesota University and Federal Reserve Bank of Minnesota.

The procedure in its essence is very simple. First, we choose variables that

should be included in the model and specify the set of equations that link them together. Second, we assign prior distribution to the parameters in the following way. The hypothesis is that the process is a random walk — the best forecast for future values is the present value. Therefore the mean of the distribution of all coefficients is zero, with the exception that the mean of distribution for parameters on most recent value of each variable is 1.

Todd [1984, p. 6] writes: *This hypothesis capitalizes on a simple statistical observation that is often a forecaster's chief source of embarrassment: many economic (and other) variables seem to behave as though changes in their values are completely unpredictable. For such a variable, the best forecast of its future values is just that they will equal its current value. Even for variables whose changes are thought to be partially predictable, these no-change forecasts can be surprisingly difficult to improve upon*.

The prior is constructed in the following way. We choose the mean and the variance of parameters on each lag of each variable. The parameter on the first lag has mean 1 and the variance σ^2 chosen by the researcher. The mean for lags higher than 1 is 0, and the variance is proportionally lower. It is useful to represent this procedure in the example.

Imagine that we have a two variable VAR of lag length k (only for simplicity we omit an intercept, which is usually included in the models):

(10) $$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = A(L) \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

where A(L) is a lag polynomial of order k:

$$A(L) = \left(A_1 + A_2 L + A_2 L^2 + \ldots + A_k L^{k-1}\right)$$

and each matrix A_i consists of four entries: $a_{i,11}$, $a_{i,12}$, $a_{i,21}$, $a_{i,22}$. Therefore we have two equations:

$$x_{1,t} = a_{1,11} x_{1,t-1} + a_{1,12} x_{2,t-1} + \ldots + a_{k,11} x_{1,t-k} + a_{k,12} x_{2,t-k} + e_{1,t}$$

$$x_{1,t} = a_{1,21} x_{1,t-1} + a_{1,22} x_{2,t-1} + \ldots + a_{k,21} x_{1,t-k} + a_{k,22} x_{2,t-k} + e_{2,t}$$

Each of them can be estimated using standard OLS together with Bayesian method described above, but it is necessary at this point to show what steps should be done if we want to assign Minnesota prior.

First we stabilize prior variances H_1 and H_2, where H_1 describes variance on own lag coefficients and H_2 describes variance on cross lag coefficients. Then we assign weights to these variances, assuming that the higher the lag the lower the variance, because the forecaster is more sure about accuracy of her guess. The weight is 1/1+k, where k is the lag number.

Therefore:

$$a_{1,11}, a_{1,22} \sim N\left(1, \frac{1}{2} H_1\right)$$

$$a_{1,12}, a_{1,21} \sim N\left(0, \frac{1}{2} H_2\right)$$

$$a_{1,11}, a_{1,22} \sim N\left(0, \frac{1}{i+1} H_1\right) \quad for \;\; i = 2...k$$

$$a_{1,12}, a_{1,21} \sim N\left(0, \frac{1}{i+1} H_2\right) \quad for \;\; i = 2...k$$

Following Litterman [1985, p. 19] we can treat σ^2 as known, taking the value of the estimate from OLS. Litterman writes: *A Bayesian solution which takes (…) a diffuse prior distribution for σ^2 leads to a normal-t posterior density for [coefficients] which would require an intractable numerical integration in order to calculate the posterior mean*.

As usually, the VAR model is estimated equation by equation. Each equation can be written in the following form:

(11) $$Y = X\beta + E$$

where
Y — is a (Tx1) matrix of all observations of a given variable
X — is a (Tx1) matrix of all observations of lags of a given variable and other variables
β — is a (px1) vector of parameters
E — is a (px1) vector of errors

Now we postulate the process for parameters of this equation in the following way:

(12) $$R\beta = r + v$$

where:
R — is a (pxp) identity matrix
β — is a (px1) vector of parameters
r — is a (px1) vector with of 1 corresponding to parameters on most recent variables, and 0 corresponding to parameters on higer order lags
v — is a (px1) vector of errors with mean 0 and variance corresponding

As many authors show (e.g. Litterman [1985, p. 15], Goldberger and Theil [1961, p. 67]) the posterior mean for the vector β is given by:

$$\hat{\beta} = \left(X'X + kR'R\right)^{-1}\left(X'Y + kR'r\right)$$

Now we can easily compute the forecasted values of the variables.

