

Białowolski, Piotr; Kuszewski, Tomasz; Witkowski, Bartosz

## Article

# Macroeconomic Forecasts in Models with Bayesian Averaging of Classical Estimates

Contemporary Economics

## Provided in Cooperation with:

University of Finance and Management, Warsaw

*Suggested Citation:* Białowolski, Piotr; Kuszewski, Tomasz; Witkowski, Bartosz (2012) : Macroeconomic Forecasts in Models with Bayesian Averaging of Classical Estimates, Contemporary Economics, ISSN 2084-0845, Vizja Press & IT, Warsaw, Vol. 6, Iss. 1, pp. 60-69, <https://doi.org/10.5709/ce.1897-9254.34>

This Version is available at:

<https://hdl.handle.net/10419/105421>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Received: 20 09 2011 | Accepted: 22 03 2012

# Macroeconomic Forecasts in Models with Bayesian Averaging of Classical Estimates

Piotr Białowolski<sup>1</sup>, Tomasz Kuszewski<sup>2</sup>, Bartosz Witkowski<sup>1</sup>

## ABSTRACT

The aim of this paper is to construct a forecasting model oriented on predicting basic macroeconomic variables, namely: the GDP growth rate, the unemployment rate, and the consumer price inflation. In order to select the set of the best regressors, Bayesian Averaging of Classical Estimators (BACE) is employed. The models are atheoretical (i.e. they do not reflect causal relationships postulated by the macroeconomic theory) and the role of regressors is played by business and consumer tendency survey-based indicators. Additionally, survey-based indicators are included with a lag that enables to forecast the variables of interest (GDP, unemployment, and inflation) for the four forthcoming quarters without the need to make any additional assumptions concerning the values of predictor variables in the forecast period. Bayesian Averaging of Classical Estimators is a method allowing for full and controlled overview of all econometric models which can be obtained out of a particular set of regressors. In this paper authors describe the method of generating a family of econometric models and the procedure for selection of a final forecasting model. Verification of the procedure is performed by means of out-of-sample forecasts of main economic variables for the quarters of 2011. The accuracy of the forecasts implies that there is still a need to search for new solutions in the atheoretical modelling.

**KEY WORDS:** Bayesian Averaging of Classical Estimates, business survey data, seasonality, automatic forecasting

**JEL Classification:** C10, C83, E32, E37

<sup>1</sup>Warsaw School of Economics, POLAND

<sup>2</sup>Univesity of Finance and Management in Warsaw, POLAND

## Introduction

The main aim of this paper is to construct a model that would accurately forecast basic macroeconomic indicators, i.e. the GDP growth rate (GDP), the consumer price inflation (CPI), and the unemployment rate (UNE) measured in line with the methodology of the Polish Labour Force Survey (BAEL). The model presented in this paper is by definition atheoretical, i.e. its specification does not intend to reflect the causal relations driven from the economic theory. Additionally, it is based on the results of business and consumer tendency surveys.

The events of the crisis period proved that Diebold's concern about macroeconomic forecasting was justified. Dynamic simultaneous equations models, general equilibrium models, dynamic stochastic general equilibrium models and vector autoregressive (VAR) models were predominantly unable to generate accurate short-term macroeconomic predictions in that period.

The concept of an atheoretical model construction was proposed by Sargent and Sims (1977) and Sims (1980). The idea was either acknowledged or criticised for being "atheoretical macroeconomics" (Cooley & LeRoy, 1985). Nevertheless, later it was distinguished by the Nobel Memorial Prize in Economic Sciences.

One of the first scholars to employ atheoretical modelling augmented with business and consumer survey

Correspondence concerning to this article should be addressed to: [bwitko@sgh.waw.pl](mailto:bwitko@sgh.waw.pl)

data in macroeconomic forecasting were Hansson, Jansson, and Lof (2003). Data from business and consumer tendency surveys were employed in the Polish forecasting models and discussed, among others, in papers by Białowolski et al., (2007) and Białowolski, Kuszewski & Witkowski, (2010). The model based on the survey data comprises opinions and forecasts of thousands of respondents answering questionnaires concerning the economic situation. Following Anuszevska (2010), it is assumed that respondents in such surveys share and shape a common knowledge on economic processes.

The earlier paper by Białowolski, Kuszewski & Witkowski, (2010) proposed specification of 6 families of forecasting models for the aforementioned variables: GDP, CPI, and UNE. The order in which regressors were explained was different for each specification and the parameters were estimated with supervised stepwise regression which enabled to estimate parameters in the models following the general to specific rule in order to avoid the phenomenon of collinearity. In this paper authors investigate whether a different approach (i.e. BACE) leads to better out-of-sample forecasts.

