QUANTITATIVE METHODS IN ECONOMICS
Vol. XIV, No. 2, 2013, pp. 200 – 209

# MEASURE OF THE LEVEL OF SOCIO–ECONOMIC DEVELOPMENT IN PROVINCES

**Aneta Sobiechowska–Ziegert, Aniela Mikulska**
Department of Economic Sciences, Gdansk University of Technology
e-mail: Aneta.Sobiechowska@zie.pg.gda.pl, Aniela.Mikulska@zie.pg.gda.pl

**Abstract:** The scope of Polish macro-economic data for assessing the level of socio-economic development of the country is largely limited because of the regional variation. Therefore there is a need for cyclic selection and the verification of criteria which allow identifying regions with a similar level of socio-economic development or those that clearly differ from the mean values. The aim of the study is to compare Poland's provinces in terms of their socio-economic development, which helps identify the most similar regions as far as the adopted criteria are concerned.

**Keywords**: socio-economic development, synthetic variables, classification of regions

## INTRODUCTION

Regional differences restrict the use of economic data to assess the level of the social-economic development of selected locations in Poland. Information about diversity can be used to analyse various topics related to social activity including the differences in the intensity of entrepreneurship in the region. The assessment of entrepreneurial activity range can be made through a variety of variables such as structure of employment, labour productivity, level of urbanization and access to infrastructure (including educational) [Strużycki 2004]. The level of socio-economic development of regions determines changes in the SME sector, because changes in the economic system and taking place in the area of efficiency and effectiveness cause the extension of production capacity, and thus they lead to the improvement of the living conditions [Strużycki 2004], which in turn result in the intensification of local entrepreneurial activities. There are many opportunities to acquire and configure variables used to assess the level of socio-economic development. Moreover, there is a real need for cyclic selection,

analysing and verifying the variables (and their groups) which allow for the indication of regions with a similar level of socio-economic development, or those that clearly deviate from the mean. The selection of variables relating to the socio-economic situation seems to be a relatively simple task, like determining the impact direction of a factor on the level of development. However, not only choice but also variable verification by means of statistical measure (e.g. variation or asymmetry coefficient) classified at the appropriate level is important. The issues of availability, integrity and comparability of data to be analysed constitute a major difficulty for researchers . This significantly determined the selection of variables taken into consideration in this study. The ultimate source of information was Local Data Bank of Central Statistical Office (CSO).

The aim of the paper is to identify the most similar regions in terms of the variables chosen for the study. A comparative analysis of provinces in terms of their socio-economic development is provided. A desire to collect information about Polish provinces which could then be used to compare the conditions necessary for their development at the enterprise level inspired the research. Studies in which entrepreneurs usually participate should be supported by an objective assessment of the conditions in the environment in which they take decisions - often of a strategic nature [Bratnicki 2011]. Many publications point to the fact that the period of the transformation of Polish economy has resulted in growing diversity in the level of the socio-economic development of provinces, which has deepened the division of the country into the eastern and western parts [eg Strużycki 2004, Bizon 2011].

## VARIABLES SELECTED FOR THE RESEARCH

The regional variation often refers to the basic indicator, which is GDP per capita. In the study presented here this indicator is not used, mainly because of the lack of available data for the year 2010, the period in which the variables were collected, but also because of the fact that the relative value of production does not translate directly and automatically into the standard of living in local communities in provinces [Bizon 2011]. Therefore, the study examines the measures that clearly determine the level of socio-economic development. The research period covers the years 2008-2010 due to the availability of CSO data and the opportunity to make a comparison with available studies.

The variables that meet the criteria of completeness, availability and comparability in the analysed period were pre-qualified for further study and subjected to further selection according to statistical criteria. Table 1 presents the variables describing the level of socio-economic development which were adopted at the initial stage of the analysis. For the sake of further study, only the variables that meet the criteria used in this kind of research were analysed [Zeliaś 2000]. These criteria are:

- sufficient variability measured by variation coefficient with threshold value ε=0,1
- positive asymmetry (in the case of stimulants) or distribution close to symmetric (the skewness index close to zero).

