Doctoral School of Economics, Business and Informatics

Collection of Theses

Zoltán Madari

Examining the development of Hungarian LAU1 districts using spatial econometric and panel econometric methods

for his Ph.D. thesis

Supervisor:
Keresztély Tibor, Ph.D
Associate Professor

Budapest, 2024

Institute of Data Analytics and Information Systems, Department of Statistics

Collection of Theses

Zoltán Madari

Examining the development of Hungarian LAU1 districts using spatial econometric and panel econometric methods

for his Ph.D. thesis

Supervisor:
Keresztély Tibor, Ph.D
Associate Professor

Contents

1. Research history and justification of the topic

My research focuses on the spatial and temporal evolution of the development of districts.I have complete data tables for districts for the period 2012-2020. My first task is to select the necessary variables, and I will therefore review relevant national and international literature. An important point of the thesis is that I not only investigate the development, but also the variables themselves in detail using descriptive statistical tools, data visualisation and the calculation of spatial autocorrelation. After identifying the necessary variables, I define development as a latent variable using structural equations. I explore spatial relationships using spatial econometric methods and indicators.

Gáspár (2013) takes a chronological look at the measurement systems of socio-economic development. The first indicator was GDP, and then factors such as education, health and the environment were added. Another way is to adjust gross national income. One such initiative was net economic welfare. The concept was introduced by Nordhaus and Tobin (1972). The modified indicator took into account leisure time, own-account work and environmental considerations. The third large group is the composite index. This includes the HDI, described in the introduction to this chapter, which is perhaps the best known index. Most of the composite indices are mainly related to life expectancy and quality of life.

I base my definition and framework of development on the work of Harcsa (2015). Development is a multi-level social and economic phenomenon. Economic, labour market, demographic, knowledge-capital dimensions are included in the measurement and definition.

According to Nemes-Nagy (2005), development is a complex concept, difficult to measure. Variables need to capture the phenomenon, but the calculation needs to be transparent. Avoid mixing absolute and relative indicators.

The primary reference for the calculation is the Central Statistical Office's complex district development indicator. In addition, a number of empirical studies have been taken into account. From these I got an idea of the commonly used variables and methods (Tóth et al., (2014), Pénzes (2015), Győri and Mikle (2017), Fertő and Varga (2014), Bella and Kazimir (2021)). The number of indicators ranges from 20-30 variables to hundreds. Tóth (2024) used 6 different methods to measure the phenomenon of territorial well-being. Based on his calculations, only the indicator of district economic power gave different results. The other indicators were almost identical. Based on the recommendations in the literature, in this research I develop a multidimensional indicator that uses fewer indicators than previous indicators. I combine the cross-sectional and time-series dimensions into a panel structure. I analyse the spatial patterns and changes not only of the development indicator but also of its components. The methodology differs from traditional dimension-reduction averaging-based indices. Similar to the study by Bella and Kazimir (2021), I use a structural equations modeling.

1.1. Research questions

The study seeks to answer 3 research questions.

- Can a latent variable measuring district-level development be created using a structural equation model?

I expect that it can. The aim of the research is to create an indicator based on a small number of variables using SEM. The method is suitable to explore the logical and causal relationship between variables. Part of the validation process is to compare the values of the latent variable with the values of the complex indicator of the KSH.

- What spatial relationships and patterns can be identified when examining the development of LAU1 districts?

I expect to obtain a spatial clustering and heterogeneous picture for both the indicators and the latent variable, over the study period. My assumption is that Central Hungary and the northern part of the Transdanubian region have a high level of development, while the eastern and north-eastern part of the country has a low level of development.

- How did the gap in the level of development of the LAU1 districts change between 2012 and 2020?

I also perform dynamic analyses. My expectation is that for the districts, development levels have converged over the study period. In underdeveloped areas, catching up has started.

1.2. Complex indicator of Development

The calculation of the complex indicator of district development and the set of variables used for it is provided for in Government Decree 290/2014 (26.XI.). As a first step, 4 major sets of indicators have been defined. A total of 23 variables were classified into these. The methodology of the indicator is based on normalisation. The normalised variables are averaged by major group and then the 4 group averages are used to form the main average, which is the district-level complex development indicator.

Districts are categorised on the basis of this indicator. These categories are in ascending order of development: to be developed with a complex programme, to be developed, beneficiary, non-beneficiary. All districts with a development value above the average are classified as non-beneficiary.

2. The methodology and data used

The methodology consists of three parts. The first is the panel analysis. This is briefly presented. The panel structure combines the cross-sectional and time-series dimensions. If you just put the two dimensions together, you get the so called pooled structure. The structure is evolvable, both time and individual effects are identified. This way we can obtain fixed effect or random effect panel models.