### Bayesian Averaging of Classical Estimators

The algorithm of Bayesian Averaging of Classical Estimators (BACE) was developed by Sala-i-Martin, Doppelhofer, and Miller (2004) and later elaborated by Stadelmann (2010). The general principle for this model can be summarized by the following assumptions: let  $Y$  be a regressor and  $X = \{X_1, X_2, \dots, X_s\}$  be an  $s$ -element set of potential predictor variables. Assuming that particular independent variables  $X_i, i=1, \dots, s$ , are not collinear, it is possible to form  $2^s - 1$  non-empty subsets of set  $X$ . It is possible to estimate a regression model for each of these subsets. Particular coefficients in the final model ( $\beta_i$ ) obtained with the BACE method are weighted averages of parameter estimates for the predictor variable  $X_i$  in all regressions estimated with the application of a classical estimator (authors provide estimates obtained with ordinary least squares). Weights employed in the calculation of averages are proportional to the “quality” of particular regressions determined by the values of the Schwarz criterion. In contrast to the classical Bayesian methods, the role of prior assumptions in the model is minimal.

As long as we treat particular potential regressors as independent, the BACE approach does not require to assume any prior distribution for all parameters but only to specify the total number of predictor variables in the “correct” model.

The aim of the BACE algorithm is to estimate the  $\beta_i$  coefficients in a regression for a given variable  $Y$  on a particular  $X_i$  without limiting the analysis to a *single* subset of  $X$ . Moreover BACE estimation allows to check the significance of the particular  $X_i$  without limiting the models taken into consideration to merely one chosen subset of  $X$ . Since it is possible to form  $2^s$  subsets of  $X$ , it is also possible to form and estimate  $2^s$  different models using one dependent variable  $Y$ . The models are marked as  $M_i$ , where  $i=1, \dots, 2^s$ . Let the number of independent variables in a set of regressors  $X_i$  of model  $M_i$  be equal to  $k_i$ . Let  $\bar{s}$  stand for the expected number of independent variables in the “correct” model. In the original BACE approach, each potential independent variable  $X_i$  is considered separately regardless of the remaining potential independent variables as it is assumed that the probability of including a variable in the “correct” model does not vary among the variables. It leads to the assumption that the probability of including a variable  $X_i$  in a model equals  $P(\beta_i \neq 0) = \frac{\bar{s}}{s}$  for each  $i=1, \dots, s$ . Therefore, the prior probability of model  $M_i$ ,  $P(M_i)$  may be written with application of binomial distribution:

$$P(M_i) = \left(\frac{\bar{s}}{s}\right)^{k_i} \left(1 - \frac{\bar{s}}{s}\right)^{s-k_i}, \quad (1)$$

where the prior probability for each model with a predefined number of independent variables  $X_i$  is identical. The posterior probability of model  $M_i$  may be estimated with the use of a Bayes formula:

$$P(M_i | D) = \frac{P(M_i)P(D | M_i)}{\sum_{j=1}^{2^s} P(M_j)P(D | M_j)}, \quad (2)$$

where  $D$  is an  $n$ -element dataset (i.e. the values of a regressant and regressors), and the probability of generating  $D$  by model  $M_i$  is:

$$P(D | M_i) = \int L(D, \theta_i) P(\theta_i | M_i) d\theta_i, \quad (3)$$

where  $L(D, \hat{\theta}_i)$  is the likelihood function of model  $M_i$ , and  $\hat{\theta}_i$  is the  $k_i$ -element vector of parameters of  $M_i$  (standing by each  $X_i$ ). Although finding the value of formula (3) is complicated in terms of calculation, it does not constitute a barrier to the utilisation of this method since (3) can be approximated with the Bayesian Schwarz's information criterion:

$$P(D | M_i) = \exp(\ln L(D, \hat{\theta}_i) - 0.5k_i \ln n), \quad (4)$$

where  $L(D, \hat{\theta}_i)$  is the value of the likelihood function of model  $M_i$  with a vector of parameter estimates  $\hat{\theta}_i$ . In estimation procedure of a model with least squares, the value of  $L(D, \hat{\theta}_i)$  may be replaced with  $-0.5T \ln SSE_i$ , where  $SSE_i$  stands for the residual sum of squares of  $M_i$ . This changes formula (2) into the following:

$$P(M_i | D) = \frac{P(M_i) n^{-k_i/2} SSE_i^{-n/2}}{\sum_{j=1}^{2^s} P(M_j) n^{-k_j/2} SSE_j^{-n/2}}. \quad (5)$$

