European Research Studies Journal
Volume XXIV, Issue 2, 2021

pp. 889-914

Male Activity Rates vs Education Levels: An Analysis Across Selected European Union Countries

Submitted 02/03/21, 1st revision 12/04/21, 2nd , accepted

Mariusz Malinowski[1], Lidia Jabłońska-Porzuczek[2]

Abstract:

Purpose: *The purpose of this paper is to present the evolution of male employment rates between 2013–2017 in selected Central and Eastern European countries (Poland, Czech Republic, Slovakia, Slovenia, Lithuania, Latvia, Estonia, Romania, Bulgaria, Croatia, and Hungary), and to analyze the interregional relationships between male activity and education levels.*
Methodology: *The TOPSIS method was used to build synthetic indicators of male activity and education. An analysis of spatial autocorrelation was carried out to determine the strength of spatial relationships between different regions in terms of male activity and education levels. Also, the dependence between synthetic indicators of male activity and education levels was empirically analyzed based on spatial regression.*
Findings: *The results of spatial regression analysis provide grounds for concluding that a 1% increase in the value of the synthetic indicator of education results, ceteris paribus, in a nearly 0,48% increase in the synthetic indicator of male activity levels in each region.*
Practical Implications: *The results of the conducted research may be indirectly used by the central and local authorities responsible for local and regional development in the context of the choice of the direction for the socio-economic restructuring of countries and local government units.*
Originality/Value: *Rarely used spatial regression together with taxonomic methods may constitute a useful tool for identification of relations between multidimensional categories.*

Keywords: *Education, linear ordering, male activity, spatial relationships.*

JEL code: *C21, J21.*

Paper type: *Research study.*

[1]*Department Economics, Faculty of Economics, University of Life Sciences, Poznań, Poland,* mariusz.malinowski@up.poznan.pl
[2]*Department Economics, Faculty of Economics, University of Life Sciences, Poznań, Poland,* lidia.jablonska-porzuczek@up.poznan.pl

1. Introduction

Many European Union countries face population ageing resulting from a decline in fertility rates and extended life expectancy. The demographic shift has had an adverse effect on the economy, especially on the labor market. The progressing demographic slowdown contributes to a decline in labor supply. The reduction in labor resources is also impacted by a decades-long policy of supporting people at near-retirement age in exiting the labor market prematurely, and by social stereotypes on the employment of certain groups (including women, the disabled and the elderly). The economic, demographic, and social situation has contributed to the growing interest in the issue of economic activity. In the recent years, European Union authorities have taken measures to increase employment to counteract the adverse developments affecting the labor market. Therefore, EU countries are taking steps under their labor market policies to raise the levels of economic activity.

Research and analyses of employment levels and determinants need to be carried out to verify the solutions in place. Indeed, the labor market is related to the demographic structure of the society and to various fields of socioeconomic activity which influence both the supply of and demand for labor. The purpose of this paper is to present the evolution of male employment rates between 2013-2017 in selected Central and Eastern European countries (namely Poland, Czech Republic, Slovakia, Slovenia, Lithuania, Latvia, Estonia, Romania, Bulgaria, Croatia, and Hungary), and to analyse the relationships between male activity and education at regional level (based on 2016 statistical data).

Multidimensional statistical analysis methods were used to quantify the activity and determine the education level of men in selected EU regions. This is because it is necessary to compare many research objects described with a broad set of variables in such an analysis. Therefore, the development level of the phenomenon concerned is difficult to express with one measurable feature. This study focuses on regions in selected EU countries who joined the Community in 2004 (Czech Republic, Estonia, Lithuania, Latvia, Poland, Slovakia, Slovenia, and Hungary) or later (Bulgaria and Romania in 2007 and Croatia in 2013) and who share a border. Therefore, Cyprus and Malta are excluded from the analysis.

Empirical data for the countries selected was analysed with measures of descriptive statistics, demographic measures (including the fertility rate, the demographic dependency index and life expectancy), and indices of the condition of the labor market (including the employment rate). An analysis of spatial autocorrelation was carried out to determine the strength of spatial relationships between different regions in terms of male activity and education levels. Also, the dependence between synthetic indicators of male activity and education levels was empirically analysed based on spatial regression. This analysis relies on Eurostat and OECD data.

2. The Demographic Situation in Selected Countries

Ageing societies are among the major economic and social challenges faced by European Union countries. The process is caused by extended life expectancy and declining birth rates. In the countries surveyed, the average life expectancy for a male born between 2013-2017 was 73 years, and the average fertility rate was 1,53. Data analysis shows that, as regards the countries surveyed, a male born in 2013 or 2017 has the longest life expectancy in Slovenia (77,2 or 78,2) and the shortest in Latvia (69,3 or 69,8) and Lithuania (68,5 and 70,7) (Table 1).

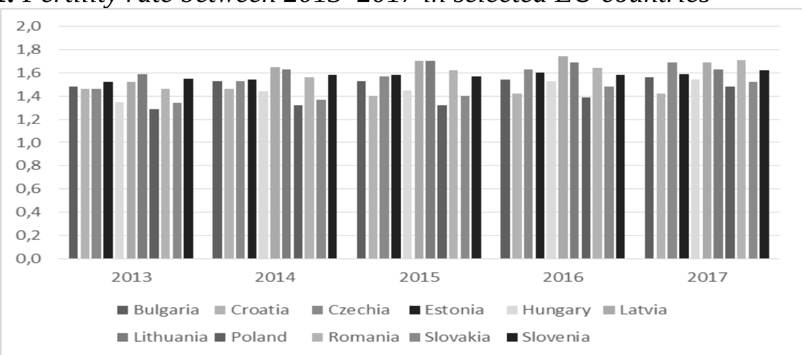
Table 1. *Life expectancy and healthy life years for a male between 2013–2017.*

	Life expectancy					Healthy life years (HLY)				
	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017
Bulgaria	71,3	71,1	71,2	71,3	71,4	62,4	62,0	61,5	64,0	62,9
Croatia	74,5	74,7	74,4	75,0	74,9	57,6	58,6	55,3	57,1	:
Czech Republic	75,2	75,8	75,7	76,1	76,1	62,5	63,4	62,4	62,7	60,6
Estonia	72,8	72,4	73,2	73,3	73,8	53,9	53,2	53,8	54,4	54,7
Hungary	72,2	72,3	72,3	72,6	72,5	59,1	58,9	58,2	59,5	59,6
Latvia	69,3	69,1	69,7	69,8	69,8	51,7	51,5	51,8	52,3	50,6
Lithuania	68,5	69,2	69,2	69,5	70,7	56,8	57,6	54,1	56,2	:
Poland	73,0	73,7	73,5	73,9	73,9	59,2	59,8	60,1	61,3	60,6
Romania	71,6	71,3	71,4	71,7	71,7	58,6	59,0	59,0	59,8	59,2
Slovakia	72,9	73,3	73,1	73,8	73,8	54,5	55,5	54,8	56,4	55,6
Slovenia	77,2	78,2	77,8	78,2	78,2	57,6	57,8	58,5	58,7	55,3

Source: *Own elaboration based on Eurostat data.*

The longest healthy life years were recorded for individuals born in Bulgaria (62,4 and 62,9 years) whereas the shortest were found in Latvia (51,7 and 50,6). Another finding is that some countries (including Latvia, Croatia, and Slovenia) have been witnessing a decline in healthy life years. The fertility rate went up between 2013–2017 in all countries surveyed except for Croatia where a minor reduction (by 0,04) was experienced (Figure 1).

Figure 1. *Fertility rate between 2013–2017 in selected EU countries*



Source: *Own elaboration based on Eurostat data.*

The highest fertility rate (1,71) and the sharpest increase (by 0,25) could be observed in Romania in 2017. Fertility depends on many factors, including socioeconomic aspects, support for families, health promotion policies, culture, customs, and reproductive awareness being of particular importance. In turn, average life expectancy tends to increase due to improvements in living conditions, greater care for health in the society and an effective functioning of healthcare services.

The demographic slowdown has an adverse effect on the old-age dependency ratio. The average ratio calculated for European Union countries grew from 27,5 to 29,9 between 2013-2017. In the group of countries considered, the highest levels were recorded in Bulgaria (28,5-31,8) and Latvia (28,1-30,8) (Table 2). Conversely, the lowest proportion of persons aged over 65 per 100 working-age population was recorded in Slovakia (18,4-21,5) and Poland (20,4-24,2).

