REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

BALÁZS LENGYEL[a] – LOET LEYDESDORFF[b]

The Effects of FDI on Innovation Systems in Hungarian Regions: Where is the Synergy Generated?

Abstract

In this study, we show how internationalization and foreign-owned firms influence synergies in the regional innovation systems of Hungary. We first distinguish three innovation system functions (knowledge exploitation, knowledge exploration, and organizational control) operating in regions and study their interactions using entropy statistics. The functions and their interactions are measured by analysing the distribution of firms in terms of geographical location, organizational size (number of employees), technologies (NACE codes of the OECD), and ownership (foreign versus domestic share in registered stock) in the 2005. Synergy is defined as mutual information among the three dimensions; a fourth dimension is added in order to bring internationalization (FDI) into the model. The factor is relevant since the four-dimensional model explains the GDP contributions to regional development in Hungary, whereas the three-dimensional model does not. We find that regional innovation systems in Hungary are self-organized differently, in relation to a relatively small number of foreign firms. These firms have a large positive effect on synergy in regions between the Hungarian capital and the Austrian border. However, FDI has negative effects on domestic synergy in the lagging eastern and southern provinces of the country.

Keywords: regional innovation systems, innovation system function, synergy, entropy, foreign firms.

Introduction

Regional innovation systems (RIS) have been analysed in order to answer policy-oriented questions such as why innovation output, technological change, and consequently regional development differs across regions (Boschma–Frenken 2006, Cooke et al. 2004 Martin–Sunley 2007). The RIS literaturemainly refers to local institutions and milieu (Lundvall et al. 2002, Tödtling–Tripple 2005), local and sticky versus global and ubiquitous knowledge (Asheim–Isaksen 2002), university-industry collaborations (Etzkowitz–Leydesdorff 2000), and governance (Braczyk et al. 1998, Cooke 2001). One assumes that a well-functioning RIS provides proper grounds for synergistic co-operation between co-located organizations, firms, etc. (Cooke–Leydesdorff 2006).

a) Centre for Economic and Regional Studies, Hungarian Academy of Sciences, H-1112 Budapest, Budaörsi út 45., Hungary and International Business School Budapest, H-1031 Budapest, Záhony utca 7, Hungary. E-mail: lengyel.balazs@krtk.mta.hu

b) Amsterdam School of Communication Research, University of Amsterdam, 1001 NG Amsterdam, P.O. Box 15793, The Netherlands. E-mail: loet@leydesdorff.net

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Yet, it was repeatedly pointed out that multinational companies can be crucial players in regional systems (Biggiero 2007, Cantwell–Iammarino 1998, Dachs et al. 2008, Ferragina–Mazzotta 2014, Majumdar 2009). This is especially the case in Central and Eastern European countries that rely extensively on the performance of such companies (Inzelt 2008, Radosevic 2002). Therefore, in a country like Hungary where the vast majority of exports are carried by multinational companies, it is crucial to understand the specific role of multinational companies in regional economic development (Acs et al. 2007). The central question of this study concerns the effect of foreign-owned companies on regional innovation systems in the Hungarian economy.

Some argue that foreign firms were general sources of knowledge spillover and their presence generated positive local externalities beneficial to domestic firms in Hungary (Halpern–Muraközy 2007). In this sense, FDI can be considered as the major engine of regional development (Lengyel–Szakálné 2014). However, other papers (Békés et al. 2009) have shown that spillover effects have occurred only among foreign-owned companies and those Hungarian domestic firms that previously performed at relatively higher levels of productivity. Thus, relatively more developed environments might have benefited more from the presence of foreign companies and less spillover can be expected from foreign firms in lagging regions (Lengyel–Leydesdorff 2011).

In this study, we calculate the effect of foreign firms on synergy in RIS using indicators derived from complex systems theory and entropy statistics (Leydesdorff et al. 2006, Leydesdorff–Fritsch 2006, Ulanowicz 2006 and 2009). A previous study focusing on Hungary and using the same methodology found regionally diverse economies, in which the capital, the western regions and eastern regions follow different development paths (Lengyel–Leydesdorff 2011). As a question for further research, we proposed the study of the role of FDI in causing this regional divergence. In this study, we had access to a new database of better quality, and used this for the four-dimensional extension of the synergy indicator in order to quantify the effects of foreign-owned firms.

We find that a North-West South-East divide prevails concerning the effect of foreign firms on the synergy in regions. Foreign firms have positive effects on the synergy in innovation systems only in some relatively developed regions; meanwhile foreign firms disturb the synergy in lagging regions. The results suggest that the innovation systems of relatively developed areas have been re-structured by foreign-owned companies that generated new synergies in the regions. However, lagging regions are found to be self-organized in terms of domestic firms; positive externalities of FDI might not prevail in these latter locations.

Theoretical background

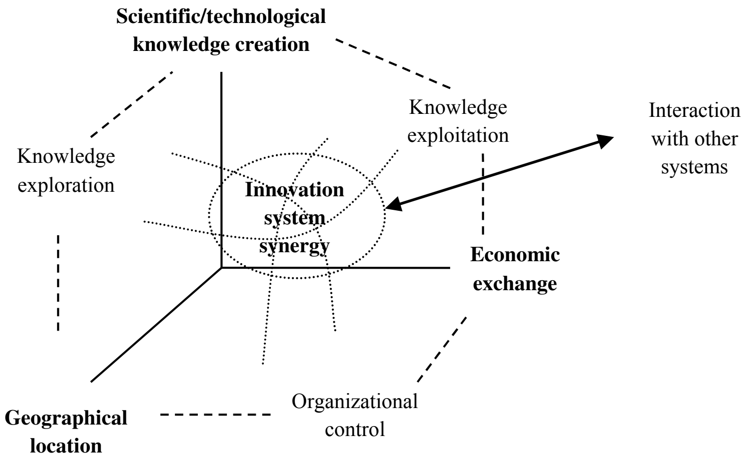
Synergy measurement in regional innovation systems derived from the triple helix model of university–industry–government collaboration has led to heuristics when looking at regional economies (Leydesdorff et al. 2006, Leydesdorff–Fritsch 2006, Strand–Leydesdorff 2013, Leydesdorff–Strand 2013, Leydesdorff–Sun 2009). The RIS synergy model follows Storper's (1997, at pp. 26 ff.) "holy trinity" theory. This argument emphasizes that interrelationships among three independent dimensions–technology, organization, and territory–shape regional economies. The three independent dimensions

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

are depicted in Figure 1, but the organizational dimension of the "holy trinity" is replaced with economic exchange relations in our model (Lengyel–Leydesdorff 2011). The reason for including the category of economic exchange instead of organization, is that the interactions with other firms and the costs of this interaction influence the size and scope of the firm (Coase 1937). We use the latter two characteristics as proxies in our model.

Figure 1

Synergy of innovation systems functions



Source: adapted from Lengyel and Leydesdorff (2011, p. 681).

In Figure 1, we consider the first-order interaction terms between geographical location, economic exchange, and knowledge creation as three innovation system functions: (*i*) knowledge exploration, (*ii*) knowledge exploitation, and (*iii*) organizational control. These functions can be defined as interactions between two dimensions of innovation systems: technological–geographical, technological–economic, and geographical–economic, respectively. The system can self-organize synergistically when the three functions operate on each other in a second-order (or three-way) interaction; the result can be called RIS synergy that we measure as *reduction of uncertainty*. One may find an analogy between this RIS function synergy and the overlap of subsystems that is used to illustrate Storper's "holy trinity" or "triple helix" collaborations. However, synergy is quantifiable, which enables the calculation of the effects of the three functions, their interactions, and the underlying distributions. The measurement technique is explained in detail in Section 3.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

The functions are:

1. *Knowledge exploitation* is associated with the reuse of existing competencies and means (March 1991); at the systems level it represents the interface between economic welfare and technological knowledge creation (Gibbons et al. 1994). Knowledge exploitation does not necessarily depend on geographical locations because economic welfare is created at the level of global markets, even if certain technologies originate in single regions.

2. *Knowledge exploration* is associated with creating new alternatives (Baum et al. 2000); this function is place-dependent rather than market-dependent, because tacit knowledge is essential in creating new knowledge, and it relates significantly to places (Jaffe et al. 1993, Acs et al. 2002).

3. *Organizational control* is defined as collective institutions and mediating organizations that increase the probability of the emergence of new knowledge (Loasby 2001); these functions operate mainly in territorial units.

Because regional advantages can more easily be constructed in regions where synergy prevails among knowledge exploitation, knowledge exploration, and organizational control than in regions without such synergy, one can expect the former regions to outperform the latter in terms of innovation capacities. Theorizing RIS in terms of innovation system function synergies can provide new types of insights into regional economic systems. In addition to the possibility to mark regions, one of the most promising features of synergy measurement in terms of entropy is the opportunity to also quantify the effects of other major forces shaping the synergy in the regions (such as FDI).

However, regional innovation systems do not stand alone, but permanently interact with other RIS, as well as organizations at the level of national and international innovation systems (Tödtling–Trippl 2005). Furthermore, the extent of the interregional interaction between systems varies across regions (Asheim–Isaksen 2002). For example, decisions taken in distant headquarters of foreign-owned firms may take innovation system functions out of the region. Therefore, the relation with other systems has to be taken into consideration when measuring innovation systems synergy.

In the following sections, we show that a four-dimensional model, in which internationalization is additionally considered by means of foreign versus domestic ownership, explains the expected relation with GDP per capita at the regional level, whereas the three-dimensional model did not. This means that foreign ownership also has to be considered when calculating RIS synergy. However, a three-dimensional RIS synergy model will first be developed as a baseline to show in the next step how foreign firms have diverse effects on regional innovation systems.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Materials and Methods

Data

To demonstrate the regionally diverse effects of FDI in Hungary, a firm level database is used that contains census type data from all sectors for 2005.[1] According to the model specified above, one-, two-, three-, and four-dimensional firm distributions will be used to calculate RIS synergy. The four dimensions are:

1. *technological specialization* indicates the characteristics of technological knowledge creation (this is represented in the data by NACE 2 codes of firms)[2];
2. *geographical location* of the company (NUTS 4 regions)[3];
3. *firm size* proxies the dimension of economic exchange (this is captured by number of employees);
4. *ownership structure* captures the dimension of internationalization (foreign or domestic ownership based on registered stock).

We pursue the analysis in terms of economic sectors, and by exclusively using data for high- and medium-tech industries and knowledge intensive services (HTMTKIS).[4] After this comparison, we focus on HTMTKIS sectors when calculating the effects of foreign firms because the majority of the foreign firms operate in these industries.

Foreign interest is attributed when 10% or more shares of the stocks of a firm are in foreign hands (HSCO 2007). We consider all these firms as foreign-owned even if domestic ownership is higher than foreign ownership in the firm. Company size reflects those categories that are generally used for economic analyses (Table 1). In 2005, companies were categorized into 20 Hungarian counties (NUTS 3 level) that contained 168 subregions (local administrative units, NUTS 4 level). We report the results at the county level.

Table 1

Number of firms in the analysis

Number of Employees	HTMTKIS		All sectors	
	Domestic	Foreign	Domestic	Foreign
0 or unknown	47,005	4,608	52,480	3,937
1– 9	59,330	2,698	100,886	6,359
10– 19	3,095	335	10,309	855
20– 49	1,800	333	5,743	757
50–249	899	322	2,567	678
250 or more	134	173	314	180
Total	***112,263***	***8,469***	***172,299***	***12,766***

Source: own compilation.

1 Differently from this study, Lengyel & Leydesdorff (2011) analysed only high- and medium-tech industries and knowledge-intensive services.

2 NACE is an abbreviation of *Nomenclature générale des Activités économiques dans les Communautés Européennes*.

3 NUTS is an abbreviation of "*Nomenclature des Unités Territoriales Statistiques*" (that is, Nomenclature of Territorial Units for Statistics). The NUTS classification is a hierarchical system for dividing up the economic territory of the EU.

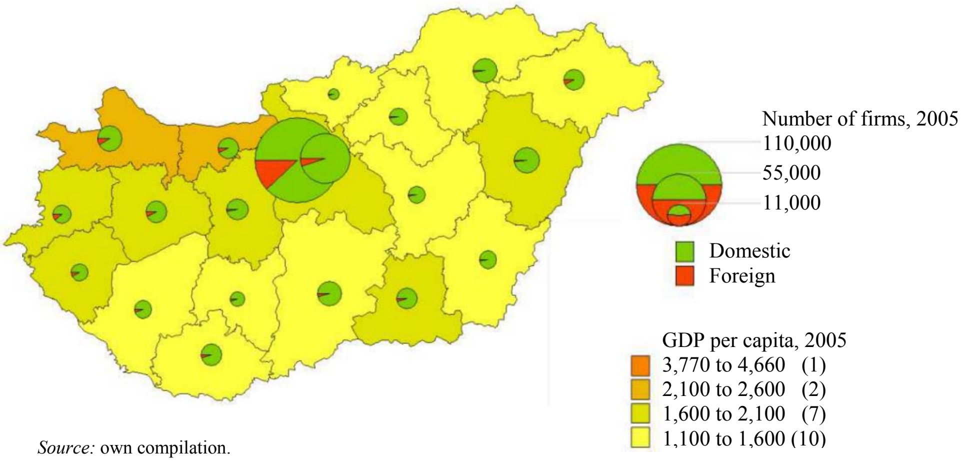
4 Nace 2 codes for high-tech manufacturing: 30, 32, 33; medium-tech manufacturing: 24, 29, 31, 34, 35; knowledge-intensive services: 61, 62, 64, 65, 66, 67, 70, 71, 72, 73, 74, 80, 85, 92.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Table 1 demonstrates the distribution of firms in the two dimensions of ownership and size. Firms without employees are left in the sample because spin-off companies can be included in this category. Firms with foreign interest are on average larger than domestic ones, and almost all large foreign-owned companies are in the high- and medium-tech industries, or knowledge-intensive services.

Figure 2

Firm distribution and GDP per capita in Hungarian counties, 2005



Source: own compilation.

The regional distribution of firms by ownership category is depicted against GDP per capita in Hungarian counties in Figure 2. The relatively developed regions constitute one block situated in the northwest of Hungary between Budapest and the Austrian border. There are only a few foreign companies in each location, and developed regions do not differ from lagging regions in this sense. The only exception is Budapest, which counts for a large rate of foreign firms. We will demonstrate that foreign firms have positive effects on RIS synergy in relatively developed regions in the Northwest and a negative impact on synergy in lagging regions in the south and east of the country.

Synergy measurement

The measurement of synergy among innovation system functions in regions is based on entropy statistics developed within a complexity approach. In such an approach, and in accordance with evolutionary economic theorizing, uncertainty is considered as a fundamental characteristic of economic progress (Foster–Metcalfe 2001, Metcalfe–Foster 2004). Complexity measurement in economics usually focuses on agents and deals with the amount of information the agent needs for decision making and for overcoming uncertainty (Delorme 2001, Louca 2001). However, economic and regional systems can be different regarding this information demand because of general trust among people, collectively accepted rules, etc. that varies across systems. Therefore, we propose a method of complexity measurement that addresses uncertainty and the reduction of it, in particular, at a system level.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Uncertainty can decrease along with growing redundancy at both the micro and macro levels, but not necessarily to the same extent. For example, when the knowledge of agents is overlapping, their expectations of each other's future actions are formed more easily; therefore, there is less uncertainty in a single action or micro level. In a similar manner, the overlap (that is, redundancy) between subsystems of RIS reduces uncertainty on a macro level because actors from different subsystems can understand each other more easily. Our model depicted in Figure 1 addresses this phenomenon with the use of innovation system functions; synergy among the three functions can reduce uncertainty prevailing in a system.

Uncertainty can be measured in bits using Shannon's (1948) formulas that were applied in economics (Delorme 2001) and geography (Johnston et al. 2000). Entropy is furthermore used for other purposes: since entropy takes a higher value if a distribution is even, it can also be used as an indicator of diversity. Furthermore, because probabilistic entropy can be fully decomposed into its components, it has also been used as the method for calculating related variety in regions (Frenken et al. 2007, Theil 1972).

Synergy in an innovation system can be measured as a reduction of uncertainty at the systems level, which is based on entropy calculation in three- and four dimensions introduced in the following subsection (Leydesdorff–Ivanova 2014, Mêgnigbêto 2014, Yeung 2008, p. 59.).

Methodology

The uncertainty contained in the distribution of a random variable x ($\sum_x p_x$) is defined by Shannon (1948; cf. Theil 1972) as $H_x = -\sum_x p_x \log_2 p_x$. Analogously, the uncertainty in a two-dimensional distribution can then be formulated as: $H_{xy} = -\sum_x \sum_y p_{xy} \log_2 p_{xy}$. This uncertainty is the sum of the uncertainty in the two variables diminished with their mutual information as a measure of their co-variation. In the case of two dimensions, the uncertainty in the two potentially interacting dimensions (x and y) is reduced with the mutual information. Transmission or mutual information that captures this reduction of uncertainty is formalized in the case of two dimensions as follows:

$$T_{xy} = (H_x + H_y) - H_{xy} \quad (1)$$

In the limiting case that the distributions x and y are completely independent, $T_{xy} = 0$ and $H_{xy} = H_x + H_y$. In all other cases $T_{xy} > 0$, and therefore $H_{xy} \leq H_x + H_y$ (Theil 1972, p. 59.). In general, two variables determine each other in their mutual information (T_{xy}), and condition each other otherwise. However, in the case of three interacting variables, there are two options: the three interacting systems may have a common interaction shared by all of them or not. In the latter case, the correlation among each two can be spuriously influenced by the third.

In the case of overlaps among three (or more) functions, synergy can be expressed by an information measure (T_{xyz}), which McGill (1954) derived from the Shannon formulas:[5]

$$T_{xyz} = H_x + H_y + H_z - H_{xy} - H_{xz} - H_{yz} + H_{xyz} \quad (2)$$

5 Both Yeung (2008, p. 59f.) and Krippendorff (2009, p. 200) noted that this information measure can no longer be considered as a Shannon-type measure because of the possible circularity in the information transfers. Shannon-type entropy measures are by definition linear and positive. Since the measure sums Shannon-type measures in terms of bits of information, its dimensionality is also bits of information, and therefore it can be used as a measure of uncertainty and uncertainty reduction, respectively.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

While the two-dimensional systems reduce uncertainty, the trilateral term in turn feeds back on this reduction and, therefore, adds another term to the uncertainty. The feedback can be so strong that the net result becomes negative, and thus uncertainty is reduced. The *configuration* of the system—distributions and their overlaps—determines the net result in terms of the value of T_{xyz} (McGill 1954). We use $T_{xyz} < 0$ as an indicator that the uncertainty is reduced to the extent that synergy is indicated. The value of T_{XYZ} measures the interrelatedness of the three sources of variance and the fit of the relations and correlations between and among them. T_{XYZ} has been used as an indicator of potential reduction of uncertainty in complex systems in many disciplines (Ulanowicz 1986, Jakulin 2005).

Analogously, it can be shown to follow for four dimensions:

$$T_{WXYZ} = H_W + H_X + H_Y + H_Z - H_{WX} - H_{WY} - i_{XY} - H_{WZ} - H_{YZ} - H_{XZ} + H_{WXY} + H_{WXZ} + H_{WYZ} + H_{XYZ} - H_{WXYZ} \quad (3)$$

It can be shown that the sign of mutual information is reversed in the case of an even number of dimensions (Leydesdorff–Ivanova, 2014). This solves a technical problem about the sign in information theory (Krippendorff 2009), but does not affect the measurement or the reasoning.

As noted, the four dimensions under study in this paper are (G)eography, (T)echnology, (O)rganization, and (I)nternationalization, and T_{GTO} and T_{GTOI} will accordingly be used to indicate the mutual information among them. Similarly to Equations 2 and 3, one can formulate as follows:

$$T_{GTO} = H_G + H_T + H_O - H_{GT} - H_{GO} - H_{TO} + H_{GTO} \quad (4)$$

$$T_{GTOI} = H_G + H_T + H_O + HI - H_{GT} - H_{GO} - H_{TO} - H_{GI} - H_{OI} - H_{TI} + H_{GTO} + H_{GTI} + H_{GOI} + H_{TOI} - H_{GTOI} \quad (5)$$

We have argued above that the synergy in innovation systems can be measured as the reduction of uncertainty among knowledge exploitation, knowledge exploration and organizational control. High negative values of T_{GTO} indicate a higher reduction of uncertainty and, therefore, a more synergistic innovation system. The same logic can be followed when interpreting the values of $T_{GTOI;}$ however, a higher positive value of the indicator means stronger synergy because the sign of the four-dimensional entropy is reversed as was explained previously.

When the basis of the logarithm is two, all values are expressed in bits of information. Therefore, our entropy measures are formal (probability) measures and thus independent of size or any other reference to the empirical systems under study. The sigma in the formula allows all information terms to be fully decomposed. This is a crucial advantage because it enables us to calculate the effects of certain subsystems (e.g., foreign firms) in the complex set of variables.[6] The calculations were done by using a computer routine described in Leydesdorff et al. (2014) and available online at http://www.leydesdorff.net/software/th4/index.htm .

6 Note that the method remains within the limitations of complexity modelling according to Allen's (2001) taxonomy: it only assumes that (1) the borders of the system are defined; (2) actors and subsystems that interact within the economic system are identified; (3) behaviour of actors are comparable to the mean of actions.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Results

Entropy values in the regions highly depend on the number of categories (e.g. NACE 2 classes, subregions, firm size categories) and on the number of firms in the region. Table 2 shows the one-dimensional entropies in order to illustrate differences among regions. The two-dimensional entropy, as well as two and three-dimensional mutual information values are normalized by the number of firms in the region.

The maximum entropy in the technological dimension is $\log_2(22) = 4.459$; and $\log_2(6) = 2.584$ in the organizational dimension (there are 22 technological and 6 categories in our sample) and it varies according to the number of subregions in terms of geographical distribution. The entropy values in both dimensions indicate a very even distribution among the regions: Hungarian counties are more or less equally diverse in terms of technological specialization (H_T) and organizational structure (H_O). However, regional economies are geographically more centralized in southern and eastern counties where regional universities and government organizations are located in centres of administration: H_G reaches 44% in Baranya (1,408/$\log_2(9)$×100), 47% in Hajdú-Bihar (1,495/$\log_2(9)$×100), 52% in Borsod-Abaúj-Zemplén (2,046/$\log_2(15)$×100) and 53% in Csongrád counties (1,497/$\log_2(7)$×100) of the maximum entropy; whereas the indicator is above 60% in all of the remaining counties. This finding resembles a study focusing on Russian innovation systems, in which similar relationships were found between state-controlled services and entropy statistics (Leydesdorff et al. 2015).

Synergy in regional innovation systems

Innovation systems synergy is defined as the interdependence between innovation system functions. These functions represent three different selection environments that operate in the system simultaneously and are measured by two-dimensional transmissions between one-dimensional firm distributions (T_{TG}, T_{TO}, T_{GO}). The value of these indicators varies across Hungarian regions (see Table 2). T_{TG} is highest values in Pest (that surrounds Budapest), Borsod-Abaúj-Zemplén (with the second largest town in the North-East), Fejér (former location of large IT multinationals), and Baranya (strong university location with relatively weak economy) counties. Budapest stands out in T_{TO}, it is followed by Pest County, Győr-Moson-Sopron (home of the automotive industry in the North-East), Borsod-Abaúj-Zemplén, and Bács-Kiskun (South of the capital). The majority of Hungarian counties with the highest values of T_{GO} also appear to be at the top of the list in terms of T_{TG}: Pest, Borsod-Abaúj-Zemplén, Győr-Moson-Sopron, Baranya.

The status of Budapest is contradictory in the hierarchy of Hungarian regions; being a capital city, it is counted as both a county and a subregion in the NUTS classification. Due to the taxonomy, H_G equals zero in Budapest and consequently the value of T_{TGO} is very close to zero. Thus, we decided to eliminate Budapest from the presentation of T_{TGO} results by setting the value equal to zero. This is a common practice in Hungarian geographical analyses and we admit that Budapest would require a separate analysis.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Table 2

Entropy measures and mutual redundancies in two, three, and four dimensions of distribution of Hungarian HTMTKIS firms

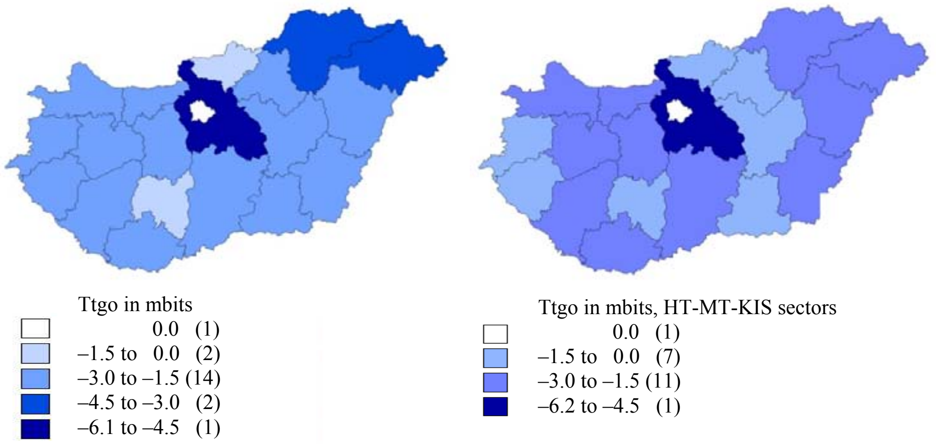
Regions (NUTS 2)	Counties (NUTS 3)	Ht	Hg	Ho	Htg	Hto	Hgo	Ttg	Tto	Tgo	Ttgo	N subreg	N firms	N HTM TKIS firms
Central Hungary	Budapest	2.816	0.000	1.360	1.282	1.881	0.619	0.039	19.594	0.009	0.002	1	111,395	54,691
	Pest	3.032	3.457	1.307	0.768	0.511	0.566	4.171	5.883	0.585	–6.190	15	38,992	14,304
Western Transdanubia	Győr-Moson-Sopron	3.015	1.785	1.360	0.166	0.149	0.109	1.804	3.709	0.416	–1.522	7	11,776	4,195
	Vas	2.963	1.830	1.371	0.084	0.075	0.056	1.374	2.236	0.309	–1.367	9	5,938	2,133
	Zala	2.975	1.936	1.343	0.102	0.089	0.069	1.279	2.299	0.194	–1.149	6	7,477	2,530
Northern Transdanubia	Fejér	3.068	2.220	1.385	0.158	0.133	0.109	2.094	2.413	0.272	–2.272	10	10,457	3,644
	Komárom-Esztergom	3.056	2.489	1.367	0.119	0.093	0.083	0.861	2.432	0.247	–2.066	7	8,014	2,598
	Veszprém	3.077	2.636	1.370	0.135	0.104	0.095	1.93	2.567	0.351	–1.885	9	8,326	2,873
Southern Transdanubia	Baranya	3.054	1.408	1.304	0.141	0.138	0.087	2.024	1.937	0.336	–1.582	9	10,744	3,851
	Somogy	3.018	2.391	1.308	0.098	0.078	0.068	1.706	1.892	0.308	–1.734	10	6,860	2,214
	Tolna	3.056	1.961	1.386	0.062	0.054	0.042	1.051	2.381	0.219	–1.189	5	4,984	1,516
Northern Hungary	Borsod-Abaúj-Zemplén	3.062	2.046	1.333	0.187	0.160	0.124	2.338	3.125	0.624	–2.758	15	13,067	4,451
	Heves	3.162	2.000	1.399	0.088	0.076	0.058	1.14	2.314	0.380	–1.128	7	6,166	2,069
	Nógrád	3.110	2.243	1.393	0.040	0.033	0.028	0.758	1.246	0.148	–1.083	6	3,182	921
Northern Great Plain	Hajdú-Bihar	3.054	1.495	1.294	0.159	0.151	0.098	1.393	2.151	0.324	–1.814	9	12,347	4,230
	Szabolcs-Szatmár-Bereg	3.075	2.115	1.361	0.112	0.095	0.076	1.902	1.775	0.294	–1.971	11	9,585	2,631
	Jász-Nagykun-Szolnok	3.061	2.126	1.379	0.095	0.080	0.065	1.247	2.125	0.279	–1.387	7	7,228	2,221
Southern Great Plain	Bács-Kiskun	3.067	2.368	1.367	0.169	0.136	0.117	1.739	2.965	0.320	–2.408	10	12,738	3,777
	Békés	3.144	2.486	1.387	0.084	0.067	0.059	1.324	1.632	0.229	–1.607	8	6,179	1,829
	Csongrád	3.114	1.497	1.323	0.133	0.127	0.082	1.517	2.691	0.244	–1.495	7	9,802	3,514

Note: Entropy values are in bits of information; mutual redundancies are reported in millibits for the sake of visualization.
Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Figure 3

Innovation systems synergy in three dimensions, the entire Hungarian economy versus HTMTKIS sectors, 2005



Source: own compilation.