Formula (5) is of key importance as it defines the degree to which data suggest the "correctness" or "relevance" of model  $M_i$ . It can be used to estimate the expected value and the variation of parameters of regression  $\beta_i$  for the particular variables  $X_i$  in a linear model including a dependent variable  $Y$  as:

$$E(\beta_i | D) = \sum_{j=1}^{2^s} P(M_j | D) \hat{\beta}_{i,j} \quad (6)$$

and

$$D^2(\beta_i | D) = \sum_{j=1}^{2^s} P(M_j | D) \cdot D^2(\beta_{i,j} | D, M_j) + \sum_{j=1}^{2^s} P(M_j | D) \cdot (\hat{\beta}_{i,j} - E(\beta_i | D))^2, \quad (7)$$

where the value of  $\hat{\beta}_{i,j}$  is an estimate of the parameter of regression  $\beta_i$  (for a given variable  $X_i$ ), which has been calculated with ordinary least squares in model  $M_j$ . Moreover, since formula (5) may be employed for the purposes of establishing the level of "correctness" of a given model, the posterior probability that variable  $X_i$  can be included in the "correct" model may be expressed by:

$$P(\beta_i \neq 0 | D) = \sum_{j=1}^{2^s} P(M_j | D) \cdot I_{\beta_{i,j} \neq 0}, \quad (8)$$

where  $I_{\beta_{i,j} \neq 0}$  is an indicator of the presence of variable  $X_i$  in  $M_j$  and it takes on value 1 if the variable  $X_i$  is included in the model  $M_j$  and 0 otherwise.

It is noteworthy that assuming independence and exogeneity of particular regressors, the only parameter which requires a prior assumption is  $\bar{s}$ . As a rule,  $\bar{s}$  is given a value that reflects the expected number of predictor variables in the "correct" model. However, it is difficult to talk about the "correct" model since the following considerations are of atheoretical nature.

Application of the BACE algorithm is time consuming as it requires a considerable number of models (precisely  $2^s$ ) to be estimated. In the estimations carried out by Sala-i-Martin, Doppelhofer, and Miller (2004), the number of variables amounted to 60, which made estimation of all possible models unattainable. The common practice in such situations is to estimate only a part of all possible equations on the basis of sets of regressors randomly selected from the initial set.

In order to employ the BACE algorithm, a set of potential regressors and the regressant must be pre-defined. Furthermore, the regressors selected for the final model are not actually orthogonal which may result in the problem of collinearity. Consequently, if we treat BACE as an algorithm of pre-selection of variables for the final model only, which is the most frequent type of approach, it is necessary to verify the possible excessive collinearity of some regressors and – probably – to modify also the list of selected variables. Consequently, sequence of activities presented below was followed:

1. It was assumed that the aim of this paper is to construct models of the following form:

$$\begin{aligned} GDP_t &= \beta_0^1 + \gamma^1 GDP_{t-1} + \mathbf{x}_{t-m}^1 \mathbf{\beta}^1 + \varepsilon_t^1 \\ UNE_t &= \beta_0^2 + \gamma^2 UNE_{t-1} + \mathbf{x}_{t-m}^2 \mathbf{\beta}^2 + \varepsilon_t^2 \\ CPI_t &= \beta_0^3 + \gamma^3 CPI_{t-1} + \mathbf{x}_{t-m}^3 \mathbf{\beta}^3 + \varepsilon_t^3, \quad 1=1, \dots, T; m \in \{1, 3\} \end{aligned} \quad (9)$$

where:

$GDP_t, GDP_{t-1}$  – is the current and one-quarter lagged GDP growth rate expressed in the year-on-year basis;  
 $UNE_t, UNE_{t-1}$  – is the current and one-quarter lagged indicator of the unemployment rate according to the Polish Labour Force Survey;

$CPI_t, CPI_{t-1}$  – is the current and one-quarter lagged indicator of the consumer price inflation;

$\mathbf{x}_{t-m}^{1,m}, \mathbf{x}_{t-m}^{2,m}, \mathbf{x}_{t-m}^{3,m}$  – are vectors of regressors selected from indices of economic tendency surveys lagged by

one or three quarters in equations describing respectively: the GDP growth rate (1), unemployment (2), and inflation (3).

From the point of view of the forecast quality, it might be expected that one-period lags in indicators of an economic situation will result in better predictions of the variables of interest. However, models with business and consumer tendency survey-based indicators lagged by three periods allow for longer forecasts without the need to model the behaviour of the indicators of the economic situation in separate models. Therefore, two three-equation models are specified. In the first model ( $m=1$ ), a regressant lagged by one period and a combination of one-period lags in indicators of an economic situation are taken into account from among predictor variables. In the second model, a regressant lagged also by one period and a combination of three-period lags in indicators of an economic situation are taken into account from among predictor variables.