Table 2. *Old-age dependency ratio between 2013–2017.*

	2013	2014	2015	2016	2017
Bulgaria	28,5	29,3	30,2	31,1	31,8
Croatia	27,1	27,6	28,3	29,0	29,8
Czech Republic	24,6	25,7	26,6	27,6	28,6
Estonia	27,2	27,9	28,7	29,3	30,0
Hungary	25,1	25,8	26,5	27,2	27,9
Latvia	28,1	28,8	29,5	30,2	30,8
Lithuania	27,2	27,5	28,1	28,6	29,3
Poland	20,4	21,2	22,2	23,1	24,2
Romania	23,9	24,3	25,2	25,9	26,7
Slovakia	18,4	19,0	19,7	20,6	21,5
Slovenia	25,0	25,7	26,6	27,6	28,6

Source: *Own elaboration based on Eurostat data.*

The demographic shift also influences the proportion of elderly people in the labor force[3]. In 2013-2017, the highest ratio for men was recorded in Croatia (0,310-0,319) and Slovenia (0,307-0,323) (Table 3). In turn, the lowest share of men aged 50-64 in the male labor force was found in Slovakia (0,271-0,279).

Table 3. *Proportion of elderly people in the male labor force between 2013–2017.*

	2013	2014	2015	2016	2017
Bulgaria	0,301	0,302	0,301	0,300	0,301
Croatia	0,310	0,312	0,313	0,315	0,319
Czech Republic	0,283	0,284	0,286	0,286	0,286
Estonia	0,278	0,281	0,281	0,282	0,283
Hungary	0,286	0,284	0,283	0,281	0,280
Latvia	0,280	0,286	0,290	0,294	0,298
Lithuania	0,280	0,286	0,291	0,297	0,303
Poland	0,292	0,292	0,291	0,290	0,288

[3]*The proportion of elderly people in the female and male labor force was calculated as the share of people aged 50-59 (for women) and 50-64 (for men) in the labor force (Kotowska, 1990).*

Romania	0,279	0,276	0,272	0,269	0,266
Slovakia	0,271	0,274	0,277	0,279	0,279
Slovenia	0,307	0,311	0,316	0,319	0,323

Source: *Own elaboration based on Eurostat data.*

The sharpest increase in that ratio could be observed in Lithuania which experienced the greatest rise in the number of men aged 50-64 over the study period. Note that between 2013-2017, all countries recorded a decline in their male populations; this is true for both the working-age population and the group aged 50-64. The exceptions are Lithuania, Slovenia, and Slovakia all of which experienced an increase in the number of men aged 50-64 by 2,2%, 2,1% and 0,7%, respectively. In turn, the greatest decline (by 4,3%) in the male population aged 50-64 was witnessed in Poland. Hence, it can be concluded that between 2013-2017, Poland was not as severely or intensively affected by the ageing of the male labor force as other countries surveyed. However, due to low fertility rates, Poland will face that process later.

3. Economic Activity of Men

European Union authorities have taken various measures under their economic, social and labor market policies to prevent a demographic crisis. The general employment rate in 2010 was to have been 70% (European Parliament, 2010) according to the Lisbon Strategy, an important Union document addressing the increase in economic activity. Note that in 2010, that level was attained only in a few Community countries (including Germany, the Netherlands, Austria, and the Nordic countries).

The next step taken to improve the condition of the labor market was the introduction of a new strategy (Europe, 2020), which set five major goals, including increasing the employment rate for the population aged 20-64 from 69% to 75%, possibly through an improvement in the economic activity of women and older people and through the integration of migrants in the labor force. That goal is supposed to be met by 2020 (European Commission, 2010). As regards the countries surveyed, the highest male employment rates were recorded in the Czech Republic (75,7-80,9%) and Estonia (71,4-77,4%) between 2013-2017, whereas the lowest rate was in Croatia (56,5-63,8%) (Table 4). The greatest increase (by 11,5 percentage points) was reported in Hungary.

Table 4. *Male employment rate (%) in 2013–2017*

	2013	2014	2015	2016	2017
Bulgaria	62,1	63,9	65,9	66,7	70,6
Croatia	56,5	59,1	60,3	61,4	63,8
Czech Republic	75,7	77,0	77,9	79,3	80,9
Estonia	71,4	73,0	75,3	75,7	77,4
Hungary	63,7	67,8	70,3	73,0	75,2
Latvia	66,8	68,4	69,9	70,0	71,9
Lithuania	64,7	66,5	68,0	70,0	70,6
Poland	66,6	68,2	69,2	71,0	72,8
Romania	67,6	68,7	69,5	69,7	71,8

Slovakia	66,4	67,6	69,5	71,4	72,0
Slovenia	67,1	67,5	69,2	68,9	72,5

Source: *Own elaboration based on Eurostat data.*

The employment rate is higher for men than for women. In 2013, the average male and female employment rates were 66% and 57%, respectively, in the countries surveyed. In turn, it grew by 6 percentage points for both men and women in 2017. Men are more economically active because they are less involved in family life. According to research, men in European Union countries spend an average of 39 hours at paid work and work for 10 hours on a non-remunerative basis per week; the corresponding figures for women are 33 and 22 hours, respectively. Differences exist between European Union countries in the sharing of domestic and professional responsibilities between women and men (European Commission, 2017). In the countries covered by this paper, women work on a non-remunerative basis 2-3 times longer per week than men. The longest and the shortest time of unpaid female work was recorded in Slovenia (32 hours) and in Slovakia and Romania (16 hours), respectively (Figure 2). In turn, the corresponding times for men were 12 hours in Estonia and 5 hours in Poland. Paid working time for women varied in the range from 36 (Poland) to 40 (Bulgaria and Romania) hours per week; the corresponding interval for men was from 40 (Poland) to 43 (in the Czech Republic and Slovenia) (European Commission, 2017). A well-designed work-life balance policy may support the economic independence of both women and men if it enables an even distribution of caretaking responsibilities between them. However, if inappropriately designed, it may contribute to the disproportion in paid and unpaid (caretaking) work between women and men and, therefore, deteriorate women's ability to engage in professional activity (European Commission, 2017).

Figure 2. *Paid and unpaid working hours by gender*

Source: *European Commission 2017, p. 12.*

Men are more economically active because of the prevailing childcare stereotypes. Women quit their economic activity more often than men to take care of their children. Taking care of a child does not result solely from women's preferences. Instead, it often results from a country-specific family or caretaking model or from limited access to institutional forms of care. "Affordable and high-quality childcare services do not only help reconciliation of work with family life, conducive to labor market participation of women and strengthening gender equality" (European Commission, 2018, p. 3). In recent years, the traditional male breadwinner model has been replaced by either the dual-earner model or the modified breadwinner model (with one partner working full-time and the other partner working part-time) (European Commission, 2018, p. 15). According to a 2015 study by Eurofound, in the respondent group, ca. 32% of European Union employees belong to a single-earner household (with 21% male earners) (Eurofound, 2017). In the context of the countries considered in this paper, the largest and the smallest share of dual-earner households is found in Hungary (73%) and Croatia (50%), respectively (Table 5). In turn, the largest number of single-earner households (with either a man or a woman being the sole breadwinner) was recorded in Croatia (27% and 20%, respectively). The largest proportion of households with the man working on a full-time basis and the woman working on a part-time basis was reported in Latvia (5%) (Eurofound, 2017).

Table 5. *Distribution of workers by household type (%).*

	Dual earners full-time	Single earner, man	Single earner, woman	Man full-time, partner part-time	Woman full-time, partner part-time
Bulgaria	59	24	13	2	2
Croatia	50	27	20	2	1
Czech Republic	71	19	7	2	1
Estonia	71	12	12	3	1
Hungary	73	15	10	2	1
Latvia	61	19	13	5	2
Lithuania	71	18	7	3	1
Poland	55	25	15	3	2
Romania	52	26	16	3	3
Slovakia	61	24	13	1	1
Slovenia	67	20	10	2	1

Source: *Eurofound 2017, p. 31.*

The level of the employment rate is impacted by the length of the economic activity cycle and by employment flexibility. The longest duration of economic activity for men in the countries surveyed between 2013-2017 was recorded in the Czech Republic (37,8 years in 2013) and Estonia (39,4 years in 2017) whereas the shortest was in Hungary (33,1 in 2013) and Lithuania (35,4 in 2017) (Table 6).

