ECONOMICS
& Sociology

Błażejowski, M., Gazda, J., Kwiatkowski, J. (2016), Bayesian Model Averaging in the Studies on Economic Growth in the EU Regions – Application of the gretl BMA package, *Economics and Sociology*, Vol. 9, No 4, pp. 168-175. **DOI: 10.14254/2071-789X.2016/9-4/10**

Marcin Błażejowski,
Toruń School od Banking,
Toruń, Poland,
E-mail:
marcin.blazejowski@wsb.torun.pl

Jakub Gazda,
Poznań University of Economics and Business,
Poznań, Poland,
E-mail: j.gazda@ue.poznan.pl

Jacek Kwiatkowski,
Nicolaus Copernicus University in Toruń,
Toruń, Poland,
E-mail: jkwiat@umk.pl

Received: March, 2016
1st Revision: October, 2016
Accepted: November, 2016

DOI: 10.14254/2071-789X.2016/9-4/10

JEL Classification: C11, O47, R11

BAYESIAN MODEL AVERAGING IN THE STUDIES ON ECONOMIC GROWTH IN THE EU REGIONS – APPLICATION OF THE GRETL BMA PACKAGE

ABSTRACT. Economic growth is again one of the most important economic issues in literature since the the 1980s. This paper falls into the mainstream of regional studies on economic growth and it tries to answer the recurring question: what are the determinants of economic growth at regional level. The objective of this article is to diagnose the determinants of economic growth among European regions on the basis of Bayesian methods applied to gretl software.

Keywords: economic growth, Bayesian methods, gretl software.

Introduction

Economic growth has been one of the most important economic issues in literature since the 1980s. Evolution of theoretical concepts and empirical studies on economic growth have resulted in a considerable broadening of the research scope, which was initially dominated by changes occurring at the level of entire economies. The interaction between theoretical concepts and empirical studies has gradually moved the research focus from the macroeconomic level to lower levels of economies' aggregation. One of these levels is the region, as seen in the NUTS (Nomenclature d'Unit´es Territoriales Statistiques) classification. The status of a region as a territorial, economic and social unit has grown along with the enlargement of the European Union and the emergence of CohesionPolicy, which promotes the rectification of disparities between regions, in particular in the countries with lower levels of development.

This paper falls into the mainstream of regional studies on economic growth and it tries to answer the recurring question: what are the determinants of economic growth at the regional level. The authors assume that the determinants of economic growth of a region are closely related to the stage of country's development level. Accurate identification of the factors that influence the pace of economic growth is among the most significant challenges contemporary theories of economics and economic policy are facing. The time frame here involves the selected years in the period 1997-2011. Thus, changes that occurred over the period from the turn of the 20th century until the global financial crunch could be identified. The selection of this period was also determined by the availability of statistical data (from the Eurostat). The foundation for the study was provided by the database developed by its authors for 222 regions of 16 economies within the EU (EU-15 plus Poland). Using the BMA (Bayesian Model Averaging) method a group of explanatory variables was proposed to determine potential factors responsible for differences in regional averages of GDP growth rate under a dynamic approach. The Bayesian approach was previously used in author's research Gazda and Puziak (2013) as well Błażejowski and Kwiatkowski (2013) and also in the economic studies by Simionescu *et al.* (2016 a) to identify the relationship between migration and economic growth and by Simionescu (2016) to select the determinants of permanent migration in Romania. The Bayesian methods are good alternatives to traditional methods used, for example, by Albu (2013) to select foreign trade and FDI as determinants of economic growth or by Albu (2006) and also Albu and Roudoi (2003) to study the relationship between economic growth, investment and interest rate. Cetin and Dogan (2015) supported the human capital-based growth hypothesis using ARDL bounds testing approach.

The Bayesian approach has the main advantage of solving the problem of small sets of data. This disadvantage might also be solved by other modern approaches. For example, Ruiz *et al.* (2016) used panel data models to analyze the relationship between economic growth and intangible capitals while Kilic and Arica (2014) used panel data models to assess the impact of economic freedom and inflation rate on economic growth. The effects on value added taxes on economic growth in the CEE countries were assessed by Simionescu and Albu (2016) using panel data models and the Bayesian approach. Moreover, Simionescu (2016 b) used Bayesian panel data models to analyze the relationship between economic growth and FDI.

