Folia Oeconomica Stetinensia
DOI: 10.2478/v10031-010-0003-9

Wydział Nauk Ekonomicznych i Zarządzania
Uniwersytetu Szczecińskiego

MODELS WITH VARYING PARAMETERS AS A TOOL TO CLASSIFY POLISH VOIVODSHIPS IN 2002–2008

Barbara Batóg, Ph.D.

Department of Econometrics and Statistics
Faculty of Economics and Management
University of Szczecin
Mickiewicza 64, 71-101 Szczecin, Poland
e-mail: batog@wneiz.pl

Katarzyna Wawrzyniak, Ph.D.

Department of Application of Mathematics in Economics
Faculty of Economics
West Pomeranian University of Technology
Janickiego 31, 71-270 Szczecin, Poland
e-mail: katarzyna.wawrzyniak@zut.edu.pl

Received 15 December 2010, Accepted 6 June 2011

Abstract

One of the often used measures of economic development is gross domestic product *per capita*. In Poland the Main Statistical Office collects the data on this variable on several levels of aggregation. The paper shows the application of panel data models in order to classify Polish voivodships according to the level of economic development. As explained variable the regional GDP *per capita* was used and such variables as structure of employees, unemployment rate or retail sales per capita were the explaining variables. As a result the groups of voivodships with similar pattern of economic development were distinguished.

Keywords: panel data models, classification, GDP *per capita*.

JEL classification: C23, R15.

Introduction

Regional analyses often compare the development degree of various objects on different aggregation levels in space and time. The basis for such comparison can be both a synthetic variable created with the use of, for instance, linear ordering methods[1], or a variable selected a priori out of a group of macroeconomic variables characterising an economic situation in the examined objects. Most often the role of such a synthetic variable is played by the gross domestic product *per capita*, the value of which is published by the national statistics authorities for the whole country, its regions, voivodships and sub-regions.

The aim of the article is to compile a classification of voivodships in the study period which both takes into consideration the diversified level of GDP *per capita* and gives the opportunity to detect common properties characterising individual voivodships as far as the impact of specific macroeconomic variables on its level is concerned.

For the GDP *per capita* the authors estimated models with a non-randomly varying intercept. When selecting the models they assumed that the mechanism of correlations between macroeconomic variables and the GDP is similar in all the voivodships, but in each of them we can observe differences in the level of a dependent variable that can be identified by varying intercept. Such an intercept varying by voivodships[2] was the base for the classification of Polish voivodships according to their economic standing. The Zachodniopomorskie (West Pomeranian) Voivodship was chosen as the reference object, then the other voivodships were subsequently analysed.

The authors suggested seven macroeconomic variables to constitute a set of explanatory variables. The study was conducted with the use of 2002–2008 statistical data. In order to examine the classification stability in time the classification results from the periods of 2002–2007 and 2002–2008 were compared.

1. Models with Varying Parameters

There are many models with varying parameters (for panel data) whose nature was described in works of such authors as Judge (1985), Johnston (1991) and Maddala (2006). Due to the aim of the study the authors used the model with intercept term non-randomly varying by objects (voivodships) (model 1).

The equation for any i-th object where the intercept term varies over the objects is:

$$y_i = \beta_{1i} j_T + x_{ki} \beta_k + e_i \quad (1)$$

where:

y_i – vector of values of endogenous variable for the object i in periods 1, 2, …, T,

e_i – random error,

j_T – vector of ones of the dimension of Tx1,

x_{ki} – matrix of explanatory variables in the model (excluding vector j) of the dimension Tx$(K - 1)$,

β_k – vector of parameters for explanatory variables (excluding the intercept term),

β_{1i} – an intercept term for the object i,

T – the number of periods,

N – the number of objects,

K – the number of explanatory variables.

We assume that the mean of random errors e_i ($i = 1, 2, ..., N$) for each object and period is zero, their variance is constant and they are not correlated over the objects.