Note that between 2013-2017, the average duration of economic activity for men in the European Union was 37-38 years. On average, men were economically active 5 years longer than women. The average European women worked for 32-33 years in the study period. Men remain economically active for a longer period due to the prevailing stereotypes about their role in the household, and because of the pension schemes in place. In some countries, the statutory retirement age differs between men and women. Women retire earlier than men (e.g., in Poland where the retirement age for women and men is 60 and 65 years, respectively). Many European Union countries introduced regulations in order both to raise the retirement age and to align it between genders.

Table 6. *Duration of economic activity in 2013-2017*

	2013	2014	2015	2016	2017
Bulgaria	33,2	33,3	33,4	33,1	34,4
Croatia	33,2	34,2	34,5	34,0	34,5
Czech Republic	37,8	38,1	38,2	38,7	38,9
Estonia	37,2	37,3	37,9	38,7	39,4
Hungary	33,1	34,2	35,0	35,9	36,4
Latvia	35,1	35,0	35,7	35,6	36,2
Lithuania	34,1	34,7	34,7	35,4	35,4
Poland	34,8	35,1	35,2	35,6	36,0
Romania	35,5	35,7	36,0	35,7	36,4
Slovakia	35,8	35,9	35,9	36,4	36,4
Slovenia	35,1	35,6	35,6	35,3	36,6

Source: *Own elaboration based on Eurostat data.*

Flexibility of employment is related to the ability to use unconventional forms of employment, e.g., part-time jobs or contracts entered for a definite period. In the countries surveyed, the share of men employed under a fixed-term contract varied greatly in the range of barely 2% to over 27%. Fixed-term contracts were most popular in Poland (from 27,2% in 2013 to 25,6% in 2017), with the opposite being witnessed in Romania (from 1,7% in 2013 to 1,4% in 2017). Men are not interested in part-time jobs. In European Union countries, the share of men employed under a part-time contract varied in the range from 1,8% in 2013 in Bulgaria to 6,4% in 2017 in Estonia, reaching the highest level of 8,4% in Romania in 2013 (Eurostat, 2019). The use of fixed-term and part-time contracts depends on legal regulations applicable in the country concerned. Nevertheless, women are more likely than men to be employed under non-standard conditions which often involve lower hourly remunerations. Therefore, this situation contributes to income disparity between genders (European Commission, 2017). In the European Union, the average share of men employed under a fixed-term contract varied in the range from 13,2% to 13,9% (vs. 14,1% to 14,8% for women) between 2013-2017. In turn, the average share of men employed on a part-time basis was 8,8% (vs. 32% for women) (Eurostat, 2019).

The level of education has an impact on economic activity, too. Between 2013-2017, the highest employment rates were recorded for men with a secondary, post-secondary and upper-secondary (non-tertiary) education, and varied in the range from 77,4-74,7% in Slovakia to 62,4-59,4% in Lithuania (Table 7). Conversely, the lowest employment rates could be observed among men with a secondary, primary, or lower education from 3,4% in Slovakia in 2013 to 19,1% in Romania in 2017.

Table 7. *Employment rate for men aged 15–64 grouped by education level in 2013 and 2017*

	Less than primary, primary, and junior secondary education		Upper secondary and post-secondary non-tertiary education		Tertiary education	
	2013	2017	2013	2017	2013	2017
Bulgaria	11,4	12,4	65,4	63,8	23,2	23,8
Croatia	11,4	8,4	67,9	68,5	20,3	23,0
Czech Republic	3,5	3,9	75,6	73,6	20,8	22,5
Estonia	11,0	13,6	59,7	56,1	29,2	30,3
Hungary	10,2	11,5	67,5	66,6	22,2	21,9
Latvia	11,6	10,9	62,6	62,0	25,7	27,1
Lithuania	5,2	5,0	62,4	59,4	32,4	35,6
Poland	7,1	5,8	68,7	67,4	24,2	26,8
Romania	19,3	19,1	64,2	63,1	16,5	17,7
Slovakia	3,4	4,4	77,4	74,7	19,2	20,8
Slovenia	10,8	8,9	64,2	63,4	25,0	27,7

Source: *Own elaboration based on Eurostat data.*

The countries surveyed have witnessed a decline in the employment rate for men with a secondary, primary (or lower), secondary, post-secondary and upper-secondary (non-tertiary) education, except for Bulgaria, Czech Republic, Estonia, Hungary, Slovakia, and Croatia. A satisfactory finding is that the employment rate for men with a tertiary education grows in nearly all the countries considered (except for Hungary). "Although tertiary education delays the start of paid employment, it substantially increases lifetime earnings and is a good investment both for the individual and for the society" (OECD, 2006).

4. Linear Ordering and Classification of Selected EU Regions by Male Activity and Education Levels

Multidimensional statistical analysis methods were used to quantify the activity and determine the education level of men in selected EU regions. This is because it is necessary to compare many research fields described with a broad set of variables in this kind of analyses. In this case, the choice of variables was largely determined by the availability of complete, up-to-date data for all objects. The sub-variables covered are indicators (rather than absolute values). Somehow this has allowed to restrict the distortions resulting from the fact that some objects demonstrate certain characteristic features (e.g., a significantly larger area than other ones). In the second phase, the discriminating capacity of variables and their capacity, i.e., the degree of correlation with other variables, was examined to obtain the final set of diagnostic variables.

When choosing the variables, specific observations must demonstrate adequate variation because a non-diversified variable is of limited analytical value. The classic coefficient of variation was used to measure the diversification of specific variables. It was assumed that the set of potential variables reflecting male activity and education levels would exclude the characteristics which demonstrate a coefficient of variation below a critical threshold arbitrarily set at 10% (such characteristics are quasi-fixed). All of them have been retained for further analysis due to high values of variation coefficients of potential diagnostic variables in the set referring to education levels. In turn, variables A1 and A2 exhibited relatively low variation (with a classic coefficient of variation at 7,54% and 2,93%, respectively) in the set referring to male activity levels. However, due to limited availability and substantive value of these statistical variables, they were retained for further analysis.

Beside variation, their correlation is an important criterion for the selection of variables. As two highly correlated variables deliver similar information, it is recommended to eliminate one of them. The inverse correlation matrix (a method for the discrimination of features depending on the correlation matrix entries) was used to assess the information value. The inverse correlation matrix was calculated for each thematic sub-group (Panek and Zwierzchowski, 2013):

$R^{-1} = \left[\tilde{r}_{jj'}\right]$, j, j'=1, 2, ..., m, with:

$$\tilde{r}_{jj'} = \frac{(-1)^{j+j'}\left|\boldsymbol{R}_{jj'}\right|}{\left|\boldsymbol{R}\right|},$$

where: $\boldsymbol{R}_{jj'}$ - is a matrix reduced by removing row *j* and column *j'*; $\left|\boldsymbol{R}\right|, \left|\boldsymbol{R}_{jj'}\right|$ - determinants of matrices $\boldsymbol{R}$ and $\boldsymbol{R}_{jj'}$, respectively.

In accordance with this method, the variable corresponding to the highest diagonal entry of the inverse correlation matrix (above the critical threshold value fixed arbitrarily, usually at r*=15) is removed where needed. Following this, the inverse correlation matrix is recalculated, and the diagonal entries are checked to see if they exceed the fixed threshold value. That procedure is continued until all diagonal entries are below or equal to that threshold. In both sets of variables, only the diagonal entry of the inverse correlation matrix corresponding to the index of the number of male pupils and students aged 25-29 (E9) and 35-39 (E11) was above the critical threshold value. Therefore, both variables were eliminated from the set of decisive variables.

The nature of each of them was specified (the effect it has on the phenomenon covered by this analysis: a stimulating effect, an inhibiting effect, or a neutral effect). In both sets, all variables were classified as having a stimulating effect (high values are desirable in the context of the characteristics of the phenomenon under consideration). One of the main requirements imposed by taxonomic methods on final diagnostic variables is their comparability (the addition postulate). A normalization procedure[4]

[4]*The objective of normalization is to ensure comparability of characteristics expressed in different units; unify the nature of variables; eliminate negative values (if any) from the calculation; and stabilize the variation (Balicki, 2009, p. 256).*

was performed with a view to ensure the comparability of characteristics expressed with different units and of different orders of magnitude. The most widely used normalization methods include standardization, unitarization and quotient transformation. The values of the variable were subject to a standardization procedure for the purposes of this paper. The objective of standardization is to obtain variables with a distribution with a mean of 0 and a standard deviation of 1. The most popular standardization formula (Młodak, 2006) is as follows:

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{s_j} \tag{2}$$

where: $\overline{x_j}$ - arithmetic mean; s_j - standard deviation of x_{ij}; i = 1, 2, …, n; j = 1, 2, …, m.