2. A group of potential regressors is preliminary selected from the set of all accessible indicators of the economic situation. The selection is based on the outcome of previous research presented by Białowolski, Kuszewski & Witkowski, (2010).
3.  $\mathbf{x}_{t-m}^{1,m}, \mathbf{x}_{t-m}^{2,m}, \mathbf{x}_{t-m}^{3,m}$  were selected with the application of the BACE method from among variables initially chosen as regressors. It was assumed that selection should be performed separately for each of the six equations (three predictor variables in models with two lag orders). Thus, the algorithm was implemented six times. A considerable number (22) of regressors in the initial set and the necessity to apply the method six times provoked the use of such a variant of an algorithm which allows to estimate only a part of all the possible models. Sets of regressors for each model were selected beforehand. For  $m=1$  over 70,000 different sets of regressors were selected and for  $m=3$  the number of selected sets amounted to 200,000.
4. Models with selected sets of variables were re-estimated and a couple of single variables were eliminated in order to avoid multicollinearity of predictor variables.
5. The final model specifications were again estimated and the estimates were used to forecast the macroeconomic indicators of interest.

The remaining part of this paper is devoted to a presentation of theoretical grounds for selecting the pre-

liminary set of predictors. Moreover, a further presentation and discussion of the results are provided.

### Indicators from economic tendency surveys in modelling the GDP, inflation and unemployment

Economic tendency surveys and the indicators derived from them are intended to provide leading information about the state of economy. Various papers focused on the applicability of qualitative indicators from economic tendency surveys to the forecasts of main economic indicators (e.g. Białowolski, Kuszewski & Witkowski, 2010). Due to a wide variety of questions in business and consumer tendency questionnaires, data from them may serve as a source of leading information on production in the manufacturing sector, construction output, turnover, prices, investments, changes in employment (with respect to the sector of companies); and analyzing the performance of the households sector allows to gain information on the demand, including the propensity to buy durable goods (OECD, 2003; European Commission, 2006; Bieć, 1996).

All data acquired in economic tendency surveys are directly related to important indices reported in national statistics and can thus be employed as leading information and allow for forecasting basic macroeconomic values – e.g. the GDP growth rate. Data from the economic tendency surveys is also acquired with significant lead in comparison to the information inferred from national statistics (concerning production, orders, and trade, etc.). It is especially important in the GDP forecasts since statistics on this indicator and its components are available with very significant delay.

The list of macroeconomic indicators that have corresponding indicators in economic tendency surveys is very long (a more detailed list of relations between qualitative and quantitative data can be found in the European Commission manual, 2006). The following analysis, however, focuses on three main areas of application of the indicators of the economic situation, namely: the forecast of the gross domestic product, forecasting inflation, and forecasting the unemployment rate. Due to the fact that the analyses are focused on the above mentioned areas, a set of time series derived from economic tendency surveys was selected. Preliminary selection was based on expert criteria, however, it was also strongly motivated by the rela-

tions identified in previous research of the authors (Białowolski, Kuszewski & Witkowski 2010).

The dependent variables in models comprise simple balances regarding an answer to a single question in a questionnaire, composite indicators combining information acquired by means of several questions, and the information inferred from national statistics. Nine composite indicators and nine balances were used in the analysis of the Polish economy. Additionally, four composite indicators concerning the state and the forecasts of the German economy were employed.

The group of composite indicators for the Polish economy comprises five indicators developed at the Research Institute for Economic Development at the Warsaw School of Economics (WSE). These indicators concern the economic situation of manufacturing (ind\_r), the economic situation of households (hhs), the economic situation of the construction industry (constr), the economic situation in trade (trade) and the economic situation in agriculture (agri) – information on the economic situation in the particular sectors of economy and the methodological basis for the construction of the indicators in these sectors is available in research papers of the Research Institute for Economic Development (RIED) at WSE respectively: Adamowicz & Walczyk, 2011; Majchrzak, 2010; Gorzelak & Zimny 2010; Białowolski & Dudek, 2010; Podgórska, 2010. The indicators contain information about the major part of economic activity performed in Poland, which is potentially important in the analysis conducted with the BACE method as a wide variety of indicators can be selected in the preliminary set.

The additional group of indicators encompasses data acquired from the BIEC which is a company oriented on providing composite indicators for the Polish economy. The following indicators were employed in the current study: the leading indicator for the economic situation (wwk), the future inflation index (wpi), and the labour market index (wrk). The proper methodological basis can be found in the literature (inter alia: Białowolski et al., 2007; Białowolski & Żochowski, 2006; Pater, 2006).

