INFLUENTIAL FACTORS OF THE HUNGARIAN TERRITORIAL INCOME

Dóra Szendi

Assistant lecturer, University of Miskolc
regszdor@uni-miskolc.hu

Summary

The distribution of the Hungarian territorial income is not equal across the local administration units (LAU1). There are huge disparities between the different parts of the country. The most developed territories can be found in the central and north-western part of Hungary, until the least developed ones are there in north-eastern Hungary and in the southern Transdanubian region. Nemes Nagy (1998) in his works emphasizes that the territorial inequalities have several aspects and because of this need complex, more-dimensional analysis. From the beginning of the 2000's several researchers analysed the inequalities from various aspects (for example the four component model of Szakálné (2011) for territorial competitiveness).
The aim of this recent research is to examine those significant factors which have a great influence onto the dispersion of the Hungarian territorial income and create optimal regression function. Because of the wide range of the indicators the use of factor analysis was strongly recommended. According to my former results (Szendi, 2015.) there is a significant medium-strong positive spatial autocorrelation in the case of the Hungarian territorial income at level LAU1. This correlation indicates the existence of spatial regression models.

THEORETICAL BACKGROUND

The analysis of the territorial social-economic inequalities is in central role of interest in Hungary (Obádovics et. al., 2011.).That is why I examined some articles related to the Hungarian territorial inequality research.
According to OTK (Hungarian Territorial Development Concept) "the income inequalities across the counties and regions are permanently growing parallel with the economic development. … The differences are increasing most dynamically between the Western and Eastern counties." (OTK, 1998. p. 1.)
Faluvégi (2000) in his work has examined the development types of the Hungarian micro-regions according nine economical-social indicators. These were: capital of the foreign enterprises per capita, income per capita, growth rate of income per capita, number of functioning economic organisations per 1000 inhabitants, growth rate of the number of functioning economic organisations, unemployment rate, migration per 1000 inhabitants, telephone stations per 1000 inhabitants, cars per 1000 inhabitants. He has made a statement that the dynamically developing territories (according the entire nine examined indicator) can be found mostly by the Budapest-Zalaegerszeg axis (including the micro-regions of the Balaton), the Budapest-Győr, the Budapest-Pécs, and Budapest-Miskolc axis. The relatively underdeveloped ones are there in Borsod-Abaúj-Zemplén, Szabolcs-Szatmár-Bereg and Hajdú-Bihar counties.

Kiss and Németh (2006) has made a statementthat in the distribution of the territorial income there exist great differences; the standard deviation of the Hoover index is significant. In the local administration unit of Tiszaújváros the value of the Hoover index is above 20%.
Hahn (2004) made a complex development analysis in which he examined the influential factors of the territorial development. According to him the most significant factors are the geographical position, foreign direct investment, accessibility of motorways, and the secondary and high school education.
Bakoset. al. (2011) has examined the territorial inequalities as a complex process and has made a statement that the migration and immigration are significant influential factors. Beside this the geographical situation, transport opportunities, infrastructural development and the qualification of the labour force are also important factors.
According to my former results there are some statements to consider in this recent analysis. The first statement is that the highest GDP values can be found along some local administrative unit axes in Hungary (Budapest-Miskolc, Budapest-Győr, Budapest-Szeged, Budapest-Keszthely és Budapest-Pécs). Second, there is a medium strong, positive linear correlation between the GDP and life expectancy, the GDP and the number of cars, and the GDP and the volume of house building. There is a strong positive linear correlation between GDP and education and a negative strong correlation between the GDP and the unemployment rate across the Hungarian local administrative units. The third important statement is that there is a medium strong positive spatial auto correlation across the territories in the case of the territorial income. (Szendi, 2015.)