1. Objective and scope of study

The objective of this article is to diagnose the determinants of economic growth among the EU regions on the basis of Bayesian methods. The study was conducted on the basis of data describing the statistical units of individual states. The analysis of economic growth determinants is to answer the question of what are the sources of economic growth among the EU regions. The main source of statistical data was the database of the European Commission (Eurostat). The amount of GDP per capita in individual regions expressed in euro was rendered in terms of fixed prices for 2000. Studying the regions of the European Union was made to the standard that has become a legally regulated taxonomic model for the EU members, namely the classification of territorial units for statistics, NUTS. As a statistical classification of classical structure NUTS is hierarchical and encompasses three ranks, named NUTS 1, NUTS 2 and NUTS 3.The ranks group those administrative units whose average size should fall within the appropriate population bracket determined for each level of NUTS. If in a given Member State there are no administrative units of a scale adequate to a given NUTS level, such a level is created by means of combining the already existing, smaller, adjacent administrative units. The assumption of the present study is that the analysis of regional economic growth was at the NUTS 2 level. The lower limit of population for these units amounts to 800,000 and the upper one is 3 mln. In the case of Poland this corresponds to the division into regions (in Polish – województwo).

2. Bayesian estimation and model selection in normal linear regression models

Let θ denotes the vector of parameters, which is the object of our interest. Let us also assume that the initial information concerning this vector can be expressed by means of prior density. Let us subsequently consider an econometric model where the observation vector $y=(y_1,...,y_N)^{'}$ has a probability distribution expressed by the density function $p(y\,|\,\theta)$. The Bayesian inference concerning the vector of parameters θ is a well-known Bayes' formula:

$$p(\theta\,|\,y)=\frac{p(\theta)p(y\,|\,\theta)}{p(y)}\propto p(\theta)p(\theta\,|\,y) \qquad (1)$$

where $p(\theta\,|\,y)$ stands for the posterior density distribution, describing a researcher's 'final' knowledge of the parameter θ, computed on the basis of initial (prior) knowledge and derived from the sample; $p(y)$ stands for the density of marginal distribution of the observation vector y, expressed for a continuous random variable as $p(y)=\int p(\theta)p(y\,|\,\theta)d\theta$; $p(y\,|\,\theta)$ is the sample density, which determines the degree of confidence concerning the values assumed by an examined phenomenon, given a set value of the parameter θ. It corresponds to the likelihood function, i.e. $l(\theta;y)=p(y\,|\,\theta)$.

Let us consider a set of mutually exclusive and competitive models $M_1,...,M_m$ and the corresponding prior probabilities $\Pr(M_1),...,\Pr(M_m)$, and $\sum_{r=1}^{m}\Pr(M_r)=1$. The posterior probability of any model M_i can be computed in the following manner on the basis of the Bayes' formula:

$$\Pr(M_i|\,y)=\frac{\Pr(M_i)p(y\,|\,M_i)}{\sum_{r=1}^{m}\Pr(M_r)p(y\,|\,M_r)}. \qquad (2)$$

Formula (2) allows the posterior probability of every model M_r $(r=1,...,m)$ to be calculated provided that we know the density of marginal distribution $P(y\,|\,M_r)$. The BMA method consists in the averaging of posterior distributions of interesting parameters, weighted by the posterior probabilities of individual specifications.

Let us assume that a researcher is interested in the parameter ψ, which is a common element of all competitive models. Since we know the posterior probability of each model, the following density of posterior distribution can be a source of information:

$$p(\psi\,|\,y)=\sum_{r=1}^{m}\Pr(M_r|\,y)p_r(\psi\,|\,y,M_r). \qquad (3)$$

Density $p(\psi\,|\,y)$ is therefore obtained by means of the weighted averaging of individual densities of posterior distributions $p_r(\psi\,|\,y)$, weighted by the posterior probabilities of competitive models. Selected moments of posterior distribution can be averaged analogically:

$$E\left(\psi^{s} \mid y\right)=\sum_{r=1}^{m} \Pr\left(M_{r} \mid y\right) E_{r}\left(\psi^{s} \mid y, M_{r}\right), \tag{4}$$

where s stands for the order of the moment $(s=1,2,...)$.