Table 1. Variables subjected to initial verification

| Variable | Description |
|---|---|
| $X_1$ | Average, monthly gross salary in PLN |
| $X_2$ | Average, monthly disposable income per person in PLN |
| $X_3$ | Average, monthly disposable income per person in employment in PLN |
| $X_4$ | Average, monthly disposable income per person in self-employment in PLN |
| $X_5$ | Registered unemployment rate in % |
| $X_6$ | Risk of poverty rate (relative poverty line in %) |
| $X_7$ | Number of physicians per 10000 inhabitants |
| $X_8$ | Share of expenditure on food and non-alcoholic beverages in the total expenditure in % |
| $X_9$ | Share of working age population as a % of total population |
| $X_{10}$ | Average useable floor space in $m^2$ |
| $X_{11}$ | Registered passenger cars per 1000 population |
| $X_{12}$ | Structure of employment by sector (agriculture, forestry) |
| $X_{13}$ | Structure of employment by sector (manufacturing) |
| $X_{14}$ | Structure of employment by sector ( services) |
| $X_{15}$ | Capital expenditures in the private sector per capita in PLN |
| $X_{16}$ | Capital expenditures per capita in PLN |
| $X_{17}$ | Expenditures on R & D per one entity in the business sector in PLN |
| $X_{18}$ | Expenditures on R & D per one inhabitant in PLN |
| $X_{19}$ | Hard surface roads in km per 100 $km^2$ |
| $X_{20}$ | Entities entered in the REGON register per 10 thousand population |
| $X_{21}$ | Share of the SME sector in sold industrial output total in % |
| $X_{22}$ | Participation of business in expenditures on R & D in % |

Source: Local Data Bank of CSO in Poland

The initial selection of variables allowed for the identification of the following factors relevant to the assessment of differences in the social dimension:

- average, monthly disposable income per person in PLN ($X_2$) [1],
- registered unemployment rate ($X_5$),
- risk of poverty rate - relative poverty line ($X_6$),
- number of physicians per 10000 inhabitants ($X_7$).

[1] This indicator reflects the purchasing power of households better than the gross salary indicator, and thus allows determining more accurately the perceived level of prosperity, which codetermines the level of socio-economic development.

Taking into consideration the variables concerning income $(X_1 \cdots X_4)$, the amount of average monthly disposable income per one inhabitant seemed the most representative. An attempt was made during the study to determine to what extent the source of income (wage labour, self-employment) was a variable that differentiated the data in the region. This variable, however, was highly correlated with average monthly disposable income per one inhabitant.

Variables approximately describing the living conditions were extended in the initial phase of the study to include other aspects: average useable floor space in $m^2$, registered passenger cars per 1000 population, share of expenditure on food and non-alcoholic beverages in the total expenditure and share of working age population as a percentage of total population. However, these were eliminated in the next stage of the analysis due to the level of variation coefficient ($\varepsilon \leq 0{,}1$) being lower than it was first assumed. Because of the same reason, the access to the Internet variable was earlier eliminated. Employment structure by sector $(X_{12} \cdots X_{14})$ was not regarded as important because of the very low variation coefficient for the service sector and negative skewness index for the industrial sector.

The criteria that also enabled the researchers to define the level of socio-economic development of the regions were as follows:

- capital expenditures per capita $(X_{16})$,
- expenditures on R & D per one inhabitant $(X_{18})$,
- hard surface roads in km per 100 $km^2$ $(X_{19})$,
- entities entered in the REGON register per 10 thousand population $(X_{20})$,
- participation of business in expenditures on R & D $(X_{22})$.

These indicators correspond to a large extend to the level of infrastructure that determines the success of business ventures. These projects, in turn, are the source and driving force behind socio-economic development in the regions. The rejected variables in this group were as follows:

- capital expenditures in the private sector per capita in PLN $(X_{15})$ due to the fact, that they do not take into account the expenditures of public funds,
- share of the SME sector in sold industrial output $(X_{21})$, due to the negative and diverging from zero asymmetry rate,
- expenditures on R & D per one entity in the business sector $(X_{17})$, due to the correlation with the expenditures on R & D per one inhabitant variable $(X_{18})$.

Nine diagnostic variables describing the socio-economic situation of the regions were finally selected. These variables are objective, measurable and represent the most important areas of socio-economic development.

## RESEARCH METHODOLOGY AND SYNTHETIC VARIABLES DETERMINATION

Variables selected for the study were divided into two groups: social variables and economic variables. They were then subjected to information capacity analysis using Hellwig's method. To do this, correlation coefficients between the variables for each year were determined and then all variables within a group were divided into subgroups containing central and isolated variables, by comparing the correlation coefficients with the assumed threshold of 0,5. This division is presented in Table 2.