T_{TGO} values calculated from Hungarian HTMTKIS firms' distribution suggest that innovation system functions overlap each other in the NUTS 3 regions of the country (see Table 2). One finds the strongest uncertainty reduction in Pest and Borsod-Abaúj-Zemplén regions, which also stood out in terms of all three innovation system functions. Regions that follow are Bács-Kiskun, Fejér that are among the top regions only in one innovation system function; and Komárom-Esztergom where the operating innovation system functions are not outstanding at all.

The national innovation system is organized around the capital's agglomeration; the synergy measured in the distribution of HTMTKIS firms is stronger in the western regions than in eastern regions; while there is no such an imbalance when synergy is measured in the distribution of firms from all economic sectors (Figure 3).

Internationalization and mutual redundancy in four dimensions

Innovation systems never operate in isolation but intensive interactions with global markets and their ability to attract investments are natural elements of success. Therefore, the dimension of internationalization is also included in our development model. Based on previous research regarding Hungarian economic transition (Lengyel–Leydesdorff 2011), one can expect that the additional dimension can be used to distinguish regions in this study. Indeed, the four-dimensional model fits better with the expectation based on the spatial distribution of GDP per capita and allows us to draw more sophisticated conclusions about RIS synergy in Hungary.

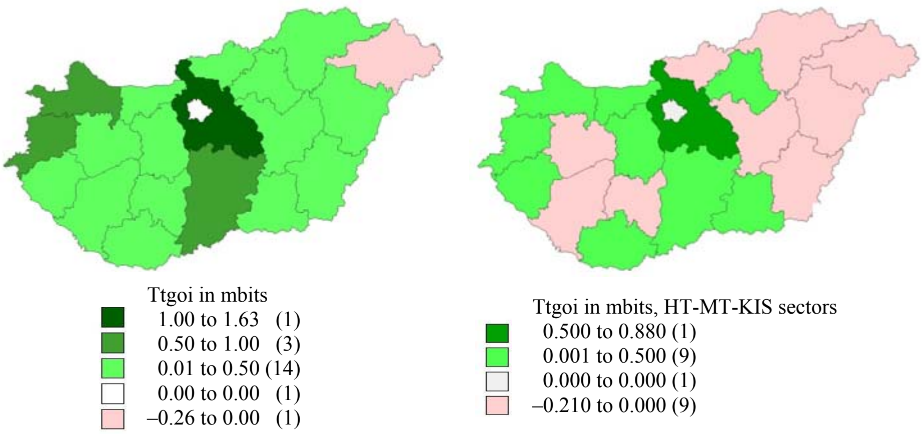
We classified firms into foreign-owned or domestic. We used the same procedure of uncertainty and redundancy calculation to measure the effect of internationalization in four dimensions as presented above for three dimensions. Contrary to the three-dimensional model, the mutual redundancy of four interacting functions (or any even number of

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

dimensions) is indicated by a positive sign of transmission (Leydesdorff–Ivanova 2014, p. 389). Accordingly, a positive value of T_{TGOI} marks a four-dimensional overlap, which confirms that the synergy is stable even after controlling for internationalization. Our results suggest that foreign-owned firms entered the region in those particular industry-location-size combinations that favoured the existing synergy. Whereas negative values of T_{TGOI} suggest that FDI and leading forces of innovation system synergy in the region did not co-occur.

Figure 4

Innovation systems synergy in four dimensions, the entire Hungarian economy versus HTMTKIS sectors, 2005



Source: own compilation.

Four-dimensional mutual redundancy of Hungarian HTMTKIS firms takes a positive value only in half of the NUTS 3 regions (Table 3). The highest value of the four-dimensional overlap is in Pest County; a negative value of T_{TGOI} occurs in one north-eastern county only. A sharp contrast between eastern and western regions emerges in this regard when looking at HTMTKIS sectors only. Similarly to the presentation of T_{TGO} results, we depict the T_{TGOI} value of Budapest separately. Note that the sign of T_{TGOI} does not depend exclusively on the share of foreign firms in the economy. In other words, these findings show the four-dimensional distribution of firms and not their FDI intensity.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Table 3

The dimension of internationalization

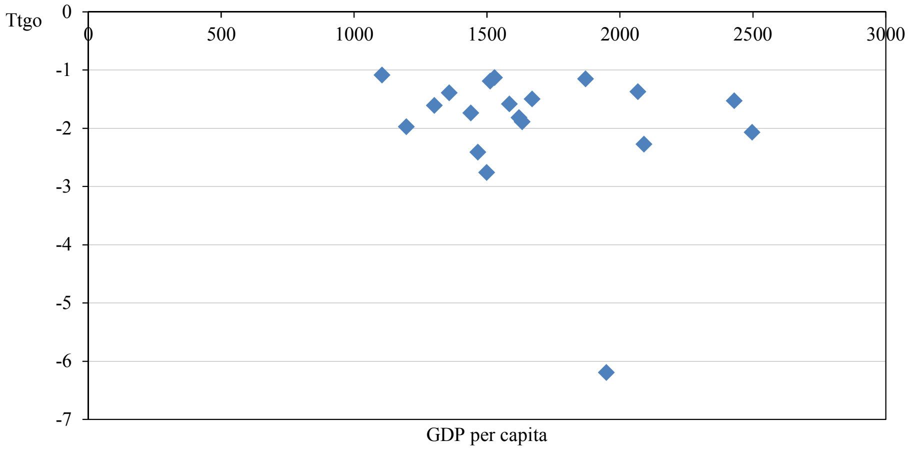
Regions (NUTS 2)	Counties (NUTS 3)	*Hi*	*Ttgo*	*Ttgoi*	N subreg	N foreign firms	N HTMTKIS firms	N HTMTKIS foreign firms
Central Hungary	Budapest	0.458	0.002	0.000	1	11,547	54,691	5,274
	Pest	0.262	–6.190	0.871	15	1,782	14,304	636
Western Transdanubia	Győr-Moson-Sopron	0.403	–1.522	0.244	7	942	4,195	336
	Vas	0.476	–1.367	0.048	9	5,318	2,133	218
	Zala	0.348	–1.149	0.032	6	586	2,530	165
Northern Transdanubia	Fejér	0.195	–2.272	0.007	10	313	3,644	110
	Komárom-Esztergom	0.272	–2.066	0.003	7	505	2,598	121
	Veszprém	0.358	–1.885	–0.139	9	571	2,873	195
Southern Transdanubia	Baranya	0.204	–1.582	0.046	9	502	3,851	123
	Somogy	0.285	–1.734	–0.066	10	373	2,214	110
	Tolna	0.219	–1.189	–0.050	5	215	1,516	53
Northern Hungary	Borsod-Abaúj-Zemplén	0.109	–2.758	–0.201	15	266	4,451	64
	Heves	0.199	–1.128	0.042	7	195	2,069	64
	Nógrád	0.207	–1.083	–0.114	6	110	921	30
Northern Great Plain	Hajdú-Bihar	0.122	–1.814	–0.062	9	204	4,230	70
	Jász-Nagykun-Szolnok	0.148	–1.387	–0.005	7	168	2,221	47
	Szabolcs-Szatmár-Bereg	0.104	–1.971	–0.053	11	670	2,631	36
Southern Great Plain	Bács-Kiskun	0.231	–2.408	0.007	10	514	3,777	142
	Békés	0.133	–1.607	–0.057	8	148	1,829	34
	Csongrád	0.188	–1.495	0.069	7	464	3,514	101

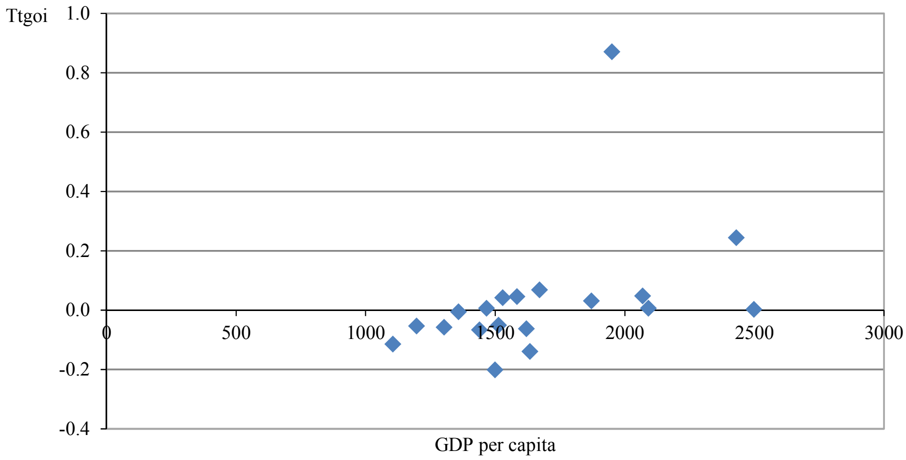
Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Figure 5

RIS synergy indicators (HTMTKIS sectors) and GDP per capita, 2005





Source: own compilation.

Figure 5 depicts the association between GDP per capita and RIS synergy indicators. Note that the expected negative relation prevails between T_{TGO} and GDP per capita, whereas the distribution of T_{TGOI} indicator is associated positively with spatial differences in GDP per capita. The outlier region is Pest County in both T_{TGO} and T_{TGOI} distributions. The quantification of the statistical relationship highlights that GDP per capita is positive and significantly related to T_{TGOI} , meanwhile there is no statistically significant relation between GDP per capita and T_{TGO} (Table 4). This major finding implies that internationalization is very important in calculating innovation systems synergy and associating the indicator with regional development.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Table 4

Correlation among RIS synergy indicators measured in HTMTKIS sectors and GDP per capita

Pearson/Spearman	T_{TGO}	T_{TGOI}	GDP per capita
T_{TGO}	1	0.122	–0.163
T_{TGOI}	–0.770[a)]	1	0.596[a)]
GDP per capita	–0.201	0.402[b)]	1

Note: Pearson correlation values are below the diagonal and Spearman correlation values are above it.
a) and b) denote statistical significance at the 0.01 and 0.05 levels, respectively.
Source: own calculation.

In summary, our results show that foreign firms have added to regional innovation system synergies in some areas but disturbed synergy in others. The map of HTMTKIS sectors is particularly interesting because the four-dimensional synergy is indicated in regions where above average development was reported in section 3 (see Figure 2). However, there is a lack of synergy – increasing uncertainty in four dimensions – in regions lagging behind in the eastern part of the country. Thus, internationalization is a very important dimension in regional innovation systems and synergy derived from four-dimensional firm distributions goes hand in hand with regional development in Hungary. Because the number of foreign-owned firms does not differ significantly across regions (as it was shown in Figure 2), the result suggests that not the presence of foreign firms per se but the synergies generated are important for regional development.

Internationalization of innovation systems and the appearance of foreign firms provide extra opportunities as well as challenges for regions. Systems synergies might be further enhanced if a region was successfully integrated into the global market by the local value chains or university relationships of present multinational companies. However, these firms can have negative effects on local synergies as well: domestic companies can be expelled from the market; talented labour might be exclusively pooled to MNEs, leaving no space for dynamic SMEs, etc. This latter phenomenon is discussed in the following subsection.

Foreign firms' effect on innovation systems synergy

Due to the decomposability of entropy measures, one can calculate the contribution of mutual information measured in sub-groups with the mutual information measured in the entire set. In order to do that, mutual redundancy in the subgroups has to be calculated in the same way as presented above and the share in the entropy measured in the whole set, weighted by the number of firms calculated. The contribution of three-dimensional mutual redundancy measured in foreign-owned and domestic subgroups to the mutual redundancy in the region is illustrated in Table 5.

Table 5

Effect of foreign-owned firms on three-dimensional synergy measured in HTMTKIS firms

Regions (NUTS 2)	Counties (NUTS 3)	All HTMTKIS firms			Foreign HTMTKIS firms			Domestic HTMTKIS firms		
		Ttgoi	Δ*Ttgo*	N	Δ*Ttgo*	N	effect on entropy	Δ*Ttgo*	N	effect on entropy
Central Hungary	Budapest	0.0000	0	54,691	0	5,274	0	0	49,417	0
	Pest	0.0073	–0.00619	14,304	–0.02458	636	297.060	–0.00582	13,668	–5.911
Western Transdanubia	Győr-Moson-Sopron	0.0070	–0.00152	4,195	–0.00661	336	333.941	–0.00142	3,859	–6.420
	Vas	0.0027	–0.00137	2,133	–0.00346	218	153.357	–0.00127	1,915	–7.060
	Zala	0.0015	–0.00115	2,530	–0.00208	165	80.557	–0.00112	2,365	–2.666
Northern Transdanubia	Fejér	0.0002	–0.00227	3,644	–0.00123	110	–45.648	–0.00235	3,534	3.544
	Komárom-Esztergom	0.0001	–0.00207	2,598	–0.00636	121	207.766	–0.00177	2,477	–14.522
	Veszprém	–0.0058	–0.00188	2,873	0.00017	195	–109.142	–0.00188	2,678	–0.162
Southern Transdanubia	Baranya	0.0014	–0.00158	3,851	–0.00143	123	–9.327	–0.00164	3,728	3.773
	Somogy	–0.0036	–0.00173	2,214	–0.00017	110	–90.011	–0.00177	2,104	2.274
	Tolna	–0.0039	–0.00119	1,516	–0.00057	53	–51.881	–0.00118	1,463	–0.814
Northern Hungary	Borsod-Abaúj-Zemplén	–0.0054	–0.00276	4,451	0.00164	64	–159.540	–0.00285	4,387	3.477
	Heves	0.0025	–0.00113	2,069	–0.00135	64	19.707	–0.00116	2,005	2.622
	Nógrád	–0.0149	–0.00108	921	0.00057	30	–152.506	–0.00108	891	–0.517
Northern Great Plain	Hajdú-Bihar	–0.0018	–0.00181	4,230	0.00013	70	–106.902	–0.00188	4,160	3.866
	Jász-Nagykun-Szolnok	–0.0002	–0.00139	2,221	–0.00066	47	–52.353	–0.00143	2,174	3.344
	Szabolcs-Szatmár-Bereg	–0.0024	–0.00197	2,631	0.00077	36	–139.017	–0.00211	2,595	6.920
Southern Great Plain	Bács-Kiskun	0.0002	–0.00241	3,777	–0.00246	142	2.005	–0.00241	3,635	0.157
	Békés	–0.0037	–0.00161	1,829	0.00025	34	–115.539	–0.00168	1,795	4.370
	Csongrád	0.0024	–0.00150	3,514	–0.00145	101	–2.986	–0.00157	3,413	5.165

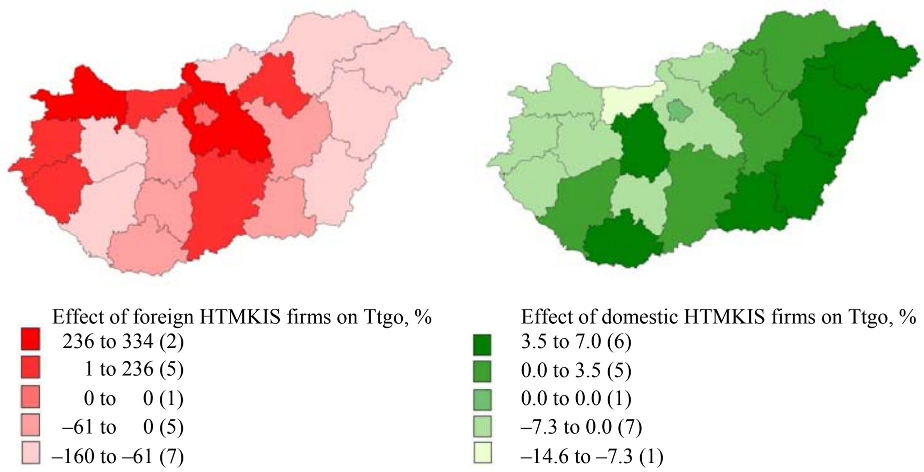
Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

We found conflicting effects of foreign and domestic firms on regional innovation systems in several counties. On the one hand, the subgroup of foreign-owned firms adds to innovation systems synergy in seven counties only, but their effect seems to disturb systems synergy in the other 12 counties. On the other hand, the domestic group contributes to synergy in 12 counties and disturbs synergy in 7 counties. It is very important to note that the effects are not always mutually exclusive: for example, both subgroups have positive effects on synergy in Bács-Kiskun and Heves counties; both subgroups disturb synergies in Veszprém, Tolna, and Nógrád counties. Firm ownership effect on innovation systems synergy is depicted in Figure 5.

Figure 6

Effect of foreign and domestic companies on synergy, 2005



Source: own compilation.

Our findings suggest that a small set of foreign firms became the leading entities in the counties in the Northwest of the country. These counties experienced a positive effect on the synergy, but the other counties were affected negatively. The first group of counties performs above the national average of regional development. However, foreign HTMTKIS companies influence RIS synergy negatively in southern and eastern counties where domestic firms play the dominant role and have a positive effect on the synergy.

Consequently, regions have followed different paths regarding the establishment of synergistic economic systems, in which the diverse effect of foreign-owned firms has been a determining factor. Foreign-owned firms have generated new synergies in a limited number of relatively developed regions where the systems do not seem to self-organize in terms of domestic companies. In other regions, foreign-owned firms have disturbed the synergy that was established by domestic companies. This result is consistent with the questions raised and conjectures specified by Lengyel and Leydesdorff (2011); however, in this study, we provide the quantitative evidence for the effect of a dual economic structure on diverse regional development in Hungary.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Conclusion and discussion

This study applied a synergy perspective to the analysis of the effect of foreign firms and FDI in regional innovation systems. We have introduced a four-dimensional model as a further development of innovation systems synergy measurement, in which the dimension of internationalization – using foreign versus domestic ownership – is taken into consideration as a new factor. We showed that RIS synergy derived from the four-dimensional model correlates with GDP per capita in regions whereas the three-dimensional model did not. In general, internationalization and the appearance of foreign firms in regions is an important factor in describing the success of innovation systems.

However, we also find that foreign firms have diverse effects on local economies. Although foreign-owned firms have become major actors for regional development in Hungary, these firms did not always have a positive effect. On the one hand, our results suggest that foreign companies have contributed to establish synergy in RIS, but only in relatively more developed regions, for example, the regions around the capital and counties in the Northwest. A small set of foreign firms have restructured the innovation systems in these areas since synergy seems to have been established as an exclusive effect of foreign companies; the larger set of domestic firms disturbs synergy in these regions. On the other hand, foreign firms disturb synergy in the relatively lagging eastern and southern parts of the country, where the system is still self-organizing itself in terms of domestic companies. Thus, we find that FDI has had a diverse effect on regional innovation systems in Hungary.

Our results are policy-relevant, because Hungary and other Central and Eastern European countries have joined the EU and therefore also reformed their innovation policies in the 2000s accordingly (Suurna–Kattel 2010), despite the warnings of several scholars (Havas 2002, Radosevic–Reid 2006, von Tunzelmann–Nassehi 2004). Indeed, European policy standards might not work in the CEE context because foreign firms might have relatively stronger or even an exclusively dominant economic role, but also interact less with domestic companies in CEE countries than in other parts of Europe. Policies would have to be tuned accordingly (Lengyel–Cadil 2009, Lengyel et al. 2013, Rama 2008).

Based on the literature, one can claim that there is a huge demand for a refined innovation policy in CEE countries, and these policies will enhance technology transfer from multinational companies to domestic firms. A major finding of this paper implies that these refined instruments have to be decentralized and adjusted to regional innovation systems. Domestic companies may benefit more from the co-location of multinationals in those regions, where foreign firms have a positive effect on innovation systems synergies. Technology transfer from multinationals to domestic firms might have a chance in these relatively more developed territories only. In other areas, where foreign firms have a disturbing effect on synergy, there might be a lower chance for efficient technology transfer, and policies should rather focus on attracting appropriate investments.

Regional innovation systems are considered in our argument as complex economic systems that develop in a self-organizing manner when adapting to external interventions. We argue that dynamics in the system are led by a synergy among three innovation system functions: knowledge exploration, knowledge exploitation and organizational control. Further theoretical research is needed.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Further studies can unfold the concordance of synergy measurement and other types of RIS theorizing. Two subsystems are distinguished in RIS literature: knowledge exploration subsystem (universities, PROs, etc.) and knowledge exploitation subsystem (companies, clusters, etc.); furthermore, their interaction, is enhanced by economic and science policies (Tödtling–Trippl 2005). A well functioning RIS equally draws on the two subsystems, and regional economic advantage can be constructed when government is involved as a third subsystem providing organizational control for the system (Cooke et al. 2004). Synergy measurement derived from the triple helix of university–industry–government model can enter the discussion at this point (Etzkowitz–Leydesdorff 2000). Since very different selection mechanisms operate in the three subsystems, the level of synergy or overlap among them can be expected to mark the innovative quality of the RIS. Furthermore, further studies may be able to specify the argument that connections to other RIS and international markets are crucial to avoid regional lock-in and too narrow local co-operations (Narula 2002, Tödtling–Tripple 2005).

More efforts are needed to connect the concept and measurement of innovation system functions to indicators in regional economics. Innovation system functions can also be explained in terms of localization economies, local competition, and urbanization economies in future studies. For example, the knowledge exploration indicator measures co-variation between geographical and technological distribution of firms; hence, it can monitor the operation of localization economies or, in other words, the extent to which companies of similar technological profiles co-locate in the region. Similarly, our indicator of knowledge exploitation measures co-variation between technologically and organizationally proximate firms in the region. This can lead to positive and negative externalities since these firms are more likely to learn from each other's complex organization-related issues. However, competition may also be more intense under the condition of intense knowledge exploitation.

The organizational control function in the innovation system organizes firms of similar sizes (regardless of technological profile) to co-locate. One can associate organizational control to urbanization economies because settlement structures may strongly affect the value of organizational control. In other words, large firms may be co-located in big cities while small firms tend to be co-located in small towns of the region, due to the size of the local market and labour force. We intend to shed light on relations between knowledge functions and localization economies, urbanization economies, and local competition in follow-up studies.

* * *

Acknowledgements

The work of Balazs Lengyel was supported by the Hungarian Scientific Research Fund (OTKA, PD106290).

REFERENCES

Acs, Z.–Anselin, L.–Varga, A. (2002): Patents and innovation counts as measures of regional production of knowledge *Research Policy* 31 (7): 1069–1085.

Acs, Z.–O'Gorman, C.–Szerb, L.–Terjesen, S. (2007): Could the Irish miracle be repeated in Hungary? *Small Business Economics* 28 (2–3): 123–142.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Allen, P.M. (2001): Knowledge, ignorance and the evolution of complex systems In: Foster, J.–Metcalfe, J.S. (Eds.) *Frontiers of evolutionary economics: competition, self-organization and innovation policy* pp. 313–350., Edward Elgar, Cheltenham-Northampton.

Asheim, B.T.–Isaksen, A. (2002): Regional innovation systems: the integration of local „sticky" and global „ubiquitous" knowledge *Journal of Technology Transfer* 27 (1): 77–86.

Baum, J. A. C.–Li, S. X.–Usher, J. M. (2000): Making the next move: How experiential and vicarious learning shape the locations of chains' acquisitions *Administrative Science Quarterly* 45 (4): 766–801.

Békés, G.–Kleinert, J.–Toubal, F. (2009): Spillovers from multinationals to heterogeneous domestic firms: evidence from Hungary *The World Economy* 32 (10): 1408–1433.

Biggiero, L. (2007): The location of multinationals in industrial district: knowledge transfer in biomedicals *Journal of Technology Transfer* 27 (1): 111–122.

Blazek, J.–Uhlír, D. (2007): Regional innovation policies in the Czech Republic and the case of Prague: an emerging role of a regional level? *European Planning Studies* 15 (7): 871–888.