DATA AND METHODOLOGY

The main goal of this recent research is to analyse the influential factors of the Hungarian territorial income. I made the analysis to the 168 local administration units of Hungary with initially 20 involved indicators. To achieve the results the regression analysis is one of the most frequent used technics. The aim of it is to explain the dispersion of a dependent variable (y) with the use of some explaining factors (x_1, x_2, x_3 …).
The linear regression analysis is based on the Ordinary Least Squares method. According to the assumptions there is a linear connection between the depending variable and the explaining factors, and the dispersion from this can be described with an error term (ε) (Szakálné, 2011.). A linear regression model can be represented in the following form:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

In this equation the β_0 parameter gives the value of the depending variable in the x=0 point (when all of the explaining factors are zero). The β_1 regression coefficient gives the value of what effect the 1% change of an indicator causes in the value of the dependent variable (in the case when every other factors are stable).
The validity and adoptability of the linear regression model has several assumptions. The explaining variables have to be:

1. independent from each other – multicollinearity (test: VIF indicator),

2. their values are fixed,
3. they do not contain measurement error,
4. they have no correlation with the error term.

For the error term there are also four assumptions to fulfil:

1. its expected value is zero (which fulfils through the Ordinary Least Squares method),
2. its standard deviation is constant – homoscedasticity (Bartlett, Goldfeld-Quandt test),
3. their values are not auto correlated (Durbin-Watson test),
4. their dispersion is normal (histogram and Chi square test).

In the analysis I used the backward method by the selection of the indicators.In the case of multivariate analysis, if there is a complex connection between the indicators (as it was in this recent analysis), it is practical to reduce the number of indicators in some factor variables. The method for this can be the factor analysis. This is a dimension reduction technic which's adoptability has several criteria. First I had to examine if the dates are appropriate to factor analysis. There are three different technics to do that: the correlation matrix analysis (strong correlation is needed with the dependent variable), anti-image matrix (MSA criterion: value above 0,5) and KMO and Bartlett test (most common method; in the case of Bartlett test the significance is important, and by KMO above 0,5 is acceptable). (Sajtos-Mitev, 2006.)

If the dates are adaptable to factor analysis, thanby determining the number of factors there can be used different methods:

1. a priori criterion: initial decision from the factor number,
2. Kaiser criterion: calculates the number of factors according to their eigenvalues (above 1 is needed),
3. variance share method: describes what share of the initial indicators can explain the new created factors,
4. Scree plot: graphic test, which makes decisions also through the eigenvalues.

To the interpretation of the factors there is a need for rotation of the factors. The varimax, quadratic rotation is one of the most frequent used technics, which I also used in my recent research.

If the values of the neighbouring territories are in connection with each other, so there is a spatial auto correlation, than the OLS estimation can give misleading results. In that case the use of spatial regression is strongly suggested. Two main types of the spatial regression are the spatial lag (which operates with spatially lagged dependent variable) and spatial error model (operates with spatially lagged error term) (Varga, 2002.). The general form of the spatial autoregressive model can be described as the following equation (Gerkman-Ahlgren, 2011.):

$$y = X\beta + \rho W y + u$$
$$u = \lambda M u + \varepsilon$$

whereρ and$\lambda$are scalarspatial autoregressive parameters. There are two special cases, if ρ=0 it is a spatial error model, and if λ=0 it is a spatial lag model. If both parameters are zero than the results of OLS model are correct. (Gerkman-Ahlgren, 2011.)

The spatial lag model contains spatially lagged dependent variable (W_y), and measures the effect of spatial spill over (Anselin-Bera, 1998.). In this model the neighbouring values of the dependent variable have direct effects onto the dependent variable. The general form of spatial lag model:

$$y = \rho W y + X\beta + \varepsilon$$

whereρ (rho) measures the value of the spatial dependency. The positive or negative sign of the rho gives the direction of the influence. If the rho is positive than the territories which neighbours have very high GDP, will also have higher GDP values. (Anselin, 2001.)
In the spatial error model the spatial dependency is also observable but in the case of the error term. It examines the spatial errors, for example missing values. The general form of spatial error model:

$$y = X\beta + u$$
$$u = \lambda W u + \varepsilon$$

whereλ (lambda) is the indicator of the spatially auto correlated error term. If the value of λ is not zero then there is a dependency by the error term across the territories.(Anselin, 2001.)
To decide which model is significant, the use of Lagrange Multiplier test is needed. The significance of the LM-lag and Robust LM-lag tests suggest the validity of spatial lag model, while the significance of the LM-error and Robust LM-error tests suggest the validity of spatial error model. The decision tree of spatial regression can be seen on following Figure 1.