2.1. Structural Equation Modeling

Formal description of the model is based on Jöreskog et al (2016) page 344.

$$\acute{\eta}=\alpha+B\acute{\eta}+\Gamma\xi+\varsigma,$$

where α is the intercept, β and Γ are coefficient matrices, and ς is the vector of random error. ή is the vector of latent dependent variable, while ξ is the latent independent variable. Γ shows the direct effect of latent independent variables on latent dependent variables. B shows the effect of latent dependent variables on each other. The error term and the latent explanatory variables are independent.

There may be observed variables x and y. These are not latent variables, they are not determined within the model.

$$y=\tau_y + \Lambda_y ή + \varepsilon \text{ és } x=\tau_x + \Lambda_x ή + \delta$$

These are simple linear regressions where the two error terms are uncorrelated with the latent variables.

The latent variables in the analysis will be the components (demographics, labour market, etc.). The observed variables are the indicators related to these components. Factors will be used to form proxy variables for the components, which will then be used by the model to perform parameter estimation. The existing parameters are used to estimate a complex indicator of development.

2.2. Spatial methods

The spatial methods are based on the work of Varga (2002) and Dusek (2004). I calculate two important indicators. These are the Moran I statistic and the Geary C indicator, which measure spatial autocorrelation.

$$[N/S_0]\ [\Sigma_{i,j}\, w_{ij}\, (x_i - \mu)(x_j - \mu)\ /\ \Sigma_i (x_i - \mu)^2]$$

Moran I statistics is a method similar to correlation. The difference is that it defines a spatial weight matrix. This is based on the distance between territorial units. It is important to use a row-standardised weight matrix.

$$\frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(X_i - X_j)^2}{2W_{ij}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Geary C is a neighbourhood quotient. It measures how similar neighbouring areas are to each other. Here again, a row-standardised weight matrix is recommended.

It is important to clarify the issue of distance measurement. According to Dusek (2004), there are several possibilities: k nearest neighbour, common boundary neighbourhood, first degree/second degree neighbourhood, neighbourhood by distance of points in a territorial unit. In my study, I use the latter. For different distances, I examine neighbourhood defined by the centre of the districts and the geographical centre of the districts.

2.3. Data

For the research I used the KSH information database and the Map Interactive Display application. In accordance with the literature, I grouped the 13 selected variables into dimensions, on the basis of which I modeled spatial development. The study period covers the years 2012 to 2020. The cross-sectional dimension includes 174 rural districts excluding Budapest.

Notation	Variable	Dimension
Y1	Proportion of older people in total society	
Y2	Live births per 1000 inhabitants	Demography
Y3	Inward migration margin per 1000 inhabitants	
X4	Job seekers as a percentage of the working age population	Labour market

X5	Personal income taxable income per resident	
Y6	Internet subscriptions per thousand permanent residents	Infratructure and transport
Y7	Length of road per 100 square kilometres	
Y8	Number of cars per thousand inhabitants	
Y9	Percentage of dwellings connected to public drinking water supply	
Y10	Percentage of dwellings connected to public sewerage	Environment, green factor
Y11	Separately collected municipal waste per capita	
X12	Number of active enterprises per thousand inhabitants	Economic factor
X13	Number of retail shops per 10 000 inhabitants	

Table 1: Variables and dimensions of the development indicator

3. Scientific results of the thesis

I report the results by answering the research questions.

Can a latent variable measuring district-level development be created using a structural equation model?

I used 13 variables from different economic and social dimensions for the SEM estimation. It was not possible to build a suitable model for 5 dimensions. Based on logical and possible causal relationships between the variables, I created one economic and one development latent variable.

Latent	Variable	Original	Standardized	Standardized
Economic	X4	1.000	1.000***	1.000
	X5	5947.285	-1.077***	-1.086***
	X12	236.840	-1.164***	-1.414***
	X13	-19.526	-0.401***	-0.524***
Development	Y1	1.000	1.000***	1.000 ***
	Y2	-0.565	-1.342***	0.205***
	Y3	4.448	2.568***	0.079***
	Y6	60.010	3.858***	-0.897***
	Y7	19.230	2.141***	0.322***
	Y8	64.478	3.855***	-0.494***
	Y9	2.288	1.170***	0.169***
	Y10	12.487	2.603***	0.095***
	Y11	21.017	2.385***	-0.185***
	Gazdasági	0.055	-3.522***	-0.444***
	Y_2013			0.312***
	Y_2014			0.324***
	Y_2015			0.332***
	Y_2016			0.345***
	Y_2017			0.327***
	Y_2018			0.338***
	Y_2019			0.326***
	Y_2020			0.301***
Comparative Fit Index		0.504	**0.721**	0.295
Tucker-Lewis Index		0.395	**0.660**	0.213
RMSEA		0.251	**0.188**	0.215