The initial set of predictors also contains a variable not derived directly from data on the economic situation but reflecting the economic situation observed by logistics managers – the PMI™ published by the HSBC bank. The index is based on individual diffusion indices which reflect changes in production, new orders,

time of delivery, and inventory of purchased goods. It is commonly recognised that the economic situation in the logistics sector is a leading indicator in the business cycle. This claim is substantiated in the book by Zarnovitz (1992) who points to an apparent tendency of new orders for suppliers to significantly lead other variables in the business cycle.

The models also utilise four series reflecting the performance of the German economy which is currently the biggest trade partner of Polish companies. The demand for Polish export depends to a large extent on the economic situation abroad. The series published by the Institute IFO in Munich (ifo\_bs – an indicator reflecting the current business situation, ifo\_be – an indicator reflecting the expectations of entrepreneurs, and ifo\_bc – reflecting the general business climate in Germany) were used as well as the indicator constructed by ZEW (zew\_ies – an Indicator of Economic Sentiment) which was employed in the study for the purposes of describing the situation in Germany. This indicator is derived from the opinions of professionals from Germany concerning the economic development in this country in consecutive six months.

In addition to synthetic indicators, the preliminary set of variables also comprised the time series selected from the survey conducted in the manufacturing industry and the State of the Households Survey. Due to the authors' objectives (i.e. creation of forecasts) only the indicators concerning expectations were taken into account in the analysis. As far as the survey on the economic situation in the manufacturing industry is concerned, the indicators relating to the forecast of orders (q2f) and the forecast of inventory (q4f) were selected. While the indicator concerning orders has a clear tendency to lead, the indicator concerning inventory has a tendency to be lagged rather than to lead in the cycle (Zarnovitz, 1992). Nevertheless, the indicators concerning inventory are very reliable and are recommended by the European Commission for creating a synthetic index of the economic situation in the manufacturing industry (European Commission, 2006). Furthermore, since this study is also devoted to forecasting inflation and unemployment, the forecasting series of price changes (q5f) and the expected employment in companies (q6f) are also taken into account as far as the survey on the economic situation in industry is concerned.

The study made use of the following questions asked in the State of the Households Survey: predictions of the general economic situation (hhs\_q4), unemployment changes forecasts (hhs\_q7), inflation expectations (hhs\_q5), and forecasts of durable goods purchases (hhs\_q9). The first three are directly related to the areas of interest in this forecast (the GDP, unemployment, and inflation). As the matter of fact, purchases of durable goods are very sensitive towards the expected variation in the economic situation. This factor is also the most volatile element of households' consumption and, therefore, the forecast of the behaviour of households in this respect constitutes an important factor influencing the forecasts of the general economic situation.

Databases contain time series of quarterly data from business and consumer tendency surveys in Poland conducted between the first quarter of 1996 and the first quarter of 2011, and time series of regressants for the period between the first quarter of 1996 and the fourth quarter of 2010.

### Estimation of the econometric models

In order to employ the algorithm of BACE in the study, it is necessary to assume  $\bar{s}$  – a number of predictor variables in the “correct” model. The value of this number is assumed arbitrary, which is especially typical of atheoretical models. Basing on the previous re-

search results by Białowolski, Kuszewski & Witkowski (2010), it has been assumed that  $\bar{s} = 6$ , however, it is noteworthy that the final number of predictor variables in particular models may differ from the prior assumption. Simultaneously, the influence of a change in the value of  $\bar{s}$  on the set of predictor variables has been tested. The research leads to the conclusion that assuming that  $\bar{s} = 6$  is adequate. Bigger models often require the excessive number of variables to be eliminated due to their collinearity. Furthermore, the values of probabilities calculated in line with (8) change proportionally to  $\bar{s}$  but the order of values (8) for each indicator of the economic situation is preserved.

The initial set of indicators of the economic situation, which are under discussion, comprises 22 variables. If we assume that  $\bar{s} = 6$ , theoretically variables whose probability (8) is over  $6/22 \approx 0.27$  should be included in the set. Table 1 contains the expected values of parameters (6), the Bayesian probabilities of inclusion (8) and the weighted empirical significance levels for test  $t$ . This paper presents only the results of estimation of models employing the three-period ( $m=3$ ) lagged indices of an economic situation. The results of estimation of models using one-period lagged indices of an economic situation ( $m=1$ ) were not discussed in the text due to the fact that the forecast acquired through the use of these models occurred highly unsatisfactory<sup>1</sup>.