Different weights were attributed to variables depending on their discriminatory and information capacity (separately in both sets of variables). The modified *BVP*[5] method was used for this purpose. It relies on a more adequate measure of information capacity than the linear correlation coefficients used in the original *BVP* which fail to take the presence of collinearity into account. The analytical form of weights may be expressed as:

$$w_j = w_j^a \cdot w_j^b,\ j = 1,2,...,m \tag{3}$$

where: w_j^a - measure of discriminatory capacity of diagnostic variable *j*; w_j^b - measure of information capacity of diagnostic variable *j*. The measure of discriminatory capacity, based on the classic coefficient of variation, is expressed as:

$$w_j^a = \frac{V(x_j)}{\sum_{j=1}^{m} V(x_j)}\text{'}\ j = 1,2,...,m. \tag{4}$$

In turn, the measure of information capacity may be defined as:

$$w_j^b = \frac{\sum_{\substack{j'=1 \\ j'\neq j}}^{m} r_{j.j'}^2}{\sum_{\substack{j=1 \\ j'\neq j}}^{m} \sum_{\substack{j'=1 \\ j'\neq j}}^{m} r_{j.j'}^2},\ j = 1,2,...,m. \tag{5}$$

where: $r_{j.j'}^2$ - squared coefficient of partial correlation between variable *j* and variable *j'*.

[5]*A modification of the method by G. Betti and V. Verma proposed by T. Panek. For a broader description, see Panek and Zwierzchowski, 2013.*

As it can be noticed, the measure of discriminatory capacity reaches the highest value for the variable with the highest coefficient of variation, whereas the measure of information capacity reaches the highest value for the variables with the highest absolute values of correlation coefficients (Table 8).

Table 8. *Weights of diagnostic variables.*

Variables	I	II	Weights	Variables	I	II	Weights	Variables	I	II	Weights
A1	0,093	0,328	0,030	E3	0,034	0,056	0,002	E8	0,227	0,120	0,027
A2	0,036	0,324	0,011	E4	0,085	0,080	0,006	E10	0,100	0,135	0,013
A3	0,871	0,347	0,302	E5	0,111	0,049	0,005	E12	0,084	0,123	0,010
E1	0,092	0,084	0,007	E6	0,098	0,143	0,014				
E2	0,037	0,064	0,002	E7	0,126	0,141	0,017				

Notes: *I: discrimination criterion, II: information criterion.*
Source: *Own study.*

As a result of performing the modified BVP routine, in the set of male activity variables, the highest weight was assigned to variable A3 (educational activity rate for men aged between 25-64). In turn, as regards the set of education level variables, the highest weights were assigned to variable E8 (number of men enrolled in doctoral programs per 1,000 population) and E7 (number of men enrolled in master's degree or equivalent programs per 1,000 population).

The classic TOPSIS (*Technique for Order Preference by Similarity to an Ideal Solution*) method was used to linearly order the regions of selected EU countries by male economic activity and education levels. According to this method, the synthetic indicator is created based on the Euclidean distance both from the positive ideal solution (pattern) and from the negative ideal solution (anti-pattern). The smaller is the distance from the positive ideal solution (and the greater is the distance from the negative ideal solution), the higher is the value of the synthetic variable. The steps of building the synthetic indicator are as follows (Hwang and Yoon, 1981):

1. Creating a normalized decision matrix.
2. In the case of weighted variables, the weight matrix and following this the weighted normalized decision matrix need to be created.
3. For the normalized features, the coordinates of the positive ideal (A^+) and the negative ideal (A^-) solution are determined; the structure of the composite object is described with the least advantageous values:

$$A^+ = (\max_i(v_{i1}), \max_i(v_{i2}), ..., \max_i(v_{iN})) = (v_1^+, v_2^+, ..., v_N^+), \tag{6}$$

$$A^- = (\min_i(v_{i1}), \min_i(v_{i2}), ..., \min_i(v_{iN})) = (v_1^-, v_2^-, ..., v_N^-), \tag{7}$$

4. Determining the Euclidean distance of each object from the positive ideal solution and the negative ideal solution:

$$s_i^+ = \sqrt{\sum_{j=1}^{N} (v_{ij} - v_j^+)^2},\ s_i^- = \sqrt{\sum_{j=1}^{N} (v_{ij} - v_j^-)^2}, i = 1,2,\ldots,M,\ j = 1,2,\ldots,N \qquad (8)$$

5. Calculating the value of the synthetic feature: $C_i = \frac{s_i^-}{s_i^+ + s_i^-}$, with $0 \le C_i \le 1$.

Table 9. *Values of the synthetic indicators of male economic activity (SMA) and education (SME) levels in selected EU countries produced by the TOPSIS method (as at 2017).*

Region	SMA	SME	Region	SMA	SME
Bratislavský kraj	0,6316	0,5164	Severoiztochen	0,2636	0,2785
București – Ilfov	0,3996	0,4024	Severovýchod	0,7088	0,2078
Centru	0,2678	0,2588	Severozápad	0,5942	0,1847
Dél-Alföld	0,4072	0,2730	Severozapaden	0,0427	0,1194
Dél-Dunántúl	0,3069	0,2246	Śląskie	0,3538	0,3745
Dolnośląskie	0,4377	0,3536	Stredné Slovensko	0,3675	0,2312
Eesti	0,6249	0,3494	Střední Čechy	0,7392	0,1990
Észak-Alföld	0,3360	0,3053	Střední Morava	0,7701	0,3397
Észak-Magyarország	0,2871	0,2732	Sud-Muntenia	0,2787	0,2318
Jadranska Hrvatska	0,3586	0,2018	Sud-Est	0,2089	0,2792
Jihovýchod	0,6663	0,2361	Sud-Vest Oltenia	0,2049	0,2841
Jihozápad	0,7329	0,2726	Świętokrzyskie	0,3435	0,2311
Kontinentalna Hrvatska	0,3083	0,2914	Vest	0,2528	0,2787
Közép-Dunántúl	0,5086	0,2163	Východné Slovensko	0,3094	0,2765
Közép-Magyarország	0,5777	0,4441	Vzhodna Slovenija	0,5898	0,2187
Kujawsko-Pomorskie	0,3770	0,2866	Warmińsko-Mazurskie	0,2728	0,2519
Latvija	0,3325	0,2851	Wielkopolskie	0,4545	0,3904
Lietuva	0,3083	0,4415	Yugoiztochen	0,2241	0,1954
Lubelskie	0,3672	0,3092	Yugozapaden	0,3622	0,3938
Lubuskie	0,3498	0,2048	Yuzhen Tsentralen	0,2145	0,2307
Łódzkie	0,3604	0,3224	Zachodniopomorskie	0,3151	0,2380
Małopolskie	0,4631	0,4020	Zahodna Slovenija	0,7090	0,3096
Mazowieckie	0,5020	0,5636	Západné Slovensko	0,4319	0,2056
Moravskoslezsko	0,6192	0,2150	**Variation**		
Nord-Est	0,3618	0,3374	Minimum	0,0427	0,1194
Nord-Vest	0,3365	0,3015	Maximum	0,7701	0,6477
Nyugat-Dunántúl	0,4562	0,2401	Arithmetic mean	0,4177	0,2957
Opolskie	0,4399	0,2007	Standard deviation	0,1672	0,0971
Podkarpackie	0,3970	0,2664	Coefficient of variation	40,04%	32,85%
Podlaskie	0,4196	0,2505	1st quartile	0,3083	0,2310
Pomorskie	0,5034	0,3443	3rd quartile	0,5047	0,3409
Praha	0,7648	0,6477	Skewness	0,5098	1,4257
Severen tsentralen	0,1679	0,3726			

Source: *Own elaboration based on Eurostat data.*

In the 2017 dataset, the coefficient of variation was over 40% and nearly 33% for the synthetic indicator of economic activity and the synthetic indicator of education, respectively (Table 9). The maximum-to-minimum ratio for SMA and SME is 18,04 and 5,42, respectively. Both synthetic indicators demonstrated right-side asymmetry which means that above-average values prevailed in most regions. However, asymmetry was much stronger[6] for the synthetic indicator of male education levels. In three quarters of the regions, the synthetic indicator of economic activity did not exceed 0,5047 (with a maximum at 0,7701 and a mean of 0,4177). In turn, in 75% of the regions considered, the synthetic indicator of education levels did not go above 0,3409 (with a maximum at 0,6477 and a mean of 0,2957). The highest values of the synthetic indicator of male economic activity levels were found in the Czech regions of Střední Morava, Praha and Střední Čechy. This was mostly due to extremely high employment rates for men aged 15-64.