In our example, taking the first equation:

(13) $$x_{1,t+1} = a_{1,11}x_{1,t} + a_{1,12}x_{2,t} + ... + a_{k,11}x_{1,t-k+1} + a_{k,12}x_{2,t-k+1} + e_{1,t+1}$$

or

$$x_{1,t+1} = \left[x_{1,t} \;\; x_{2,t} \;\ldots\; x_{1,t-k+1} \;\; x_{2,t-k+1}\right] \hat{\beta} + e_{1,t+1}$$

In general, once we have estimated the whole BVAR, the j-step ahead forecast may be computed in easier way by transforming the model.

Let

$$\hat{A}(L)$$

be the matrix polynomial of order k estimated by the Bayesian method presented

above, therefore we obtain the following model:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \hat{A}(L) \begin{bmatrix} x_{1,t-1} \\ x_{2,t-2} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \quad \textbf{(14)}$$

or

$$X_t = \hat{A}_1 X_{t-1} + \hat{A}2X_{t-2} + \ldots + \hat{A}_k X_{t-k} + E_t$$

Which can be transformed into:

(15)

$$\begin{bmatrix} X_t \\ X_{t-1} \\ \ldots \\ X_{t-k} \end{bmatrix} = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 & \ldots & \hat{A}_k & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \ldots \\ X_{t-k-1} \end{bmatrix} + \begin{bmatrix} E_t \\ 0 \\ \ldots \\ 0 \end{bmatrix}$$

$$\gamma_t = \hat{\Theta} \gamma_{t-1} + V_t$$

The forecast is computed in the same way as in the simple AR(1) model:

$$\gamma_{t+j} = \hat{\Theta}^j \gamma_t + \sum_{k=1}^{j-1} \hat{\Theta}^k V_{t+j-k} \quad \textbf{(16)}$$

3. Time varying parameters and Kalman filter

In the spirit of Lucas [1976] which argued that traditional macroeconometric models based on estimated time invariant coefficients can't be used for the examination of economic policies, Sims [1982] suggested the use a time varying version of the Litterman VAR. Christopher Sims in 1982 paper argued that a complete rejection of reduced form models as a way of policy analysis is not justified.

In general, rational expectations revolution questioned usefulness of the large scale econometric models due to the fact that they do not reflect the structural changes that the policy introduces in the economy. For example, if the policymaker considers various actions she may be trapped by the fact that a change in the policy would change parameters of the model on which that policy rely. As a response to that problem economists started to think about policy actions as policy rules — permanent changes in the behaviour of policymaker and its structural implications. Sims [1982, p. 109] writes: *It is claimed, policy analysis should be formulated as choice among rules of behaviour for the policy authorities and estimates should be made of the stochastic properties of the economy under each proposed rule to choose the best [...]. I argue that it is a mistake to think that decisions about policy can only be described, or even often be described, as choice among permanent rules of behaviour for the policy authorities*.

A proposition to answer to questions raised by econometricians in 1970s is a use of Bayesian models which allow for time varying parameters. *Disputes about the optimal rule are no more important in principle than disputes about how to implement the existing "rule" as it emerges from existing institutions or interests* (Sims [1982, p. 139]). Making things simple, Sims argues that in the historical data we may discover changes in the policy. Bayesian models which allow for param-

eters variation in time may be a good tool to assess such processes.

In this paper I limit to only present the model of Doan, Litterman and Sims [1986], but detailed discussion of Kalman filter used to evaluate such a model will be based on a simpler model.

Doan et al. analyze the forecasting procedures stemming from estimation of vector autoregressive model of the following form:

$$X_t = A_t(L)X_{t-1} + C_t + V_t \quad (17)$$

The prior — as in the example discussed above — is imposed in such a way that the best guess is a random walk process with drift:

$$X_t = X_{t-1} + C + V_t$$

The system is estimated equation by equations, therefore all parameters in an equation are gathered in a vector ϑ which follows a process:

$$\vartheta_t = \prod \vartheta_{t-1} + (1-\prod)\overline{\vartheta} + \mu_t$$

The prior is assigned to the value ϑ_0 which is distributed:

$$\vartheta_0 \sim N\left(\overline{\vartheta}, \Sigma_0\right)$$

The parameter π controls the rate of decay towards a prior mean, whereas μ — the random change in the parameter vector — is assumed to be drawn from a distribution with zero mean and covariance matrix proportional to Σ_0.

Having specified the probability model, we apply the Kalman filter to each equation to obtain recursively posterior modes $\hat{\vartheta}_t$ for ϑ_t based on data through t-1. When we have passed through the full sample this way, we end up with a value for the likelihood of the sample and with a full-sample estimate of the parameter vector applying at the first postsample date (Doan et al. [1986, p. 7]). Therefore we are able to make a forecast of one step ahead.