Table 2: Parameter estimation of SEM models

Of the models run on original variables, standardised variables and standardised variables with time fixed effects, the second one was the best on the basis of model diagnostics. Unfortunately, none of the commonly used model qualification indicators (CLI, TLI and RMSEA) reached the required level. Therefore, the latent variable was compared with the KSH complex development indicator. This validates the latent variable result, with a correlation between the two indicators of 0.952. There is also a high degree of overlap in the classification of the 4 development categories. Thus, I have accepted the latent variable as an indicator of territorial development. I would like to point out that the methodology used allowed to describe development with only 13 variables. This is very low compared to the number of variables reported in the literature. In the complex development indicator, 23 indicators were used. Latent variable and SEM estimation can capture and describe development in a similar way with 13 variables. So relevant number of redundant variables can be filtered out.

What spatial relationships and patterns can be identified when examining the development of LAU1 districts?

In my work, I have deliberately sought to use variables not only to create an index or indicator, unlike previous literature. I wanted to analyse all components of the development indicator.

In the case of demographic indicators, the problem of an ageing society is also clearly visible at district level. The proportion of elderly people has increased in almost all districts. Some districts had an average annual growth rate of over 5%. The spatial distribution cannot be considered as random, significant positive autocorrelation can be identified. A slight improving trend can be identified for the birth rate. However, the Moran I shows moderate and positive autocorrelation, it has increased significantly over the study period. In the case of the migration differential, the trends are clearly marked. Migration was directed towards the developed districts.

In terms of labour market trends, the national labour shortage is heterogeneous across districts. However, this has had a steadily decreasing impact on the value of spatial autocorrelation until 2019. As regards incomes, they have risen significantly in all areas of the country over the observed period, but the differences between districts have not decreased.

For internet access, autocorrelation first increased and then significantly decreased from the middle of the study period. The low-development districts started to catch up. There was no reordering for cars. The indicator improved for all districts, but regional differences did not decrease.

A tourism impact could be identified for separately collected waste. The variable is the main indicator of environmental awareness. The spatial autocorrelation behaved very hectically. Small decreases and increases alternate.

In terms of economic factors, there was an improving trend for all districts in the case of active enterprises, but the districts improved compared to their own level, with no catching-up. In the case of retail shops, there was almost no spatial autocorrelation in the initial period, the distribution of districts was completely random. However, the weight of traditional retailing started to decrease with online ordering and retailing, resulting in a weak positive autocorrelation by the end of the observation period.

In the case of the district development indicator I have calculated, the processes described above are summarised. The latent variable is typically high in Central Hungary (mainly Pest county) and in the northern part of Transdanubia. The southern and eastern border districts and the north-eastern areas are the lagging districts. This suggests that the spatial autocorrelation takes a positive value. The spatial autocorrelation is positive and moderate according to the Moran I statistic and the Geary C indicator.

How did the gap in the level of development of the LAU1 districts change between 2012 and 2020?

Looking at the changes over time, we see slightly lower values for 3 years after 2012, and then a significant increase in the Moran I statistic from 2016 onwards. The indicator reaches its highest values in 2020. In all periods there is a moderately strong positive spatial autocorrelation. The spatial autocorrelation did not decrease during the observed period, and the indicator reached its highest value at the end of the period. A similar picture emerges for the Geary C indicator. There is a slight weakening between 2013 and 2015, followed by a slight increase in autocorrelation from 2016 onwards. The strongest value is again reached in 2020. This suggests that for the latent variable, development, the spatial differences have not changed substantially over the observed period.

The descriptive statistics for the latent variable also paint a similar picture. Looking at the indicators, it can be seen that the value of development follows a nearly symmetric distribution. It can be seen that the median and mean values are almost identical in each year. For both indicators a continuous increase can be seen. The level of development has increased.

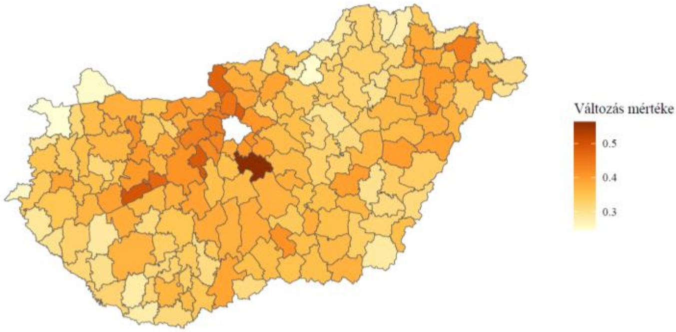

Figure 1: Absolute change in the development index between 2012 and 2020

This is also shown on the map describing the absolute change, with no decrease in any of the districts. Very similar patterns are shown by the deviation indicators as the Moran I statistic and Geary C indicator. After 2012, the range decreased, so the gap between the most and least developed districts decreased. This trend reversed after 2016, turning into an increase. In 2020, both the interquartile range and the total range are larger than in 2012. So the gap between low and high developed districts has increased somewhat. This is definitely a negative change. During the observed period, the level of development increased mainly in the economically strong areas of Pest county, the Budapest agglomeration and the central part of Transdanubia, around Lake Balaton. The lagging districts could not keep pace with these areas in terms of development growth. Thus, by the end of the study period, the development levels of the districts had not converged, but the gap between them had slightly widened.