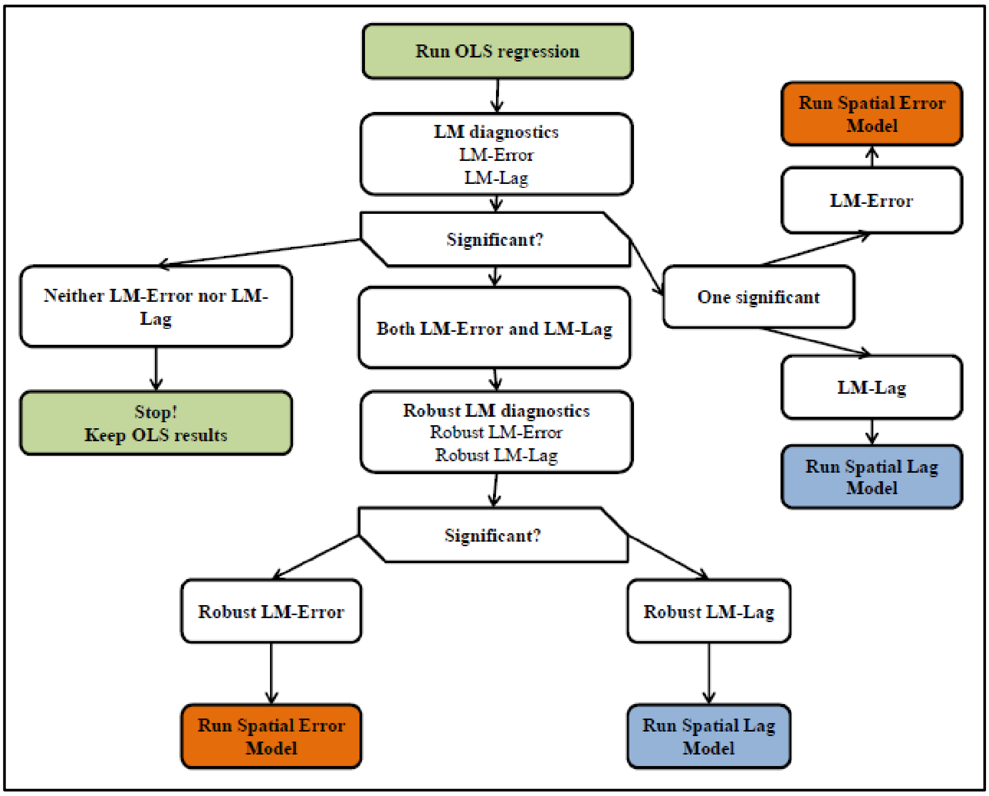

Figure 1. Spatial regression decision tree
Source: author's own calculation according to Anselin (2005)

RESULTS OF MY ANALYSIS

To examine the influential factors of the Hungarian territorial income, first I have built a database of variables which I supposed that can make an influence onto the income distribution. The data sources of my research were the dates of the Hungarian Central Statistical Office, the TEIR database and the Census data of 2011. Initially I have put 20 indicators into the analysis:

- life expectancy by birth,
- educational attainment,
- functioning hospital beds per 1000 persons,
- cars per 1000 persons,
- ground-space of new built housesper 1000 persons,
- unemployment rate,
- traffic accidents on public roads per 1000 persons,
- number of reg. crimes per 1000 persons,
- number of functioning enterprises in the processing industry per 1000 persons,
- number of functioning enterprises employing 10-19 persons per 1000 persons,
- number of functioning enterprises employing 1-9 persons per 1000 persons,
- number of functioning enterprises employing 250-499 persons per 1000 persons,
- number of functioning enterprises employing above 500 persons per 1000 persons,
- number of commercial quarters per 1000 persons,
- number of guest nights per 1000 persons,
- from the inhabitants transferred waste per 1000 persons,
- migration per 1000 persons,
- immigration per 1000 persons,
- number of live births per 1000 persons,
- number of death per 1000 persons,
- number of marriages per 1000 persons.

In the regression analysis I have tested the initial criteria of the error term and the explanatory variables, and have made a statement that in the model the expected value of the error term is zero, its standard deviation is stable (homoscedasticity criterion). The values of the error term are not autocorrelated (tested by Durbin-Watson test). But in the case of the explanatory variables there exists several times disturbing multicollinearity (according to the VIF indicator). Because of this there was a need of dimension reduction.