Model (1) can be expressed as the following equation (2), including all the $N \cdot T$ observations and all the explanatory variables

$$Y = \left[I_N \otimes j_T \ \ x_k \right] \begin{bmatrix} \beta_1 \\ \beta_k \end{bmatrix} + e \tag{2}$$

where:

Y – vector of a values of endogenous variable for all objects and periods,

I_N – unit matrix of the N degree,

x_k – matrix of explanatory variables in the model (excluding vector j) of the dimension $T \cdot N \times (K - 1)$,

e – vector of random errors,

β_1 – vector of intercept terms consisting of components β_i,

$\otimes$ – the Kronecker product.

The model (2) form shows that the intercept term β_i for a given object is a parameter for a zero-one variable that adopts the value of 1 in case of the object i and the value of 0 in case of other objects.

Prior to estimating the model (2) its re-parametrization has been performed due to which the model takes the following form:

$$Y = \left[j_{NT} \begin{bmatrix} I_{N-1} \\ o^T \end{bmatrix} \otimes j_T x_k \right] \begin{bmatrix} \delta \\ \beta_k \end{bmatrix} + e \tag{3}$$

where $\delta = [\delta_1, \delta_2, ..., \delta_N]'$.

Thus the re-parametrization is performed through deleting from the model (2) the zero-one variable for the object N and replacing it with a variable consisting of $N \cdot T$ ones. The relation between the model (2) and model (3) parameters is defined by the formula (4).

$$\begin{aligned} \beta_{1N} &= \delta_1 \\ \beta_{1i} &= \delta_1 + \delta_{i+1} \end{aligned} \tag{4}$$

where $i = 1, 2, ..., N - 1$.

2. The Characteristics of the Data

The authors took advantage of statistical data from the Local Data Bank of the Central Statistical Office of Poland (GUS). The statistical population consisted of 16 Polish voivodships in the period of 2002–2008. The dependent variable describing the voivodships' economic standing was the gross domestic product *per capita* (PKB1), while the set of potential explanatory variables included the following macroeconomic variables[3]:

- the unemployment rate according to BAEL (%) – SB,
- the share of agriculture workers to the workers' total (%) – UPR,
- the share of industry workers to the workers' total (%) – UPP,
- the share of services workers to the workers' total (%) – UPU,
- an average gross wages in the enterprise sector (PLN) – WB,
- gross disposable income *per capita* (PLN) – DB1,
- retail sales *per capita* (PLN) – SD1.

The variables used in the study are the indicators of volume, structure and dynamics (excluding the gross salary). Such form of variables has enabled the authors to compare the voivodships in space and time.

Figure 1 presents the dynamics of the (PKB1) variable in individual voivodships in the period of 2002-2008. The presentation has been made in four quartile groups defined basing on the 2008 data according to the following rule:

group 1 – voivodships with PKB1 $\leq Q_1$,

group 2 – voivodships with $Q_1 <$ PKB1 $\leq Q_2$,

group 3 – voivodships with $Q_2 <$ PKB1 $\leq Q_3$,

group 4 – voivodships with PKB1 $> Q_3$.

We can see that the graphic presentation of this variable is similar in all the voivodships, the only difference being in its level (the Mazowieckie Voivodship evidently stands apart from the others).

The selected quartile groups could end the process of voivodship classification if their economic situation was assessed solely on the basis of their GDP *per capita*. Further in the article the authors will demonstrate that the same regularities are typical of individual voivodships if we do not presuppose the number of groups and if we take into consideration the impact of specific macroeconomic variables on the GDP *per capita* level. Since the authors applied models with varying intercept term, they simultaneously assumed that the impact of specific macroeconomic variables on the GDP *per capita* level is identical in each voivodship.