4. Main references

Bella, K. M., - Kazimir, I. R. (2021). A structural Equation Model for Measuring Relative Development of Hungarian Counties in the Years 1994–2016. Journal of Official Statistics (JOS), 37. évf. 2. szám, 261-287. o. DOI: http://dx.doi.org/10.1515/JOS-2021-0012

Dusek, T. (2004): A területi elemzések alapjai. Regionális Tudományi Tanulmányok 10., ELTE Regionális Földrajzi Tanszék, Budapest. ISSN 1585-1419

Fertő, I. - Varga, Á. (2014). A jóllét területi különbségei Magyarországon: egy lehetséges térségfejlettségi index alkalmazása. Statisztikai Szemle, 92. évf. 10 szám, 874-891. o.

Gáspár, T. (2013): A társadalmi-gazdasági fejlettség mérési rendszerei. Statisztikai Szemle, 91. évf. 1. szám, 77-92. o.

Győri, R. – Mikle, Gy. (2017). A fejlettség területi különbségeinek változása Magyarországon, 1910–2011. Tér és társadalom, 31. évf. 3. szám, 143-165. o. DOI: https://doi.org/10.17649/TET.31.3.2866

Harcsa, I. (2015a): A területi fejlettség és egyenlőtlenségek lehetséges értelmezései - kritikai értékelés és kutatási eredmények I.. Statisztikai Szemle, 93. évf. 5. szám, 460-480. o.

Jöreskog, K. – Olsson, U. – Wallentin, F. (2016). Multivariate Analysis with LISREL. Springer International Publishing, Svájc.

Nemes-Nagy, J. (2005): Összetett jelenségek, osztályozás, regionalizálás. In: Nemes-Nagy, J. (szerk.): Regionális elemzési módszerek. Regionális Tudományi Tanulmányok. 11. Eötvös Loránd Tudományegyetem. Budapest. 169–215. o.

Nordhaus, W. – Tobin, J. (1972): Is Growth Obsolete?. Economic Research: Retrospect and Prospect, 5. évf. 1-80. o.

Pénzes, J. (2015): A kedvezményezett térségek lehatárolásának aktuális kérdései. Területi Statisztika, 55. évf. 3. szám, 206-232. o.

Tóth, G. (2024): A társadalmi innovációs potenciál és a területi jóllét számszerűsítésének lehetőségei. Hungarian Statistical Review/Statisztikai Szemle, 102. évf. 7. szám, 679-713. o. DOI: https://doi.org/10.20311/stat2024.07.hu0679

Tóth, I. J. (2014): Fejlődő és leszakadó járások – Magyarország társadalmi-gazdasági profilja, MKIK műhelytanulmány

Varga, A. (2002): Térökonometria. Statisztikai szemle 80. évf. 5. szám, 354-370. o.

5. Publications related to the topic

Máténé Bella, K., - Madari, Z. (2020): Differences in elders' habits among European countries. KÖZGAZDASÁG, 15. évf. 2. szám, 108–115. o.

Cserháti, I., - Keresztély, T., - Madari, Z. (2020): A jövő munkahelyével kapcsolatos elvárások/várakozások alakulása. In Munkapiaci trendek – demográfiai és jóléti kihívások. 281–297. o.

Keresztély, T., - Madari, Z. (2021): A háztartások eltitkolt jövedelmének becslése inverz keresleti függvénnyel. KÖZGAZDASÁGI SZEMLE, 68. évf. 7–8. szám, 736–752. o.

Madari, Z. (2021): A magyar élelmiszeripari KKV szektor helyzete és kilátásai. KÖZ-GAZDASÁG, 16. évf. 1. szám, 39–49. o.

Hartvig, Á.D. et al. (2023): A digitalizáció hatása a vállalati hozzáadott értékre Magyarországon.

KÖZGAZDASÁGI SZEMLE, 70. évf. 6. szám, 672–689. o.

I have also been a topic leader in the field of spatial analysis for 4 years. I have been supervisor for more than 10 institutional TDK and OTDK prize-winning theses. Two of my students have won Pro Scientia gold medals, Áron Dénes Hartvig in 2021 and Petra Eszter Török in 2023.