Let us assume that we have data derived from $i=1,...,N$ objects. The vector of observations $y=(y_1,...,y_N)'$ refers to the dependent variable. Let us also assume that we have *K* potential explanatory variables related to a dependentvariable. The matrix with dimensions $N \times K$ contains observations on the explanatory variables. Let M_r stand for $r=1,...,m$ regression models, where *m* stands for a maximum number of combinations of independent variables, i.e. $m=2^K$

The regression model has the following form:

$$y=\alpha l_N + X_r \beta_r + \varepsilon, \tag{5}$$

where l_N means an $N \times 1$ vector of ones, X_r is an $N \times k_r$ matrix related to model M_r, and containing some (or all) columns of matrix X, β_r is a $k_r \times 1$ vector of structural parameters, $\alpha$is an intercept coefficient, common for all regression models, a random errorε is a vector of dimensions $N \times 1$with normal distribution $N\left(0, h^{-1} I_N\right)$, parameter h is an inverse variance of random error, i.e. $h=\frac{1}{\sigma^2}$, and the symbol I_N stands for an identity matrix of size *N*.

Let us assume that we have initial information on regression coefficientsβ_r, and some knowledge on common parameters, i.e. h and α:

$$\beta_r \mid h \sim N\left(0_{k_r}, h^{-1}\left[g_r X_r' X_r\right]^{-1}\right) \tag{6}$$

and

$$p(h) \propto \frac{1}{h},\ p(\alpha) \propto 1. \tag{6a}$$

Symbol $N(a,B)$ stands for a multidimensional normal distribution with mean *a*, and variance *B*, g_r stands for a constant defined as follows (Fernandez *et al.*, 2001; Zellner, 1986):

$$g_r=\begin{cases} \frac{1}{K^2} & dla \quad N \le K^2 \\ \frac{1}{N} & dla \quad N > K \end{cases}. \tag{7}$$

Using Bayes' formula (1) we obtain the posterior distribution of the parameters we are interested in. It can be demonstrated that in this case, the posterior distribution of the vector of regression coefficients β_r is a multivariate Student-*t* distribution with the following vector of means:

$$E(\beta_r \mid y, M_r)=\left[(1+g_r) X_r' X_r\right]^{-1} X_r' y. \tag{8}$$

The matrix of posterior covariance has the following form:

$$\operatorname{var}(\beta_r \mid y, M_r) = \frac{Ns_r^2}{N-2}\left[(1+g_r)X_r^{'}X_r\right]^{-1}, \tag{9}$$

where $s_r^2 = \dfrac{\dfrac{1}{g_r+1}y^{'}P_{X_r}y + \dfrac{g_r}{g_r+1}(y-\bar{y}l_N)^{'}(y-\bar{y}l_N)}{N}$.

Given the above-mentioned assumptions, the sample density after the analytical integration of parameters in model r is as follows:

$$p(y \mid M_r) \propto \left(\frac{g_r}{g_r+1}\right)^{\frac{k_r}{2}}\left[\frac{1}{g_r+1}y^{'}P_{X_r}y + \frac{g_r}{g_r+1}(y-\bar{y}l_N)^{'}(y-\bar{y}l_N)\right]^{-\frac{N-1}{2}}, \tag{10}$$

where $P_{X_r} = I_N - X_r\left(X_r^{'}X_r\right)^{-1}X_r^{'}$.

We might be also interested in the estimates of posterior inclusion probability $Pr(i \mid y)$ (PIP) i.e., the probability that, conditional on the data, but unconditional with respect to the model space, thevariablex_i is relevant in explaining the dependent variable y.The posterior inclusion probability is calculated as the sum of the posterior model probabilities for all of the models including variable x_i.

3. MC^3 sampling algorithm

Let us now discuss the foundations of the MC^3 algorithm. It facilitates easy ‘capturing’ of the models with the greatest explanatory power. Its main task is to sample in the regions where the most likely models occur, while neglecting the areas with the least likely models. The MC^3 algorithm, developed by Madigan, York and Allard (1995), is a special case of a numerical procedure, referred to in the literature on the subject as the Metropolis-Hastings method, which in turn is a special case of the Monte Carlo method, based on Markov chains. It simulates a chain of models M^i for $i = 1, \dots, T$ to find the equilibrium distribution $Pr(M_r|y)$ of the posterior model probabilities. We do it as follows. We set a candidate model from the set of models, including the previously accepted model $M^{(i-1)}$, all models which delete one independent variable from $M^{(i-1)}$ and all models which add one independent variable to $M^{(i-1)}$. The chain is then constructed by drawing a candidate model M^*.