Braczyk, H.J.–Cooke, P.–Heidenreich, M. (1998): *Regional Innovation Systems: The Role of Governance in a Globalized World* UCL Press, London.

Boschma, R.–Frenken, K. (2006): Why is economic geography not an evolutionary science? Towards an evolutionary economic geography *Journal of Economic Geography* 6 (3): 273–302.

Cantwell, J.–Iammarino, S. (1998): MNCs, technological innovation and regional systems in the EU: some evidence from the Italian case *International Journal of the Economics of Business* 5 (3): 383–408.

Coase, R.H. (1937): The Nature of the Firm *Economica* 4 (16): 386–405.

Cooke, P. (2001): Regional innovation systems, clusters, and the knowledge economy *Industrial and Corporate Change* 10 (4): 945–974.

Cooke, P.–Heidenreich, M.–Braczyk, H.J. (2004): *Regional Innovation Systems. Second Edition, The role of governance in a globalized world* Routledge, London and New York.

Cooke, P.–Leydesdorff, L. (2006): Regional development in the knowledge-based economy: the construction of advantage *Journal of Technology Transfer* 31 (1): 5–15.

Dachs, B.–Ebersberger, B.–Lööf, H. (2008): The innovative performance of foreign-owned enterprises in small open economies *Journal of Technology Transfer* 33 (4): 393–406.

Delorme, R. (2001): Theorizing complexity In: Foster, J.–Metcalfe, J.S. (Eds.) *Frontiers of evolutionary economics: competition, self-organization and innovation policy* pp. 80–108., Edward Elgar, Cheltenham-Northampton.

Etzkowitz, H.–Leydesdorff, L. (2000): The Dynamics of Innovation: From National Systems and 'Mode 2' to a Triple Helix of University-Industry-Government Relations *Research Policy* 29 (2): 109–123.

Ferragina, A.M.–Mazzotta, F. (2014): FDI spillovers on firm survival in Italy: absorptive capacity matters! *Journal of Technology Transfer* 39 (6): 859–897.

Foster, J.–Metcalfe, J.S. (2001): *Frontiers of evolutionary economics: competition, self-organization and innovation policy* Edward Elgar, Cheltenham-Northampton.

Frenken, K.–van Oort, F.–Verburg, T. (2007): 'Related Variety, Unrelated Variety and Regional Economic Growth' *Regional Studies* 41 (5): 685–697.

Gibbons, M.–Limoges, C.–Nowotny, H.–Schwartzman, S.–Scott, P.–Trow, M. (1994): *The new production of knowledge: The Dynamics of Science and Research in Contemporary Societies* SAGE, London.

Halpern, L.–Muraközy, B. (2007): Does distance matter in spillover? *Economics of Transition* 15 (4): 781–805.

Havas, A. (2002): Does innovation policy matter in a transition country? The case of Hungary *Journal of International Relations and Development* 5 (4): 380–402.

Hungarian Central Statistical Office (2007). *National Accounts of Hungary, 2004-2005* Hungarian Central Statistical Office, Budapest. http://portal.ksh.hu/pls/ksh/docs/hun/xftp/idoszaki/monsz/monsz0405.pdf (downloaded from: 01. 08. 2015)

Inzelt, A. (2008): The inflow of highly skilled workers into Hungary: a by-product of FDI *Journal of Technology Transfer* 33 (4): 422–438.

Ivanova I.A.–Leydesdorff, L. (2014): A simulation model of the Triple Helix of university-industry-government relations and the decomposition of redundancy *Scientometrics* 99 (3): 927–948.

Jaffe A.B.–Trajtenberg M.–Henderson R. (1993): Geographic localization of knowledge spillovers as evidenced by patent citations *Quarterly Journal of Economics* 108 (3): 577–598.

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

Jakulin, A. (2005): *Machine learning based on attribute interactions* University of Ljubljana, Ljubljana.

Johnston, R. J.–Gregory, D.–Pratt, G.–Watts, M. (2000): *The Dictionary of Human Geography* Blackwell, Oxford.

Krippendorff, K. (2009): Ross Ashby's information theory: a bit of history, some solutions to problems, and what we face today *International Journal of General Systems* 38 (2): 189–212.

Lengyel, B.–Cadil, V. (2009): Innovation policy challenges in transition countries: foreign business R&D in the Czech Republic and Hungary *Transition Studies Review* 16 (1): 174-188.

Lengyel, B.–Leydesdorff, L. (2011): Regional innovation systems in Hungary: The failing synergy at the national level *Regional Studies* 45 (5): 677–693.

Lengyel, B.–Sebestyén, T.–Leydesdorff, L. (2013): Challenges for regional innovation policies in CEE countries: spatial concentration and foreign control of US patenting *Science and Public Policy* 42 (1): 1–14.

Lengyel, B.–Szakálné Kanó, I. (2014): Regional growth in a dual economy: ownership, specialization and concentration in Hungary *Acta Oeconomica* 64 (3): 257–285.

Leydesdorff, L.–Dolfsma, W.–van der Panne, G. (2006): Measuring the knowledge base of an economy in terms of Triple Helix relations among 'technology, organization, and territory' *Research Policy* 35 (2): 181–199.

Leydesdorff L.–Fritsch M. (2006): Measuring the knowledge base of regional innovation systems in Germany in terms of a Triple Helix Dynamics *Research Policy* 35 (10): 1538–1553.

Leydesdorff, L.–Ivanova, I. A. (2014): Mutual Redundancies in Inter-human Communication Systems: Steps Towards a Calculus of Processing Meaning *Journal of the Association for Information Science and Technology* 65 (2): 386–399.

Leydesdorff, L.–Park, H. W.–Lengyel, B. (2014): A Routine for Measuring Synergy in University-Industry-Government Relations: Mutual Information as a Triple-Helix and Quadruple-Helix Indicator *Scientometrics* 99 (1): 27–35.

Leydesdorff, L.–Perevodchikov, E.–Uvarov, A. (2015): Measuring Triple-Helix Synergy in the Russian Innovation Systems at Regional, Provincial, and National Levels *Journal of the Association for Information Science and Technology* 66 (6): 1229–1238.

Leydesdorff, L.–Strand, Ø. (2013): The Swedish system of innovation: regional synergies in a knowledge-based economy *Journal of the Association for Information Science and Technology* 64 (9): 1890–1902

Leydesdorff, L.–Sun, Y. (2009): National and International Dimensions of the Triple Helix in Japan: University-Industry-Government versus International Co-Authorship Relations *Journal of the American Society for Information Science and Technology* 60 (4): 778–788.

Loasby B. J. (2001): Time, knowledge and evolutionary dynamics: why connections matter *Journal of Evolutionary Economics* 11 (4): 393–412.

Louca, F. (2001): Measuring complexity: puzzles and tentative solutions In: Foster, J.–Metcalfe, J.S. (Eds.) *Frontiers of evolutionary economics: competition, self-organization and innovation policy* pp. 278–306., Edward Elgar, Cheltenham-Northampton.

Lundvall, B.-Å.–Johnson, B.–Andersen, E.S.–Dalum, B. (2002): National Systems of Production, Innovation and Competence Building *Research Policy* 31 (2): 213–231.

Majumdar, S.K. (2009): Technology transfer by foreign firms and the utilization of competencies within Indian industry *Journal of Technology Transfer* 34 (1): 95–117.

March, J.G. (1991): Exploration and exploitation in organizational learning *Organization Science* 2 (1): 71–87.

Martin, R.–Sunley, P. (2007): Complexity thinking and evolutionary economic geography *Journal of Economic Geography* 7 (5): 573–601.

McGill, W. J. (1954): Multivariate information transmission *Psychometrika* 19 (2): 97–116.

Mêgnigbêto, E. (2014): Efficiency, unused capacity and transmission power as indicators of the Triple Helix of university-industry-government relationships *Journal of Informetrics* 8 (1): 284–294.

Metcalfe, J. S.–Foster, J. (2004): *Evolution and Economic Complexity* Edward Elgar, Cheltenham–Northampton.

Narula, R. (2002): Innovation systems and 'inertia' in R&D location: Norwegian firms and the role of systemic lock-in *Research Policy* 31 (5): 795–816.

Radosevic, S. (2002): Regional Innovation Systems in Central and Eastern Europe: Determinants, Organizers and Alignments *Journal of Technology Transfer* 27 (1): 87–96.

Radosevic, S.–Reid, A. (2006): Innovation policy for a knowledge-based economy in Central and Eastern Europe: driver of growth or new layer of bureaucracy? In Piech, K.–Radosevic, S. (Eds.) *The knowledge-based*

REGIONAL STATISTICS, 2015, VOL 5, No1: 3–24; DOI: 10.15196/RS05101

economy in Central and Eastern European countries: countries and industries in a process of change pp. 295–313., Palgrave McMillan, London.

Rama, R. (2008): Foreign investment innovation: a review of selected policies *Journal of Technology Transfer* 33 (4): 353–363.

Shannon, C.E. (1948): A Mathematical theory of communication *Bell System Technical Journal* 27: 379–423 and 623–656.

Storper, M. (1997): *The Regional World – Territorial Development in a Global Economy* Guilford Press, New York.

Strand, Ø.–Leydesdorff, L. (2013): Where is synergy indicated in the Norwegian innovation system? Triple-Helix relations among technology, organization, and geography *Technological Forecasting and Social Change* 80 (3): 471–484.

Suurna, M.–Kattel, R. (2010): Europeanization of innovation policy in Central and Eastern Europe *Science and Public Policy* 37 (9): 646–664.

Theil, H. (1972): *Statistical Decomposition Analysis* North-Holland, Amsterdam and London.

Tödtling, F.–Trippl M. (2005): One size fits all? Towards a differentiated regional innovation policy approach *Research Policy* 34 (8): 1203–1219.

von Tunzelmann, N.–Nassehi, S. (2004). Technology policy, European Union enlargement, and economic, social and political sustainability *Science and Public Policy* 31 (6): 475–483.

Ulanowicz, R. E. (1986): *Growth and Development: Ecosystems Phenomenology* Springer-Verlag, New York.

Ulanowicz, R. E. (2009): The dual nature of ecosystem dynamics *Ecological modelling* 220 (16): 1886–1892.

Yeung, R. W. (2008): *Information Theory and Network Coding* Springer, New York.

FEDERICO BENASSI[a)] – MIRELA DEVA[b)] – DONATELLA ZINDATO[c)]

Graph Regionalization with Clustering and Partitioning: an Application for Daily Commuting Flows in Albania*

Abstract

The paper presents an original application of the recently proposed spatial data mining method named GraphRECAP on daily commuting flows using 2011 Albanian census data. Its aim is to identify several clusters of Albanian municipalities/communes; propose a classification of the Albanian territory based on daily commuting flows among municipalities/communes. Starting from 373 local units, we first applied a spatial clustering technique without imposing any constraining strategy. Based on the input variables, we obtained 16 clusters. In the second step of our analysis, we impose a set of constraining parameters to identify intermediate areas between the local level (municipality/commune) and the national one. We have defined 12 derived regions (same number as the actual Albanian prefectures but with different geographies). These derived regions are quite different from the traditional ones in terms of both geographical dimensions and boundaries.

Keywords: GraphRECAP, regionalization, daily commuting flows, census data, Albania, territorial imbalances.

Introduction

In the last decade, there has been a growing interest in modelling and understanding commuting behaviours (Eliasson-Lindgren-Westerlun 2003, Schwanen-Dieleman-Dijist 2004, Champion 2009) and the derived urban spatial structure (Ding 2007, Knox–McCarthy 2005). Particular interest has been devoted to the use of work commuting flows as a base to construct functional areas and to propose more efficient delimitations of the territory (Cövers–Hensen-Bongaerts 2009, Landré–Håkansson 2010, Rain 1999).

For the first time in Albanian history, the 2011 Population and Housing Census has collected data on commuting from home to work. Based on the use of these data at the municipality/commune level of analysis, this paper aims to i) identify several clusters of municipalities/communes, similar in terms of commuting profiles; ii) propose a

a) Department for Censuses, Statistical and Administrative Archives, Italian National Institute of Statistics, Rome, 00144, Italy. E-mail: benassi@istat.it

b) Cartography and GIS sector, National Institute of Statistics, Tirana, 1004, Albania. E-mail: mdeva@instat.gov.al

c) Department for Censuses, Statistical and Administrative Archives, Italian National Institute of Statistics, Rome, 00144, Italy. E-mail: zindato@istat.it

* The authors would like to thank Professor Nicola Salvati (University of Pisa) for reading a preliminary version of the paper and for providing us with many useful considerations.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

classification of the Albanian territory based on daily commuting flows among municipalities/communes.

To achieve our research objectives (i and ii), we applied a recently proposed spatial data mining method, named GraphRECAP (Guo 2010). GraphRECAP – Graph Regionalization with Clustering and Partitioning – is a toolkit for partitioning spatially embedded graphs (such as country-to-country migrations or commuting flows) and deriving spatially contiguous regions based on graph connections (Guo 2010). Readers are referred to Guo's publication for computational, technical and methodological details (Guo 2008, 2009, 2009a, 2010, 2010a).

Graph Regionalization with Clustering and Partitioning: the first exploratory analysis without imposing a spatial constraining strategy

We computed ten indicators for each of the 373 Albanian local units (municipalities/ communes). These indicators plus the spatial attributes of each municipality/commune are the input variables of GraphRECAP. The ten indicators are: 1) surface (square meters); 2) daily outflow degree: the number of destinations that have daily flows from each municipality; 3) daily outflow: the total daily volume (i.e. total daily commuters) going outside each municipality/commune; 4) adjusted daily outflow ratio: the ratio of the daily outflow and the usually resident population; 5) adjusted daily net flow ratio: the ratio of the daily net flow (daily inflow-daily outflow) and the usually resident population; 6) adjusted daily inflow ratio: the ratio between daily inflow and the usually resident population; 7) daily inflow degree: the number of destinations that have daily flows to each municipality/commune; 8) usually resident population; 9) daily inflow: the total daily volume (i.e. the total daily commuters) coming to each municipality/commune; 10) daily net flow: daily inflow – daily outflow.

On the basis of the standardized input variables, and thanks to the use of the SOM (Self-Organizing Map) technique, GraphRECAP identified 16 clusters of municipalities/ communes. This number is due to the dimension of the SOM (4*4), which was chosen out of the various alternatives, since it proved to ensure the best results in terms of cluster differentiation (each cluster has the highest possible inner homogeneity while heterogeneity among clusters is the highest possible). As mentioned previously, SOM stands for Self-Organizing Map and was first described by Kohonen (1995). A SOM is an artificial neural network that is trained by using an unsupervised learning process. The dimension of the SOM $(X \times Y)$ is normally defined by the user; we just recall that mathematically a 2D-SOM is a 3D-matrix of dimension with n the dimension of the input space. The SOM is composed of a neuron that is a cell characterized by its position in the plane of the SOM. The dimension of the neuron is n. The SOM is initialized randomly with a uniform distribution within the range of the input data. For each iteration, the algorithm randomly selects an input vector from the input space. It then computes the Euclidean distance between the input vector and each neuron of the map. The neuron with the minimal distance, which is called the Best Matching Unit (BMU), is therefore selected. After these two steps, the map is modified with the following formula:

$$M(t+1) = M(t) + \alpha(t) \times \theta(t, \beta_1, \beta_2) \times (V - W_{ij}(t)) \quad 1 \le i \le X, 1 \le j \le Y \qquad (1)$$

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

where $M(t)$ and $M(t+1)$ are, respectively, the maps at iteration *t* and at iteration *t+1*; $\alpha(t)$ is the learning rate that weights the effect of the input vector during the training process; $\theta(t,\beta_1,\beta_2)$ is the neighbourhood function that regulates the influence of Best Matching Unit (β_1,β_2) on the neighbouring neurons; and finally, $(V-W_{ij}(t))$ $1\le i\le X, 1\le j\le Y$ is a linear adjustment of the weights. In particular, the neighbourhood function is:

$$\theta(t,\beta_1,\beta_2)=\exp\left(-\frac{(i-\beta_1)^2+(j-\beta_2)^2}{2\sigma^2(t)}\right) \qquad (2)$$

In (2), σ it is the radius of the neighbourhood function, and it is equal to:

$$\sigma(t)=(\sigma(t_i)-\sigma(t_f))\times\exp\left(-\frac{t}{\lambda}\right)+\sigma(t_f) \qquad (3)$$

while α, the learning rate, is equal to:

$$\alpha(t)=(\alpha(t_i)-\alpha(t_f))\times\exp\left(-\frac{t}{\lambda}\right)+\alpha(t_f) \qquad (4)$$

The iterative procedure continues since each elementary unit is assigned to a certain cluster. The results are finally presented in a U-matrix (Unified distance matrix). This matrix is the mean Euclidean distance of each neuron with their *n* neighbours. That is to say:

$$U=\frac{1}{n}\sum_{\mu\, inN(v)}d(v,\mu) \qquad (5)$$

From a graphical perspective, each SOM node (cluster) is represented by a circle, whose size (area) is proportional to the number of elementary units (municipalities/ communes) that it contains. As was mentioned, SOM uses Euclidean distance to assess the multivariate similarity between spatial objects. Therefore, nearby clusters are more similar to each other than those far away (Guo–Gahegan–MacEachren–Zou 2005, Kohonen 1995). Behind the SOM nodes, there is a U-Matrix layer where hexagons are shaded to show the multivariate dissimilarity between neighbouring nodes, with darker tones representing greater dissimilarity (Guo 2010a).

Furthermore, a Parallel Coordinate Plot (PCP) is used to reveal the meaning of each municipality/commune assigned to each cluster by SOM, since on this plot we can observe the clusters' statistical profile related to the statistical indicators used in the analysis, and their level of dissimilarity. Finally, the results are related to each other and visualized on a multivariate/interactive map.

In Figure 1, we can see the general result of this first step of the analysis. In the multivariate mapping (a), municipalities/communes with the same colours belong to the same cluster while, in the clustering with SOM (b), node hexagons (clusters of municipalities/communes) with similar colours present a lower level of dissimilarity. This difference can also be seen in the PCP (c), where lines of colours very different from each other refer to clusters that present divergent values of the input indicators (that is to say a

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

divergent "profile"). It should be noted that the thickness of the lines in (c), as the circumference of the node hexagons in (b), is proportional to the size of clusters.

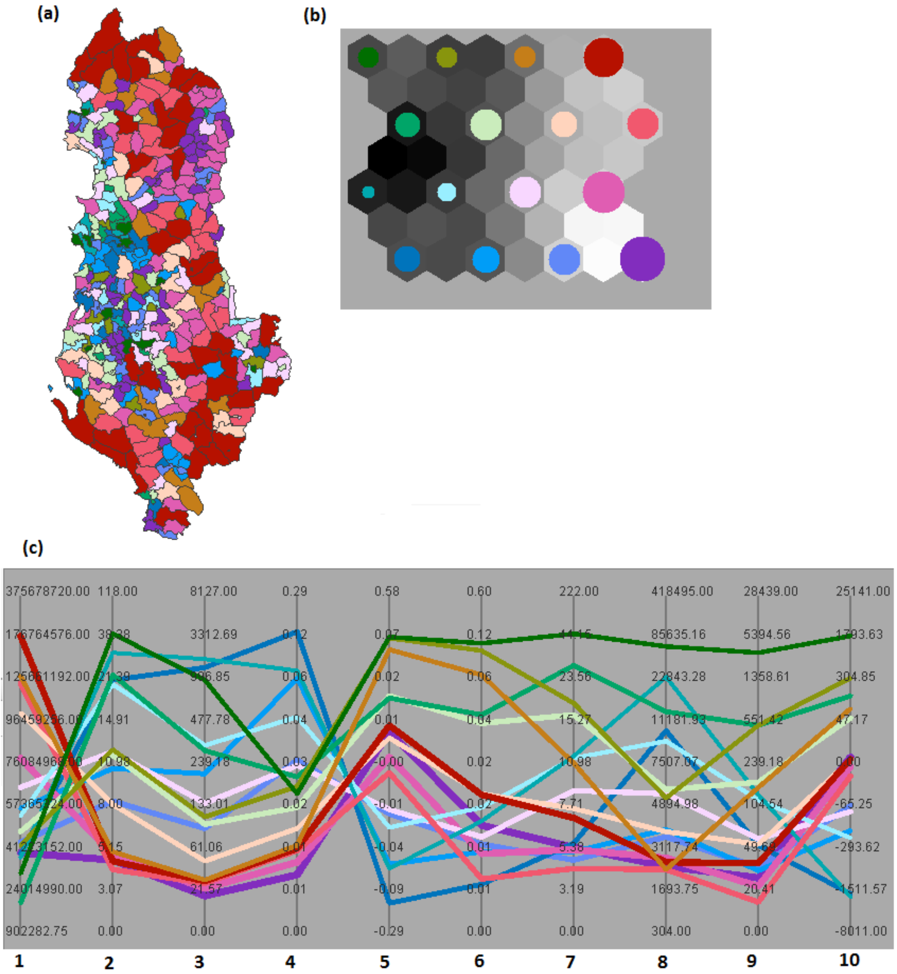
PCP is made up of as many parallel axes as the indicators used in the analysis. Every axis is scaled using the nested means method, which puts the mean value at the centre of the axis, thus making comparable axes defined by different units and different data ranges (Guo 2010a). This scaling method can alleviate overlapping problems in PCP for skewed data distributions. More specifically, a nested means method is a non-linear scaling that recursively calculates a number of mean values (and sub means) and uses these values as break point to divide each axis into equal-length segments.

An explorative analysis was carried out based on these first general results, with the aim of obtaining a further classification of the 16 clusters (and therefore of the basic units that belong to each cluster) in order to identify some primary groups. Based on the properties of the SOM (i.e. clusters with similar colours present low levels of dissimilarity among each other), the 16 clusters have been further grouped into 6 groups. We will now describe and discuss the main features of each of them. The first, Group 1, is characterized by clusters of municipalities/communes with a comparatively high value of indicator 1; low value of indicators 2, 3, 4, 8 and 9; medium values (except for the brown cluster) of indicators 5, 6 and 10. This group is therefore composed of clusters of municipalities/ communes with a large territory and a small usually resident population, characterized by a low level of daily spatial interactions (the levels of daily inflow and daily outflow are comparatively low). Looking at the spatial distribution of the municipalities/communes belonging to this first group, we can clearly see that almost all of them are located in the rural and mountain areas of Albania and on the coastal areas of the southwestern part of the country. It should be noted that the municipalities/communes belonging to this group are often territorially contiguous, thus showing the existence of spatial patterns. We can define this group as "Big/Peripheral" since it is composed of clusters of municipalities/communes that are big in terms of surface, yet peripheral with regard to their territorial location but also with regard to the role played in the daily interactions spatial system of Albania (Figure 2).

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

Figure 1

General results. (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)

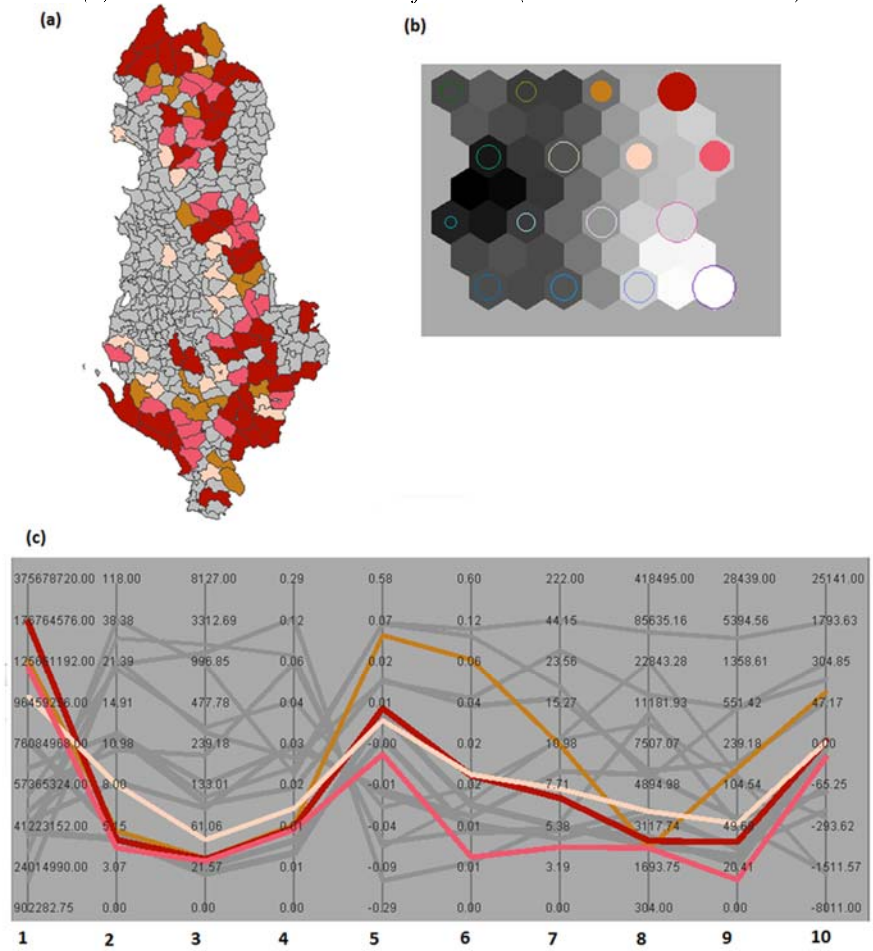


Source: own processing on Instat data, 2011 Population Census.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

Figure 2

Group 1, "Big/Peripheral". (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

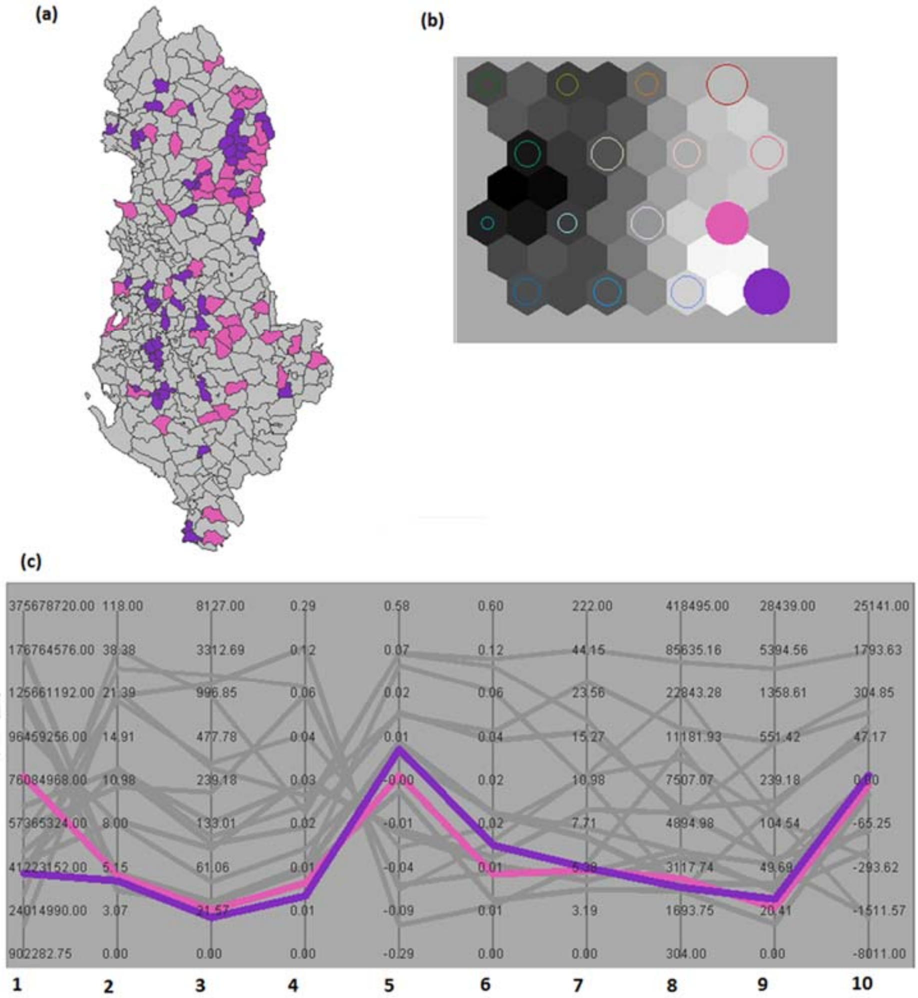
The second group, Group 2, is quite similar to the first one, except the values recorded for indicator 1 (surface). Municipalities/communes belonging to this second group are characterized by a small territorial area, by comparatively low levels of indicators 2, 3, 4, 6, 7, 8 and 9, and by medium values of indicators 5 and 10. We can define this group as "Small/Peripheral" since it is composed of municipalities/communes with a low level of daily spatial interactions, and which are comparatively small in terms of usually resident population and territory. The level of spatial contiguity among municipalities/communes belonging to this group is lower compared to that of municipalities/communes belonging

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

to group 1, but even in this group some municipalities/communes are territorially contiguous (namely, municipalities/communes located in the north-eastern part of the country). In terms of localization, this group may be divided into two categories; the first category is composed of municipalities/communes located in the north-eastern part of the country, with a certain level of spatial concentration and spatial contiguity among them. The second category, on the contrary, is composed of municipalities and communes, which are quite scattered and located mainly in the central and in the southern part of the country (Figure 3).