The procedure is described on the slightly different version of the dynamic model, which can be found in the lecture notes of Cifarelli and Muliere [1989]. The system is described by two equations:

$$\underline{y}_t = F_t \underline{\vartheta}_t + \underline{v}_t \quad (18)$$

where

$\underline{y}_t$ — *is (mx1) vector of the observed variables of the process*
$\underline{\vartheta}_t$ — *is (nx1) vector of parameters of the process at time t*
F_t — *is (mxn) matrix of independent variables noted at time t*
v_t — *is (mx1) vector of stochastic independent errors with mean zero and cov matrix* V_t

The equation describing the evolution of parameters is following:

$$\underline{\vartheta}_t = G_t \underline{\vartheta}_{t-1} + \underline{w}_t \quad (19)$$

where

G_t — *is (nxn) matrix, describing how parameters evolve in time*
$\underline{w}_t$ — *is (nx1) vector of stochastic errors with covariance matrix* W_t

With the property:

$$\underline{\vartheta}_{t+1} \perp \underline{\vartheta}_{t-1} | \underline{\vartheta}_t$$

We assign a prior distribution to the first observation of the parameters vector:

$$\underline{\vartheta}_0 \sim N(m_0, C_0)$$

Then step by step we assign probability distribution to next time parameters using following procedure.

We know the distribution of

$$\underline{\vartheta}_{(t-1)} \Big| \underline{y}_{(t-1)}$$

Therefore it is straight forward to determine the distribution of

$$\underline{\vartheta}_t \Big| \underline{y}_{(t-1)}$$

which is given by:

$$\underline{\vartheta}_t \Big| \underline{y}_{(t-1)} \sim N\left(G\underline{m}_{t-1}, G_t C_{t-1} G_t^{'} + W_t\right)$$

with

$$G_t C_{t-1} G_t^{'} + W_t \equiv R_t$$

We can treat this as a prior for time t inference. Now we know that:

$$\underline{Y}_t \Big| \underline{\vartheta}_t \sim N(F_t \underline{\vartheta}_t, V_t)$$

Therefore, combining prior and likelihood function we obtain posterior distribution for $\underline{\vartheta}_t$

(20) $$\underline{\vartheta}_t \sim N(\underline{m}_t, C_t)$$

where:

$$\underline{m}_t = \left(R_t^{-1} + F_t^{'} V_t^{-1} F_t\right)^{-1} + \left(F_t^{'} V_t^{-1} \underline{y}_t + R_t^{-1} G_t \underline{m}_{t-1}\right)$$

$$C_t = \left(R_t^{-1} + F_t^{'} V_t^{-1} F_t\right)^{-1}$$

Now we are able to compute forecast of the variable y for the time t+1, having all the information available at time t:

(21)

$$\underline{y}_{t+1} \Big| \underline{y}_t \sim N\left(F_t G_t \underline{m}_{t-1}, F_t R_t F_t^{'} + V_t\right)$$

4. Structural models as a source for prior information

This part of the paper could be easily placed into above discussion about BVAR models and Minnesota prior, but I decided to present the issue separately, because it deserves some attention. The simple forecast procedure associated with already estimated VAR has been already described, but here we discuss another method of assigning prior distribution to the parameters. It combines structural approach, which pursuits to discover structural relations in the economy and represent them through the model, with the pure statistical procedures used to forecast future values of economic variables. The procedure below is described after Dejong and Dave [2007].

The VAR model is the same as above:

(22)

$$\gamma_t = \vartheta \gamma_{t-1} + V_t$$

The structural model in the form of stochastic difference equations, coming from solution of the first order conditions and log-linearization of equations, is the following:

(23) $$x_t = F(\mu)x_{t-1} + e_t$$

(24) $$X_t = H(\mu)x_t$$

The first dynamic equation describes the evolution of unobserved predetermined variables (like capital stock, technology), the second static equation describes relation between unobserved and observed variables (like output, interest rate, unemployment). We assume that the prior distribution over parameters μ has been specified by a researcher. We can also assume that X_t and Y_t coincide.