The acceptance probability is as follows:

$$\alpha\left(M^{(i-1)}, M^*\right) = \min\left\{\frac{p\left(y|M^*\right)p\left(M^*\right)}{p\left(y|M^{(i-1)}\right)p\left(M^{(i-1)}\right)}, 1\right\}, \tag{11}$$

where densities $p\left(y \mid M^*\right)$and $p\left(y \mid M^{(i-1)}\right)$ are calculated from the formula (10).

Posterior characteristics of selected parameters can be obtained by means of weighted averaging of individual posterior distributions (formula (3)) or Rao and Blackwell approach (Koop, 2003).

Summing up, it should be emphasized that Bayesian inference provides tools that describe the uncertainty related to the selection of a model in a strictly probabilistic manner. The above-mentioned MC^3 algorithm is in turn an efficient technique providing for sampling in the areas where the most likely models occur, while neglecting those where models with very small explanatory power emerge.

4. Empirical results

All numerical computations were carried out in the gretl's BMA package (see Błażejowski and Kwiatkowski, 2013). Authors specify the following entries in the GUI BMA window: model prior = 'binomial', prior average model size = '12.5' (we set the models priors to the uniform distribution), number of the top ranked models = '15', g-prior type = 'Benchmark prior', total number of replications = '1000000', percentage of burn-in draws = '10'. *Table 1* present the estimation results. They report the posterior means, standard errors of regressors and PIPs variables. Authors used the following variables:

1. Percentage of population with upper secondary and post-secondary education in 1997 (Upper secondary and post-secondary);
2. Natural logarithm of gross domestic product *per capita* in 1997 (Ln of GDP pc 1997);
3. Ratio of gross fixed capital formation to GDP from 1997 to 2011 (GFCF ratio);
4. Average number of bed places in 2002 (Number of bed places);
5. Average number of non-residents arriving from 1997 to 2011 (Arrivals of non-residents);
6. Feminization ratio in a region in 2011 (Feminization ratio);
7. Percentage of population with higher education in 2011 (Education);
8. Average number of high-tech patents per one million citizens in a region from 1997 to 2000 (Patents per million of inhabitants);
9. Average share of higher education sector in outlay for research and development in a region in relation to GDP from 1997 to 2011 (R&D higher education);
10. Average outlay for research and development in a region from 1997 to 2011 (R&D all sectors);
11. Percentage of people employed in industry in 1997 (Industry employment);
12. Average number of nights spent in hotels by non-residents from 1997 to 2011 (Nights spent by non-residents);
13. Average share of private sector in outlay for research and development in a region in relation to GDP from 1997 to 2007 (R&D private enterprise);
14. Average size of an agricultural farm in a region from 1997 to 2011 (Area of agricultural farm);
15. Average number of nights spent in hotels by residents from 1997 to 2011 (Nights spent by residents);
16. Population in a region in 1997 (Population);
17. Average number of high-tech patents per one million citizens in a region from 1997 to 2000 (High-tech patents);
18. Share of service sector generating gross value added in 1997 (Service sector);
19. Average number of ICT patents per one million citizens in a region from 1997 to 2000 (ICT patents);
20. Professional activity rate in 1997 (Professional activity rate);
21. Percentage of people employed in the service sector in 1997` (Services employment);
22. Average share of government sector in outlay for research and development in a region in relation to GDP from 1997 to 2007 (R&D government sector);
23. Share of industry sector generating gross value added in 1997 (Industry sector).

Table 1. PIPs, posterior means and standard errors for the regression coefficients obtained in BMA analysis