In turn, the lowest SMA was found in Bulgarian and Romanian regions (Severozapaden, Severen Tsentralen, Sud-Vest Oltenia) characterized by exceptionally low employment rates for men aged 15–64. The highest values of the synthetic indicator of education levels were identified in regions which include the capital cities of the Czech Republic, Poland, and Slovakia (Praha, Mazowieckie and Bratislavský kraj). This was primarily due to extremely high indicators of the number of men enrolled in bachelor or equivalent programs and the number of men enrolled in master's degree or equivalent programs. Conversely, the lowest values of that indicator were observed in Bulgarian (Severozapaden and Yugoiztochen) and Czech (Severozápad) regions which exhibited low indicators of the number of pupils and students aged 40 or more and of the number of men enrolled in doctoral programs.

An analysis based on the Pearson linear correlation coefficient was performed to verify the relationship between the education levels and economic activity of men. The calculated coefficient of correlation between the defined synthetic indicators was 0,2664, which suggests a relatively weak relationship between the aspects covered by the analysis and allows to conclude that the correlation coefficient was significant at $p < 0{,}05$.

It is worth noting in this regard that "all developments taking place in space are interrelated and the relationship is an inverse function of distance," as noted by Bivand (1980). Therefore, spatial relationships (referred to as spatial autocorrelation) may exist between neighboring units. Spatial autocorrelation is defined as the correlation degree between the identified value of a variable in a specific location and the value of the same variable in another location. This means the values of the variable under consideration determine, and are determined by, the corresponding values recorded in

[6] *|0.0-0.4| – extremely weak asymmetry in the distribution; |0,4–0,8| – weak asymmetry in the distribution; |0,8–1,2| – moderate asymmetry in the distribution; |1,2–1,6| – strong asymmetry in the distribution; above |1,6| – extremely strong asymmetry in the distribution; after: Pułaska-Turyna, 2008, p. 87.*

other locations. There are two variants of spatial correlation, positive autocorrelation, and negative autocorrelation. Positive autocorrelation means spatial accumulation of high or low values of a variable. In turn, negative autocorrelation means that high and low values are adjacent to each other (resulting in a chessboard-like arrangement) (Suchecki, 2010). In such analyses, another problem is to address the impacts of the existing spatial structure. To do that, neighborhood structures are specified with the use of spatial weights. They are based on the distance or neighborhood matrix (the weights are non-zero if two locations share a border or are separated by a predefined distance). The approach used in this paper considers a shared border to be the proximity criterion. This is the most widely adopted neighborhood modeling method which uses a binary matrix as the starting point, 1 means that the areas share a border, 0 means they do not. This is a symmetric square matrix. Defined as above, the binary matrix is standardized by rows so that the sum of all entries is equal to 1 (Anselin, 2003; LeSage and Pace, 2009).

The global Moran's *I* was used to analyze the spatial interactions between the values of synthetic indicators of male activity (and of education levels) in specific regions and the corresponding values recorded in neighboring regions. It enables determining the strength and nature of correlations throughout the study area, and is calculated as follows (Bivand *et al.*, 2008; Suchecki, 2010):

$$I = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{9}$$

where: x_i, x_j - values observed in locations i and j ($i, j = 1, 2, \ldots, n$), $\bar{x}$ - average value in all areas under consideration, w_{ij} - entries of the spatial weight matrix.

A value of 0 means spatial randomness, i.e., lack of spatial autocorrelation (the numeric characteristics of the phenomenon in one territory do not depend upon the characteristics of adjacent territories). Positive and significant *I* value indicate the existence of positive autocorrelation (i.e., similarity of the examined objects). Conversely, negative *I* values mean negative autocorrelation (i.e., differentiation of the examined objects). Positive autocorrelation means that objects with similar values are grouped into clusters, while negative autocorrelation is interpreted as "*hot spots*," i.e., isolated areas where distinctly different values are recorded (Kopczewska, 2007).

The global Moran's *I* statistic falls within the interval [-1, 1]. Its statistical significance is verified by testing the following hypotheses:

H_0: The observed values of the variable under consideration are distributed randomly across different locations.
H_1: Spatial autocorrelation exists: the observed values of the variable under consideration are not distributed randomly across different locations.

The significance of Moran's *I* may be tested based on theoretical moments or on the permutation approach. Normalized Z_I is used as the test statistic:

$$Z_I = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1) \tag{10}$$

The functional forms and the values of moments in the sample depend on the assumptions made in the study. Under the hypothesis of normality (the sampled objects are values of an independent random variable having a normal distribution), the expected value and variance are expressed with the following formulas:

$$E(I) = -\frac{1}{n-1}, \quad \mathrm{Var(I)} = \frac{\mathrm{n^2 S_1 - nS_2 + 3S_0^2}}{\mathrm{(n^2-1)S_0^2}} - \frac{1}{\mathrm{(n-1)^2}} \tag{11}$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}, \; S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2, \; S_2 = \sum_{i=1}^{n}(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji})^2,$$

With *n* - the number of areas covered by the study.

Conversely, under the hypothesis of randomness, the expected value is identical to that specified above while variance may be expressed as:

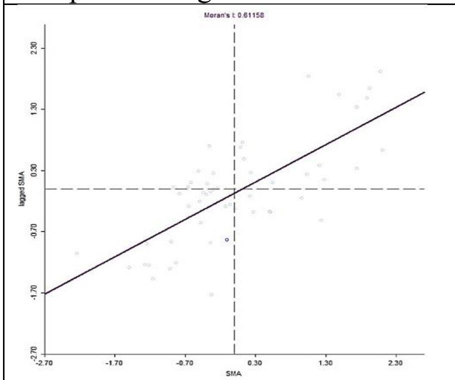
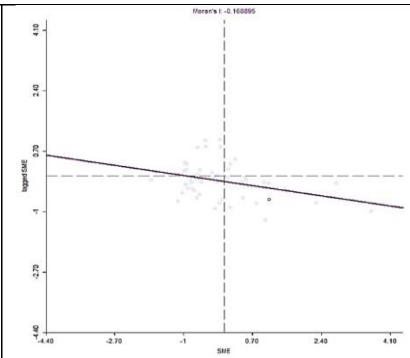
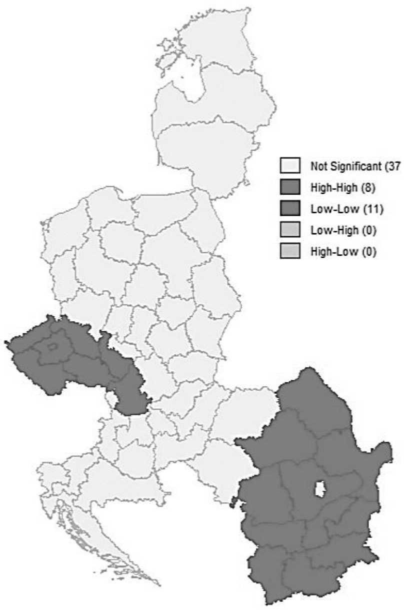
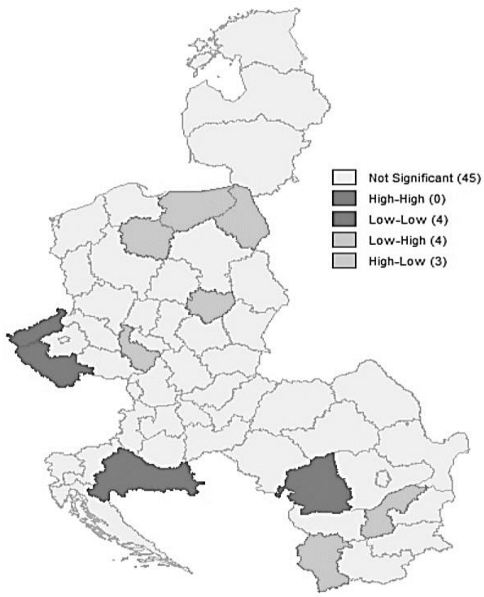
$$Var_R(I) = \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2] - k[(n^2 - n)S_1 - 2nS_2 + 6S_0^2]}{(n-1)(n-2)(n-3)S_0^2} - \frac{1}{(n-1)^2}, \tag{12}$$

where: *k* - fourth moment divided by the second moment squared.