By the dimension reduction I used a factor analysis. From the database I put out the most correlating four indicators (traffic accidents on public roads per 1000 persons, number of functioning enterprises in the processing industry per 1000 persons, number of functioning enterprises employing 10-19 persons per 1000 persons, and number of functioning enterprises employing 1-9 persons per 1000 persons), and I have checked if the factor analysis can be carried out. The applied methods were the Kaiser-Meyer-Olkin test, the Bartlett test and the anti-image matrix. The basic hypothesis, that there must be at least a 10 times difference between the number of indicators and the number of territories, was realized (168 territories/ 16 indicators = 10,5).

Table 1. KMO and Bartlett test

Kaiser-Meyer-Olkin test		0,764
Bartlett's Test	Chi square	1718,981
	degree of freedom	136
	significance	0,000

Source: author's own calculation

According to the results the factor analysis can be carried out, hence the value of the KMO is 0,764 (which is higher than the 0,5 goalpost), and the significance level of the Bartlett test is zero (Table 1.) The covariance and correlation of the anti-image matrix also underlines the factor analysis (the MSA criterion was realized). I made the analysis with one of the most frequent used method, the Principal components. The eigenvalues, and also the explained variance have suggested a five factor solution (Figure 2.).

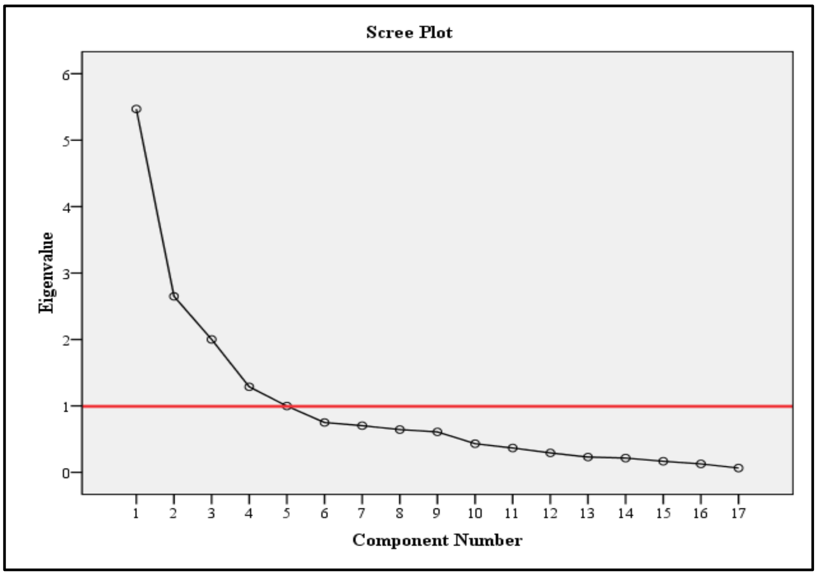

Figure 2. Number of factors based on the eigenvalues
Source: author's own calculation

To calculate the 5 factors I used the Varimax quadratic transformation method. The 5 factors determine 72,96% of the indicators variance (Table 2.). According the rotated component matrix I have calculated the following 5 factors, in which names I used the most dominant indicators.

1. **factor:***life quality* (unemployment rate, cars per 1000 persons, life expectancy by birth, number of marriages per 1000 persons, ground-space of new built housesper 1000 persons, educational attainment),
2. **factor:***tourism* (number of commercial quarters per 1000 persons,number of guest nights per 1000 persons, from the inhabitants transferred waste per 1000 persons),

3. **factor:***economic environment* (number of functioning enterprises employing 250-499 persons per 1000 persons,number of functioning enterprises employing above 500 persons per 1000 persons, functioning hospital beds per 1000 persons, number of reg. crimes per 1000 persons),
4. **factor:***migration* (migration per 1000 persons,immigration per 1000 persons),
5. **factor:***natural reproduction* (number of live births per 1000 persons,number of death per 1000 persons).