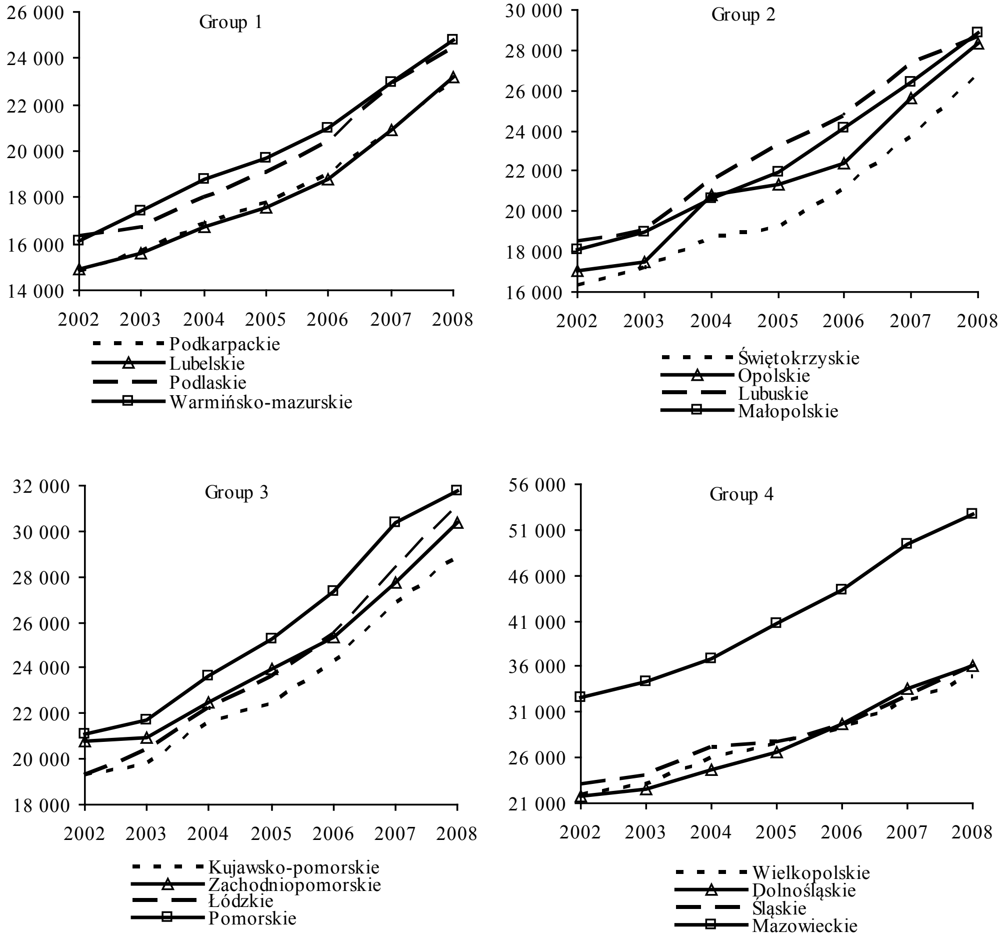

Fig. 1. Gross Domestic Product *per capita* (PKB1) in the Polish voivodships in 2002–2008 in the quartile groups

Source: the Local Data Bank of GUS.

3. Results of Estimation of the Applied Models

The authors suggested including in the set of explanatory variables 7 macroeconomic variables which (in line with the theory of economics) influence the level of the gross domestic product. According to the econometric modeling process, after we have selected 'candidate' variables and collected statistical data, the next step to specify variables for the model is the selection of explanatory variables which satisfy the definite formal and statistical criteria. This is why before estimating the ultimate model forms the authors calculated the coefficients of correlation among the proposed variables (see Table 1).

Since the coefficients of correlation among the variables in the period of 2002–2007 differed only slightly from those of 2002–2008, Table 1 contains just the latter ones.

Table 1. Coefficients of correlation among the variables
(correlation coefficients statistically significant at the level of 0.05 are marked in bold)
in the period of 2002–2008

Variables	PKB1	SB	UPR	UPP	UPU	WB	DB1	SD1
PKB1	1.00	**–0.52**	**–0.46**	0.18	**0.56**	**0.94**	**0.96**	**0.82**
SB		1.00	–0.14	0.09	0.14	**–0.67**	**–0.56**	**–0.35**
UPR			1.00	**–0.82**	**–0.88**	**–0.31**	**–0.46**	–0.18
UPP				1.00	**0.45**	0.03	**0.24**	–0.16
UPU					1.00	**0.46**	**0.52**	**0.41**
WB						1.00	**0.91**	**0.74**
DB1							1.00	**0.70**
SD1								1.00

Source: own calculations.