Figure 3

Group 2, "Small/Peripheral". (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

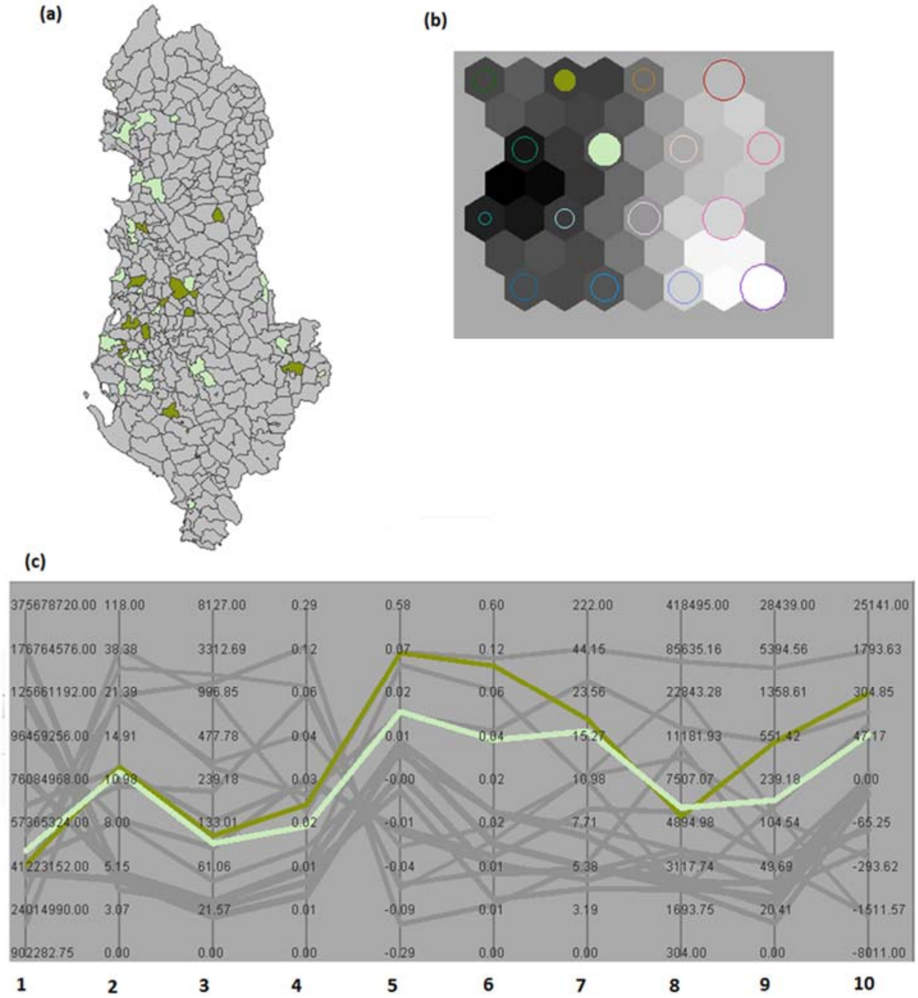
The third group, Group 3, is quite different from groups 1 and 2. We have defined this group as "Medium/Semi Central" since the municipalities/communes that belong to this group are characterized by almost medium values of indicators 1 and 3, medium values of indicators 2, 4 and 8, and by comparatively high values of indicators 5, 6, 7, 9 and 10. In other words, the municipalities/communes that belong to this group are characterized by a system of spatial daily interactions where the level of daily inflows is higher than the level of daily outflow (for this reason, the values of daily net flow and adjusted net flow ratio are quite high). The number of municipalities/communes from where the flows originate is quite high, while almost all other indicators reveal a medium situation, especially in terms of surface and usually resident population. The municipalities/communes of this group are therefore medium in terms of these two dimensions (surface and usually resident population) and play a semi-central role in the system of spatially daily interactions of Albania (Figure 4).

Group 4 is very different from those described until now. It is composed of municipalities/communes with a very low level of indicator 1, a medium value of indicator 4, and very high level of indicators 2, 3, 5, 6, 7, 8, 9 and 10. Consequently, the municipalities/communes belonging to this group are characterized by a very dynamic system of daily spatial interactions, where the level of the daily inflow is higher than the level of daily outflow. The number of municipalities/communes both originating and providing a destination for daily movements is very high, as well as the size of the usually resident population. On the contrary, the municipalities of this group are very small in terms of surface. We have defined this group as "Small/Central (Prey)". The municipalities of this group are, in fact, small in terms of surface, but play a crucial role in the system of daily spatial interactions of Albania. The term "prey" is adopted taking into account in a broad sense the logic and definitions of the prey-predator model elaborated by Lotka (1925) and Volterra (1926). This is because the municipalities/communes of this group (as we will see when describing the profile and characteristics of the next two groups) are predated by a number of other municipalities/communes that are very close to them in terms of spatial location, but, at the same time, they present divergent profiles compared to the them. In terms of spatial location, we can clearly see that, in this case, the condition of spatial contiguity is not confirmed. Municipalities belonging to this group are quite scattered in terms of spatial location; finally, it should be noted that the main Albanian municipalities (Tirana, Durrës, Vlorë, Elbasan and Shkodër) belong to this group (Figure 5).

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

Figure 4

Group 3, "Medium/Semi Central". (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)

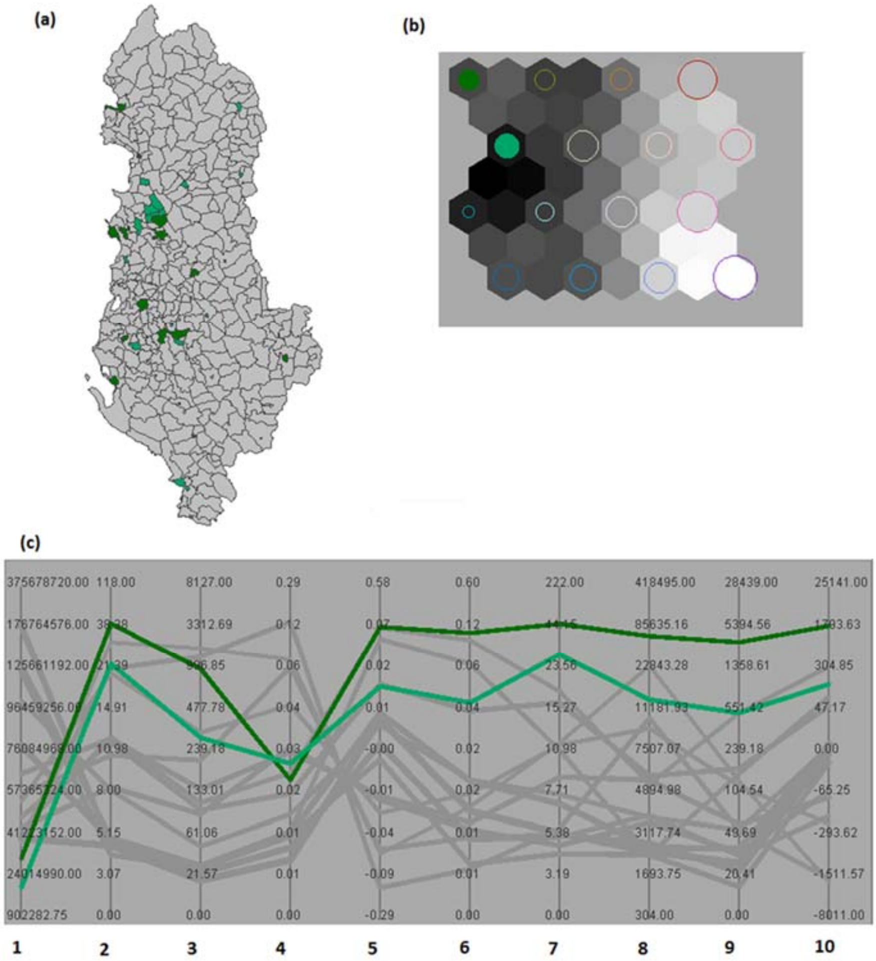


Source: own processing on Instat data, 2011 Population Census.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

Figure 5

Group 4, "Small/Central (prey)". (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

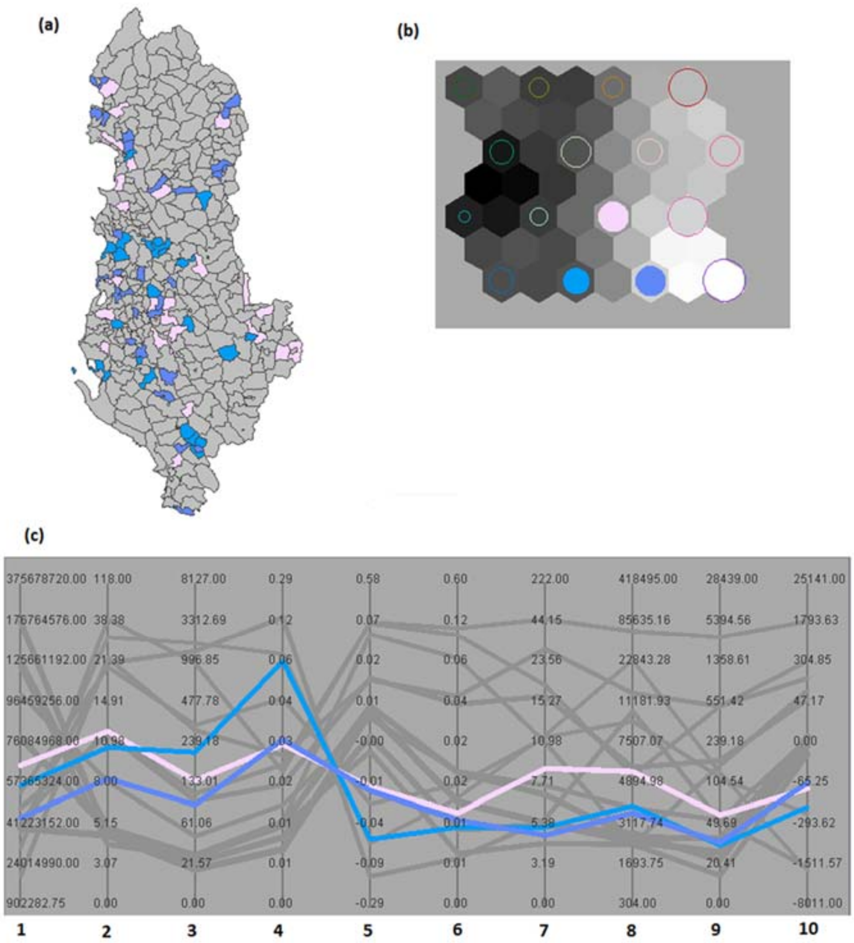
Group 5, "Medium/Central (semi-predators)", is composed of municipalities/ communes that have medium values of the indicators 1 and 2; medium/high values of the indicators 3 and 4 and, finally, comparatively low values of indicators 5, 6, 7, 8, 9 and 10 (with the exception, in some cases, of the pink cluster of municipalities/communes). That is to say that this group is characterized by a low level of daily inflow and a comparatively high level of daily outflow. This is also why the level of daily net flow and adjusted daily net flow ratio is comparatively low. Looking at the spatial location of the municipalities/communes that belong to this group, we can see from the map that they are often close to the municipalities belonging to group 4 (prey). For such a reason, we defined

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

this group as "Medium/Central (semi predators)". The municipalities belonging to this group have a medium-small dimension in terms of usually resident population and in terms of surface. They play a central role in the Albanian system of daily spatial interactions. A role that can be defined as that of "semi-predators" in that they present a quite low level of daily inflows and a relatively high level of outflows together with a spatial distribution that underlines that they are usually not so far from the prey (Figure 6).

Figure 6

Group 5, "Medium/Central (semi predators)". (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

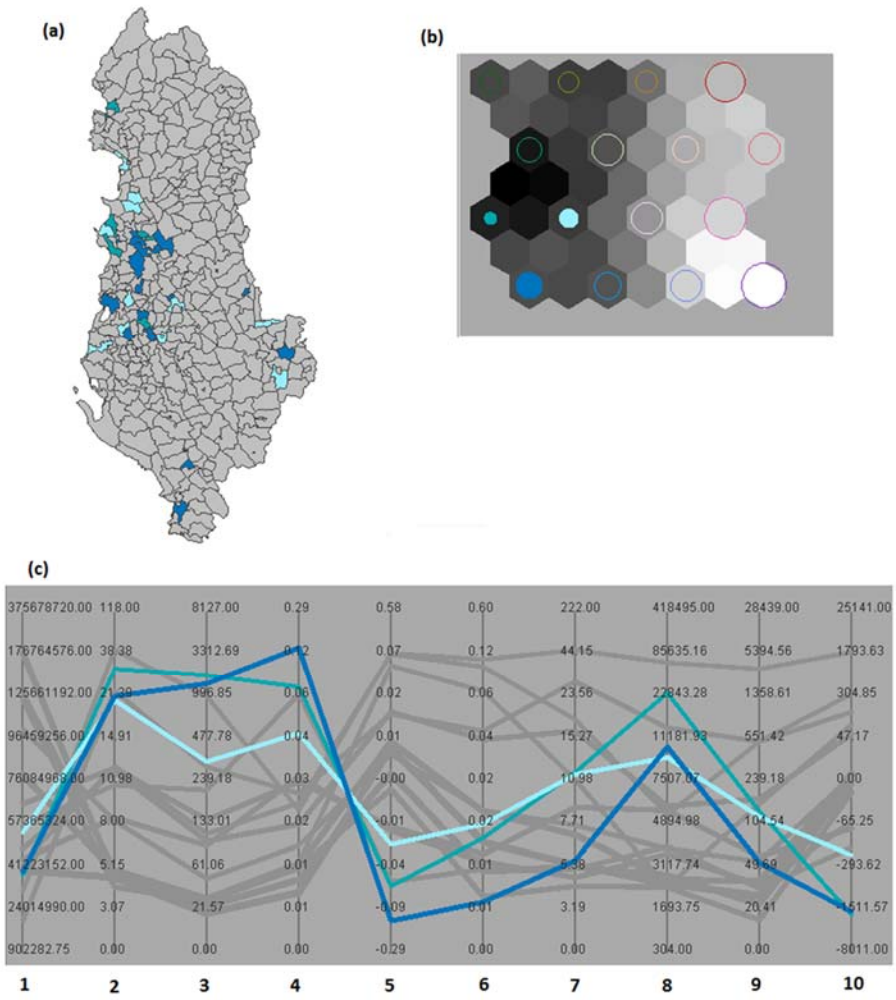
The last group, "Small/Central (predators)", clarifies the previously mentioned concepts of prey and predators. This group is characterized by very high levels of indicators 2, 3, 4 and 8; by a low level of indicators 1, 5, 6, 9, 10 and by medium/low levels of

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

indicators 7. That is to say that, in the municipalities belonging to the clusters that constitute this group, the level of total daily flow (daily inflow + daily outflow) is relatively high, but with a clear prevalence of daily outflows. The level of daily inflow is, in fact, lower compared to the level of daily outflow. Municipalities of this group are relatively big in terms of usually resident population and relatively small in terms of territory. Looking at the spatial location of the municipalities belonging to this group, we can clearly see that they are mainly located close to the prey municipalities, with a high level of spatial contiguity. It is especially the case of Durrës, Tirana and Shkodër (Figure 7).

Figure 7

Group 6, "Small/Central (predators)". (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

Graph Regionalization with Clustering and Partitioning: Identifying Natural Regions By Applying a Spatial Constraining Strategy

In the second step of our analysis, we used a spatially constrained graph partitioning technique to identify a hierarchy of natural regions defined by spatial interactions. The natural (derived) regions so identified, under a set of constrained parameters, are composed of clusters of municipalities/communes that are both spatially contiguous and homogeneous in terms of the characteristics of their daily spatial interactions. At a territorial level, the natural regions are intermediate areas between the local level (municipalities/communes) and the national one.

This kind of spatial statistical analysis belongs to the class of regionalization methods. As defined in Guo (2008), regionalization is a process that divides a large set of spatial objects into a number of spatially contiguous regions, while optimizing an objective function, typically a homogeneity (or heterogeneity) measure of the identified regions. Therefore, regionalization is a special kind of spatial clustering where the condition of spatial contiguity between spatial objects plays a key role.

As recalled by Bernetti–Ciampi–Sacell–Marinelli (2011), regionalization processes play an important part in many research sectors, finding applications in areas like climatic zoning (Fovell–Fovell 1993, Wang–Zhang–Li-Song 2010), environmental analysis (Henderson 2006, Romano–Balzanella–Verde 2010), landscape analysis (Long–Nelson–Wulder 2010), the interpretation and organization of Census data (Openshaw–Rao 1995) and public health data (Haining–Wise–Blake 1994, Osnes 1999), the analysis of socio-economic phenomena (Assuncão–Neves–Câmara–Da Costa Freitas 2006), the analysis and interpretation of demographic and urban/regional dynamics (Behnisch–Ulsch 2010, Benassi–Bocci–Petrucci 2013) and the analysis of migration flows (Guo 2009). The concept of regionalization hypothesized and applied to socio-economic entities by Openshaw (1977) results in the creation of geographic objects formed by combining contiguous elements sharing one or more characteristics and it is closely connected with spatial statistics (Bernetti–Ciampi–Saccelli–Marinelli 2011).

The starting point is that, following Guo (2010, 2010a), spatial interactions naturally form a network/graph, where each node is a location (or area) and each link is an interaction between two nodes (location). Such spatial interaction networks (e.g. municipalities/communes to municipalities/communes daily commuting flows) normally consist of: *S*, a set of locations (nodes), in our case the municipalities/communes of Albania; *F*, a set of flows (links) between locations, in our case the daily spatial commuting flows (direct) among Albanian municipalities/communes; and *Vf*, a set of variables for each flow. From this perspective, regionalization can reduce spurious data variations caused by uneven sizes or small base populations, and generalize (i.e. find general rules in) large spatial interactions data to discover general flow patterns. The key requirement is that the regionalization process should allow major patterns in the network to be preserved while suppressing details (Guo 2010, 2010a).

Coming back to our study, our aim is to identify *n* spatial areas (natural regions), intermediate between the local and the national level, that, under a constrained strategy, will minimize inner heterogeneity (within regions) and maximize external heterogeneity (between regions) with regard to daily spatial interactions.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

The key challenge of this operation is to identify regions based on commuting flows (this is why we call them natural or derived regions) instead of using pre-defined political or administrative boundaries.

After computing a contiguity matrix, which specifies which items (municipalities/communes) are neighbours in space, in order to complete the regionalization process, we have to define a constraining strategy and a set of parameters. Referring to the work of Guo for in-depth methodological details (2008, 2010, 2010a), we will now describe the constraining strategy and parameters adopted. As a regionalization method, the Full Order-ALK method has been chosen, which is a combination of the agglomerative clustering method named ALK (Average Linkage Clustering) and the spatial constraining strategy named Full Order. The ALK method derives natural regions in two steps. It first constructs a hierarchy of clusters from the bottom by iteratively merging the most connected clusters. Therefore, the method needs a contiguity matrix as input. The output is a spatially contiguous tree, where each edge connects two geographic neighbours and the entire tree is consistent with the cluster hierarchy. Second, the spatially contiguous tree is partitioned from the top by finding the best edge to remove. By repeating this step for each new region, a hierarchy of regions is constructed. During this partitioning process, additional constraints may be enforced; for example, we may want to impose a minimum population size for each region (Guo 2010a). Guo proves in his work (2008) that this method derives regions of significantly better quality (in terms of the objective function value) than other existing methods.

As a flow expectation model, we chose the Expectation SI_FLOW; this model calculates an expected flow value for each pair of spatial objects based on the total in and out flows of each object (in our case the total in and out daily flows of each Albanian municipality/commune).

Finally, to derive regions from the spatial interactions flows, a measure of similarity of the strength of connection has to be defined for each pair of locations (or regions). Following the work of Guo (2010a), in this paper we adopt the concept of modularity measure, which is defined by the following equation:

$$Modularity = Actual \ Flows - Expected \ Flows \qquad (6)$$

Different statistical models can be used to calculate expectation flows. In this paper, the simplest model is used, which assumes that interactions among locations are random and proportional to the origin and destination populations (Guo 2010a). In our case, we assume that each individual has the same probability to commute and the choice of the destination is proportional to the population of the destination place:

$$Expected \ Flows \ (A, B) = P_A P_B F \ / \ P_S^2 \qquad (7)$$

where P_s is the total population for all locations S, P_A is the population of the region $A \subseteq S$, P_B is the population of $B \subseteq S$, $A \cap S = \emptyset$, and F is the total flow among all locations (including flows within the same location). In this way, we ensure that the total expected flows are the same as the total actual flows (Guo 2010a).

In addition to these parameters, we impose two additional constraints during the partitioning process. We have fixed a maximum number of regions equal to 12 (the actual number of Albanian prefectures) and a minimum size of the usually resident population

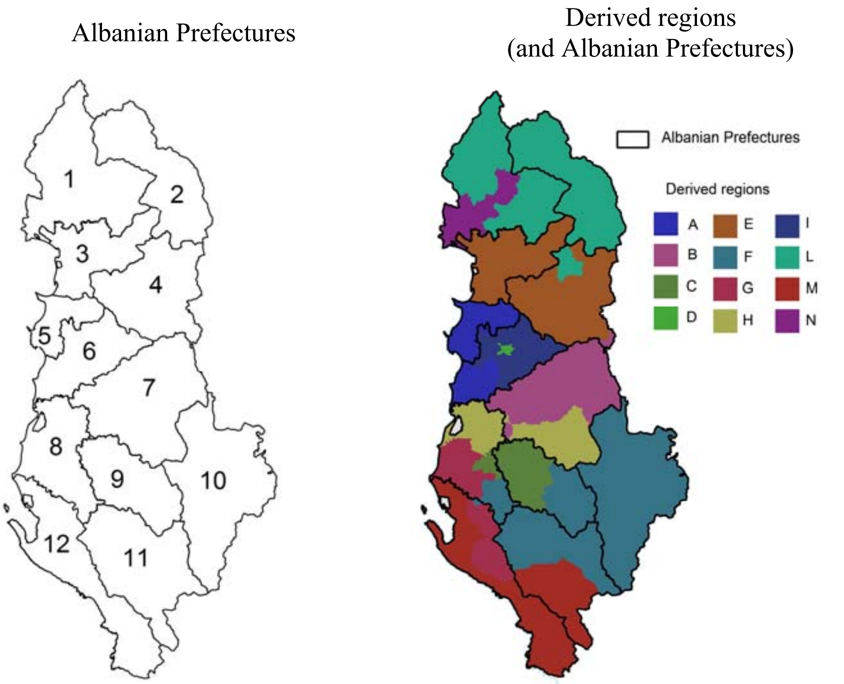
REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

per region equal to 150,000 units. With these parameters and under the constraint of spatial contiguity, we identified 12 natural regions (Figure 8).

As can clearly be observed from Figure 8, the derived regions are quite different from the traditionally used Albanian prefectures, both in terms of geographical dimension (some are smaller, others are bigger) and, obviously, in terms of territorial boundaries. Combining the geography of the derived regions with the commuting profiles of the elementary units of which they are composed (municipalities/communes), we can define the nature of each derived region inside the system of daily spatial interactions of Albania.

Figure 8

Albanian prefectures and derived regions



Note: Albanian prefectures: 1 Shkodër; 2 Kukës; 3 Lezhë; 4 Dibër; 5 Durrës; 6 Tirana; 7 Elbasan; 8 Fiër; 9 Berat; 10 Korçë; 11 Gjirokastër; 12 Vlorë.