Table 2. Division of variables into central and isolated ones

| Year | Social variables | | Economic variables | |
|---|---|---|---|---|
| | Central | Isolated | Central | Isolated |
| 2008 | $X_5$ | $X_2$, $X_7$ | $X_{16}$ | $X_{19}$, $X_{22}$ |
| 2009 | $X_2$ | $X_5$, | $X_{16}$ | $X_{19}$, $X_{22}$ |
| 2010 | $X_2$ | $X_5$, $X_7$ | $X_{16}$ | $X_{19}$, $X_{22}$ |

Source: own calculations

The final decision on the selection of variables for further analysis was based on the incidence of variables in each subgroup (central or isolated) and fulfilment of all the statistical criteria used, the positive asymmetry in particular. Therefore, for further analysis the following variables were adopted:

- average monthly disposable income per person ($X_2$),
- registered unemployment rate ($X_5$)[2],
- capital expenditures per capita ($X_{16}$),
- hard surface roads in km per 100 km$^2$ ($X_{19}$),
- participation of business in expenditures on R & D in ($X_{22}$)

To examine the level of socio-economic development of provinces and to achieve the targets, methods that allow finding similar regions in terms of the level of development (and the selected variables) were used, in particular synthetic variable method. In order to determine a synthetic variable, diagnostic variables were first divided into stimulants and destimulants. It was assumed that among the selected variables there was only one destimulant: registered unemployment rate ($X_5$). The other variables were considered as stimulants. Variable $X_5$ was therefore transformed into a stimulant using for this purpose the weighted average rate of unemployment for Poland according to formula (1) [Zeliaś 2000]:

$$x_{ij}^S = 2\overline{x}_j - x_{ij}^D \quad (1)$$

[2] In the group of variables characterizing social development, it was decided to remove $X_7$ variable because of the growing negative asymmetry rate.

Next, the synthetic variable was calculated as the average of standardized diagnostic variables according to two standardisation variants – alternating and optimal pattern. In the first variant, the average value of diagnostic variable in a given year period was used as a reference point, in the second one - the optimal value in a given year, which is the maximum value in the case of stimulants. In order to compare the results, the synthetic variables were transformed into taxonomic measures according to formula (2):

$$z_i' = \frac{z_i}{\max\limits_i z_i} \tag{2}$$

The formula allows obtaining values in the range $\langle 0,1 \rangle$ . The provinces for which taxonomic measures are close to one, will have a better level of socio-economic development in terms of business development. The synthetic variables as well as taxonomic measures are presented in table 3. Next, the provinces were sorted out in terms of socio-economic development. The selected and verified criteria for assessing the level of socio-economic development of regions resulted in achieving groups of regions characterized by a similar level of development. In comparison with individual indicators they better showed regional differences as far as entrepreneurship, innovation, the development of knowledge-based economy, development in the cultural dimension or the purely social one are concerned.

Table 3. Synthetic variables and taxonomic measures according to the adopted standardisation options

| Province | 2008 | | | | 2009 | | | | 2010 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Variant1 | | Variant2 | | Variant1 | | Variant2 | | Variant1 | | Variant2 | |
| | z | z' | z | z' | z | z' | z | z' | z | z' | z | z' |
| ŁÓDŹ | 0,995 | 0,755 | 0,620 | 0,782 | 0,977 | 0,670 | 0,629 | 0,706 | 1,015 | 0,797 | 0,647 | 0,810 |
| MAZOVIA | 1,242 | 0,942 | 0,775 | 0,976 | 1,253 | 0,859 | 0,807 | 0,906 | 1,237 | 0,971 | 0,799 | 1,000 |
| MAŁOPOLSKA | 1,182 | 0,897 | 0,719 | 0,906 | 1,147 | 0,786 | 0,720 | 0,808 | 1,113 | 0,874 | 0,700 | 0,877 |
| SILESIA | 1,318 | 1,000 | 0,794 | 1,000 | 1,460 | 1,000 | 0,891 | 1,000 | 1,274 | 1,000 | 0,790 | 0,988 |
| LUBLIN | 0,744 | 0,564 | 0,475 | 0,598 | 0,757 | 0,518 | 0,497 | 0,558 | 0,808 | 0,634 | 0,523 | 0,655 |