The gross domestic product *per capita* is strongly correlated with such variables as the average gross salary in the enterprise sector, the gross income *per capita* and the retail sales *per capita*. Moreover, these variables are strongly interrelated, therefore set of explanatory variables in the final version of the model includes the unemployment rate and the share of services workers to the workers' total[4]. Also, the gross domestic product *per capita* is significantly correlated with the share of agriculture workers, but due to the strong correlation of this share with the share of services workers the model includes only the latter variable[5].

The estimation of the model (3), where the zero-one (dummy) variable for the object *N* has been dropped, is presented in Tables 2 and 3 for the periods of 2002–2007 and 2002–2008, respectively where z_i denotes the zero-one variable for the object *i*. Thus, the parameter estimations corresponding to the z_i variables are the estimations of the parameters δ_i (i = 2, ..., N). The estimation of the common intercept term is equivalent to the estimation of the δ_1 parameter. The voivodships have been numbered alphabetically, so the missing parameter is the intercept term for the Zachodniopomorskie Voivodship.

Table 2. The estimation results for the model with varying parameters for the period of 2002–2007

R^2 = 0.97; F (17.78) = 137.38, p < 0,000; standard error of estimation: 1,218.5				
Variable	Parameter estimates	Standard errors of parameter estimates	t(94)	p level
Common intercept term	–107,110.13	17,791.96	–6.02	0.000
SB	–648.48	31.06	–20.88	0.000
UPU	22,9471.2	28,124.45	8.16	0.000
z_1	11,134.45	1,188.66	9.37	0.000
z_2	21,619.33	3,066.23	7.05	0.000
z_3	33,386.77	5,456.90	6.12	0.000
z_4	6,510.94	1,278.23	5.09	0.000
z_5	27,876.79	3,902.22	7.14	0.000
z_6	11,436.81	2,286.72	5.00	0.000
z_7	13,208.43	770.07	17.15	0.000
z_8	17,543.55	3,008.61	5.83	0.000
z_9	26,495.92	4,519.09	5.86	0.000
z_{10}	31,026.08	5,001.49	6.20	0.000
z_{11}	5,213.88	1,057.72	4.93	0.000
z_{12}	14,120.34	1,766.51	7.99	0.000
z_{13}	36,403.87	5,313.16	6.85	0.000
z_{14}	15,608.66	2,628.19	5.94	0.000
z_{15}	28,346.45	3,644.35	7.78	0.000

Source: own calculations.

Table 3. The estimation results for the model with varying parameters for the period of 2002–2008

R^2 = 0.9705; F (17.94) = 181.60, p < 0.000; estimation standard error: 1,268.3				
Variable	Parameter estimates	Standard errors of parameter estimates	t(94)	p level
1	2	3	4	5
Common intercept term	–99,666.42	14,623.39	–6.82	0.000
SB	–619.22	25.20	–24.57	0.000
UPU	216,344.25	22,921.84	9.44	0.000

1	2	3	4	5
z_1	10,947.22	1,031.67	10.61	0.000
z_2	20,358.46	2,529.33	8.05	0.000
z_3	31,148.47	4,459.85	6.98	0.000
z_4	6,349.12	1,153.92	5.50	0.000
z_5	26,407.59	3,211.30	8.22	0.000
z_6	10,778.89	1,911.15	5.64	0.000
z_7	13,999.83	714.56	19.59	0.000
z_8	16,521.19	2,491.18	6.64	0.000
z_9	24,520.48	3,698.30	6.63	0.000
z_{10}	28,882.91	4,105.87	7.03	0.000
z_{11}	5,046.82	957.83	5.27	0.000
z_{12}	13,989.65	1,514.00	9.24	0.000
z_{13}	34,102.24	4,333.04	7.87	0.000
z_{14}	14,183.48	2,178.80	6.51	0.000
z_{15}	27,061.74	2,999.99	9.02	0.000

Source: own calculations.