**Table 1.** The results of the predictors selection process with application of the BACE algorithm in models with three-period lagged business and consumer sentiment indicators ( $m=3$ )

Regressant →	GDP			UNE			CPI		
	$E(\beta)$	P(X)	t(X)	$E(\beta)$	P(X)	t(X)	$E(\beta)$	P(X)	t(X)
ifo_bc	-0.040	0.115	0.185	-0.017	0.075	0.438	-0.012	0.100	0.321
ifo_bs	-0.029	0.043	0.217	-0.009	0.084	0.445	-0.038	0.084	0.273
ifo_be	-0.052	0.208	0.065	-0.012	0.059	0.507	0.047	0.128	0.207
zew_ies	-0.003	0.027	0.286	-0.003	0.111	0.412	-0.001	0.064	0.511
Pmi	-0.001	0.040	0.193	-0.001	0.111	0.210	-0.002	0.656	0.015
biec_wwk	-0.002	0.043	0.679	-0.049	0.505	0.049	-0.009	0.080	0.296
biec_wpi	-0.025	0.027	0.405	0.118	0.626	0.009	-0.077	0.176	0.186
biec_wrp	-0.021	0.037	0.343	0.038	0.109	0.256	0.020	0.105	0.336
ind_q1f	-0.020	0.101	0.120	0.025	0.151	0.132	0.016	0.119	0.264
ind_q2f	0.001	0.022	0.470	-0.001	0.061	0.508	-0.003	0.359	0.143



Table 1. (continued)

Regressant →	GDP			UNE			CPI		
	$E(\beta)$	P(X)	t(X)	$E(\beta)$	P(X)	t(X)	$E(\beta)$	P(X)	t(X)
ind_q4f	0.040	0.033	0.342	0.078	0.177	0.088	-0.100	0.331	0.060
ind_q5f	-0.019	0.146	0.172	-0.008	0.196	0.549	-0.008	0.093	0.541
ind_q6f	0.006	0.022	0.490	0.0128	0.100	0.363	0.001	0.058	0.456
ind_r	0.001	0.018	-0.013	0.001	0.112	0.567	0.004	0.803	0.009
hhs_q4	0.009	0.032	0.417	-0.013	0.129	0.320	0.018	0.151	0.216
hhs_q6	-0.043	0.992	0.001	-0.011	0.174	0.281	-0.006	0.091	0.509
hhs_q7	-0.006	0.084	0.435	0.004	0.079	0.475	-0.021	0.282	0.098
hhs_q9	0.003	0.020	0.591	-0.033	0.772	0.009	0.038	0.726	0.010
Hhs	0.009	0.028	0.401	-0.028	0.193	0.276	0.009	0.152	0.222
Trade	-0.018	0.027	0.271	0.043	0.848	0.003	-0.009	0.074	0.564
Agri	-0.014	0.032	0.344	-0.054	0.916	0.002	-0.061	0.539	0.017
Constr	-0.003	0.051	0.595	-0.012	0.086	0.251	0.016	0.253	0.145
Y(t-1)	0.858	-	0.000	0.841	-	0.000	0.844	-	0.000

Note: Y(t-1) denotes a one-quarter lagged variable, i.e. GDP, UNE or CPI respectively;  $E(\beta)$  denotes parameters estimated along with the formula (6), P(X) those estimated along with formula (8), t(X) is an average of the empirical significance level of a given variable with weights estimated in line with formula (5).

The BACE algorithm assumes independence of regressors. In fact, it would be hard to find an area in macroeconomics where a relationship between variables considered as regressors were inexistent. Therefore, the set of regressors finally included in models under consideration was slightly modified by means of excluding the variables responsible for an increase in collinearity from the set. Additionally, in one case, the set of variables was expanded by one variable whose Bayesian probability of inclusion was slightly below 0.27. If strict rules governing the algorithm of BACE had been maintained in this case, the equation under consideration would contain one regressor only.

Table 1 contains also the average empirical significance levels for test  $t$ . They have been estimated as weighted averages of the empirical significance levels for the particular variables in each model where weights are calculated in line with formula (5). It is noteworthy that an almost identical set of regressors  $\mathbf{x}_{t-m}^{1,m}, \mathbf{x}_{t-m}^{2,m}, \mathbf{x}_{t-m}^{3,m}$  would be selected if the average empirical significance levels for test  $t$  were taken into account (as with the use of the Bayesian probabilities of inclusion).

After having selected the variables with the use of the BACE and having eliminated the variables causing statistical collinearity, the final forms of models were estimated and applied for the purposes of forecasting afterwards. Table 2 contains the results of estimation of models.

The Final forms of models underwent standard statistical verification. The models occurred to be well-adjusted to empirical data. The series of residuals of models were found to be stationary, the normality of residuals in all models could not be rejected, no phenomenon of heteroscedasticity of residuals was observed, and the statistical significance of the set of predictor variables was confirmed.