| Province | 2008 | | | | 2009 | | | | 2010 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Variant1 | | Variant2 | | Variant1 | | Variant2 | | Variant1 | | Variant2 | |
| | z | z' | z | z' | z | z' | z | z' | z | z' | z | z' |
| PODKARPACIE | 1,078 | 0,818 | 0,619 | 0,780 | 0,995 | 0,682 | 0,608 | 0,682 | 1,107 | 0,869 | 0,665 | 0,832 |
| PODLASIE | 0,946 | 0,718 | 0,578 | 0,729 | 0,863 | 0,591 | 0,557 | 0,625 | 0,814 | 0,639 | 0,529 | 0,662 |
| ŚWIĘTOKRZYSKIE | 0,891 | 0,676 | 0,540 | 0,680 | 0,975 | 0,668 | 0,608 | 0,682 | 0,951 | 0,746 | 0,599 | 0,749 |
| LUBUSKIE | 1,012 | 0,768 | 0,605 | 0,762 | 0,892 | 0,611 | 0,566 | 0,635 | 0,989 | 0,776 | 0,630 | 0,788 |
| WIELKOPOLSKA | 1,084 | 0,822 | 0,675 | 0,850 | 1,004 | 0,688 | 0,652 | 0,731 | 1,020 | 0,801 | 0,659 | 0,825 |
| WEST POMERANIA | 0,739 | 0,561 | 0,477 | 0,602 | 0,788 | 0,540 | 0,513 | 0,576 | 0,814 | 0,639 | 0,522 | 0,653 |
| LOWER SILESIA | 1,115 | 0,846 | 0,680 | 0,857 | 1,073 | 0,735 | 0,682 | 0,765 | 1,142 | 0,896 | 0,713 | 0,892 |
| OPOLE | 0,950 | 0,721 | 0,592 | 0,746 | 1,107 | 0,758 | 0,688 | 0,772 | 0,939 | 0,737 | 0,596 | 0,746 |
| KUJAWY-POMERANIA | 0,944 | 0,717 | 0,568 | 0,716 | 0,847 | 0,580 | 0,543 | 0,609 | 0,898 | 0,705 | 0,568 | 0,711 |
| POMERANIA | 1,158 | 0,879 | 0,703 | 0,886 | 1,229 | 0,842 | 0,769 | 0,863 | 1,194 | 0,938 | 0,737 | 0,923 |
| WARMIA-MASURIA | 0,602 | 0,457 | 0,388 | 0,489 | 0,632 | 0,433 | 0,414 | 0,465 | 0,686 | 0,539 | 0,444 | 0,556 |

Source: own calculations

## MEASURMENT OF SOCIO-ECONOMIC DEVELOPMENT – RESULTS

After arranging the regions using taxonomic measures $z_i'$, it was examined by means of Spearman's rank correlation coefficient whether the arrangement depended on the method of standardization of diagnostic variables. It turned out

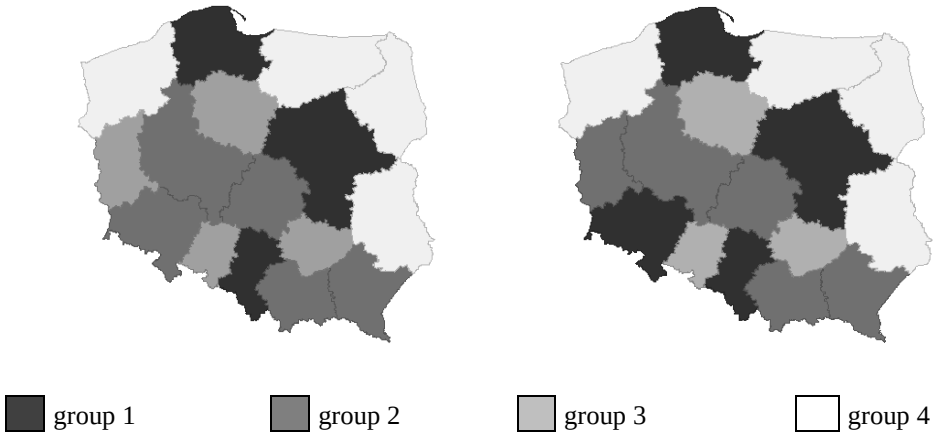
that for each analysed year the correlation coefficients were statistically significant and equal: 0,988; 0,997; 0,994 respectively, which suggests that the province arrangement according to the level of socio-economic development can be considered compatible in both adopted versions. Due to the fact, that the adopted version of standardization of diagnostic variables did not affect significantly the results of province ranking, the optimal variant (variant 2) was chosen and two methods - the standard deviation method and the division of the variation range into four pre-determined classes – were used to classify regions in 2010. The visualization of both classifications is shown in Figure 1. In the first method, the provinces were divided into four groups, including those regions for which synthetic indicator $z_i$ obtained values in the following range ($s_z$ – standard deviation):