A researcher draws 10 000 realizations from the prior distribution of μ. For each realization the following procedure is repeated 10 times. First, T random shocks e_t are simulated, which together with the value x_0 are fed to equation (23). Then T realizations of X_t are obtained. After this, a researcher estimates the parameters ϑ from the VAR by standard OLS regression. These results allow to compute the prior distribution of parameters ϑ and Σ — covariance matrix of V_t. Having specified their means, variances and other properties (ex. Degrees of freedom in Inverted-Wishart distribution), we can write the multivariate prior distribution for parameters (note that from here below the small theta denotes vec(ϑ) from above, whereas big theta denotes matrix; hats stand for OLS estimators):

(25) $$P(\vartheta|\Sigma) \sim N\left(\beta^*, \Sigma \otimes N^{(*-1)}\right)$$

— multiv.normal distribution

(26) $$P(\Sigma) \sim IW\left(H^*, \tau^*, n\right)$$

— inverted — Wishart distribution

The posterior distribution is given by:

(27)

$$P(\vartheta|\Sigma) \sim N\left(\vartheta^P, \Sigma \otimes \left(\gamma'\gamma + N^*\right)^{(-1)}\right)$$

(28)

$$P(\Sigma) \sim IW\left(H^P, \tau + \tau^*, n\right)$$

where:

$$\vartheta^P = \left[\Sigma^{-1} \otimes \left(\gamma'\gamma + N^*\right)\right]^{-1} \left[\left(\Sigma^{-1} \otimes \gamma'\gamma\right)\hat{\vartheta} + \left(\Sigma^{-1} \otimes N^*\right)\vartheta^*\right]$$

$$H^P = \tau^* H^* + \hat{H} + \left(\hat{\Theta} - \Theta^*\right) N^* \left(\gamma'\gamma + N^*\right)^{-1} \left(\gamma'\gamma\right) \left(\hat{\Theta} - \Theta^*\right)$$

Conclusions

Bayesian procedures in forecasting are not very different from classical ones, but the underlying difference is that they base

on different assumptions. The key point is that in bayesians statistics parameters are stochastic. We have to chose a model for their distribution together with its mean and variance (so called prior distribution), and then combine this distribution with a likellihood function for the data. In economic applications the most common model for the prior distribution is multivariate normal.

What is important, we can distinguish two very general approaches when it comes to selection of the prior distribution. First, a researcher can attribute a prior by himself, basing on his intuition, knowledge and some standard procedures, like with Minnesota prior, where a mean of the distribution is assumed to be one for the first lag and zero for higher order lags. Second, a researcher can use a structural models to form a prior.

In general, it can be said that in Bayesian approach intuition of a researcher plays greater role than in the classical approach. The subjective factor is certainly more important. Is it an advantage or a drawback? I doubt if this question would be answered unanimously.

References

Cifarelli D.M., Muliere P. [1989], *Statistica Bayesiana*, Gianni Iuculano Editore.

DeJong D., Dave Ch. [2007], *Structural Macroeconometrics*, Princeton University Press.

Doan T., Litterman R., Sims Ch. [1986], *Forecasting and Conditional Projection Using Realistic Prior Distribution*, Staff Report 93, Federal Reserve Bank of Minneapolis.

Goldberger A.S., Theil H. [1961], *On Pure and Mixed Statistical Estimation in Economics*, „International Economic Review", Vol. 2, no. 1.

Hamilton J.D. [1994], *Time Series Analysis*, Princeton University Press.

Litterman R. [1985], *Forecasting With Bayesian Vector Autoregressions — Four Years of Experience*, Staff Report 95, Federal Reserve Bank of Minneapolis.

Lucas R. [1976], *Econometric Policy Evaluation: A Critique*, Carnegie-Rochester Conference Series on Public Policy, Elsevier, Vol. 1(1).

Sims Ch. [1982], *Policy Analysis with Econometric Models*, „Brookings Papers on Economic Activity", No. 1.

Todd R. [1984], I*mproving Economic Forecasting With Bayesian Vector Autoregression*, „Federal Reserve Bank of Minneapolis Quarterly Review", Vol. 3., Fall.

Summary

Bayesian methods are widely used by central banks' researchers to estimate forecasting models. The main difference between Bayesian approach and classical statistical approach is that the former treats data as given and parameters as stochastic variables, whereas the latter looks at sample data as one from various possible realizations, given a single vector of parameters. Therefore the source of uncertainty is different when we analyze the economy using the Bayesian methods. In this paper I present the basic procedures associated with Bayesian approach in forecasting. I show the simplest application to the stationary autoregression process of order one and more developed applications to VAR models. The aim of the paper is to give to a reader a comprehensive guide to Bayesian forecasting methods, which can be treated as a foundation to more advanced studies.

Key words: Bayesian statistics, Bayesian methods, forecasting