Although as a rule there is no need for substantive interpretation of the carried out estimations and sets of predictor variables selected with the use of the method of BACE, several facts are noteworthy. In line with the initial intention, all equations contain predictor variables which are one-period lagged dependent variables. The equation concerning changes in the gross domestic product takes into account the influence of the external environment on the GDP (ifo\_be)



and the expectations of households with respect to prices (hhs\_q6). The equation concerning the unemployment rate unexpectedly takes into account the balance of answers referring to the purchase of durable goods acquired through the survey on the economic situation of households (hhs\_q9), the leading indices of the economic situation and inflation (biec\_wwk and biec\_wpi), and the results of a test on the economic situation in agriculture and trade. The equation concern-

ing changes in the consumer price inflation takes into account, among the regressants, the balance of answers to questions asked in the households' survey regarding the unemployment forecast (hhs\_q7), purchases of durable goods (hhs\_q9), and expectations concerning the inventory of manufacturing companies (ind\_qf4). However, no variables regarding the perception of the processes on the market of consumption goods are included.

**Table 2.** Results of econometric models with three-period lagged business and consumer survey indicators ( $m=3$ )

Regressant →	GDP		UNE		CPI	
Predictor variables and values of t Student statistics	GDP (t-1)	16.20	UNE(t-1)	18.47	CPI(t-1)	17.82
	ifo_be	-2.59	biec_wwk	-1.80	pmi	-2.16
	hhs_q6	-4.68	biec_wpi	2.98	ind_q4f	-2.59
			hhs_q9	-3.97	ind_r	4.64
			trade	3.64	hhs_q7	-2.06
			agri	-7.01	hhs_q9	2.06
					agri	-4.13
					constr	1.73
Adj. R-squared	0.8612	0.9790	0.9776			

**Table 3.** Out-of-sample forecasts of GDP, CPI and UNE for the quarters of 2011 based on econometric models with three-period lagged business and consumer survey indicators ( $m=3$ )

Quarters	GDP	UNE	CPI
I	4.5	9.5	3.4
II	3.8	8.7	3.4
III	2.5	7.9	3.0
IV	1.2	8.1	2.9

## Forecasts and conclusions

The lags of predictor variables in the econometric models were selected in such a way so as to enable the forecast of GDP, UNE, and CPI four quarters ahead without the need to make additional assumptions concerning the values of predictor variables in the forecast period. As already mentioned, the last observation in the time series of regressants was reported for the fourth quarter of 2010 and the last observation in the time series of data acquired from the economic tendency surveys was reported for the first quarter of 2011.

Since regressants lagged by one period are included as predictor variables, it is necessary to perform the forecast for four successive quarters in a stepwise manner. The values of regressors – variables derived from the tendency surveys performed in the past – are known for the whole period of forecast. The predicted values of the GDP, UNE, and CPI for a given quarter will become the values of predictor variables in the consecutive quarter. Such a procedure is faulty since it leads to accumulation of forecast errors. Table 3 presents the results of the forecasting process for the four quarters of 2011.

The values acquired in the forecasts point to the necessity of consideration of further improvement of forecasting models. According to the Central Statistical Office, the variable GDP took the value of 4.4 and 4.3 in the first two quarters of 2011, the variable UNE took the values of 10.0 and 9.5 in the same period, and the variable CPI took the values 3.8 and 4.6 in these periods. The forecasts draw attention to a rapid decrease in the rate of growth of the gross domestic product, which was not in line with the expectations of analysts. On the contrary, the estimates of the state of the labour market and the consumer price inflation seem to be too optimistic.

Further research in the field of forecasting models of basic macroeconomic variables employing survey data will encompass the inclusion of dynamic factor models into the set of models under analysis (Boivin & Ng, 2005; Forni & Reichlin, 1998; Stock & Watson, 2002). Furthermore, the possibilities raised by the forecasting models whose parameters are estimated on the basis of time series with removed seasonal effects should be considered further, especially the methods of deseasonalization taking different phases of the economic situation into account (Canova & Ghysels, 1994).