The model estimation results presented in the Tables 2 and 3 match very well the empirical data – the value of the determination coefficient is close to 1 and the F statistics is very high. In the both periods of 2002–2007 and 2002–2008 all the estimates of the β_k and δ_i parameters were statistically significant as well. It proves that it was justified to construct a model with a varying intercept term because its level in individual voivodships clearly differentiates their GDP *per capita*, providing the identical impact of the unemployment rate and of the share of residents employed in the service sector. It is worth noting that the parameter estimates for the selected macroeconomic variables have signs consistent with the theory of economics, i.e. the growing unemployment rate lowers the GDP *per capita*, while the rise in the share of people employed in the service sector results in the growth of the GDP *per capita* value.

4. Results of the Voivodship Classification

Since the authors have estimated models of the form (3), it could be observed statistical significance of differences among the intercept terms attributed to individual voivodships in relation to the common intercept term, which in turn made the authors examine the question if there were any significant differences among the intercept terms attributed to individual voivodships in relation to one another. Therefore, the model (3) was estimated 15 times for each of the two periods (2002–2007 and 2002–2008), every time a different one-zero variable (one for each estimation) was dropped. In the new models the estimations of δ_i parameters are different, but the β_{1i} (i = 1, ..., N) parameters' values determined by means of formulas (4) are the same. The values of the β_{1i} parameters are to be found in Table 4.

Table 4. Values of the β_{1i} parameters estimated for the period of 2002–2007 and 2002–2008

Voivodships	2002–2007	Voivodships	2002–2008
zachodniopomorskie	–107,110.00	zachodniopomorskie	–99,666.40
pomorskie	–101,896.12	pomorskie	–94,619.58
lubuskie	–100,599.06	lubuskie	–93,317.28
dolnośląskie	–95,975.55	małopolskie	–88,887.51
małopolskie	–95,673.19	dolnośląskie	–88,719.18
mazowieckie	–93,901.57	śląskie	–85,676.75
śląskie	–92,989.66	mazowieckie	–85,666.57
warmińsko-mazurskie	–91,501.34	warmińsko-mazurskie	–85,482.92
opolskie	–89,566.45	opolskie	–83,145.21
kujawsko-pomorskie	–85,490.67	kujawsko-pomorskie	–79,307.94
podkarpackie	–80,614.08	podkarpackie	–75,145.92
łódzkie	–79,233.21	łódzkie	–73,258.81
wielkopolskie	–78,763.55	wielkopolskie	–72,604.66
podlaskie	–76,083.92	podlaskie	–70,783.49
lubelskie	–73,723.23	lubelskie	–68,517.93
świętokrzyskie	–70,706.13	świętokrzyskie	–65,564.16

Source: own calculations based on the results of the estimation in Tables 2 and 3.

The intercept terms β_{1i} determined on the basis of the 2002–2008 period are higher than the 2002–2007 ones. Note, however, that the order of the voivodships listed according to the intercept term value has changed just a little, namely two pairs of voivodships neighbouring on the list changed positions: Małopolskie replaced Dolnośląskie and Śląskie replaced Mazowieckie.

The β_{1i} value makes it possible to put the voivodships in order, but it is rather difficult to distinguish groups of them (to classify them) – you cannot explicitly indicate the differences between individual values of β_{1i}. This is why in order to classify the voivodships the authors used the approach where subsequent 15[6] models were estimated (in line with the formula 3) and they checked which of the δ_i parameters were not statistically significant. The lack of statistical significance of the δ_i parameter proves that there is no difference between the intercept term for the ith voivodship and the intercept term for the voivodship, for which the zero-one variable has been dropped in the currently estimated model. In this way the authors distinguished groups of voivodships that were similar as far as the level of the β_{1i} intercept term was concerned. The groups are presented in Table 5, in the order from the least to the biggest value of the β_{1i} intercept term.

Table 5. Groups of voivodships distinguished on the basis of the models (3) estimated for the periods of 2002–2007 and 2002–2008

Groups of voivodships – 2002–2007	Groups of voivodships – 2002–2008
zachodniopomorskie	zachodniopomorskie
lubuskie, pomorskie	lubuskie, pomorskie
dolnośląskie, małopolskie, mazowieckie	dolnośląskie, małopolskie, mazowieckie
opolskie, śląskie, warmińsko-mazurskie	opolskie, śląskie, warmińsko-mazurskie
kujawsko-pomorskie	kujawsko-pomorskie
łódzkie, podkarpackie, podlaskie, wielkopolskie	łódzkie, podlaskie, wielkopolskie
lubelskie	podkarpackie
świętokrzyskie	lubelskie
	świętokrzyskie

Source: own calculations.