Table 2. Explained variance

Total explained variance						
	Initial eigenvalues			Rotated eigenvalues		
Component	Sum	In % of variance	Total %	Sum	In % of variance	Total %
1	5,469	32,171	32,171	4,243	24,960	24,960
2	2,649	15,585	47,756	2,403	14,133	39,093
3	2,001	11,768	59,524	2,108	12,398	51,491
4	1,287	7,571	67,095	1,874	11,023	62,514
5	0,997	5,868	72,962	1,776	10,448	**72,962**
6	0,750	4,410	77,372			
7	0,702	4,132	81,504			
8	0,644	3,786	85,290			
9	0,609	3,581	88,870			
10	0,431	2,535	91,406			
11	0,367	2,156	93,562			
12	0,292	1,720	95,282			
13	0,229	1,350	96,632			
14	0,214	1,262	97,893			
15	0,167	0,980	98,873			
16	0,127	0,747	99,620			
17	0,065	0,380	100,000			
Principal components method.						

Source: author's own calculation

I made the regression analysis of the Hungarian territorial per capita income using the above mentioned five factors. The regression function fulfils the requirements of the error term and explaining variables, there is no disturbing multicollinearity. According the Durbin-Watson test the error term is not auto correlated, and the significance of F statistic proves the reliability of the model. (Table 3.)

Table 3. Dates of the model, ANOVA and coefficients

Model	R	R^2	Corrected R^2	Standard error	Durbin-Watson
1	0,844[a]	0,713	0,704	156,94509	1,898

	Sum of squared	degree of freedom	F-statistic	Significance
Regression	9962287,9	5	80,890	,000[a]
Residuals	4014977,2	163		
Total	1,398E7	168		

a. Explaining variables: constant, natural reproduction, migration, economic environment, tourism, life quality factors

b. Depending variable: territorial income

	Unstandardized coefficients		Collinearity	
	B	Standard error	Tolerance	VIF
constant	1603,675	12,073		
life quality	215,925	12,109	1,000	1,000
tourism	-6,514	12,109	1,000	1,000
economic environment	78,873	12,109	1,000	1,000
migration	15,359	12,109	1,000	1,000
natural reproduction	78,590	12,109	1,000	1,000

Source: author's own calculation

Beside these the model also fulfils the requirements of the normality, which is represented on Figure 3.

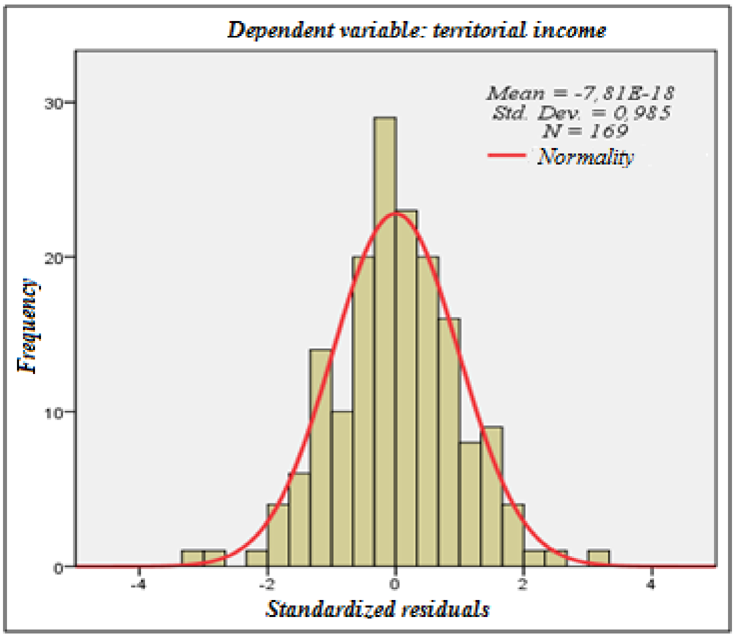

Figure 3. Normality of the territorial income
Source: author's own calculation

The optimal regression function can be determined by the following equation:

$$\boldsymbol{Y = 1603,675 + 215,925X_1 - 6,514X_2 + 78,873X_3 + 15,359X_4 + 78,59X_5}$$

It means that if every factors are zero, then the territorial income per capita is 1603,675 thousand Forint. If the life quality indicator increases by 1% then the territorial income per capita will be increasing also, with 215,9 thousand Forint (in the case when every other factors remain stable). As an effect to the increase of the economic environment and the natural reproduction factor the territorial income will be growing with about 78 thousand Forints. The positive effect of migration is relatively lower, while the tourism factor makes a negative effect onto the territorial income per capita.