Source: own processing on Instat data, 2011 Population Census.

From this perspective, we found 5 (E, F, G, L, M) derived regions characterized by a peripheral commuting profile; 4 derived regions (B, C, H, N) characterized by a semi-central commuting profile and finally, 3 derived regions (A, D, I) characterized by a central profile (Table 1). The derived regions with a peripheral commuting profile are almost all located on the mountains and the border areas of the country and, in general, in the less urbanized part of Albania. They are composed of medium/big municipalities in terms of surface, with comparatively small populations, which play a marginal role in terms of attraction and repulsion of commuting flows. The derived regions with a semi-central commuting profile are, on the contrary, located in areas with a high level of urbanization

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

in the coastal part but also on the western, central and northern part of Albania. Finally, the derived regions with a central commuting profile are located in areas with a comparatively high level of urbanization. We refer in particular to Durrës and its surrounding area, and to Tirana and its surrounding area. These derived regions are characterized by high levels of daily commuting inflows and outflows and qualify themselves as primary players in the Albanian system of daily spatial interactions.

Table 1

Synoptic table: derived regions, traditional components of the derived regions, commuting profiles of the derived regions, geographical locations of the derived regions

Derived regions	Traditional components of the derived regions	Commuting profile of the derived regions	Geographical locations of the derived regions
A	Durrës (5) + part of Tirana (6)	Central profile	Western/coastal areas in the central part of Albania
B	Part of Elbasan (7)	Semi central profile	Eastern/ central areas of Albania
C	Part of Fier (8) + part of Berat (9)	Semi central profile	Western coastal and central part of Albania
D	Part of Tirana (6)	Central profile	Western/central part of Albania (municipality of Tirana)
E	Lezhë (3) + part of Dibër (4)	Peripheral profile	Eastern and Western part of Albania/Central-North areas
F	Korçë (10) + part of Fier (8) + part of Berat (9) + part of Gjirokastër (11)	Peripheral profile	Southern/east and Southern/central part of Albania
G	Part of Vlorë (12) + part of Fier (8)	Peripheral profile	Southern/west coastal area in south part of Albania
H	Part of Fier (8) + part of Berat (9)	Semi central profile	Western coastal/central part of Albania
I	Part of Tirana (6)	Central profile	Central part of Albania (surrounding areas of Municipality of Tirana)
L	Kukës (2)+ part of Shkodër (1)	Peripheral profile	Eastern and Western North part of Albania/mountain areas
M	Part of Vlorë (12) + Part of Gjirokastrër (11)	Peripheral profile	South/western coastal part of Albania
N	Part of Shkodër (1)	Semi central profile	North Western coastal part

Conclusions

The application of the recently proposed method GraphRECAP on daily commuting flows of Albania has obtained interesting results.

Starting from 373 local units (municipalities/communes of Albania) and, therefore, from a 373*373 square matrix of daily commuting flows, we first applied a spatial clustering technique, without imposing any constraining strategy, and obtained 16 clusters.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

These clusters were then further classified into sub-groups defined by a number of demographic and territorial dimensions and by the role played in the commuting system of Albania. We thus identified 6 primary sub-groups: 1) Big/Peripheral; 2) Small/Peripheral; 3) Medium/Semi central; 4) Small/Central (Prey); 5) Medium/Central (semi-predators); 6) Small/Central (predators).

In the second step of our analysis, we imposed a set of constraining parameters to identify intermediate areas between the local level (municipality/commune) and the national one. We have defined 12 derived regions, the same number as Albanian prefectures. Although, these derived regions are quite different from the administrative ones both in terms of geographical dimensions and boundaries.

In our opinion, the derived regions (as well as the 6 subgroups) effectively represent the territorial and demographic imbalances that characterize today's Albania. Here, mountainous and less developed areas are composed of a comparatively large number of municipalities/communes that are quite vast in terms of surface but with a very low demographic density and with a marginal role in the commuting system. These municipalities/communes are characterized by a depopulation process where the population migrates to other areas and in particular, to the bigger urban centres (Tirana and Durrës) and their surrounding areas. This is why these areas (Tirana and Durrës) have been identified as specifically derived regions with a high level of commuting activities. The surrounding areas of these are characterized by a semi-central commuting profile and by an intermediate situation between the two described.

The largest municipalities (Tirana, Durrës, Vlorë, Elbasan, Shkodër, Fier, Korçë) act both as attraction poles and as poles from which a daily redistribution of commuting workers takes place. Altogether, large municipalities and municipalities/communes in their surrounding areas form complex systems of daily mobility, which in some cases are also linked to each other (Tirana/Durrës; Tirana/Elbasan; Tirana/Fier).

The leading role of large municipalities is confirmed and further clarified by the results of the multivariate spatial analysis of daily commuting flows. More precisely, they appear as playing a crucial central role, but also as "prey" predated by a number of other municipalities/communes (that we could define "predators"), which are very close to them in terms of spatial location and that present quite divergent profiles in terms of the indicators chosen for the analysis (e.g. they are characterized by a low level of the daily inflow and by a comparatively high level of the daily outflow; they are also medium/small size in terms of usually resident population and in terms of surface). From a commuting perspective of analysis, it seems reasonable to define Albania as a dual system. A system in which one part plays a primary role (A, B, C, D H, I, N derived regions) and the other part, which seems to play a very marginal role (E, F, G, L, M, derived regions). In the author's opinion, these dynamics could reinforce territorial imbalances and have negative repercussions on Albania's sustainable development.

Summary

This paper presents an original application of GraphRECAP, a recently proposed method of Graph Regionalization with Clustering and Partitioning elaborated by Guo (2010). The application concerns the study of daily spatial interactions (work commuting flows) among

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

the Albanian municipalities/communes, based on the use of Population Census data (2011), which for the first time investigated this phenomenon.

The study firstly defines clusters of municipalities/communes similar in terms of commuting profiles and then, through the imposition of a set of constrained strategies, proposes a new kind of regionalization of the Albanian territory. The results clearly show the actual territorial and demographic imbalances of Albanian society. These imbalances seem to be reinforced by the spatial patterns of the commuting flows. A dual territorial space comes to light, in which largest municipalities act both as attraction poles and as poles from which a daily redistribution of commuting workers takes place, while the smallest and peripheral municipalities/communes play the role of origin areas of commuting flows.

REFERENCES

Assuncão, R. M.–Neves, M. C.–Câmara, G.–Da Costa Freitas, C. (2006): Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees *International Journal of Geographical Information Sciences* 20 (7): 797–811.

Behnisch, M.–Ultsch, A. (2010): *Are there clusters of communities with the same dynamic behavior?,* Classification as a tool for research, Springer, Verlag.

Benassi, F.–Bocci, C.–Petrucci, A. (2013): *Spatial data mining for clustering: an application to the Florentine Metropolitan Area using RedCap*, Classification and Data Mining, Springer, Berlin/Heidelberg/New York.

Bernetti, I.–Ciampi C.–Sacelli, S.–Marinelli, A. (2011): The planning of agro-energetics districts. An Analysis model for Tuscany Region *Italian Journal of Forest and Mountains Environments* 66 (4): 305–320.

Champion, T. (2009): Urban-Rural Differences in Commuting in England: A Challenge to the Rural Sustainability Agenda? *Planning Practice & Research* 24 (2): 161–183.

Cövers, F.–Hensen, M.–Bongaerts, D. (2009): Delimitation and Coherence of Functional Administrative Regions *Regional Studies* 24 (2): 161–183.

Ding, C. (2007): *Urban Spatial Planning: Theory, Method and Practice* Higher Education Press, Beijing.

Eliasson, K.–Lindgren, U.–Westerlund, O. (2003): Geographical Labour Mobility: Migration or Commuting *Regional Studies* 37 (8): 827–837.

Fovell, R.G.–Fovell, M.Y.C. (1993): Climate zones of the conterminous United States defined using cluster analysis *Journal of Climate* 6 (11): 2103–2135.

Guo, D.–Gahegan, M.–MacEachren, A. M.–Zhou, B. (2005): Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach *Cartography and Geographic Information Science* 32 (2): 113–132.

Guo, D. (2008): Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP) *International Journal of Geographical Informative Sciences* 22 (7): 801–823.

Guo, D. (2009): Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data *Transaction on Visualization and Computer Graphics* 15 (6): 1041–1048.

Guo, D. (2009a): *Greedy Optimization for Contiguity-Constrained Hierarchical Clustering*, Proc., 4th International Workshop on Spatial and Spatio Temporal Data Mining, Miami, Florida.

Guo, D. (2010): *GraphRECAP: A Toolkit for Spatially Constrained Graph Partition* User Manual [Online]. Available: www.SpatialDataMining.net

Guo, D. (2010a): Flow Mapping with Graph Partitioning and Regionalization. User Manual [Online]. Available: www.SpatialDataMining.net

Haining, R. P.–Wise, S. M.–Blake, M. (1994): Constructing regions for small area analysis: material deprivation and colorectal cancer *Journal of Public Health Medicine* 16 (4): 429–438.

Henderson, B. (2006): Exploring between site in water quality trends: a functional data analysis approach *Environmetrics* 17 (1): 65–80.

REGIONAL STATISTICS, 2015, VOL 5, No1: 25–43; DOI: 10.15196/RS05102

Knox, P. L.–McCarthy, L. (2005): *Urbanization: An Introduction to Urban Geography* 2nd Edition, Pearson, Toronto.

Kohonen, T. (1995): *Self-Organizing Maps* Springer, New York.

Landré, M.–Håkansson J. (2010): Rule versus interaction Function: Evaluating Regional Aggregations of Commuting Flows in Sweden *EJTIR* 13 (1): 1–19.

Long J.–Nelson T.–Wulder, M. (2010): Regionalization of landscape pattern indices using multivariate cluster analysis *Environmental Management* 46 (1): 134–142.

Lotka, A.J. (1925): *Elements of Physical Biology* Williams & Willikins, Baltimore.

Openshaw, S. (1977): A geographical solution to scale and aggregation problems in region building, partitioning and spatial modelling *Transaction of the Institute of British Geographers* 2 (4): 259–472.

Openshaw, S.–Rao, L. (1995): Algorithms for reengineering 1991 Census Geography *Environmental and Planning* 27: 425–446.

Osnes, K. (1999): Iterative random aggregation of small units regional measures of spatial autocorrelation of cluster localization *Statistics in Medicine* 18 (6): 707–725.

Rain, D. R. (1999): Commuting Directionality, a Functional Measure For Metropolitan and Non Metropolitan Areas Standards *Urban Geography* 20 (8): 749–767.

Romano, E.–Balzanella, A.–Verde, R. (2010): *A new regionalization method of spatially dependent functional data based on local variogram models: an application on environmental data* Proc., 45th Scientific Meeting of the Italian Statistical Society, Padua, Italy.

Schwanen, T.–Dieleman, F. M.–Dijst, M. (2004): The Impact of Metropolitan Structure on Commute Behavior in the Nederland's: A Multilevel Approach *Growth and Change* 35 (3): 304–333.

Volterra, V. (1926): *Variazioni e Fluttuazioni del Numero d'Individui in Specie Animali Conviventi* Regia Accademia dei Lincei, Roma.

Wang, H.–Zhang, X.–Li, S.–Song, X. (2010): Spatial clustering for the regionalization of maize cultivation in China and its outlier analysis *Transactions on Informative Science and Applications* 7 (6): 860–890.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

NORBERT SZABÓ[a)]

Methods for regionalizing input-output tables

Abstract

The present paper introduces the most common methods of regionalizing national input-output tables. First we describe the different groups of methods based on our review of the international literature regarding regionalization. Then we focus on particular methods that can be applied for Hungarian counties highlighting their advantages and disadvantages and synthetize the empirical results of them again based on the literature. On the basis of these experiences we attempt to create a complex method fitted to the available Hungarian regional data. For better understanding in the end we apply our method on an illustrative example consisting of three regions with hypothetical sectors and data.

Keywords: input-output table, regionalization, interregional trade.

Introduction

The aim of current study is to introduce estimation methods of the regional input-output framework. Our starting point is the I-O table itself, which is capable of depicting the interindustry relationships of the whole economy. Roots of the theory can be dated back to the 1930s, to Leontief's (1934, 1941) pioneer work. Thanks to the rapid development of information technology it became possible for his contribution (the table and other derived indicators) to be applied to the analysis of different problems worldwide. The approach quickly gained a degree of popularity among researchers; today, more and more examinations have been conducted with new extensions (energy and environmental block), at different levels (national, regional, county).

The literature of national level research is enormous but because of spreading regional science, the focus of I-O research has been shifted from the national level to the smaller local level. The greatest disadvantage of purely national analysis is that they hide certain regional differences; they consider the whole economy as an aspatial entity. However, regional differences are not negligible at all. Moreover, the interregional trade connecting the regions of the economy plays a significantly different role in periphery and central regions. The consequence of taking into account aspects like these was that the initially national methods' regionalized versions soon appeared. In principle, there were no significant obstacles but during practical applications, researchers faced many problems. Since most European countries publish only national I-O tables, there is no detailed information about the inter-industry relations of regions, which would serve as the basis of much regional research. This is the main reason several regionalization methods were

a) MTA-PTE Innovation and Economic Growth Research Group H-7622, Pécs Rákóczi út 80., Hungary. E-mail: szabon@ktk.pte.hu

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

elaborated that were capable of generating estimates of regional I-O relations. Since their first appearances, countless examples can be found in the literature. Most of them engaged with estimating regional tables, comparing them with actual empirical data and evaluating different methods highlighting their advantages and disadvantages.

This study has three aims: 1) introduce the regionalization methods of national input-output tables, highlighting those that are not too resource intensive and can also be applied to Hungary; 2) introduce the different possibilities of estimation of interregional trade flows; 3) propose a suitable method for generating multiregional input-output table in Hungary.

Introduction of the I-O table, the core of input-output analysis

The aim of I-O tables is to depict the interrelated system of economic processes, production, consumption and savings. The input-output table is able to "depict the relationship between different sectors of the national economy, and the structural connection of production and final demand in a consistent manner" (KSH, 2005, p. 5). A basic requirement of the table is symmetry, which means that sectoral output and use have to be equal. The fundamental component is the intermediate transaction table, which depicts the flow of output between producer and user sectors. In Figure 1, it is matrix *Z* where the rows indicate the distribution of output produced by different sectors among other industries. On the other hand, columns indicate the combination of inputs required by each industry to produce its output.

Besides this, one can find additional rows and columns in the table. Supply can be used not only by industries (as intermediate use) but by final users as well. These transactions are not used directly in production processes as inputs (such as final demand by households and government, investment, export). In Figure 1, it is depicted by matrix *FD*. The additional rows consist of different value-added factors, and the further input requirements of industries, which are not provided by other industries (such as labor, capital inputs, indirect taxes, import).

All cells in the table can be interpreted as a flow of product in forints (millions) during a single year. In principle it is possible to assemble a physical input-output table that would be a better indicator of material use; However, the data requirements of such a table are enormous and the comparison between different industries would be difficult (small city car vs. truck).

Figure 1

The scheme of an input-output table

	Industries (as users)	Final demand	Total output
Industries (as producers)	**Z** "inner square"	**FD** "side wing"	**x**
Value added	**WA** "lower wing"		
Total input	**x**		

Source: Zalai (2012) p. 177.

So the mentioned matrices together are called an input-output table. The national tables can be classified into groups on the basis of many criteria. One of them is the way tables

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

depict imports. Imports and domestic products can be represented together. Thus, all domestic supply is considered in the table (type *A* and *C*). On the other hand, the distribution of the pure domestic production can be represented. In this case, imports are depicted in a separate row in the lower part of the table. Thus the three basic types of matrices are:

- **Type A:** domestic activities are not distinguished from imports. International trade is represented as a net export.
- **Type B:** rows depict only the pure domestic supply, and imports are accounted for in a separate row (by user industries) with no further detailed information about the sectorial source of the imports.
- **Type C:** in matrix *Z,* rows of production can separately contain both imports and domestic activities.

Regionalization of input-output tables

The aim of the application of input-output tables and models was the estimation of expected economic effects of national shocks. To date, this approach has maintained its popularity. However due to development of science in the second half of 20th century, especially from the 1960s, the intensifying interest for regional applications and for other different related regional fields of regional science has led to modifications and extensions of the original I-O framework. One of the most important points is the adjustment of the methodology to smaller territorial units (region, county) by which it became suitable for depicting local economic characteristics. In most countries, this was (is) necessary because only national input-output tables are available, and the application of national coefficients in regional studies can severely distort results. In connection with this, there are two important points to be noted:

1) The national input coefficients can be interpreted as an average of regional technical coefficients. Thus, in principle, it is possible that the input structure of a region is identical to the country's as a whole, but can also be completely different. For example, electric energy production in Pécs is based on a thermal power plant and in Paks it is based on a nuclear power plant. Obviously the two technologies are significantly different.

2) A valid thumb rule is that the smaller a territorial unit (country, region or county,), the greater its dependence on external territories through trade. The self-supplying ability of regions is shrinking, which means, in input-output relations, that the intraregional coefficients are smaller because a part of the inputs are purchased from other regions via interregional imports. Thus a new element, the interregional trade (export and import) has to be accounted for in the table.

At this stage, it is important to distinguish two kinds of coefficients that can be the result of regionalization (Miller and Blair, 2009):

1) regional input (A^{rr}) and
2) regional technical coefficients (A^{r}).

The input coefficient depicts the level of inputs used by industries of a region that were produced by industries in the same region. The technical coefficient represents the level of inputs used by local industries regardless of the source of the input. From now on, the current study focuses on input coefficients.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

There are several well-known variations of regionalization methods in the international literature. We cannot undertake the complete introduction of all methods; however, we will endeavor to describe more thoroughly the most important and most applied techniques. During the introduction, we will proceed from higher levels toward narrower groups of approaches to finally arrive at particular techniques.

Based on the underlying methodology, techniques can be divided into three fundamental groups (Greenstreet, 1989):

1) Survey-,
2) Non-survey and
3) Hybrid methods.

In case of survey methods, companies in the sample provide information about their sales (towards other companies and consumers) and their purchases (from other companies) within and outside the region. Industries in different regions can produce different products or use different inputs (like the case of thermal and nuclear power plant) for their production processes. Surveys can help researchers to take into consideration such special local characteristics. Thereby, the approach can create a precise picture of the technological and trade structure of the region; although, it has its drawbacks. 1) Usually, it is a very time and resource intensive process to conduct a survey. 2) The assembling of the sample is crucial; any mistakes at this point can lead to significant distortions in the final results. 3) Often, data gathered from companies is not balanced, and the results of balancing methods are sometimes not convergent (especially in case of smaller samples).

By applying non-survey methods, researchers can save time and resources. In this case, different estimation methods can generate regional tables (using national ones as a starting point). In many cases, the precision of procedures is enhanced by using additional regional data. A clear advantage of the non-survey approach is that these techniques can be implemented relatively easily if secondary data is available. Thus they are very cost-effective. Nevertheless, there is still no consensus about the "best" procedure to date; partially as a result of the questionable performance of different approaches. Furthermore, the lack of appropriate detailed regional data restricts the variety of potential methods.

The aim of the hybrid (or partial survey) approach is to combine the advantages of the other two mentioned approaches and to moderate their potential drawbacks. These procedures consist of several steps. The first of them is usually a non-survey model, which is augmented in further steps with survey data, expert estimates or information from other databases. Consequently, the hybrid approach is less resource demanding but still preserves the credibility of results.

The focus of science in this field has changed continuously during the last 60 years. During the 50s, survey, and later non-survey methods gained remarkable popularity; then, in the 60s, tables were mostly estimated by survey techniques. After that, the 70s saw an increasing number of non-survey studies conducted, which was the consequence of intensifying regional input-output research. Later, the step-by-step emphasis moved towards a hybrid approach.

By the millennium, it became clear that mainly hybrid and ready-made methods could generate reliable regional tables for a reasonable cost. The latter will be discussed later. Although hybrid techniques are characterized by cost-efficiency, relative precision and simplicity, non-survey methods did not disappear from regional research. In most

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

countries, hybrid methods are not applicable simply because there is not enough regional data available to run sophisticated models, or there are barriers that prevent the conducting of representative surveys. Fortunately, several new studies are promoting the potential further development of non-survey approaches.

Furthermore, methods can be classified by their choice of local unit:

1) Single region
2) Multiregional or Interregional methods

The detailed description of single region, multi- and interregional models are not part of this study; international literature has already discussed the topic in detail.

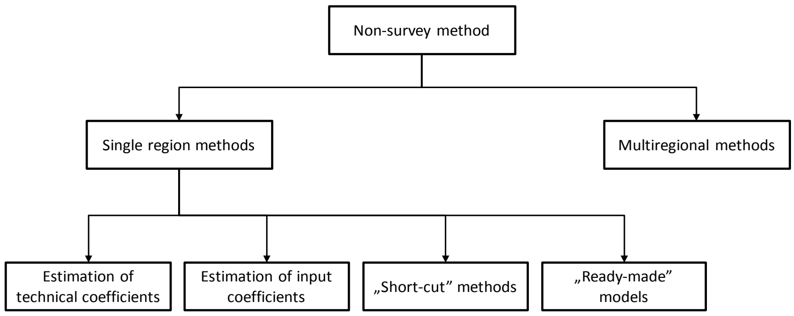
To choose the appropriate table, it is vital to match the aim of the research and model that will serve that goal. During our research, the initial intention was to estimate an interregional table that can depict how different industries are interconnected in a given region, moreover, how they are interconnected with industries in other regions. If the production of a region is affected by a policy shock, it will affect other regions through interregional linkages because regions sell to and buy from each other. Using such tables, we can take these spillover effects into account (Miller–Blair 2009).

Single region methods concentrate their attention on a specific local territory of a country. In this case, the region examined is separated from the rest of country. Suppose that the aim of our research is to estimate impacts of different policy investments in South-Transdanubia, Central Hungary and Western Transdanubia. The predicted impact of the single region models would fall short of the real effects. The reason is that regional interventions would affect the economy of other regions through interregional trade. Furthermore, the intervention would have further positive effects even in regions not affected by a policy shock (spillover effect). Thus, single region models are not capable of depicting the bilateral interconnections whilst a multiregional approach breaks down a nation into smaller regions and analyses both inter- and interregional linkages to give a more realistic picture of the economy.

During our research (for antecedents see: Varga et al., 2013) we did not consider the survey approach as a reasonable choice because of its high demand for time and resources. Based on the same reasoning, we did not choose partial non-survey methods either. In the rest of study, the focus is narrowed to purely non-survey methods. After Bonfiglio (2005), we can distinguish the following groups of non-survey methods:

Figure 2

The system of non-survey methods



Source: Bonfiglio (2005) p. 18

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

The first group consists of those methods that do not distinguish the flow of inputs between regions and within the region (technical coefficients). This group was dropped from the research because the aim was to depict interregional trade flows between regions. Since our interest is to estimate intra- and interregional flows as well, the second group of methods that seek to estimate regional input coefficients can be considered as an ideal choice. The purpose of short-cut methods is to predict the regional multiplicators without estimating the actual input-output coefficients. One of the most well-known examples is the RIMS (Regional Industrial Multiplier System) model, which was elaborated by Drake (1976). Our research requires the assembling of an input-output table because our further aim is to extend an estimated table with additional blocks (e.g. environmental). Finally, ready-made models are available as an integrated part of a purchasable software package. These packages consist of valuable regional data, a regionalization method, and a ready input-output table. The software themselves can estimate regional I-O tables, multiplicators, and forward and backward linkages. The most well-known examples are IMPLAN, REMI, RIMS II (Bonfiglio 2005).

In the next chapter, we will concentrate on those non-survey methods that are capable of estimating regional input coefficients. First we will describe single region methods, and then we will continue with methods of interregional trade. We will also place more emphasis on approaches that are suitable for Hungarian conditions.

Single region non-survey methods

The basis of most non-survey methods is a national input-output table since there is no other available source of data. In the first step, we assume that technology is the same across regions and likewise in the nation as a whole. Then there is a wide variety of available methods to generate an estimate of regional coefficients. In general, these methods modify the production technology of regions (rows). There is no competition between local producers and firms in other regions. Interregional trade occurs when the region is not able to cover its needs for a given regional activity; the opposite is also true if the supply of a region can meet its needs, then any potential surplus must be an interregional export to other parts of country. Such a net approach of interregional trade can result in a possible underestimation of interregional trade coefficients and an overestimation of intraregional input coefficients (Bonfiglio–Chelli 2008).

Regional Supply Percentage

Early works focused on using national coefficients with slight adjustments. An example of the earliest pioneer works is the regional supply percentage, which expresses the proportion available output that originated within the region and output regardless of its origin. The following equation is a general form of *RSP* index:

$$rsp_i^r = \frac{(x_i^r - e_i^r)}{(x_i^r - e_i^r + m_i^r)},$$

where the numerator depicts the difference of locally produced products (x_i^r) and regional exports (e_i^r), in the denominator one can find the amount of available products (including

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

imports(m_i^r)) in region *r*. The intraregional coefficient matrix can be derived by the simple multiplication of the diagonal matrix of sectorial *RSP*s and the national coefficient matrix:

$$A^{rr} = \widehat{rsp^r} A^n.$$

Modification across rows is a strict restriction, which means that we do not make a difference between national and regional technology. In other words, the proportion of input *i* used in production is the same at both territorial levels. A special variety of the method is the regional purchase coefficient (*RPC*), which is originated from Stevens et al. (1983). The calculation is almost the same as *RSP* but the biggest difference is the source of data. In this case, data is generated through an econometric estimation. Again exact details are not described here due to length limitations.

Since most non-survey methods estimate regional trade as a residue after consumption and intermediate use, which can generate serious distortion, the *RPC* method can be considered as a promising approach. Stevens et al. (1983) defined *RPC*s as a function of transport cost and unit cost of products. Despite its theoretical advantages, the approach did not gain popularity due to its high data requirements, which usually cannot be satisfied at regional level.

Methods based on location quotient

Another early group of methods is based on the well-known location quotient. The aim of the approach is to use *LQ* as an indicator of sectorial regional specialization and to use it to transform national technology into a regional one. There are plenty of different *LQ* variants in the international literature. In the next sub-chapters, we will describe the most important features of different *LQ* methods.