The permutation approach does not require making an a priori assumption on the distribution because the frequency distribution in the sample is determined using the permutation method. The Monte Carlo simulation is used to capture the empirical variation. The permutation test consists in a random allocation of feature values to different locations. Each time, the test statistic is calculated. The number of permutations specifies the maximum significance level; several thousand permutations must be drawn to improve the accuracy of the inference process (Kopczewska, 2007; Suchecki, 2010). In this paper, the global statistical significance test was based on the analysis of permutation histograms for the randomization test. The hypothesis was verified based on the pseudo-significance level. The number of permutations was 9999.

The global Moran's *I* statistic calculated for the synthetic indicator of male economic activity was positive (0,6116) and, most importantly, statistically significant. A dot plot of the global Moran's *I* statistic was produced for a more in-depth analysis (Figure 3). The slope of the regression line plotted on the graph is equivalent to the value of the global Moran's *I* statistic. As most of the points are in the third quadrant of the graph of the global Moran's *I* statistic, it may be assumed that a low level of the synthetic indicator of male activity is the clustering criterion for most of the EU regions considered. In turn, the global Moran's *I* statistic calculated for the synthetic indicator of education level was -0,1689 and was statistically significant. The dot plot of the global Moran's *I* statistic for the synthetic indicator of education levels suggests and provides grounds for concluding that a low level of the indicator was the usual clustering criterion for the regions, just as in the case of male activity levels.

Figure 3. *Dot plot of the global Moran's I statistic for the synthetic indicator of male economic activity and education levels; local Moran's I statistics maps.*

<table>
<tr><td>Synthetic indicator of male activity</td><td>Synthetic indicator of education level</td></tr>
<tr><td colspan="2">Dot plot of the global Moran's I statistic</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Maps of values of the local Moran's I statistics by cluster type</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Source: *Own study.*

Calculating the local statistics for spatial autocorrelation enabled identifying clusters of areas with similar levels of the feature under consideration and areas characterized by different values of the synthetic indicators of male activity and education levels. As regards the synthetic indicators of male activity, a large compact cluster was identified comprising 8 high-high areas (a high value of the indicator surrounded by high values) located in the Czech Republic (7 regions) and Slovakia (1 region). The structure of districts analyzed also included 11 *low-low* areas (reporting low values of the variable under analysis) which formed a large compact area extending over Romanian and Bulgarian territory. In turn, the maps of local Moran's *I* statistics created for synthetic indicators of education levels allowed to identify 4 low-low areas (2 neighboring areas in the Czech Republic, 1 in Croatia and 1 in Romania). Also, 7 objects that need to be regarded as outliers were observed, including 4 *low-high* areas (4 Polish regions adjacent to the Mazowieckie voivodeship, the largest and wealthiest Polish region) and 3 spatially dispersed high-low areas (high values of the indicator surrounded by low values): Střední Morava (Czech Republic) and Yugozapaden and Severen Tsentralen (Bulgaria).

Table 10. *Values of local Moran's I_i statistics calculated for the synthetic indicator of male economic activity and education levels.*

Region	SMA	SME	Region	SMA	SME
Bratislavský kraj	0,2017	-1,7042	Podkarpackie	0,0346	-0,0279
București – Ilfov	0,0895	-0,7226	Podlaskie	-0,0038	-0,4593*
Centru	0,7701**	0,0402	Pomorskie	-0,1925	-0,0206
Dél-Alföld	0,0159	0,0028	Praha	3,9901*	-3,6074
Dél-Dunántúl	-0,0095	0,3055	Severen tsentralen	1,9114**	-0,6097*
Dolnośląskie	0,0907	-0,1659	Severoiztochen	1,1975**	0,0243
Eesti	-0,6312	-0,0602	Severovýchod	2,3297**	0,3084
Észak-Alföld	0,0996	0,0182	Severozápad	1,9519**	0,8148*
Észak-Magyarország	0,0616	0,0077	Severozapaden	2,3545**	0,2052
Jadranska Hrvatska	-0,2492	0,2238	Śląskie	-0,0333	-0,2393
Jihovýchod	2,3015**	0,3208	Stredné Slovensko	-0,0797	-0,0160
Jihozápad	2,8048**	0,2181*	Střední Čechy	3,1694**	-0,1441
Kontinentalna Hrvatska	-0,0237	0,0297*	Střední Morava	1,3455*	-0,2798*
Közép-Dunántúl	0,0595	0,1536	Sud-Muntenia	1,0053**	0,0653
Közép-Magyarország	-0,1346	-0,6453	Sud-Est	1,1176*	-0,0001
Kujawsko-Pomorskie	-0,0014	-0,0761*	Sud-Vest Oltenia	1,5761**	0,0906*
Latvija	-0,1490	-0,1119	Świętokrzyskie	0,0276	-0,5291*
Lietuva	0,2974	-0,5127	Vest	0,6693	0,0295
Lubelskie	0,0039	0,0458	Východné Slovensko	0,1840	0,0002
Lubuskie	0,0370	-0,3044	Vzhodna Slovenija	0,2483	0,2855
Łódzkie	0,0121	0,1282	Warmińsko-Mazurskie	-0,0226	-0,3784*
Małopolskie	-0,1030	-0,2229	Wielkopolskie	-0,0265	-0,1716
Mazowieckie	-0,1836	-0,5804	Yugoiztochen	1,6992**	0,4828
Moravskoslezsko	0,4694	0,0787	Yugozapaden	0,5737**	-1,2545**
Nord-Est	0,2930*	-0,0702	Yuzhen Tsentralen	1,5108**	0,4101

Nord-Vest	0,2685	-0,0031	Zachodniopomorskie	-0,0668	-0,1066
Nyugat-Dunántúl	0,0622	0,0996	Zahodna Slovenija	0,5888	-0,1261
Opolskie	0,0649	-0,3712	Západné Slovensko	0,0586*	-0,1568

Notes: ** statistically significant at p < 0,05; ** statistically significant at p < 0,01.*
Source: *Own study.*

The analysis of local Moran's I_i statistics calculated for the synthetic indicator of male activity provides grounds for concluding that the statistic is positive and statistically significant in 19 regions, which means these areas are adjacent to territories with similar levels of the activity indicator. As regards other regions, values of the local Moran's I_i statistics calculated for the variable under consideration were positive in 21 cases and negative in 16 cases (Table 10). However, as the negative values were not statistically significant, no particular attention should be paid to these results. In turn, as regards the calculated synthetic indicators of education levels, the local Moran's I_i statistics were statistically significant and positive in 4 regions, and statistically significant and negative in 7 regions. In other districts, most (24) of the local statistics were negative but statistically insignificant.

A spatial regression analysis was carried out to assess the strength of spatial relationships between the synthetic indicators of male activity and the education level. Spatial models can be considered an extension of "classical"[7] econometric models supplemented with additional variables to address the spatial interactions. This is important in that the values of the variable under consideration determine, and are determined by, the corresponding values recorded in other locations (observations in different locations are not independent of one another). The presence of spatial relationships contributes to changing the properties of structural parameters in models estimated with the least squares' method. The purpose of spatial modeling is to enhance the specifications of the econometric model[8]. Generally, spatial interactions can involve the following (Suchecki, 2010) the explained variable (this means spatial autoregression, i.e., a situation where the values of the endogenous variable recorded in other locations affect the values of that variable in location i), the explanatory variable (if the endogenous variable in location i is affected by values of exogenous variables recorded in other locations), the random effect (if the model does not or cannot include certain spatially autocorrelated variables).