According the five factors the following Figure 4.can be described. In the case of life quality factor there is a developed northern - north-western zone. Beside this there is observable a highly developed Budapest - Győr – Sopron, Budapest – Szeged, Budapest – Pécs, Budapest – Miskolc axis, and also the micro-regions of the Balaton area show this characteristics.

In the case of tourism factor the micro-regions of the most visited tourism attractions appear like hot spots in the space. For example the dominance of the Balaton-area, Tokaj-Hegyalja, Eger, Debrecen, Székesfehérvár and Szombathely is huge.

The factor map of the economic environment is more heterogenic, the most developed territories can be found in the area of the county centres and in the bigger towns which are centres of greater enterprises.

According to the results of migration factor two main paths can be described. First there are huge values near the country borders, where both directions of the migrationare observable. In the case of Northern Hungarian and Southern Transdanubian small villages the migration is relatively dominant. The case of the Budapest agglomeration shows the tendencies of suburbanisation.

By the natural reproduction factor the highest values can be found in the Budapest agglomeration and the northern – north-eastern part of the country.

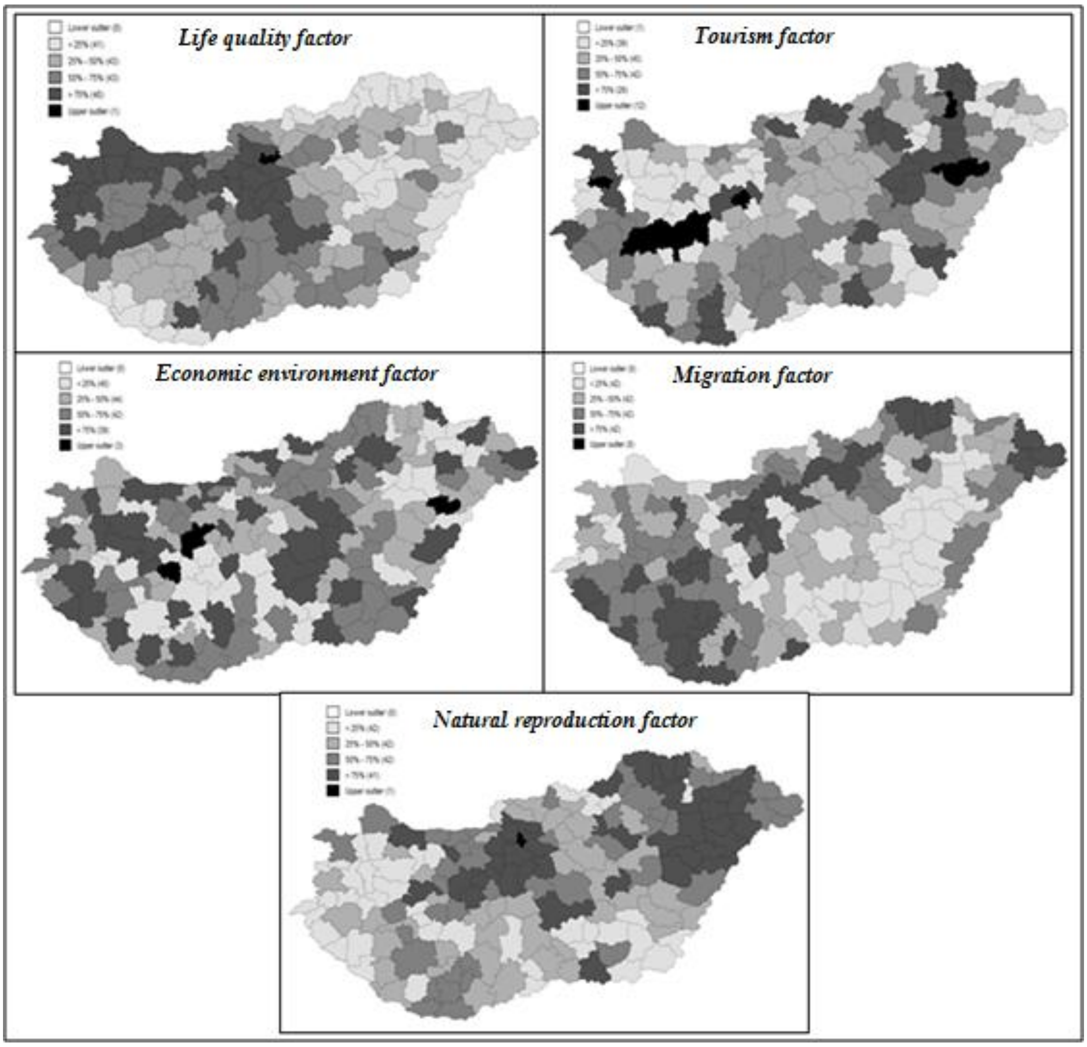

Figure 4. Factor maps related to the Hungarian territorial income
Source: author's own calculation

The spatial effects can play important role also by the regression analysis that is why I examined the role of neighbourhood effects in the case of territorial income. By the classic OLS (Ordinary Least Squares) regression I have built the neighbourhood matrix into the model to examine the validity of spatial models. My question was whetherone of the spatial models (spatial lag – lagged value of the dependent variable or spatial error – lagged error term) is validin the case of Hungarian territorial income. The parameters of spatial regression are summarized by the following Table 4.

Table 4. Spatial regression test – Hungarian territorial income

Test	Moran I/ degree of freedom	Value	Significance
Moran I	0,3747	8,4344	0,0000
Lagrange Multiplier (lag)	1	51,9052	0,0000
Robust LM (lag)	1	7,9929	0,0047
Lagrange Multiplier (error)	1	60,5130	0,0000
Robust LM (error)	1	16,6006	0,00004
Lagrange Multiplier (SARMA)	2	68,5058	0,0000