Simple location quotient

The first and most simple variant is the unmodified traditional location quotient. This ratio can be calculated as follows:

$$LQ_i^r = \left(\frac{\frac{x_i^r}{x^r}}{\frac{x_i^n}{x^n}} \right),$$

where x_i^r is the output of industry *i* in region *r,* x^r is the total output in region *r*. In the denominator x_i^n and x^n stand for the same reason but for the national level. To begin with, it is worth reviewing the background of the quotient. The numerator / denominator expresses to what extent the region/nation is specialized in production of industry *i*. Thus the ratio can be interpreted as the relative specialization of the region compared to the nation as a whole. If $LQ = 1.5$ it means that the share of output *i* in regional output in *r* is 1.5 times higher than at the national level. Thus the region is assumed to be more specialized in industry *i*. If LQ=0.8 then the opposite is true. The region is assumed to be less specialized in industry *i* compared to national level. In the regionalization we consider *LQ* as an indicator of regional self-sufficiency. As long as $LQ > 1$ the region is relatively specialized and it can cover its own demand locally; we do not adjust the national coefficients (a^n_{ij}). If $LQ < 1$ the region needs to import; in this case the regionalization is carried out by the following equation (Miller and Blair, 2009):

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

$$a_{ij}^{rr} = \begin{cases} (LQ_i^r)a_{ij}^n, & ha\ LQ_i^r < 1 \\ a_{ij}^n, & ha\ LQ_i^r \geq 1 \end{cases}.$$

The LQ method can also be used (in a similar manner) for estimation of final demand block. Of course, additional data is required (e.g. total regional final demand, total value added). In the absence of these, the use of simplistic assumptions is necessary to estimate these blocks too. Ultimately, if the total estimated intermediate use and final demand exceed / fall short of the regional output, further adjustments are required (see later: *RAS*). However, the method has several weaknesses. On the one hand, the quotient is asymmetric. If the value of the quotient is smaller than one, coefficients in the given row of the table should be adjusted. On the other hand, if the value is bigger than one, they are not adjusted at all. However, in reality, some regional coefficients can be bigger than their national counterparts since the national coefficients can be considered as an average of regional ones. Furthermore, the adjustment should be done across rows (producing sectors). In other words, less specialized sectors will be scaled down. Thus, the input structure will be changed due to changes in the volume of rows. However, if the input structure is different in regions, then there is a chance that the structure of sales may also be different; this aspect is neglected by the LQ approach. Further, the relative size of a region is not taken into consideration in the quotient, however, as we noted above, the relative size is in strong connection with the self-supplying ability of regions.

Since rows should be modified in the first place, the structure of inputs will also be changed. This could be a strong assumption for a lot of transnational companies like Coca-Cola, which produces the same product in all regions using the same recipe. However, it could be a reasonable assumption for companies producing different varieties in different regions (like car manufacturers).

The interregional trade can be accounted for as a net export since LQ tends to underestimate interregional trade. Thus, an industry is only able to export or import. Simultaneous export and import (cross-hauling) is not allowed in this framework.

Purchases-only location quotient (PLQ)

In this approach, only those industries are taken into consideration that use product *i* as an input in their production processes.

$$PLQ_i^r = \left(\frac{\frac{x_i^r}{x^{*r}}}{\frac{x_i^n}{x^{*n}}} \right).$$

This equation is almost equivalent to the previous one. The small differences can be seen in the nominator and denominator. Where x^{*r} and x^{*n} do not denote total output in region *r* and in the nation, they denote the total output of those industries that provide inputs for industry *i*. The logic behind is that if industry *j* does not supply inputs for another industry i then it has no influence on regional self-supplying ability of industry *i*. Thus it is not necessary to include it in PLQ. On the other hand if industry j does supply inputs for industry i then it can have a real effect the self-supply ability. Thus, industry j must be accounted for in the quotient. In principle it would be a superior method compared to traditional approach but empirical results (see later) did not confirm it.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

Cross-industry location quotient (CILQ)

This measure finally relaxes the strict assumption of modifying coefficient only by rows. *CILQ* takes into account that size of both purchaser and producers industries can affect the self-supplying ability of region *r*. Due to cell-by-cell modification, this approach can describe I-O relations more realistically than its predecessors. The *CILQ* formula can be written as follows:

$$CILQ_{ij}^{r} = \left(\frac{\frac{x_i^r}{x_i^n}}{\frac{x_j^r}{x_j^n}} \right)$$

$$a_{ij}^{rr} = \begin{cases} (CILQ_{ij}^{r})a_{ij}^{n}, \ ha \ CIQ_{ij}^{r} < 1 \\ a_{ij}^{n}, \ ha \ CIQ_{ij}^{r} \geq 1 \end{cases}.$$

It is worthwhile to look at the logic of *CILQ*. If the share of industry *i* (producer) in region *r* (compared to the national level) is higher than the share of industry *j* (purchaser) in the same region (again compared to the national level), then the region can satisfy industry *j*'s input requirements in industry *i* ($CILQ_{ij}^{r} > 1$). Likewise if the share of industry *i* is smaller than the share of industry *j,* the region is not able to cover its own needs, the coefficients in the table must be adjusted according to the equations above. It should be noted that *CILQ* can be expressed as the ratio of two location quotients $\left(CILQ_{ij}^{r} = \frac{LQ_i^r}{LQ_j^r} \right)$. Furtermore it is also true that in the diagonal (*i=j*) *CILQ* = 1. So in this case *CILQ* should be replaced by the traditional *LQ*.

Flegg location quotient (FLQ)

FLQ is one of the latest modifications of the cross-industry quotient that can be attributed to Flegg et al. (1995). Factors that can affect the self-supplying ability of a region expand by the relative size of the region. Thus, *FLQ* takes into account relative specialization, the size of the purchaser and producer sectors, and also the size of the region. The Flegg formula can be expressed as follows:

$$FLQ_{ij}^{r} = (\lambda^{r})CILQ_{ij}^{r},$$

where λ stands for the relative size of a region and it can be calculated by using the next equation:

$$\lambda^{r} = \{log_2[1 + (x_E^r/x_E^n)]\}^{\delta}, 0 \leq \delta < 1.$$

As one can see, the quotient is not modified directly by the relative size, but by the logarithmic value of it. Thereby, the scaling is not so intensive, and the whole expression is then raised to the power of δ, which is a sensitivity parameter. The larger the value of δ the stronger the adjustment in *FLQ*. The determination of δ is crucial to generate realistic results, but the way how it can be carried is not so evident. There can be found several attempts to estimate δ (0.15 < δ < 0.30) in international literature but there still no absolute consensus. Finally, the quotient can be applied like its predecessor:

$$a_{ij}^{rr} = \begin{cases} (FLQ_{ij}^{r})a_{ij}^{n}, \ ha \ FLQ_{ij}^{r} < 1 \\ a_{ij}^{n}, \ ha \ FLQ_{ij}^{r} \geq 1 \end{cases}.$$

In case of diagonal (*i=j*) elements, traditional *LQ* can substitute *FLQ* again.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

Other LQ variations

The literature is remarkably rich in further modifications of the original method (semi-logarithmic LQ, semi-logarithmic $CILQ$, augmented FLQ). However, these variants could not gain significant popularity in comparison with previous examples. Further details can be found in Miller and Blair (2009).

Nevertheless, the work of McCann and Dewhurst (1998) drew attention to an important aspect of regionalization. The authors emphasized that the values of regional coefficients could exceed their national counterparts because they can be considered as an average of regional ones. This aspect was not considered in any of the previous examples. Flegg and Webber (1998) published a modification of FLQ to overcome this shortcoming. The augmented quotient ($AFLQ$) improved by an additional logarithmic factor, which stands for the relative specialization of a region:

$$AFLQ_{ij}^{r} = FLQ_{ij}^{r} log_2(1 + LQ_j) = (\lambda^r)CIQ_{ij}^{r} log_2(1 + LQ_j).$$

Consequently if region r can be considered as specialized (LQ>1), it will use its inputs more intensively, thus the intraregional coefficients can be higher than the national ones. It should be noted that this approach simply neglects the possibility of increasing returns to scale that will occur with increasing specialization and will lead to decreasing intensity of input use. Besides, empirical results did not prove it to be more realistic.

Commodity balance method

The commodity balance (or supply-demand pool /SDP/) is simply the difference between regional supply and demand (Bonfiglio 2005):

$$\tilde{x}_i^r = \sum_j a_{ij}^n x_j^r + \sum_f c_{if}^n f_f^r.$$

In the first step, regional demand $\tilde{x}_i^r$ is estimated using national input coefficients (a_{ij}^n), regional sectorial output (x_j^r), national final demand coefficients (c_{if}^n) and total regional final demand (f_f^r). In the second step the balance (b_i^r) can be calculated:

$$b_i^r = x_i^r - \tilde{x}_i^r.$$

If b is greater than 0 then region r is capable of supplying its need, thus coefficients remain unadjusted. Of course, if supply is exceeded by demand then region r must rely on imports, and coefficients need to be scaled down. The correction is carried out in a similar manner as before in case of LQ methods:

$$a_{ij}^{rr} = \begin{cases} (x_i^r/\tilde{x}_i^r\, a_{ij}^n, & if\ b_i^r < 0 \\ a_{ij}^n, & if\ b_i^r \geq 0 \end{cases}.$$

Round (1972) showed that both $CILQ$ and CB methods share common advantages and shortcomings.

The role of simultaneous import and export (cross-hauling)

Obviously, the net approach of interregional trade is too strict an assumption. Still, most non-survey methods cannot overcome this problem. However, several attempts were made to shed light on this matter; there are only a few recommendations on how to assemble

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

input-output framework with cross-hauling ((first Jackson (1998) then Kroenenberg (2007)).

The introduction of cross-hauling is not straightforward due to data unavailability. The first potential ad hoc estimation procedure was suggested by Jackson (1998). He introduced k_i^r coefficient as the share of cross-hauling as a percentage of total sectoral output in region *r*. If there is available data on interregional transportation or trade, one can derive *k,* thus the calculation of regional cross-hauling can be completed ($CH_i^r = k_i^r x_i^r$). At this stage using net interregional trade and cross-hauling, gross interregional trade volumes can easily be derived.

Later, Kroenenberg (2007) reformed Jackson's idea to a more concrete empirical example. The procedure is known as the Cross-Hauling Adjusted Regionalization Method (CHARM), which is a non-survey method using international trade data to estimate product heterogeneity and re-export. First, it is assumed that cross-hauling is a function of product heterogeneity because interregional trade is motivated by heterogeneous products. In a fictional world where all products are homogenous, there would be no intention for simultaneous export and import. Furthermore, Kroenenberg also noted that re-export would also be affected by the size of region; his formulation can be expressed as follows:

$$CH = f(\varepsilon, X, Z, D), \text{ in simpler form: } CH = \varepsilon(X + Z + D),$$

where cross-hauling (*CH*) is a function of product heterogeneity (*ε*), total output (*X*), total intermediate use (*Z*) and total demand (*D*). ε can be expressed from the equation on the right-hand side:

$$\varepsilon = \frac{V - |B|}{2(X + Z + D)},$$

where *V* is the total gross trade (import and export) and *B* is trade balance. The expression in the nominator stands for the difference between gross trade volume and the absolute value of the trade balance, which is in fact twice the cross-hauling (that is the reason for the multiplication by 1/2). *X*, *Z* and *D* can be calculated easily for region *r* using any non-survey regionalization method. According to Kroenenberg, ε should be estimated using national data, thus ε is only national and not region specific. Of course, the problem is more complex, but it is necessary to make such assumptions to keep mathematical and statistical calculations simple.

Bi-proportional methods (*RAS*, entropy and mathematical programming models)

The second group of non-survey methods relies on national input-output tables and supposes that only the framework of a regional table is available; thus all cells within the table have to be estimated. For example, in case of intermediate consumption, usually the national table, the national total intermediate use and sales (as a frame) is known. Furthermore, total regional intermediate use and sales (as a regional frame) is also available. The aim is to generate a regional table that will preserve the national structure as much as possible and at the same time will satisfy the available regional frames as constraints (Lahr–Mesnard 2004). These procedures can be categorized, at the least, into two groups:

1) Bi-proportional methods – *RAS* and alternative techniques,
2) Mathematical programming (optimization) models.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

RAS is the most well-known bi-proportional method. Stone (1961) first recognized the potential application of *RAS* in regionalizing problems. *RAS* is considered as a partial or a non-survey method due to its low data requirements. Unfortunately, I-O tables are not published in each year, therefore, for years between the publications of two tables, researchers have to generate estimated tables. Initially, *RAS* was considered as such an updating technique for national I-O tables (especially for intermediate use matrix). Later (as a consequence of greater interest for regional sciences), more attention was drawn to *RAS* as a "spatial updating" (regionalizing) method (Pigozzi and Hinojosa, 1985). In case of regionalization, the application is quite straightforward (just like in the case of updating). The initial matrix is always the national table (or sometimes the I-O table of another region's table /that is similar in an economic sense/). In the first step, the regional table is assumed to be identical to the national one ($\boldsymbol{Z^0} = \boldsymbol{Z^n}$), which obviously cannot satisfy the equality criteria between the total of rows, columns and regional frames. Thus further adjustment is needed, which will be carried out in two steps. First, rows need to be scaled down by a certain ratio to satisfy equality of the regional frame and total supply in *Z*. Second, the same procedure has to be run for columns. The row scaling ratio (column vector) is $r_i^1 = \frac{z_i^r}{\sum_j z_{ij}^{r1}}$, where the numerator is actual regional data and the denominator is the sum of the estimated table by *j*. If $r_i^1 < 1$, then elements in row *i* of the estimated table are higher than they should be and vice versa. Thus the rows of the estimated table will satisfy constraints by multiplying the table by this vector (diagonal matrix). But, at this stage, column totals will differ from the regional column frame (column constraint). Thus, the same procedure also has to be applied for columns. The column scaling ratio (row vector) is $s_j^1 = \frac{z_j^r}{\sum_i z_{ij}^{r1}}$. If $s_j^1 < 1$, then elements in the estimated table are higher than they should be, thus they need to be scaled down by s_j^1 to achieve consistency by columns ($\boldsymbol{Z}^2 = \hat{\boldsymbol{r}}^1 \boldsymbol{Z}^0 \hat{\boldsymbol{s}}^1$). At this stage, it is likely that the rows will not satisfy regional constraints again, thus the procedure has to be started again. The sequential repeat of step 1 and step 2 will adjust the initial table to be constrained by regional frames. Usually the procedure is convergent and after a few iterations, estimated values will be very close to regional frames.

Although RAS is the most accepted and applied method, it is not the only bi-proportional technique. The literature is rich in other similar alternative scaling methods (e.g. *DSS*) and improved *RAS* methods (*CRAS*, *GRAS*, additive *RAS* – for additive *RAS* see Révész 2011).

Where *GRAS* is a specialized procedure that can adjust both negative and positive elements by separating them in two different matrices (one with non-negative elements and one with absolute values), then the adjustment of rows and columns will be carried out for the sum of two matrices, and in the final step, the results will be adjusted by negative elements (Junius and Oosterhaven 2002).

A further evolution of *RAS* is the two-staged *CRAS* (cell-corrected *RAS*). In the first stage, traditional RAS is run; thus the matrix will be fitted to regional constraints. In the second stage, the matrix is modified by using distributions calculated from tables of other regions. The modification is carried out by a constrained optimization problem, which can take spatial variations into consideration (Mínguez et al. 2009).

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

The last group of non-survey methods consists of entropy methods that are close to RAS techniques (Jaynes 1957, Wilson 1970). These models are based on information theory, and their aim is to preserve as much information from the initial table as possible by maximizing entropy (a well-known concept in thermodynamics). The following model (*Minimum Sum of Weighted Cross-Entropy*) illustrates the basic concept of the approach. Suppose that c_{ij} is the initial coefficient matrix, and regional frame (x_i^r and x_j^r) is also available:

$$c_{ij} = \begin{bmatrix} 0{,}714 & 0{,}2 & 0{,}060 \\ 0{,}143 & 0{,}6 & 0{,}176 \\ 0{,}143 & 0{,}2 & 0{,}765 \end{bmatrix}.$$

In this approach, entropy is defined by the following equation:

$$H = \sum_i \frac{x_j^r}{\sum_j x_j^r} \sum_j b_{ij} log(b_{ij}/c_{ij}),$$

where b_{ij} is objective variable. If the structure of this estimated matrix is precisely identical to the initial matrix c_{ij} then entropy is zero (H_{MSCE}=0), which means that the preservation of the initial matrix was a 100 % success. If there is a slight change in b_{ij} (b_{11} = 0.8) then H_{MSCE} = 0.039, which means preservation of information is not complete. The final model searches for the maximum of entropy (minimal H) that will satisfy additional row and column constraints:

$$\sum_j b_{ij} x_{rj} = x_{ri} \text{ and } \sum_i b_{ij} = 1.$$

At this stage, one can show (by deriving Lagrangian of the problem and the first order conditions (McDougall, 1999)) that both entropy and *RAS* methods will generate the same result.

Besides these, there are several other mathematical optimization methods in the literature that cannot be simplified to bi-proportional techniques. These models use different special objective functions to find the optimal values of regional I-O tables. The objective function can be formulated in many different ways (for example, the sum of traditional, weighted, normalized absolute or squared differences). It should be noted that not all methods are capable of preserving the sign of coefficients, conequently, negative values can appear in estimated I-O tables even if all elements of the initial matrix were nonnegative. Based on literature review, we can state that methods based on entropy and squared differences are the most widely used approaches.

Empirical results of non-survey methods

Fortunately, the literature is rich in empirical studies. Most of them focus on evaluating the precision and reliability of different regionalization methods. They can be carried out only in regions where regional survey generated tables are available as a benchmark. Since tables consist of hundreds of cells, it is difficult to compare initial and estimated matrices. Therefore several error formulas (such as squared differences) that can be applied to compress the information regarding the difference between coefficients (Lahr 2001).

Non-survey methods are still recommended, especially where there is a lack of appropriate regional data. Usually, these studies compare the performance of different *LQ* methods. Based on findings of early studies, it can be stated that traditional *LQ* generated the most precise results (Morrison–Smith 1974) by exceeding other developed *LQ* indices

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

(for example *PLQ, CILQ, RLQ*). However, the recent notion of Flegg et al. (1995) has gained remarkable popularity in the last two decades. Furthermore, it can also be seen that *FLQ* generally performed better than its predecessors and proved to be one of the best non-survey alternatives to estimate regional I-O relationships (Miller and Blair 2009, Bonfiglio és Chelli 2008, Kowalewski 2012, Lindberg 2011, Riddington et al. 2006, Swaminathan 2008, Tohmo 2004, Bonfiglio 2005). The only shortcoming of *FLQ* is the estimation of δ, the sensitivity parameter, which, at best, should be carried out by econometric estimation, although, the lack of data makes it difficult. Thus, the optimal range of δ can be determined by reviewing the literature. Based on studies of Flegg and Tohmo (2013) in Finland, Kowalewski (2012) in Germany, Flegg and Webber (2000) in Scotland and Bonofiglio and Chelli (2008), we can state that δ is between 0.1 and 0.3.

Furthermore, as we have seen, *RAS* is the alternative (augmentation) to quotient-based regionalization. *RAS* can be applied alone, but most hybrid methods use *RAS* as an intermediate step in the procedure. In general, *RAS* produced satisfactory results (Miller and Blair 2009, Morrison and Smith 1974, Jalili 2005, Riddington et al. 2006, Round 1983, Harris and Liu 1997). Flegg and Tohmo (2013) suggested the combined use of *FLQ* and *RAS* to improve the accuracy of estimation. In the first step, national coefficients are modified by standard *FLQ;* then the table is constrained to be given a regional frame by *RAS*. The resulting regional tables are considered as a reasonably good representation of true regional I-O relations (compared to other non-survey methods).

This combination was chosen in our previous research in which we examined the possible effects of introducing a Blue Economy innovation in the Transdanubian region. First we carried out the two-stage regionalization (*FLQ+RAS*), then the table was integrated into the multisectoral block of a spatial computable general equilibrium model that was capable of describing the expected effects of different policy scenarios. The complete impact analysis was carried out by running off the full GMR model. For additional information, please see our published study in Regional Statistic (Varga et al. 2013).

The other part of the literature completely refuses non-survey methods because, despite all the advantages, they cannot produce reliable results that would justify their application. Nevertheless, non-survey methods are widely applied, and their significance should not be underestimated. Lahr (1993) emphasized that most hybrid methods (in their initial step) are based on non-survey techniques, which are then later enhanced by additional regional data. One of the earliest methods (Morrison and Smith 1974) comparing different methods suggests the use of hybrid techniques, although, since then a lot of other studies have also reinforced these findings: (Brand (2012) in Finland, Kowalewski (2012) and Lindberg (2011) in Sweden, Patriquin et al. (2002) in Western-Middle Alberta, Ralston and Hastings (1986) in Delaware, Harris and Liu (1997) in Porthmouth, Bonfiglio (2005) in Marche region, Oosterhaven et al. (2003) in the Netherland, Jiang et al. (2012) in China and Round (1983)). By examining several methods (*LQ, , PLQ, RLQ, FLQ, AFLQ, SDP, RSP, RAS*), the authors concluded that even if improved methods (like *FLQ*) can generate relatively accurate results, the assumptions of non-survey methods are generally too strict to reflect true input-output relations.

Others came to even stricter results; McMenamin and Haring (1974) and Kipnis (1984) found that even hybrid input-output tables are not accurate enough. Thus, the only reliable

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

way of estimating regional tables is the survey method. Later, Harris and Liu (1997) found that hybrid methods can excel non-survey techniques; however, they are still too resource and input demanding. The improvement in accuracy of results is not in line with the increased cost of the procedure, thus, it is worthwhile investing a bit more resources and generating pure survey based tables. On the other hand, Riddingtion et al. (2006) revealed some interesting aspects; they found that hybrid methods were able to generate misleading, distorted results in the case of relatively small regions. Nevertheless, in bigger regions their reliability is improved greatly.

Based on the above-mentioned studies, we can conclude that there is no consensus in the literature. It did not become clear whether the pure non-survey method can provide reliable results or if they should only be used as a step in hybrid methods. Despite this, most studies agreed that non-survey or hybrid methods are clearly efficient ways of estimating regional tables, and they can generate a good approximation of sectorial relationships.

Estimation of interregional trade

In case of multiple regions, it is crucial to estimate transactions between regions; thus both intra- and interregional data are required. In practice, usually there is no appropriate available data to assemble a table of interregional trade. Furthermore, by increasing the number of regions, the number of tables will increase faster. In case of two regions, four tables need to be estimated, in case of three regions nine, in case of four regions sixteen tables. In the next sections, different interregional methods will be presented.

Extended LQ-method

One of the simplest approaches is the well-known *LQ* approach for multiple regions. The basic concept is the same as before:

$$a_{ij}^{rr} = \begin{cases} (LQ_i^r)a_{ij}^n, \ if \ LQ_i^r < 1 \\ a_{ij}^n, \ if \ LQ_i^r \geq 1 \end{cases}.$$

For transactions from other regions the following is true:

$$a_{ij}^{sr} = \begin{cases} (1 - LQ_i^r)a_{ij}^n, \ if \ LQ_i^r < 1 \\ 0, \ if \ LQ_i^r \geq 1 \end{cases}.$$

The same can be carried out in case of region *s*. These trade flows have to be balanced by each other, which means that imports from region *s* in region *r* should be equal to exports from *s* to *r*. As mentioned previously, the approach can calculate only net exports and imports, which means that if region r exports product *I,* then *s* can only import it.

Next, consider an example consisting of three regions. Our aim is to estimate an interregional input-output table for the three regions system. First, the usual *LQ* method is used for all regions to generate an intraregional input-output table, with net exports to and imports from the rest of country, thus we can calculate the input-output table for rest of country (Z^{NN}) as well. For region 1 it can be represented by:

$$\begin{bmatrix} Z^{11} & Z^{1N} \\ Z^{N1} & Z^{NN} \end{bmatrix},$$

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

where Z^{11} is the intraregional matrix, Z^{1N} is the total outflow from region (to region 2 and 3), Z^{N1} is total inflow to region from rest of country, and finally Z^{NN} is the total intraregional transactions outside the region. The same procedure can be carried out for the other regions, but it is worth noting that the meaning of Z^{rN}, Z^{Nr} and Z^{NN} will be different because different regions will form the rest of country. In the end the following matrices will be calculated:

$$\begin{bmatrix} Z^{11} & \blacksquare & \blacksquare & \boldsymbol{Z^{1N}} \\ \blacksquare & Z^{22} & \blacksquare & \boldsymbol{Z^{2N}} \\ \blacksquare & \blacksquare & Z^{33} & \boldsymbol{Z^{3N}} \\ \boldsymbol{Z^{N1}} & \boldsymbol{Z^{N2}} & \boldsymbol{Z^{N3}} & 0 \end{bmatrix},$$

where diagonal elements are intraregional input-output tables. Black squares stand for unknown interregional transactions we would like to estimate. Red vectors represent the estimated total outflow in each region. In the last step, these values are distributed amongst black matrices. It can be performed in many different ways (for simplicity, assume equal distribution). At this stage, the complete interregional system takes the following form (without intraregional transactions):

$$\begin{matrix} 0 & Z^{12} & Z^{13} & \boldsymbol{Z^{1N}} \\ Z^{21} & 0 & Z^{23} & \boldsymbol{Z^{2N}} \\ Z^{31} & Z^{32} & 0 & \boldsymbol{Z^{3N}} \\ \boldsymbol{Z^{N1}} & \boldsymbol{Z^{N2}} & \boldsymbol{Z^{N3}} & 0 \end{matrix}.$$

The last task is to achieve consistency in the system. The estimated interregional trade flow data is not guaranteed to satisfy constraints in the frame, but *RAS* provides a simple way of ensuring full consistency within the system.

Gravity method

The concept of the gravity model is based on Newton's law in physics to estimate interregional trade flows using available regional data. In this approach, if two regions are close to each other in space and large in an economic sense, then it is more likely that they will trade with each other. The basic generalized equation can be expressed as follow:

$$z_i^{rs} = \frac{(c_i^r x_i^r)(b_i^s x_i^s)}{(d^{rs})^{e_i}} = (k_i^{rs}) \frac{x_i^r x_i^s}{(d^{rs})^{e_i}}.$$

It can be seen that the magnitude of interregional trade (z_i^{rs}) is affected by two factors: the size of output in regions (x_i^r, x_i^s) and distance (d^{rs}). Of course the latter has a negative effect on trade. The other output-sensitivity parameters (c_i^r, b_i^s, k_i^{rs}) need to be estimated in order to generate reliable results. Finally e_i is the sensitivity of distance, which also needs to be estimated outside the model.