Group 1: $\langle \bar{z} + s_z; \max_i z_i \rangle$
Group 2: $\langle \bar{z}; \bar{z} + s_z)$
Group 3: $\langle \bar{z} - s_z; \bar{z})$
Group 4: $\langle \min_i z_i; \bar{z} - s_z)$

According to this method the provinces were classified as follows:

- Group 1: Mazovia, Pomerania, Silesia,
- Group 2: Lower Silesia, Łódź, Małopolska, Podkarpacie, Wielkopolska,
- Group 3: Kujawy-Pomerania, Lubuskie, Opole, Świętokrzyskie,
- Group 4: Lublin, Podlasie, Warmia-Masuria, West Pomerania.

Figure 1. Classification of provinces in 2010 according to standard deviation method (the map on the left) and division of the variation range (the map on the right)




Source: own calculations

The classification results using the division of the variation range into the four classes were as follows:

- Group 1: Lower Silesia, Mazowia, Pomerania, Silesia,
- Group 2: Lubuskie, Łódź, Małopolska, Podkarpacie, Wielkopolska,
- Group 3: Kujawy-Pomerania, Opole, Świętokrzyskie,
- Group 4: Lublin, Podlasie, Warmia-Masuria, West Pomerania.

## SUMMARY

When a set of diagnostic variables was determined, five variables from two groups were selected and the synthetic variables were used for the analysis of socio-economic development in provinces. Regardless of the standardization option, the synthetic variables allowed ranking the provinces according to the adopted criteria, from the best to the least developed ones. During the whole examined period the most developed regions were Mazovia Province and Silesia Province, and the least: Lublin Province and Warmia–Masuria Province.

Using the synthetic variables, regions with the same level of socio-economic development were grouped. The clustering methods led to very similar results. In the most developed regions group were: Mazovia Province, Pomerania Province and Silesia Province, and in the least developed regions group were: Lublin Province, Podlasie Province, Warmia-Masuria Province and West Pomerania Province. The difference in grouping can be seen when it comes to Lower Silesia Province, which, depending on the grouping belongs either to the first or the second group, and Lubuskie Province, which is either in the second or in the third group. There were no differences in the grouping of the other provinces.

The possibilities of using the analysed set of variables, which allowed assessing the level of socio-economic development in regions, are very broad. The conclusions of the study and the classification of regions can be used for further research. In the authors' opinion the selection of regions with a similar level of socio-economic development can be used to do research on entrepreneurship, its scale in relation to the generally understood conditions of life and economy. The described results can be the basis for analyzing the impact of different variables on economic conditions, conditions for conducting entrepreneurial activities and the intensity of the innovation process.

## REFERENCES

Bizon W. (2011) Dobrobyt społeczno-ekonomiczny oraz gospodarka oparta na wiedzy w kontekście historycznych podziałów na Polskę A i B, E-mentor nr 4 (41).

Bratnicki M. (2011) Sprawdzanie teorii przedsiębiorczości, Przedsiębiorstwo przyszłości, Wyższej Szkoły Zarządzania i Prawa im. Heleny Chodkowskiej, nr 3 (8) lipiec, s. 9-18.

Dąbrowa M. (2011) Badanie poziomu życia – metodologia konstrukcji wybranych wskaźników, Zeszyty Naukowe MWSE w Tarnowie, Matematyka nr 1(17), s. 67-82.

Łapczyński M. (2005) Wpływ aktywności mieszkańców na poziom życia w gminach woj. małopolskiego, www.statsoft.pl/czytelnia/artykuly/Wplyw_aktywnosci.pdf

Malina A. (2004) Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski wg województw, Wydawnictwo AE w Krakowie.

Strużycki M. red. (2004) Małe i średnie przedsiębiorstwa w gospodarce regionu, PWE, Warszawa.

Winiarczyk–Raźniak A., Raźniak P. (2011) Regional differences in the standard of living in Poland (based on selected indices), Procedia Social and Behavioral Sciences 19, s.31-36.

Zeliaś A. (red.) (2000) Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym, Wydawnictwo AE w Krakowie, Kraków.