Source: author's own calculation

According the above both of the LM spatial lag and error models are significant, so the validity of spatial model can be proved. The type of the spatial regression is determined by the significance of Robust LM tests (spatial regression decision tree, Figure 1.).
As both of the Robust LM tests are significant, I have chosen the model with higher significance level. According to this in this case of the Hungarian territorial income the spatial error model can be verified, which contains a lagged error term. I made the analysis of the spatial error model, and got the following results compared to the OLS.

Table 5. Results of OLS and spatial error model – Hungarian territorial income

Coefficients	OLS	Spatial error
constant	1603,675**	1607,922**
life quality	215,925**	197,21**
tourism	-6,514	20,72
economic environment	78,873**	82,94**
migration	15,359	-4,39
natural reproduction	78,590**	42,81**
Lambda	-	0,7276**
R^2	72,96	82,66
Akaikecriterion	2182,59	2119,88
Schwarz criterion	2201,33	2138,62

** significant (p<0,001)
Source: author's own calculation

The spatial error model can better explain the dispersion of the Hungarian territorial income compared to the OLS, because the explanatory power of the spatial error model is 82,66%, while the OLS only 72,96%. (Table 5.)In the spatial error model the value of the Akaike and Schwarz information criterion is lower which also underlines the validity of the model.
According to the spatial error model the territorial income is higher compared to the OLS (1608 thousand Forint) in the case if every other factors are zero. The 1 % increase of the life quality factor has a positive effect on the income, it increases the income by 197,21 thousand Forints. The tourism, economic environment, and natural reproduction factors make positive effects on the income while the change of the migration factor causes a decrease in the income (which is more closed to the reality). The coefficient of the spatial error term (lambda) is 0,7276 which means that it has a significant, positive effect. The spatial error model can be described as follows:

$$\boldsymbol{Y = 1607{,}922 + 197{,}21X_1 + 20{,}72X_2 + 82{,}94X_3 - 4{,}39X_4 + 42{,}81X_5 + \varepsilon}$$
$$\boldsymbol{\varepsilon = 0{,}7276\, W\varepsilon + \xi}$$

whereε is the error term, W matrix of the neighbourhood connections, andξvector of uncorrelated error terms.