The most obvious advantage of the gravity approach is its ability to estimate interregional trade on a gross basis compared to *LQ*. On the other hand, further data requirements make it difficult to calculate all parameters of the model (Black 1972). Furthermore, it can be seen that the gravity method is only capable of generating trade between regions in commodity *i*. It cannot show which industry is going to use this product in its production procedure. For this purpose, the extended *LQ* method offers a potential alternative.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

Mathematical programming models

Besides, other methods approach the estimation of interregional trade and regional input-output table as a constrained mathematical optimization problem. In this case, different parts of the estimated regional table are restricted by given constraints (regional data as frames). Clearly, a great advantage of mathematical models is that the convergence of the procedure is guaranteed while for example in case of *RAS,* it always depends on the initial table and regional data (especially in case of interregional trade).

Based on Canning and Wang (2005), we can reformulate the problem as follows:

$$\sum_{s=1}^{g}\sum_{j=1}^{n} z_{ij}^{rs} + \sum_{s=1}^{g}\sum_{k=1}^{h} y_{ik}^{rs} + e_i^r = x_i^r.$$

In this sense, total output in region r (x_i^r) is allocated between intraregional (z_{ij}^{rr}), interregional intermediate use (z_{ij}^{rs}) and intra- (y_{ik}^{rr}), interregional final consumption (y_{ik}^{rs}) and international export (e_i^r). On the expenditure side, the following is true:

$$\sum_{r=1}^{g}\sum_{i=1}^{n} z_{ij}^{rs} + \sum_{i=1}^{n} m_{ij}^{s} + v_j^s = x_j^s.$$

Output of region s is produced using local (z_{ij}^{ss}) and interregional intermediate use (z_{ij}^{rs}), internation import (m_{ij}^s) and value added (v_j^s). For final demand the constraint is the following:

$$\sum_{h=1}^{k}\sum_{s=1}^{g} y_{ih}^{sr} + \sum_{h=1}^{k} m_{ih}^{r} = y_i^r.$$

In this case regional final demand (y_i^r) can be satisfied by local production (y_{ih}^{rr}), interregional trade (y_{ih}^{sr}) and international imports (m_{ih}^r). Finally intermediate use is described by the last equation:

$$\sum_{r=1}^{g}\left(\sum_{s=1}^{g} z_{ij}^{sr} + m_{ij}^{r}\right) = z_{ij}.$$

This is an important constraint because it requires the sum of intraregional intermediate use (z_{ij}^{rr}), interregional trade (z_{ij}^{sr}) and imports (m_{ij}^r) to be perfectly fitted to national intermediate use. After that, the remaining constraints will ensure equality between the sum of regional variables and their national counterparts (e.g. $\sum_{r=1}^{g} m_{ij}^r = m_{ij}$). Then a complex objective function that measures the difference between estimated regional and national data will be minimized. For European (including Hungarian) applications, please see Thissen et al. (2010).

The proposed method and its introduction through a numerical example

The estimation of a full interregional table on the basis of pure non-survey methods requires the combination of several techniques. First, based on its empirical results, *FLQ* can modify the national table to incorporate special regional characteristics. By applying *FLQ,* it is possible to estimate the total sectorial interregional in- and outflow of the region, although, it can be applied only in single region cases. Thus, it is necessary to extend the method to estimate region-to-region economic relationships, which can be accomplished by the idea of the standard extended *LQ* method. The next step is to achieve full consistency of the table. By applying *FLQ,* the rows and columns of the initial table are modified so there is no guarantee that it will be constrained by actual regional data. These inconsistencies can be eliminated by bi-proportional matrix updating methods (like *RAS*). For ease of understanding, in the following, we can see a numerical example of the method

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

where our starting point is the national intermediate transaction table (*Zn*, and its coefficient table *An*):

Table 1

An example for national intermediate use matrix (absolute values and coefficients)

Zn	Agr	Ind	Ser	An	Agr	Ind	Ser
Agr	50	20	10	Agr	0.56	0.17	0.05
Ind	10	60	30	Ind	0.11	0.50	0.16
Ser	10	20	130	Ser	0.11	0.17	0.68

Source: author's own editing.

Furthermore, total regional sectorial output is also available; thus it is possible to calculate *FLQ*s:

Table 2

An example for FLQ in three regions

FLQ1	Agr	Ind	Ser
Agr	1.48	1.78	1.69
Ind	0.56	0.83	0.95
Ser	0.59	1.05	0.88

Region1	Output
Agr	40
Ind	30
Ser	50
Total	120

FLQ2	Agr	Ind	Ser
Agr	0.95	1.00	0.90
Ind	1.00	0.95	0.90
Ser	1.11	1.11	1.05

Region2	Output
Agr	30
Ind	40
Ser	70
Total	140

FLQ3	Agr	Ind	Ser
Agr	0.63	0.53	0.60
Ind	1.88	1.19	1.13
Ser	1.66	0.88	1.05

Region3	Output
Agr	20
Ind	50
Ser	70
Total	140

Source: author's own editing.

Considering the calculated values of the quotient, we can see that certain cells are below 1, which means that for example in region 1, industry and service cannot satisfy the intermediate demand of agriculture, thus the region needs to import. On the other hand, agriculture is capable of supplying both industry and service and even exporting to other regions (*FLQ*>1).

Using traditional rules of the LQ approach, we can now adjust cells of the national table using *FLQ*. If *FLQ*<1, cells should be scaled down or otherwise unadjusted. The result can be seen below in the diagonal of the matrices:

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

Table 3

An example for an interregional intermediate transaction matrix

Interregional intermediate use		Region1			Region2			Region3			Total
		Agr	Ind	Ser	Agr	Ind	Ser	Agr	Ind	Ser	
Region1	Agr	22.2	5.00	2.63	0.40	0.61	0.48	0.48	2.35	0.94	41,59
	Ind	2.01	12.50	6.04	0.31	0.48	1.45	0.00	0.00	0.43	22,55
	Ser	2.12	4.24	30.01	0.16	0.33	0.00	0.00	1.16	0.00	37,99
Region2	Agr	0.00	0.00	0.00	15.87	5.44	2.72	2.03	2.35	0.94	25,93
	Ind	1.22	1.25	0.93	2.72	19.05	8.16	0.00	0.00	0.43	32,45
	Ser	1.16	0.38	1.25	3.01	6.01	47.89	0.00	1.16	0.00	60,80
Region3	Agr	0.00	0.00	0.00	0.40	0.61	0.48	7.05	3.63	1.81	13,75
	Ind	1.22	1.25	0.93	0.31	0.48	1.45	2.22	25.00	10.20	44,14
	Ser	1.16	0.38	2.10	0.16	0.33	0.00	2.22	6.01	47.89	60,80
	Total	*31,11*	*25.00*	*44.74*	*23.33*	*33.33*	*62.63*	*15.56*	*41.67*	*62.63*	*340.00*

Source: author's own editing.

The next step is generating an estimate of interregional trade (off-diagonal matrices). On a net basis, *FLQ* can be used to calculate trade flows between regions using the following equation:

$$T_{ij}^{roc,r} = \left(1 - FLQ_{ij}^{r}\right) \cdot a_{ij}^{n} \cdot x_{j}^{r},$$

where $T_{ij}^{r,roc}$ is net trade flow entering the region (from rest of country). In this hypothetical example, we distributed these imports equally between regions (for example in case of interregional industrial imports from regions 2 and 3 to agriculture in region 1 /1.22-1.22/). But again, this should be carried out by estimation based on further regional data. At this stage the problem is that the estimated table is not consistent. The columns satisfy regional intermediate use constraints (row total) (because of the import side distribution of *FLQ* generated imports), but in the rows, intermediate supply is not guaranteed to be consistent. Using export-oriented distribution of trade ($T_{ij}^{r,roc} = \left(FLQ_{ij}^{r} - 1\right) \cdot a_{ij}^{n} \cdot x_{j}^{r}$), the desired sum of rows can be generated that will be not in line with the actual sum of the rows (e.g. in first row: 35.71 < 41.59). Otherwise, total intermediate supply can also be approximated by subtracting (estimated) final demand from regional industrial output. Thus in the final step, the matrix needs to be balanced by a matrix updating technique such as *RAS*.

Summary

The current study first attempted to introduce the most important methods of estimating symmetric regional input-output tables in Hungary. However, even this concise review indicates how enormous the international literature is and how widely spread the applications are.

In the absence of clear consensus in regionalization, considering the advantages and shortcoming of different approaches, we can conclude that the hybrid method is still the most suitable choice. Hybrid results are relatively accurate and are not as time and resource

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

intensive,athough, it is essential to acquire further regional data to enhance accuracy (i.e. surveys or experts)

In case of scarce regional data, non-survey methods offer alternatives for describing interregional interindustry relationships. Although, without a regional reference table, results cannot be evaluated, there are still numerous examples in international literature with reliable results confirming the possibilities of non-survey approaches.

Based on these experiences, we recommend the application of a combined method, which is attributable to several reasons. First, these methods pursue simplicity, second, they do not differ from the mainstream approach of international literature and third, they can be considered as reliable and efficient methods (based on international experiences due to a lack of domestic references). Our starting point is *FLQ* because its theoretical ground is more solid than its predecessors, and its empirical results are satisfactory. The missing δ parameter can be estimated based on literature, or the average of previous deltas can be used as a rule of thumb. After the estimation of interregional trade and the intraregional table, there are multiple further options. First, simultaneous exports and imports (cross-hauling) can be incorporated based on Jackson (1998) and Kroenenberg (2007). Then using these trade data (either net or gross), the gravity method can be applied to estimate interregional flows. If results of this approach are not detailed enough, the extended LQ method can generate a four-dimensional (region-to-region, industry-to-industry) interregional table. In the final step, we need to ensure consistency between and within tables by using all available regional data. By preserving the basic structure of tables, *RAS* can modify them to achieve full consistency between tables and constraints.

* * *

Acknowledgements
The current study was supported by MTA-PTE Innovation and Economic Growth Research Group. I would like to express thanks to Professor Randall Jackson and Professor Attila Varga for their valuable comments and advice.

REFERENCES

Black, William R. (1972): Interregional commodity flows: Some experiments with the gravity model *Journal of Regional Science* 12 (1): 107–118.

Bonfiglio, A. (2005): *A Sensitivity Analysis of the Impact of CAP Reform. Alternative Methods of Constructing Regional I-O Tables* PhD Dissertation. Polytechnic University of Marche Ancona, Italy.

Bonfiglio, A.–Chelli, F. (2008): Assessing the Behaviour of Non-Survey Methods for Constructing Regional Input–Output Tables through a Monte Carlo Simulation *Economic Systems Research* 20 (3): 243–258.

Brand, S. (2012): *A Note on Methods of Estimating Regional Input-Output Tables: Can the FLQ Improve the RAS Algorithm?* The Business School with Plymouth University, Working Paper., Plymouth.

Canning, P.–Wang, Z. (2005): A flexible mathematical programming model to estimate interregional input-output accounts *Journal of Regional Science* 45 (3): 539–563.

Európai Bizottság (2010): A forrás-felhasználás táblák és az input-output keretrendszer In: *Javaslat az Európai Parlament és a Tanács Rendelete az Európai Unióban alkalmazandó nemzeti és regionális számlák európai rendszeréről* 2010.12.20. COM(2010) 774 végleges. A. melléklet/9. fejezet, Brüsszel.

Flegg, A. T.–Webber, C. D. (2000): Regional Size, Regional Specialization and the FLQ Formula *Regional Studies* 34 (6): 563–569.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

Flegg, A. T.–Tohmo, T. (2013): Regional Input−Output Tables and the FLQ Formula: A Case Study of Finland *Regional Studies* 47 (5): 713–721.

Flegg, A. T.–Webber, C. D.–Elliott, M. V. (1995): On the Appropriate Use of Location Quotients in Generating Regional Input–Output Tables *Regional Studies* 29 (6): 547–561.

Greenstreet, D. (1989): A Conceptual Framework for Construction of Hybrid Regional Input-Output Models *Socio-Economic Planning Sciences* 23 (5): 283–289.

Harris, R. I. D.–Liu, A. (1997): Input-Output Modelling of the Urban and Regional Economy the Importance of External Trade *Regional Studies* 32 (9): 851–862.

Jackson, R. W. (1998): Regionalizing National Commodity-by-Industry Accounts *Economic Systems Research* 10 (3): 223–238.

Jiang, X.–Dietzenbacher, E.–Los, B. (2012): Improved Estimation of Regional Input–Output Tables Using Cross-regional Methods *Regional Studies* 46 (5): 621–637.

Junius, T.–Oosterhaven, J. (2003): The Solution of Updating or Regionalizing a Matrix with both Positive and Negative Entries *Economic System Research* 15 (1): 87–97.

Kipnis, B. A. (1984): Input-Output Tables for Medium-Sized Cities: Survey Coefficients or Short-cut Methods? A Case Study in Brazil *Journal of Regional Science* 24 (3): 443–450.

Kowalewski, J. (2012): *Regionalization of national input-output tables: empirical evidence on the use of the FLQ formula* Hamburg Institute of Regional Economics, HWWI Research Paper No. 126., Hamburg.

Központi Statisztikai Hivatal (2005): *Az Ágazati Kapcsolatok Mérlegének matematikai feldolgozása, 2000* KSH, Budapest.

Kronenberg, T. (2007): *How Can Regionalization Methods Deal With Cross-hauling?* Institut für Energieforschung (IEF), Systemforschung und Technologische Entwicklung (STE), Working Paper 2007/14., Jülich.

Lahr, M. L. (1993): A Review of the Literature Supporting the Hybrid Approach to Constructing Regional Input-Output Models *Economic Systems Research* 5 (3): 277–293.

Lahr, M. L. (2001): A strategy for producing hybrid regional input-output tables In: Lahr M. L.–Dietzenbacher, E. *Input-Output Analysis: Frontiers and Extensions* pp. 1–31., Palgrave, London.

Lahr, M. L.–de Mesnard, L. (2004): Biproportional Techniques in Input–Output Analysis: Table Updating and Structural Analysis *Economic System Research* 16 (2): 115–134.

Leontief, W. (1936): Quantitative Input-Output Relations in the Economic System of the United States *Review of Economics and Statistics* 18: 105–125.

Leontief, W. (1941): *The Structure of American Economy 1919–1939* Oxford University Press, New York.

Lindberg, G. (2011): *On the appropriate use of (input-output) coefficients to generate non-survey regional input-output tables: Implications for the determination of output multipliers* ERSA Conference Paper, No. ersa10p800.

Mínguez, R–Oosterhaven, J.–Escobedo, F. (2009): Cell-corrected RAS method (CRAS) for updating or regionalizing an input-output matrix *Journal of Regional Science* 49 (2): 329–348.

McDougall, R. A. (1999): *Entropy Theory and RAS are Friends* GTAP Working Paper No. 06., Pardue University West Lafayette, USA.

McCann, P–Dewhurst, J. HH (1998): Regional Size, Industrial Location and Input-Output Expenditure Coefficients *Regional Studies* 32 (5): 435–444.

McMenamin, D. G.–Haring, J. E. (1974): An Appraisal of Nonsurvey Techniques for Estimating Regional Input-Output Models *Journal of Regional Science* 14 (2): 191–205.

Miller, R. E.–Blair, P. D. (2009): *Input-Output Analysis: Foundations and Extensions* Cambridge University Press, New York.

Morrison, W. I.–Smith, P. (1974): Nonsurvey Input-Output Techniques at the Small Area Level: An Evaluation *Journal of Regional Science* 14 (1): 1–14.

Oosterhaven, J.–van der Knijff, E. C.–Eding, G. J. (2003): Estimating interregional economic impacts: an evaluation of nonsurvey, semisurvey, and full-survey methods *Environment and Planning A* 35 (1): 5–18.

Patriquin, M. N.–Alavalapati, J. R.R.–Wellstead, A. M.–White, W. A. (2002): A comparison of impact measures from hybrid and synthetic techniques: A case study of the Foothills Model Forest *The Annals of Regional Science* 36 (2): 265–278.

REGIONAL STATISTICS, 2015, VOL 5, No1: 44–65; DOI: 10.15196/RS05103

Pigozzi, B. WM.–Hinojosa, R. C. (1985): Regional Input-Output Inverse Coefficients Adjusted from National Tables *Growth and Change* 16 (1): 8–12.

Ralston, S. N.–Hastings, S. E. (1986): Improving Regional I-O Models: Evidence Against Uniform Purchase Coefficients Across Rows *The Annals of Regional Science* 20 (1) 65–80.

Révész, T. (2011): A Magyar Gazdaság 2010. Évi Ágazati Kapcsolatok Mérlegeinek Becslése. ENERGIAKLUB Szakpolitikai Intézet és Módszertani Központ, Budapest.

Riddington, G.–Gibson, H.–Anderson, J. (2006): Comparison of Gravity Model, Survey and Location Quotient-based Local Area Tables and Multipliers *Regional Studies* 40 (9): 1069–1081.

Round, J. I. (1972): Regional Input-Output Models in the UK: A Reappraisal of Some Techniques *Regional Studies* 6 (1): 1–9.

Round, J. I. (1983): Nonsurvey Techniques: A Critical Review of the Theory and the Evidence *International Regional Science Review* 8 (3): 189–212.

Stevens B. H.–Treyz, G. I.–Ehrlich, D. J.–Bower, J. R. (1983): A New Technique for the Construction of Non-Survey Regional Input-Output Models and Comparison with Two Survey-Based Models *International Regional Science Review* 8: 271–286.

Stone, R. (1961): *Input-Output and National Accounts* Organization for Economic Cooperation and Development, Paris.

Swaminathan, A. M. (2008): *Methods for generation of a regional input-output table for the state of Maharashtra: A comparative analysis* Working Paper No.29, UDE Dr. Vibhooti Shukla Unit in Urban Economics & Regional Development, University of Mumbai, Mumbai.

Thissen, M.–van Oort, F.–Diodato, D.–Ruijs, A (2010): *Regional Competitiveness and Smart Specialization in Europe: Place-based Development in International Economic Networks* Edward Elgar Publishing, Cheltenham.

Tohmo, T. (2004): New Developments in the Use of Location Quotients to Estimate Regional Input-Output Coefficients and Multipliers *Regional Studies* 38 (1): 43–54.

Varga, A.–Hau-Horváth, O.–Szabó, N.–Járosi, P. (2013): A GMR–Európa-modell alkalmazása kék gazdaság-típusú innovációk hatásvizsgálatára. *Területi Statisztika* 53 (5): 411–434.

Zalai, E. (2012): *Matematikai Közgazdaságtan II.: Többszektoros modellek és makrogazdasági elemzések* Akadémia Kiadó, Budapest.

GYÖRGY CSOMÓS[a)]

The ranking of cities as centres of the Hungarian economy, 1992–2012

Abstract

This analysis has relied on an empirical method to identify and rank cities as centres of the Hungarian economy from 1992 to 2012. After the change in the political and economic system of Hungary, a new economic climate emerged (e.g. Hungary joined the European Union, foreign direct investments appeared in the economy, special taxation regulations were introduced), which changed the position of cities. During this two-decade-long transformation, the dominance of the capital, Budapest, and its agglomeration considerably increased; the east-west dichotomy became more pronounced, while the economic role of the traditional industrial centres and that of some large cities weakened. Because of these processes, new types of economic centres emerged, which did not have a significant role in the national economy. Cities that became crucial economic actors because of the offshore operation of foreign multinational corporations are exceptional in this.

The ultimate goal of this study is to introduce a straightforward urban hierarchy, establish a classification based on the economic profile of cities, and address the typical anomalies after the change in the political system.

Keywords: urban hierarchy, industry profile, transformation, economic power, Hungary.

Introduction

In recent decades, a number of studies have been concerned with the role of cities in the Hungarian economy. For example, Enyedi (2000), Barta (2001) Beluszky & Győri (2004), Rechnitzer et al. (2004), Kukely (2006), Koltai (2007), and Lux (2013) represent this line of research, which is fundamentally void of a survey of the economic power of cities, as well as an urban hierarchy that falls from a mathematical model. However, the need for this kind of empirical research is obvious. In the past two decades, Hungary has been influenced by external and internal economic effects, which have changed the position of cities in the economic system. The most significant effects are as follows:

– Before the change in the political system in 1989, cities were arranged in a straightforward manner in the relatively closed Hungarian economy. The most important economic centre was the capital, followed by the county seats, whose development was of great importance for the government, and some industrial cities. In state socialism, the economic development of cities was centrally

a) University of Debrecen Faculty of Engineering, H-4028 Debrecen, Ótemető u. 2-4., Hungary. E-mail: csomos@eng.unideb.hu

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

controlled while unplanned economic processes – if there were any – could take place only locally (Bajmócy–Hegedűs 2008). The change in the political system also altered Hungary's economic system: Central economic planning vanished, and foreign direct investments entered the markets, which then became the most important actors in the economy.

– The former state-owned large companies were often privatized by multinational corporations (Barta 2001), which were directly linked to their foreign headquarters. This is also confirmed by the TOP500 rankings published annually by the World Economy Weekly (HVG): the majority of the largest companies/banks operating in Hungary are owned by foreign entities. Thus, they are not commanded from Hungary, but from the headquarters of their respective parent companies.
– Initially, most of the foreign multinational corporations that appeared in Hungary established their national (or in some cases even Central European) headquarters in Budapest or its agglomeration. There were a number of production companies (e.g. Opel, Audi, Suzuki, IBM, and Nokia) that implemented green field investments to set up factory units in Western Hungary, where the infrastructure had already been well developed. In contrast, some large cities where establishing production facilities had demanded significant financial resources during state socialism lost their significance in the national economy. This was due to the new owners of these facilities that had been privatized, initiated cutbacks and in some cases, the closure of factories (Diczházi 1997).
– The east–west dichotomy that had characterized Hungary's spatial structure for a century was further aggravated, meaning that the gap between the settlements of East Hungary and their counterparts in Western Hungary further widened (Rechnitzer 2004). However, certain large cities in East Hungary (e.g. Debrecen, Szeged, and Kecskemét) that had a relatively developed infrastructure and skilled labour also became a target area for foreign investments, which in turn made these cities stand out from their environments.
– The economic performance of a number of small and medium-sized cities is primarily determined by individual, large, foreign-owned multinational companies, which are economically dominant despite the fact that these cities have a wide range of businesses. Typical examples are the national centres of the automobile industry, such as Szentgotthárd (Opel) and Esztergom (Suzuki). It has been demonstrated that positive and negative effects of the global economy can potentially influence the local economy of these cities to the same extent. In addition, these cities respond to these effects in the same way by either sinking into recession or growing.
– The taxation regulations associated with business associations have been gradually changing since the early 1990s. Some offshore companies have been registered mainly in small towns near Budapest, albeit outside the agglomeration. This was because these towns benefited these companies by waiving local taxes and have thus become seemingly crucial actors of the Hungarian economy in spite of the fact that the offshore companies do not pursue any domestic economic activities.

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Consequently, the economic potential of Hungarian cities has changed significantly over the past 20 years: some cities that occupied outstanding positions before the change in the political system have weakened, and have been replaced by newly emerging economic centres. The analysis below examines how the economic performance of cities altered from 1992 to 2012, and why these changes occurred. I have classified these cities based on their similar economic structures and concluded that they respond to global economic changes in the same manner.

Data and methods

In most analyses that focus on examining the economic performance of cities, it is crucial to use well-defined data and methods.

In previous analyses (see Csomós 2013, Csomós–Derudder 2014), we used data from Forbes 'The Global 2000' to define the world's leading command and control centres through the financial performance of large corporations headquartered in cities. In this paper, we have used data from the National Tax and Customs Administration of Hungary (NTCA) database from the years 1992 to 2012. These data correspond to the main financial data (e.g. net income, sales, total assets, equity and value-added) of firms headquartered and registered in Hungary. Of course, most of these firms, even the largest ones, are subsidiaries of foreign multinational corporations; however, with regards to taxation, they can be regarded as domestic ones. The most typical example is the Audi Hungária Motor Kft. (headquartered in Győr), one of the largest taxpayers and employers in Hungary. Nevertheless, the NTCA data used in the current analysis are slightly different from data of Forbes since NTCA combines specific financial data of all firms registered in cities into a single data.

The level of the economic performance of cities is expressed by the Economic Power (EP). $EP_{x,y}$ of a given city x in a given year y is calculated as follows:

$$EP_{x,y} = \sqrt{\frac{(NI_{x,y}+S_{x,y}+A_{x,y}+E_{x,y})}{4} * VA_{x,y}},$$

where $NI_{x,y}$ = the proportion of net income in the total dataset; $S_{x,y}$ = the proportion of sales in the total dataset; $A_{x,y}$ = the proportion of total assets in the total dataset; $E_{x,y}$ = the proportion of equity in the total dataset; $VA_{x,y}$ = the proportion of value added in the total dataset.

EP is a cumulative measurement that integrates net income, sales, total assets, equity, and value added of all firms headquartered in cities in a specific way, which enables us to economically rank Hungarian cities. The first part of the formula refers to the power of the city. However, some offshore companies (even some Hungarian companies) have impressive financial parameters while their value added is zero or almost zero; therefore, their headquarters cities do not contribute to Hungary's GDP. If value-added had not been used as a multiplier in the formula, those cities would have had very significant economic power; which, they do not have. Given how EP has been calculated, its yearly value concerning Hungary is 100 and is equal to the combined value of all selected cities and towns.

Among settlements included in the NTCA database (1074 settlements in 1992, 2736 settlements in 2012), 1000 settlements with the largest EP values were selected. We did

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

this for two reasons: 1. The change in settlement position can be realistically compared if the database features the same number of elements for every year. 2. The rest of the settlements following the TOP1000 (all of them are small villages) have negligible economic performance.