If spatial effects are identified, the spatial regression model is estimated in a manner to minimize their impact on the model's discriminatory capacity. Generally, three models of spatial interaction processes can be identified (Suchecki, 2010):

[7]*The "classical" regression analysis includes creating models (based on a series of possible modeling methods) which provide a quantitative description of relationships between the dependent variable and the set of one or multiple dependent variables.*
[8]*Maximum likelihood is the most widely used method in estimating the parameters of spatial regression models. The instrumental variables method or, in some cases, the generalized method of moments are also employed.*

- SSAR (Simultaneous Spatial Autoregression), SAR/SEM (Spatial Autoregression/Spatial Error Models),
- SMA (Spatial Moving Average, with the endogenous variable's expected value being zero),
- SEC (Spatial Error Components, with the random variability being decomposed into two random components). The first component is interpreted as the regional effect whereas the second one stands for effects specific to different locations.

In accordance with a widely employed strategy for the selection of spatial regression models, the classical least squares method was used to estimate the structural parameters of the linear regression model in the first step of this study. The results of the Jarque–Bera test do not allow for the hypothesis of normal distribution of the random effect being rejected. Therefore, the values of asymptotic Lagrange multiplier tests can be calculated, and the maximum likelihood method can also be used. This is important because if spatial autocorrelation exists, the classical estimator based on the least squares' method can be incompatible (or at least inefficient) with SEM models, for instance (Table 11).

Table 11. *Estimation results for the male economic activity model, the classical model and the spatial error model.*

Models	**Classical model**	**SEM**
	Estimation	
λ	-	0,7347 (0,0000)
Intercept	0,2821 (0,0002)	0,2878 (0,0000)
SME	0,4586 (0,0472)	0,4779 (0,0001)
AIC	-42,4975	-84,3799
SC	-38,4468	-80,3292
Log likelihood	23,2488	44,1899
Normality test		
JB	4,2444 (0,1198)	-
Heteroscedasticity tests		
BP	0,9160 (0,3385)	0,0503 (0,8225)
KB	1,0673 (0,3016)	-
Spatial autocorrelation tests		
Moran I_{error}	7, 6282 (0,0000)	-
LM_{SAR}	46,3522 (0,0000)	-
RLM_{SAR}	2,2494 (0,1337)	-
LM_{SEM}	51,5102 (0,0000)	-
RLM_{SEM}	7,4073 (0,0065)	

Notes: *Significance levels for the rejection of the null hypothesis are put in brackets.*
W_SMA: lag of the explained variable (SMA); AIC: Akaike Information Criterion; SC: Schwarz Criterion; Log likelihood: logarithm of the likelihood function; JB: Breusch–Pagan test; KB: Koenker–Bassett test; Moran I_{error}: Moran's I error significance test; LM_{SAR}: Lagrange multiplier test for SAR; LM_{SEM}: Lagrange multiplier test for SEM, RLM_{SAR}: robust Lagrange multiplier test for SAR; RLM_{SEM}: robust Lagrange multiplier test for SEM.
Source: *Own calculations.*

As shown by the calculations, spatial autocorrelation exists between residuals (as reflected by the low *p-value* for the Moran's *I* statistic calculated for the regression residuals). Hence, spatial estimation methods need to be used in the estimated model. As mentioned earlier, there are two groups of models which specify this type of spatial relationships, SAR, SLM (spatial lag models) and SEM (spatial error models).

The spatial lag model includes what is referred to as the spatially lagged endogenous variable, which makes it an autoregression model. In turn, the spatial error model assumes the existence of spatial autocorrelation between residuals. The existence of spatial autocorrelation in the error term of the model may result from the omission of non-observed variables which could be spatially correlated (Kopczewska, 2007).

The general form of spatial lag models (SLM) is as follows:

$$y = \beta X + \rho W y + u, u \sim IID\ N(0,1), \tag{13}$$

where: X - matrix of independent variables; β - vector of coefficients; W - matrix of spatial weights; ρ - spatial autocorrelation coefficient; u - model's error term; Wy - spatial lag of dependent variable (interpreted as the level of dependent variable y in neighboring regions). The general form of spatial error models (SEM) is as follows:

$$y = \beta X + u, u = \lambda W u + \varepsilon, \varepsilon \sim IID\ N(0,1), \tag{14}$$

where: λ - spatial autocorrelation parameter; Wu - spatially lagged error term (mean error in neighboring locations); ε - model's independent error term.

The Lagrange multiplier tests were used to determine the type of spatial interaction: LM_{SEM} (for the autocorrelation of the random effect) and LM_{SAR} (for the autoregression of the explained variable). The residuals of the model developed based on the least squares method and on the weight, matrix standardized by rows were used to verify the null hypothesis on the absence of spatial autocorrelation in the error term. The following statistic was employed for that purpose:

$$LM_{SEM} = \left(\frac{1}{T_1} \left(\frac{\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{e}}{\hat{\sigma}^2} \right)^2 \right) \overset{as}{\sim} \chi^2_{(1)},\ T_1 = \mathrm{tr}[(\boldsymbol{W}^T + \boldsymbol{W})\boldsymbol{W}] \tag{15}$$

where: tr[.] - trace of a matrix; e - vector of residuals of the model estimated using the least squares method; $\hat{\sigma}^2$ - residual variance estimator.

In the second case, the null hypothesis (with H_0: $\rho = 0$, $H1$: $\rho \neq 0$) is verified with the following statistic:

$$LM_{SAR} = \left(\frac{1}{T_2} \left(\frac{\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{y}}{\hat{\sigma}^2} \right)^2 \right) \overset{as}{\sim} \chi^2_{(1)}, \; T_2 = T_1 + \frac{(\boldsymbol{WX\hat{\beta}})^{\boldsymbol{T}} \boldsymbol{M_x} (\boldsymbol{WX\hat{\beta}})}{\hat{\sigma}^2} \tag{16}$$

where: $M_x = \frac{e^T W e}{e^T e} - X(X^T X)^{-1} X^T$, $\hat{\boldsymbol{\beta}}$ - estimator expressed as $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$.

In accordance with the approach proposed by L. Anselin[9], the robust tests were used because of the significance of both tests in their basic form:

$$RLM_{SEM} = \left(\frac{1 - T_2 T_1}{T_1} \left(\frac{\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{y}}{\hat{\sigma}^2} - \frac{T_1}{T_2} \frac{\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{e}}{\hat{\sigma}^2} \right)^2 \right) \overset{as}{\sim} \chi^2_{(1)}, \tag{17}$$

$$RLM_{SAR} = \left(\frac{1}{T_2 - T_1} \left(\frac{\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{y}}{\hat{\sigma}^2} - \frac{\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{e}}{\hat{\sigma}^2} \right)^2 \right) \overset{as}{\sim} \chi^2_{(1)}. \tag{18}$$

The comparison of the values of RLM_{SAR} and RLM_{SEM} tests (calculated *ex ante* based on the residuals of the model estimated with the least squares method) decided upon the use of the spatial lag model ($RLM_{SEM} > RLM_{SAR}$[10]).

The last column of Table 11 presents the results of the SEM model estimation based on the maximum likelihood method. In its structural form, the estimated model may be written as:

$$\begin{aligned} SMA &= 0{,}2888 + 0{,}4779 \cdot SME + u, \\ u &= 0{,}7347 \cdot W_u + \varepsilon \end{aligned} \tag{19}$$

where: *SMA* - synthetic indicator of male economic activity; *SME* - synthetic indicator of male education in selected EU regions; *Wu* - spatially lagged error term (mean error in neighboring locations); ε - model's independent error term.

All regression coefficients are statistically significant. This means the variables included affect the synthetic indicator of male economic activity in the EU regions considered. The statistical significance of λ (-0,7347, p-value=0,0000) implies the existence of spatially autocorrelated extra-model factors that affect the economic activity. This means that the model fails to take account of some non-observed (e.g., non-measurable or random) variables which can be spatially correlated. In turn, according to Lechman (2015), a statistically significant λ value can be interpreted as the existence of spatial autocorrelations caused by random non-modellable factors

---

[9] *For more information, Anselin L. 2003.*

[10] *Additionally, the test value in the extended approach was not statistically significant (p<0.05).*

and/or measurement errors. It may also be assumed that the exogenous shock in each region within the spatial error model will affect not only the situation prevailing in that very region but also the condition of neighboring regions due to the presence of a spatial dependency of errors (Kopczewska, 2007).