## References

- Anuszevska I. (2010). *Sondaże – zwierciadło społeczeństwa. Rytuały komunikacyjne a kreowanie wiedzy wspólnej* [Surveys – the Reflection of Society. Communication Rituals and Formation of Common Knowledge]. Warszawa: Wydawnictwo Fachowe CeDeWu.pl.
- Adamowicz, E., & Walczyk, K. (2011). *Koniunktura w przemyśle – Polska, styczeń 2011* [The Economic Situation in Manufacturing Industry– Poland, January 2011]. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- Białowolski P., Drozdowicz-Bieć M. (Ed.), Lada K., Pater, R., Zwiernik P., Żochowski D., (2007). *Forecasting with composite coincident and leading indexes and the CLIMA model*. Warsaw: Warsaw School of Economics.
- Białowolski, P., Kuszewski, T., Witkowski, B. (2010). Prognozy kombinowane wskaźników makroekonomicznych z użyciem danych z testów koniunktury [Combined Forecasts of Macroeconomic Indicators with Business Survey Data]. *Współczesna Ekonomia*, 4(16), 41-58.
- Białowolski, P., Kuszewski, T., Witkowski, B., (2011). *Prognozy podstawowych wskaźników makroekonomicznych z użyciem danych z testów koniunktury* [Forecast of Basic Macroeconomic Indicators with the Use of Business Survey Data], IV Konferencja Naukowa Modelowanie i Prognozowanie Gospodarki Narodowej. Uniwersytet Gdański, Wydział Zarządzania, Gdańsk.
- Białowolski, P., & Dudek, S. (2010). *Kondycja gospodarstw domowych – badanie okresowe II kwartał 2010* [The State of the Households – Periodic Survey: 2nd Quarter of 2010]. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- Białowolski, P., & Żochowski, D. (2006). *Analiza własności prognostycznych komponentów WPI. W: Wskaźniki wyprzedzające* [The Analysis of Forecast Properties of the Future Inflation Index. In: Leading Indicators]. Prace i Materiały Instytutu Rozwoju Gospodarczego SGH. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- Bieć M. (1996). *Test koniunktury – metody, techniki, doświadczenia* [Economic Tendency Surveys – Methods, Techniques, Experiments]. Prace i Materiały Instytutu Rozwoju Gospodarczego, zeszyt 48. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- Boivin, S. & Ng, S. (2005). Understanding and comparing, factor-based forecasts. *International Journal of Central Banking*. December, 117-151.
- Canova F., & Ghysels E. (1994). Changes in seasonal patterns. Are they cyclical? *Journal of Economic Dynamics & Control*, 18, 1143-1171.
- Cooley, T.F. & LeRoy, S.F. (1985). Atheoretical macroeconomics: a critique. *Journal of Monetary Economics*, 16(3), 283-308.
- Diebold F.X. (1998). The past, present and future of macroeconomic forecasting. *The Journal of Economic Perspectives*, 12(2), 175-192.
- European Commission (2006). *The joint harmonised EU programme of business and consumer surveys*. European economy – Report and studies No 5. Bruksela: Belgia.
- Forni, M. & Reichlin, L. (1998). Let's get real: a factor analytical approach to disaggregate business cycle dynamics. *The Review of Economic Studies*, 65, 453-473.

- Gorzelak, E., & Zimny, Z. (2010). *Koniunktura w rolnictwie. Badanie okresowe - II kwartał 2010* [The Economic Situation of Agriculture. Periodic Study for the 2nd Quarter of 2010]. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- Hansson J., Jansson, M., & Lof M. (2003). *Business survey data: do they help in forecasting the macro economy?* Working Paper 84. Stockholm: The National Institute of Economic Research,
- Majchrzak, K. (2010). *Koniunktura w handlu. Badanie okresowe - II kwartał 2010* [The Economic Situation of Trade. Periodic Study for the 2nd Quarter of 2010]. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- OECD (2003). *Business tendency surveys – a handbook*. Paris: OECD.
- Pater, R. (2006). *Analiza przebiegu i wskazań wskaźnika rynku pracy. W: Wskaźniki wyprzedzające* [The Analysis of Labour Market Index. In: Leading Indicators], Prace i Materiały Instytutu Rozwoju Gospodarczego SGH. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- Podgórska, M. (2010). *Koniunktura w budownictwie. Badanie okresowe - II kwartał 2010* [The Economic Situation of Construction Industry. Periodic Study for the 2nd Quarter of 2010]. Warszawa: Instytut Rozwoju Gospodarczego SGH.
- Sala-i-Martin X., Doppelhofer G., Miller R. (2004). Determinants of long-term growth: a bayesian averaging of classical estimates (BACE) approach. *American Economic Review*, 94, 813-835.
- Sargent, T., & Sims, C.A. (1977). Business cycle modeling without pretending to have too much a-priori economic theory. In: C. Sims (Ed.), *New methods in business cycle research*. Minneapolis, MN: Federal Reserve Bank of Minneapolis.
- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 48, 1-48.
- Stadelmann D. (2010). Which factors capitalize into house prices? A bayesian averaging approach. *Journal of Housing Economics*, 19(3), 180-204.
- Stock, J. & Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97, 1167-1179.
- Zarnovitz, V. (1992). *Business cycles: theory, history, indicators and forecasting*. Chicago, IL: National Bureau of Economic Research.

## Notes

- 1 It is probable that the superiority of models with  $m=3$  is a consequence of a certain seasonal quality of the series under analysis.