CONCLUSION

In my recent research I have examined the influential factors of the Hungarian territorial income. As a result of the analysis I have calculated a model where the distribution of the territorial income is explained by five complex factors and in the model also the spatial connections have significant effects.The five influential factors are life quality, tourism, economic environment, migration and natural reproduction. It is remarkable that by most of the factors there is observable a highly developed Budapest - Győr – Sopron, Budapest – Szeged, Budapest – Pécs, Budapest – Miskolc axis. Beside this also the micro-regions of the Balaton area show great development tendencies.In the case if every influential factors are zero than the value of the territorial income is 1608 thousand Forint. The greatest positive effect makes the life expectancy factor onto the income distribution, while the only one factor which has negative effect is the migration. In the case of the Hungarian income the spatial error model can be proved. This model can explain through the above mentioned factors more than 82% of the Hungarian territorial income distribution.The coefficient of the spatial error term (lambda) is 0,7276 which indicates significant, positive effect.

LIMITATION

Further on I would like to expand this research to other countries (for example Germany), to analyse the most influential factors of territorial GDP. In my opinion to adopt the model for other countries it has some limitations. First, the applied indicators are varying across countries, and time periods. The indictors of this model are fitted only for the Hungarian territorial income in 2011. Other limitation can be found in the applied database; hence the availability of the indicators at territorial level is very limited. (Probably there are other significant variables influencing the territorial income, but the dates availability is not ensured.) The third relevant limitation can be found in the methodology of the spatial regression, because I suppose that the border regions of the neighbouring countries might have an influence onto the results of regression, so a cross-regional map would be useful. So before expanding the model it is important to analyse these limitations.

REFERENCES

ANSELIN L. (2001): SpatialEconometrics*Acompaniontotheoreticaleconometrics*, pp. 310-330.

ANSELIN L. (2005): *ExploringSpatial Data withGeoDaTM: A Workbook*Center forSpatiallyIntegratedSocial Science

ANSELIN L. &BERAA. K. (1998): SpatialdependenceinLinearRegressionModelswith an IntroductiontoSpatialEconometrics*Statisticstextbooks and monographs*, Vol. 155., pp. 237-290.

BAKOS N., HIDAS ZS. &KEZÁNA. (2011): Területi különbségek Magyarországon A főbb társadalmi és gazdasági folyamatok az ezredforduló után *Területi Statisztika*, Vol. 1.(51.)., Nr. 4., pp. 335-357.

FALUVÉGI A. (2000): *A magyar kistérségek fejlettségi különbségei*Előadás a Magyar Statisztikai Társaság Területi Statisztikai Szakosztálya Vándorülésén, Balatonöszöd.

GERKMAN L. &AHLGREN N. (2011): *PracticalProposalsforSpecifyingk-NearestNeighboursWeightsMatrices*HankenSchool of Economics, WorkingPapers 555.
HAHN CS. (2004): A térségi fejlődést befolyásoló tényezők Magyarországon *Területi Statisztika*, Vol. 6., pp. 544-563.
KISS J. P. & NÉMETH N. (2006): *Fejlettség és egyenlőtlenségek, Magyarország megyéinek és kistérségeinek esete* Budapesti Munkagazdaságtani Füzetek, BWP – 2006/8
NEMES NAGY J. (1998): *A tér a társadalomkutatásban*Hilscher Rezső Szociálpolitikai Egyesület, Budapest.
OBÁDOVICSCS.,BRUDER E. & KULCSÁR L. (2011): A gazdasági és szociális helyzet területi egyenlőtlenségei a vidéki Magyarországon – hasonlóságok és eltérések *Gazdaság és Társadalom* 2011/2.
ORSZÁGOS TERÜLETFEJLESZTÉSI KONCEPCIÓ (1998)
SAJTOS L. &MITEVA. (2006): *SPSS Kutatási és adatelemzési kézikönyv*Alinea Kiadó, Budapest
SZAKÁLNÉ K. I. (2011): A gazdasági aktivitástérbeli eloszlásának vizsgálati lehetőségei*Statisztikai Szemle,* vol. 89. (1), pp. 77-100.
SZENDI DÓRA: Az életminőség, mint komplex indikátor területi differenciái a német és magyar gazdaság esetében; Fiatal Regionalisták IX. Konferenciája, Győr, 2015. – *underpublishing*
VARGA A. (2002): Térökonometria*Statisztikai szemle,* 80. évf. 4. sz., pp. 354–370.