In the case of the spatial error model, the parameter of the explanatory variable can be interpreted as a partial derivative, i.e., a change in the explained variable triggered by a change in the explanatory variable with other explanatory variables remaining constant (Pander *et al.*, 2014). The results of SEM estimation give grounds for concluding that a 1% increase in the value of the synthetic indicator of education results, *ceteris paribus*, in a nearly 0,48% increase in the synthetic indicator of male activity levels in each region. The model which includes the mean error in neighboring locations may be concluded to be better than the one based on classical least squares, according to the Akaike and Schwarz information criteria. Similar conclusions may be drawn based on the likelihood logarithm (the model with the higher value is the better one).

In the spatial regression models, in addition to spatial interactions in the form of autoregression or autocorrelation of the random effect, the analysis should also extend over spatial heterogeneity, i.e., the instability in the space of relationships (which can be, for instance, of an economic nature). This can result from many factors, including the asymmetry of the relationship between central and remote areas. Spatial heterogeneity can also be the consequence of an erroneous specification of the model; if so, the spatial distribution of model errors is the same as the distribution of the variable not included in the model.

Econometric models may address these differences in two ways (Suchecki, 2010), based on variability in variance of the random effect (if the problem of heteroscedasticity of the random effect emerges in the economic model; this is caused by the failure to take significant explanatory variables into account or by other errors in the specifications) or based on variability in structural parameters (if the parameters of the regression model vary across locations).

Heteroscedasticity may be verified with the Breusch–Pagan (BP) test. The test statistic is expressed as:

$$BP = \tfrac{1}{2}[g^T Z(Z^T Z)^{-1} Z^T g - N] \overset{as}{\sim} \chi^2_{(k)} \qquad (20)$$

where g is the vector created based on residuals of the model developed with the least squares method: entries of the vector are $g_i = \dfrac{e_i^2}{(\frac{e^T e}{N})}$, Z is the complete matrix of explanatory variables, N is the total number of objects (e.g., districts).

Another routine which may employed is the Koenker–Bassett test with the test statistic expressed as:

$$KB = \frac{1}{Var(\varepsilon^2)}(u - \bar{u}i)^T Z(Z^T Z)^{-1} Z^T (u - \bar{u}i) \tag{21}$$

where: $Var(\varepsilon^2) = \frac{1}{N}\sum_{i=1}^{N}(e_i^2 - \frac{e^T e}{N})^2$, $u = [e_1^2\ e_2^2 \dots e_N^2]$, $\bar{u} = \frac{e^T e}{N}$.

Neither the Breusch–Pagan nor the Koenker–Bassett test (see Table 11) allowed to reject the null hypothesis on the homoscedasticity of the random effect (at p > 0,05). The linear regression model developed could have been used in the analysis without the need to introduce any enhancing variables (reflecting the parameters' variability, e.g., east/west). If the model's residuals proved to be heteroscedastic, binary modeling of spatial regimes would be the right solution.

5. Conclusions

Research suggests that men enjoy a more favorable situation than women in the labor market. They have a higher employment rate and a longer employment period (by ca. 10 percentage points and ca. 5 years, respectively). Men are more economically active because they are less involved in housekeeping and childcare responsibilities. An average European man spent an average of 39 hours in paid employment and worked for 10 hours on a non-remunerative basis per week; the corresponding figures for an average European woman are 33 and 22 hours, respectively. Also, men remain economically active for a longer period due to the prevailing stereotypes about their role in the household. Usually, men are not interested in part-time jobs. In the countries surveyed, the average share of men employed on a part-time basis was 4,5% (vs. 9% for women). In addition to socioeconomic factors, the economic activity of men is also impacted by demographic aspects (including place of residence and education level).

As regards the countries covered by this analysis, the highest and the lowest employment rates were recorded in the Czech Republic (75,7-80,9%) and Croatia (56,5-63,8%), respectively. When it comes to education, the highest employment rates were recorded for men with a secondary, post-secondary and upper-secondary education (59,7-77,4%), whereas the lowest were found for men with a vocational, primary and lower education (3,4-19,3%). Note also that the indicators follow a downward trend. Conversely, there was an increase in the employment rate for men with a tertiary education (by 1,7 percentage points, on average). The results of the spatial regression analysis give grounds for concluding that (as at 2017) a 1% increase in the synthetic indicator of male education level results in a nearly 0,48% increase in the synthetic indicator of activity level at regional level (under the assumption that other factors remain constant).

In further research, it is worth considering other spatial statistics, including both global (e.g., Gamma, Geary's *C*) and local ones (e.g., the Getis–Ord *G* statistic). Alternatively, another spatial neighborhood structure (higher-order neighborhood) could be employed. Obviously, the level of data aggregation and the area of territories surveyed also had an impact on the results of spatial analyses. Therefore, it would also be useful to analyze smaller spatial units or put the analysis in an international context. However, the smaller extent of potential diagnostic variables available at that level of aggregation is a major problem.

References:

Anselin, L. 2003. Spatial Econometrics. In: Baltagi, B.H. eds. A Companion to Theoretical Econometrics. Blackwell Publishing Ltd., Oxford.

Balicki, A. 2009. Statystyczna analiza wielowymiarowa i jej zastosowanie społeczno-ekonomiczne. Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.

Bivand, R. 1980. Autokorelacja przestrzenna a metody analizy statystycznej w geografii. In: Chojnicki Z., eds. Analiza regresji w geografii. PWN, Warszawa-Poznań.

Bivand, R., Pebesma, E.J., Gómez-Rubio, V. 2008. Applied spatial data analysis with R. Springer, New York.

Eurofound. 2017. Sixth European Working Conditions Survey-Overview report (2017 update). Publications Office of the European Union, Available at: https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef1634en.pdf.

European Commission. 2017. Report on equality between women and men in the EU. Available at: https://eeas.europa.eu/sites/eeas/files/2017_report_equality_women_men_in_the_eu_en.pdf.

European Commission. 2010. Communication from the Commission Europe 2020 A strategy for smart, sustainable and inclusive growth, 3.3.2010 COM(2010) 2020 final. Available at: https://www.eea.europa.eu/policy-documents/com-2010-2020-europe-2020.

European Commission. 2018. Report from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the development of childcare facilities for young children with a view to increase female labour participation, strike a work-life balance for working parents and bring about sustainable and inclusive growth in Europe (the "Barcelona objectives"). COM(2018) 273 final. Available at: https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:52018DC0273.

European Parliament. Directorate General for Internal Policies Policy. Department Economic And Scientific Policy Employment And Social Affairs. 2010. The Lisbon Strategy 2000 – 2010. An analysis and evaluation of the methods used and results achieved. Final report. Available at: http://www.europarl.europa.eu/document/activities/cont/201107/20110718ATT24270/20110718ATT24270EN.pdf.

Eurostat. 2019. Statistic date, Available at: https://ec.europa.eu/eurostat/data/statistics-a-z/abc.

Hwang, C.L., Yoon, K. 1981. Multiple Attribute Decision Making: Methods and Applications. Springer-Verlag, Berlin - Heidelberg - New York.

Kopczewska, K. 2007. Ekonometria i statystyka przestrzenna z wykorzystaniem programu R-Cran. CeDeWu, Warszawa.
Kotowska, I. 1990. Starzenie się zasobów pracy w Polsce, Studia Demograficzne, 3(101), 41-61.
Lechman, E. 2015. Analiza zależności przestrzennych między wdrożeniem ICT a wzrostem gospodarczym na przykładzie krajów Ameryki Łacińskiej w latach 2000-2013. Politechnika Gdańska, Gdańsk.
LeSage, J.P., Pace, R.K. 2009. Introduction to Spatial Econometrics. Boca Raton, Taylor & Francis Group.
Młodak, A. 2006. Analiza taksonomiczna w statystyce regionalnej. Difin, Warszawa.
OECD. 2006. Women and men in OECD countries. Available at: https://www.oecd.org/sdd/37962502.pdf.
Pander, W., Rauzer, A., Stawicki, M., Sycz, P., Wojnicka-Sycz, E. 2014. Wyznaczanie, monitoring i ewaluacja inteligentnych specjalizacji. Wydawnictwo Maciej Stawicki, Warszawa.
Panek, T., Zwierzchowski, J. 2013. Statystyczne metody wielowymiarowej analizy porównawczej: Teoria i zastosowania. Oficyna Wydawnicza SHG, Warszawa.
Pułaska-Turyna, B. 2008. Statystyka dla ekonomistów. Difin, Warszawa.
Suchecki, B. 2010. Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych. Wydawnictwo C.H. Beck, Warszawa.