Name of variable	PIP	Mean	Std. Dev.
Upper secondary and post-secondary	1.00000	0.02665	0.00397
Ln of GDP pc 1997	0.99999	-0.01308	0.00274
GFCF ratio	0.99988	0.06478	0.01271
Number of bed places	0.93993	0.00000	0.00000
Arrivals of non-residents	0.88694	0.00000	0.00000
Feminization ratio	0.74682	0.19425	0.13662
Education	0.49597	0.03003	0.03469
Patents per million of inhabitants	0.30945	0.00001	0.00001
R&D higher education	0.29866	0.00135	0.00237
R&D all sectors	0.20242	0.00022	0.00056
Industry employment	0.17676	0.00288	0.00754
Nights spent by non-residents	0.14583	0.00000	0.00000
R&D private enterprise	0.13073	0.00013	0.00049
Area of agricultural farm	0.10932	0.00000	0.00000
Nights spent by residents	0.09691	0.00000	0.00000
Population	0.09534	0.00000	0.00000
High-tech patents	0.08898	0.00000	0.00001
Service sector	0.07592	0.00053	0.00263
ICT patents	0.07276	0.00000	0.00001
Professional activity rate	0.07134	0.00001	0.00004
Services employment	0.05701	0.00013	0.00326
R&D government sector	0.04966	0.00006	0.00073
Industry sector	0.04513	-0.00012	0.00199

Conclusions

In the case of the analysis carried out for the EU 15 plus Poland for the period 1997-2011 the results suggest that the average GDP growth in the regions was correlated mainly with variables which in the literature are widely recognized as being responsible for economic growth. Education took the first place what is very good evidence of the real use of the budget and from EU programs. The next place in the ranking was occupied by the GDP per capita in the first year of the study. This is consistent with the theory of convergence. High place GFC (expenditure on gross capital) illustrates importance of growth processes in the regional GDP growth. On top places of the ranking there are also variables associated with tourism. This shows an even distribution of economic policy priorities among the NUTS 2 regions in the process of creating economic growth.

References

Albu, L. L. (2006), Trends in the interest rate-investment-GDP growth relationship, *Journal for Economic Forecasting*, *3*(3), pp. 5-13.

Albu, L. L., & Roudoi, A. (2003), Factors and Mechanisms of Economic Growth in Transition Economies of Different Types (Case of Romania), *Journal for Economic Forecasting*, (4), pp. 50-64.

Albu, L. L. (2013), Foreign Trade and FDI as Main Factors of Growth in the EU, *Journal for Economic Forecasting*, (2), pp. 7-17.

Błażejowski, M., Kwiatkowski, J. (2013), Bayesian Model Averaging and Jointness Measures for gretl, *MPRA Paper*, 44322.

Cetin, M., & Dogan, I. (2015), The impact of education and health on economic growth: Evidence from Romania (1980-2011), *Journal for Economic Forecasting*, (2), pp. 133-147.

Fernandez, C, Ley, E, Steel, M. F. J. (2001), Benchmark Priors for Bayesian Model Averaging, *Journal of Econometrics*, 100(2), pp. 381-427.

Gazda, J., Puziak, M., (2013), The sources of economic growth in Polish regions, *International Journal of Academic Research*, 5, pp. 245-250.

Kilic, C., Arica, F. (2014), Economic Freedom, Inflation Rate and their Impact on Economic Growth: A Panel Data Analysis, *Journal for Economic Forecasting*, 17(1), pp. 160-176.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons, New Jersey,USA.

Madigan, D., York, J., Allard, D. (1995), Bayesian Graphical Models forDiscrete Data, *The International Statistical Review*, 63(2), pp. 215-232.

Ruiz, V. R. L., Navarro, J. L. A., & Peña, D. N. (2016), Economic Growth and Intangible Capitals: An International Panel Data Model Applied in the 21st Century, *Journal for Economic Forecasting*, (2), pp. 102-113.

Simionescu, M. (2016a), Macroeconomic Determinants of Permanent Emigration from Romania: a Bayesian Approach, *Journal of International Studies*, 9(2), pp. 170-180.

Simionescu, M., & Albu, L. L. (2016), The impact of standard value added tax on economic growth in CEE-5 countries: econometric analysis and simulations, *Technological and Economic Development of Economy*, 22(6), pp. 850-866.

Simionescu, M., Ciuiu, D., Bilan, Y., Strielkowski, W. (2016), GDP and Net Migration in Some Eastern and South-Eastern Countries of Europe. A Panel Data and Bayesian Approach, *Montenegrin Journal of Economics*, 12(2), p. 161.

Simionescu, M. (2016b), The relation between economic growth and foreign direct investment during the economic crisis in the European Union, *Zbornik radova Ekonomskog fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu*, 34(1), pp. 187-213.

Zellner, A. (1986), On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions, In: P. Goel and A. Zellner (ed.), *Bayesian Inferenceand Decision Techniques: Essays in Honor of Bruno de Finetti*, Amsterdam, Holland, Elsevier.