In relation to the 2002–2007 period only one change took place, namely an additional group appeared which consisted of just one Podkarpackie Voivodship. When we take into consideration the data from the period of 2002–2008, we can say that Świętokrzyskie, Lubelskie, Podkarpackie, Łódzkie, Podlaskie and Wielkopolskie are the voivodships that, being on the same level of response variables identical for all the voivodships, would reach the highest GDP *per capita*.

Conclusions

The conducted study shows that in case of panel data the models with an intercept term varying in a non-random manner over the objects can be a useful tool for classifying objects. Thanks to these models it is possible to classify objects with reference to both the level of a dependent variable as well as the influence of other economic variables on this particular one. Distinguished groups of objects are similar regarding the level of the β_{1ii} intercept term and, additionally, the higher the value of this parameter, the better operational efficiency of objects belonging to a given group. In this case better efficiency is understood as a higher level of a dependent variable when the level of explanatory variables is the same. Such interpretation refers to the dependent variable being a stimulant. In case of a destimulant the better efficiency would mean a lower level of the dependent variable with the level of explanatory variables being the same.

The article demonstrates that the voivodships with the real GDP *per capita* level from the first and the second quartile groups (excluding the Wielkopolskie Voivodship belonging to the fourth quartile group) would achieve higher GDP *per capita* than other voivodships, provided their unemployment rates and the share of the services workers in all the voivodships were the same.

Notes

1 See Pociecha et al. (1988), Siedlecki (2000), Malina (2004), Rozkrut (2006).

2 Compare the results presented in Batóg, Mojsiewicz, Wawrzyniak (2010). The Authors applied models with varying intercept to classify the households on the insurance market. The intercept varied by objects in the regard to the education level and place of residence.

3 Besides the names of variables the authors placed the abbreviations which are used in the paper.

4 The examination of the dependency between GDP per capita and the shares of workers in economy sectors to the workers' total in European regions was conducted by Strahl, Markowska (2006).

5 The model with the following explanatory variables: the unemployment rate (SB), the share of agriculture workers to the workers' total (UPR) and the share of services workers to the workers' total (UPU) was also estimated, but the parameter for UPR was not statistically significant.

6 The model for Zachodniopomorskie Voivodship is presented in Table 2 and Table 3.

References

Batóg, B., Mojsiewicz, M., Wawrzyniak, K. (2010). *Klasyfikacja gospodarstw domowych pod względem popytu potencjalnego i zrealizowanego na rynku ubezpieczeń w Polsce,* Prace Naukowe UE we Wrocławiu nr 107, Wrocław: Wydawnictwo UE we Wrocławiu, 153–160.

Johnston, J. (1991). *Econometric Methods*, McGraw-Hill Book Company.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., Lee, T-C. (1985). *Theory and Practice of Econometrics*, John Wiley and Sons.

Maddala, G.S. (2006). *Ekonometria*, Warszawa: Wydawnictwo Naukowe PWN.

Malina, A. (2004). *Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw*, Kraków: Wydawnictwo Akademii Ekonomicznej w Krakowie.

Pociecha, J., Podolec, B., Sokołowski, A., Zając, K. (1988). *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, Warszawa: PWN.

Rozkrut, M. (2006). *Badanie poziomu rozwoju gospodarczego powiatów Polski w kontekście podziału na jednostki terytorialne NUTS 1*. Prace Naukowe AE we Wrocławiu nr 1126, Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 343–352.

Siedlecki, J. (2000). *Równowaga a wzrost gospodarczy*, Warszawa: Wydawnictwo Naukowe PWN.

Strahl, D., Markowska, M. (2006). *Sektorowa struktura pracujących w UE jako podstawa klasyfikacji przestrzeni regionalnej*. Prace Naukowe AE we Wrocławiu nr 1126, Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 308–316.