Classification of economic centres and changes in their positions

In 1992, only three years after the change in the political and economic system, settlements were ranked in the hierarchical system that had evolved in the era of state socialism. Hungary's major economic centre was the capital, Budapest, followed by large and medium-sized cities of complex industrial profiles (county seats that had been intensely developed during socialism), as well as several special industrial nodes. By 2012, two decades later, this historically well-defined structure had evolved into something more complicated while the emergence of new types of economic centres had been triggered almost solely by external economic mechanisms. Budapest retained and even reinforced its unquestionable leading role, but the formerly homogeneous group of cities with complex industrial profiles broke up; the positions of the traditional industrial centres changing depending on the successfulness of privatization. However, the intensive foreign direct investments (FDI) that had been launched after the change in the political system elevated new actors among the economic nodes including settlements that had previously not had any significant economic function. As a special element, Hungary saw the emergence of settlements that accommodated the offshore entities of foreign multinational corporations (MNCs). These settlements of typically less than two thousand inhabitants became – at least seemingly – very important actors in the Hungarian economy despite the fact that the offshore companies did not pursue domestic economic activities at all. Therefore, in 2012, the following settlement groups could be found based on their economic functions:

1) Metropolis with a complex industrial profile;
2) Large- and medium-sized cities with a complex industrial profile;
3) Traditional industrial cities;
4) New economic centres;
5) Tourism cities;
6) Offshore cities/towns;
7) Other cities, towns and villages with insignificant economic power.

Table 1 shows the classification and EP values of the 25 leading economic centres while Figure 1 reflects changes in the EP values between 1992 and 2012.

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Table 1

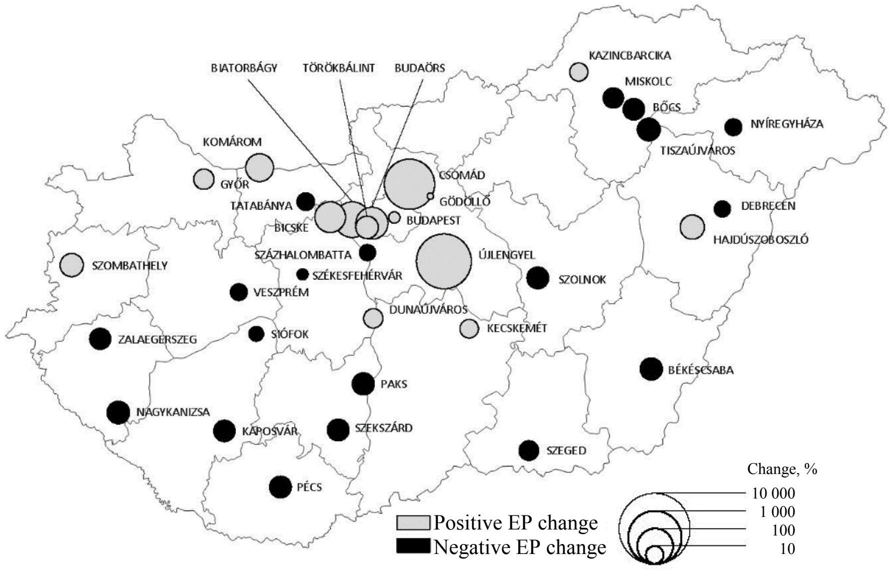
TOP25 economic centres, 1992/2012

Rank	City/town	Type	EP 1992	City/town	Type	EP 2012
1	Budapest	Metropolis	46.68	Budapest	Metropolis	51.07
2	Győr	Complex industrial profile	2.20	Győr	Complex industrial profile	3.37
3	Debrecen	Complex industrial profile	1.90	Szombathely	Complex industrial profile	1.53
4	Pécs	Complex industrial profile	1.79	Debrecen	Complex industrial profile	1.48
5	Szeged	Complex industrial profile	1.61	Budaörs	New economic centre	1.42
6	Miskolc	Complex industrial profile	1.58	Hajdúszoboszló	Tourism cities and towns	1.23
7	Paks	Traditional industrial cities	1.52	Dunaújváros	Traditional industrial cities	1.10
8	Székesfehérvár	Complex industrial profile	1.14	Székesfehérvár	Complex industrial profile	1.05
9	Szombathely	Complex industrial profile	0.80	Újlengyel	Offshore cities/towns	1.00
10	Nyíregyháza	Complex industrial profile	0.79	Szeged	Complex industrial profile	0.89
11	Nagykanizsa	Complex industrial profile	0.78	Kecskemét	Complex industrial profile	0.84
12	Dunaújváros	Traditional industrial cities	0.77	Miskolc	Complex industrial profile	0.77
13	Tiszaújváros	Traditional industrial cities	0.75	Bicske	New economic centre	0.69
14	Szolnok	Complex industrial profile	0.74	Nyíregyháza	Complex industrial profile	0.59
15	Bőcs	Traditional industrial cities	0.65	Csomád	Offshore cities/towns	0.56
16	Kecskemét	Complex industrial profile	0.62	Pécs	Complex industrial profile	0.55
17	Hajdúszoboszló	Tourism cities and towns	0.60	Törökbálint	New economic centre	0.47
18	Békéscsaba	Complex industrial profile	0.59	Százhalombatta	Traditional industrial cities	0.43
19	Zalaegerszeg	Complex industrial profile	0.56	Gödöllő	New economic centre	0.42
20	Százhalombatta	Traditional industrial cities	0.56	Kazincbarcika	Traditional industrial cities	0.41
21	Veszprém	Complex industrial profile	0.55	Veszprém	Complex industrial profile	0.40
22	Kaposvár	Complex industrial profile	0.53	Komárom	New economic centre	0.40
23	Szekszárd	Complex industrial profile	0.48	Paks	Traditional industrial cities	0.40
24	Siófok	Tourism cities and towns	0.46	Siófok	Tourism cities and towns	0.38
25	Tatabánya	Complex industrial profile	0.45	Biatorbágy	New economic centre	0.37

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Figure 1

Change of the Economic Power of the TOP25 Hungarian cities, 1992/2012



This chapter discusses the characteristic groupings of the centres of the Hungarian economy. The economic mechanisms that influence the positions of settlements and give rise to the various groups are also described. It will clearly demonstrate that the open Hungarian economy is moved mostly by external global effects instead of endogenous forces. Positive economic decisions made abroad can quickly raise any settlement to the top ranks or, alternatively, negative investment decisions can cause a decline at the same swift pace.

Metropolis with a complex industrial profile: Budapest and its agglomeration

According to Sassen (2006, p. 63.), after the change in the political system, the major Eastern European cities – particularly the capitals – regained their pre-war significance in the economic life of the region. In her opinion, Budapest proved to be the best example of this. Sassen supported his view by referring to the fact that the value of foreign direct investments coming to the Eastern European economies in 1992–1997 was the greatest in Hungary, with the majority of it targeting Budapest. Sassen claims, that at the beginning of the 1990s, one of the underlying reasons was because the Hungarian capital functioned as an international business enclave; one that featured Western patterns in Eastern Europe, and offered business and tourism services that put the city ahead of its regional competitors. It is not surprising that the Globalization and World Cities Research Network (GaWC) ranked Budapest among the highly prestigious group of Gamma world cities, in the same category as Atlanta, Berlin and Shanghai[1] (Beaverstock et al. 1999). By the early 2000s,

1 According to GaWC's 2012 classification, Budapest came to belong to Beta+ world cities, although among the regional competitors, Prague, Warsaw and Vienna overtook it: all three of them were grouped among Alpha– world cities.

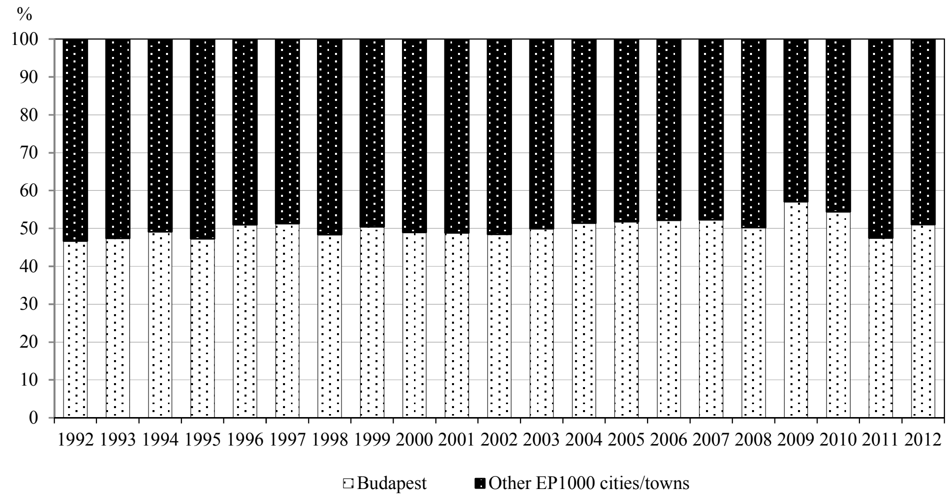
REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Budapest was followed by other cities such as Győr, Székesfehérvár and Debrecen in becoming parts of the regional economy (Szemző–Tosics 2005).

In spite of these later changes, Hungary's most serious problem in terms of spatial structure (for about a century now) has been and still is the Budapest versus the countryside dichotomy (Enyedi 2000, Cséfalvay 2001). Budapest's dominance can be observed not only in economic performance (yet it is especially striking), but also in its extension into all facets of life, which in certain cases (e.g. financials, governance, R&D) is nearly exclusive (Gál 1998, Wágner 2004, Barta et al. 2007). The capital's economic performance after the change in the political system can, in fact, be regarded as steady with minor fluctuations; in the two decades from 1992 until 2012, it surpassed almost 50% of the EP1000 cities/towns. Figure 2 shows that the EP values of the 999 settlements following Budapest do not often reach the corresponding figure of the capital; moreover, it is indeed the annual fluctuation of Budapest's value that influences variances in the EP values of all the other settlements.

Figure 2

Change of the Economic Power of Budapest and the next 999 cities/towns, 1992–2012



Budapest's extreme economic dominance is clearly reflected by the fact that, in 1992, when the minimum EP value for the capital was recorded, the following 613 settlements could together outbalance the economic potential of the capital. However, throughout the 11 years of the reviewed period – due to its EP value in excess of 50% – even the 999 settlements were not able to show such performance (Figure 2). The capital-oriented nature of the economy is properly demonstrated by the fact that while its population constitutes only 17% of Hungary's total population, its EP value averages 50.27% of the national figure; moreover, the former rate dropped by 14% in comparison with 1992, whereas the latter figure remained virtually unchanged.

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Budapest's massive dominance in the economy is further intensified by the role of 81 settlements belonging to the Budapest Agglomeration[2] (e.g. Budaörs, Gödöllő, Fót, Törökbálint), as they also belong to the group of settlements that have recently aspired to the ranks of the country's leading economic centres. The aggregate EP value of the agglomeration in 1992 was 51.19, while in 2012, it was recorded as 57.80; in 1992, 91% of the overall EP value had been provided by Budapest; this had slid to 88% by 2012. This means that, with the relative stability of the capital's EP value, the settlements of the agglomeration consistently strengthened, i.e., the economic potential of the Budapest Agglomeration is becoming increasingly dominant. One of the best examples is Budaörs, which in 1992 ranked 45th among the settlements, but in 2012 was in 5th position after a nearly 500% rise in its EP value (the associated reasons will be revisited in Section 3.5). The tendencies observed in Budaörs are obviously not at all unique. After the change in the political system, those settlements that offered favourable conditions for tax payments to foreign companies contemplating the creation of local bases were able to snatch numerous greenfield investments away from Budapest (Sági 2000). Additionally, as noted in the analysis of Koós (2004), from the middle of the 1990s, relocation also became a common process in Hungary, meaning that many foreign companies transferred their seats from the capital (for reasons differing from company to company) to the surrounding settlements (e.g. Szentendre, Budaörs, Törökbálint, Érd, Szigetszentmiklós). Budaörs hosts foreign companies, such as British American Tobacco, Kaiser + Kraft, Tchibo, Metro, Tesco and Auchan, whereas Törökbálint hosts the Hungarian headquarters of e.g. Telenor and Johnson & Johnson.

Large- and medium-sized cities with a complex industrial profile

From the early 1970s, a key element of regional planning became the moderation of Budapest's dominant position in the Hungarian urban system, as well as the reinforcement of the larger regional centres by the decentralization of certain economic functions (Perczel 1989; Rechnitzer 1998; Csomós 2009a). In fact, Hungary imported and adapted the growth pole strategy worked out by Perroux (1955) and subsequently refined by Boudeville (1966) and Lausén (1969). In this context, so-called counterpole cities (Debrecen, Győr, Miskolc, Pécs, and Szeged) were designated, and their development was prioritized for decades.

Table 1 shows that in 1992 – immediately after the change in the political and economic system – Budapest was followed by five counterpole cities in the ranking. Nevertheless, the combined EP value of the five counterpole cities – indicative of the inefficiencies of the intensive economic development programs launched in the 1970s – did not reach 20% of the capital's EP value. From 1992, regional centres took a considerably different course of development. As is apparent from Table 1, in both years, Győr was Hungary's most significant economic centre in the countryside, primarily due to its developed automobile industry (Lengyel 2012). Audi Hungaria Motor Kft made its appearance in Győr in 1993 (one of Hungary's largest companies as measured by sales), and undoubtedly added to the increase in Győr's EP value, although, it did not substantially influence the city's position. In the period under review, only Győr had a steady position (only left behind by

2 Act LXIV of 2005 on Spatial Planning in the Agglomeration of Budapest.

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Székesfehérvár in 1997 and 1998), while the other regional centres lost more (Szeged, Miskolc, and Pécs) or less (Debrecen) from their respective EP values and positions.

After the change in the political system, large cities in Western Hungary – especially Székesfehérvár and Szombathely – enjoyed extremely favourable positions (Baráth et al. 2001, Kukely 2006), as they attracted many more foreign companies (and foreign direct investments) than did their counterparts in the eastern part of the country. According to Table 1, Szombathely ranked 9th in 1992, while by 2012, it occupied 3rd position with its EP value having nearly doubled. Szirmai et al. (2003: 37) suggest that by 2002, Székesfehérvár had come to follow Győr as the most important city in the countryside to host foreign direct investments, as its EP value had consistently increased since 1992 (the city even ranked ahead of Győr in 1997 and 1998). On the other hand, after 2002, some of the major domestic and foreign companies stopped their production operations in Székesfehérvár (IBM and Kenwood in 2002, Ikarusbus-Irisbus in 2003, Parmalat Hungária in 2004, Cornexi in 2008); therefore, it is not surprising that the drop in its EP value was accompanied by a deteriorating position.[3] In 2012, Székesfehérvár was one of the largest Hungarian economic centres (8th rank), yet some of the over-developed segments (e.g., manufacturing of car parts) were considered crisis areas.

Among the regional centres, Pécs was the most sorely affected by economic processes following the change in the political system; this downturn was reflected by the fact that by 2012, its EP value had dropped by nearly 70%, with the city falling 12 places. In the case of Szeged and Miskolc, EP values also decreased materially, yet to a smaller extent in comparison with Pécs (45 and 51%), and they became just less low-ranked. The reasons are extremely complex, and the associated explanations are sometimes ambiguous at best. Most researchers agree that Debrecen, Pécs and Szeged are unexceptionable regional centres, quite closely followed by Miskolc and Győr (Beluszky–Győri 2004, Rechnitzer et al. 2004). It is clear, however, that there is no significant correlation between the regional functions of the cities and their economic performance (which is also supported by this analysis). The former aspect is much more complex and influenced by the role of the given city in public administration (Pálné Kovács 2001), its functions in the healthcare system, or higher education (Beluszky 2003, Csomós 2009b). Lengyel (1999) claims that Miskolc, Debrecen and Szeged could claim relatively well-trained labour resources – owing to their universities of national significance – but to no avail, as the private sector did not appreciate them; there was no demand for them. In light of the survey made by Koltai (2007), ranking Hungarian cities as business sites on the basis of questions answered by companies, the most competitive cities in the countryside were Győr, Székesfehérvár and Sopron, although this last settlement has never been one of the leading 25 localities in the EP rank. In Koltai's (2007) competitiveness ranking, cities in Western Hungary (Győr, Székesfehérvár, and Sopron) come ahead of the most populous cities in the countryside, Debrecen, Pécs and Szeged. As did Koltai, Kukely (2006) also mentions Győr, Székesfehérvár and the other regional centres (Debrecent, Pécs and Szeged) as Hungary's most quickly developing cities in the countryside, though he calls attention to the

3 When in 2002 IBM – having employed 7000 workers from 1997 to 2000, and in 2000 having an 8% share in Hungary's overall exports (UNCTAD, 2002) – closed down its facilities, it was a major event in the national economy, and Székesfehérvár's position weakened by three spots.

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

remarkable progress of certain settlements belonging to the Budapest Agglomeration (e.g. Budaörs). Given the economic performance of these cities, a key issue is the concentration of foreign direct investments (FDI). FDI is an important factor in the economy of Hungary because the majority of domestic companies pursuing large-scale production operations are owned by foreign multinational corporations and banks (Barta 2001, Antalóczy et al. 2011). Studies examining the territorial implications of FDI, have found that Pécs has relatively poor capital-attracting abilities (Juhász–Schottner 2003) while, according to Antalóczy & Sass (2005), Szeged, Debrecen and especially Miskolc can be characterized by growth in their FDI volumes, they also supersede the national average. Therefore, it is not by mere chance that Szeged, Debrecen and Miskolc (only after Győr, Székesfehérvár and Szombathely, of course) hold more favourable positions in the EP ranking than the not-so-dynamically developing Pécs, which attracts less FDI.

In general, recent EP values and positions of other county seats that do not function as regional centres weakened in comparison with 1992. The only exception was Kecskemét (1992: 16th position, 2012: 11th position), which has reached the ranks of the leading cities of the countryside by its competitiveness (Koltai 2007). With respect to the positive effects of the giant investments of the automobile industry on the economic performance of the cities – as implied by the presence of Audi in Győr, Suzuki in Esztergom and Opel in Szentgotthárd (see Rechnitzer–Smahó eds. 2012) – it is expected that the Mercedes-Benz project, launched in 2012, will further improve Kecskemét's current position. At the same time, Nyíregyháza, considered one of the northeastern Hungarian focal points of FDI (Barta 2001), is witnessing promising capital expenditures; in 2014, LEGO Manufacturing Kft., which settled in the Industrial Park of Nyíregyháza in 2008, further increased its production output.

Traditional industrial cities

Unlike regional centres featuring complex economic structures, the positions of traditional industrial cities are usually determined by some larger individual companies performing special activities. Particular cities belonging to the top ranks include Dunaújváros, Tiszaújváros, Jászberény, Bőcs and Kazincbarcika. Cities in this group had two features in common; even before the change in the political system, they had considerable production capacities; while after privatization, many of them took relatively successful development paths. Understandably, the economic performance of dominant companies exercised strong positive or negative influence on the achievements of these cities. Among these cities, Dunaújváros has the largest EP value. Not only its population makes it stand out from this category but also its economic structure approximates those of the complex centres. The city owes its current position primarily to ISD Dunaferr having Ukrainian owners (it is indirectly owned by the Russian Vnesheconombank), as well as the Hungarian subsidiary of South Korean tyre-manufacturing Hankook Tire launched in 2007. Yet, the EP value of Dunaújváros shows strong fluctuations: in 1992, it ranked 9th; in 2011 it dropped back to 18th position, but in 2012 it stood at 7th position. This fluctuation evidently results from changes in Dunaferr's operations, and because of the consistent weakening of the company's international position from the 1990s, the situation of the city became increasingly critical (which could not be counterbalanced even by new investments). Then,

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

following state-of-the-art developments at the company, its aspiration to open to the Russian market generated positive changes. The EP value of Tiszaújváros fell to nearly one-fifth of its corresponding value in 1992, and the city reached its nadir ranking 56th in 2012. With its population of less than 16 thousand, the position of Tiszaújváros is fundamentally determined by the presence of Mol Nyrt's subsidiary, TVK Nyrt., which attracted such a global enterprise to the city as the American Jabil Circuit, currently 7000 employees strong. The city's position was negatively influenced (2010: 25th position, 2012: 56th position) by the American AES Corporation closing its steadily loss-making Tisza II Power Plant of Tiszaújváros in 2012 and the consequential dismissal of three-quarters of the workforce.

Unlike Dunaújváros and Tiszaújváros, which boast relatively complex economic structures, the positions of most traditional industrial centres are determined by only one company such as the Swedish Electrolux in Jászberény, BorsodChem Zrt. owned by the Chinese Wanhua Industrial Group in Kazincbarcika, and Borsodi Sörgyár Zrt., a subsidiary of Molson Coors Brewing in Bőcs. During the two decades under review, among the leading economic centres, only Százhalombatta had a stable position (ranging from 20th to 30th) thanks to two companies. Those companies were 1) the Dunamenti Power Plant (owned by the French GDF Suez), the biggest power plant in Hungary, consisting of six blocks with a combined capacity of more than 2000 MW, and 2) Hungary's only oil refinery, the 8.1 million ton capacity Duna Refinery run by Mol Nyrt.

A common characteristic of chemical, industrial (Tiszaújváros, Kazincbarcika, and Százhalombatta) and metallurgic centres (Dunaújváros) is that the energy demands of the factories are satisfied by smaller or larger power plants. On the other hand, there are certain extraordinary cases when solely the power plant determines the position of the city. The most important power plant city is Paks, with Hungary's only nuclear power plant. Although in 2012 this four-block power plant of 2000 MW capacity produced a record 45% of the country's overall electric power output, the position of Paks weakened; its EP value dropped to nearly one-quarter of its former value, and the city lost 16 points in rank. This can be explained by the fact that, while the operating costs of the power plant steadily increase, because of its full state ownership, the price of the generated energy must be kept low. However, the projected expansion of the nuclear power plant in Paks has the potential to improve the position of the city again. A special example is Visonta, with less than 1100 inhabitants, as its position (2012: 46th place) is almost exclusively determined by Mátrai Hőerőmű Zrt., with a 950 MW capacity plant owned by the German RWE and EnBW energy companies, as well as lignite mines integrated with the power plant.

Tourism cities and towns

According to the National Tourism Development Strategy[4], Hungary's tourist attractions with the most significant international appeal are the capital, Budapest, and the Balaton region. In view of the geographic distribution of guests in commercial accommodation units, Behringer and Kiss (2004) concluded that Hungary can be described as having a very

4 National Tourism Development Strategy of Hungary 2005-2013 (http://neta.itthon.hu/szakmai-oldalak/strategiai-dokumentumok/nemzeti-100112)

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

strong concentration around Budapest/Lake Balaton but with very different annual preferences. According to the analysis provided by Sulyok and Kiss (2006), Budapest receives tourists at nearly the same intensity throughout the year. However, in the case of Lake Balaton, significant summer (from June to August) seasonality can be detected; with nearly two-thirds of the guest nights purchased for this part of the year. Besides Budapest and Balaton, only the spa cities and towns (e.g. Bükfürdő, Gyula, Hajdúszoboszló, Hévíz, and Zalakaros), which are less affected by seasonal effects, are tourist attractions of international appeal.

In all respects, Budapest can be regarded as Hungary's most significant tourist destination; however, it cannot boast of such an outstanding position that it can be assumed based on its EP value. In 2012, the capital's EP value exceeded the combined EP value of the next 999 settlements (Figure 2), but – for instance – with respect to the number of guest nights, the following 20 settlements (Table 2) exceeded that of Budapest. Among the leading tourist cities and towns, the bathing towns (e.g. Hévíz, Hajdúszoboszló, Bük, Sárvár, Zalakaros, Gyula, Harkány, and Egerszalók) have outstanding roles, including the settlements around Lake Balaton (e.g. Siófok, Balatonfüred, and Balatonszemes) and the large cities (e.g. Sopron, Győr, Eger, Szeged, Debrecen, Miskolc, and Pécs) offering complex tourist attractions.

Table 2

Most visited cities and towns in Hungary by overnight stays in hotels, campsites and other collective accommodation establishments, 2012

City/town	Rank by overnight stays	Overnight stays	Population	Proportion (Hungary = 100%)	EP Rank	EP
Budapest	1	7,412,561	1,735,711	33.99	1	51.07
Hévíz	2	1,004,622	4,663	4.61	157	0.02
Hajdúszoboszló	3	712,764	23,988	3.27	6	1.23
Bük	4	635,181	3,454	2.91	295	0.01
Siófok	5	625,333	25,441	2.87	24	0.38
Balatonfüred	6	479,711	13,313	2.20	117	0.04
Sárvár	7	453,000	14,812	2.08	80	0.08
Zalakaros	8	403,133	1,849	1.85	211	0.02
Sopron	9	369,103	60,528	1.69	34	0.25
Győr	10	357,916	128,567	1.64	2	3.37
Eger	11	304,187	54,867	1.40	36	0.24
Gyula	12	296,690	31,199	1.36	101	0.05
Szeged	13	250,649	161,837	1.15	10	0.89
Debrecen	14	248,397	204,333	1.14	4	1.48
Miskolc	15	243,622	162,905	1.12	12	0.77
Pécs	16	203,138	147,719	0.93	16	0.55
Visegrád	17	180,914	1,795	0.83	212	0.02
Harkány	18	163,625	4,087	0.75	233	0.02
Balatonszemes	19	149,602	1,827	0.69	–	0.00
Egerszalók	20	137,828	1,927	0.63	356	0.01
Total		*14,631,976*		*67.10*		*60.50*

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

A comparison of the EP ranking (Table 1) with the list compiled based on the number of guest nights (Table 2) highlights the fact that settlements with mainly tourism profiles carry outstanding significance, yet their positions tend to be weak. Nine settlements of the leading tourism centres could not manage the TOP100 in the EP ranking (Balatonszemes was not in the TOP1000, either), while their combined EP value was only 0.27, corresponding to Kaposvár's EP value (45th position). This means that tourism – no matter the outstanding significance this segment of the national economy is attributed in Hungary (Tóth 2009) – influences the positions of the individual settlements to a negligible extent. In the case of large cities, the role and significance of tourism can be observed even to a smaller extent.

However, the EP ranking includes two medium-sized cities whose tourism attitudes have a crucial influence on their leading positions, although they both host a manufacturing firm and utility company with a determining nationwide role. One of these medium-sized cities is Siófok (2012: 24th position), the most important node of tourism in the Balaton region. According to Beluszky & Győri (2004), beside its role in tourism, Siófok has already established typical urban institutions, while its economic function is strengthened by the presence of the German Eckes-Granini Group-owned Sió-Eckes Kft., Hungary's largest producer of fruit juices. The other important tourism centre of the EP ranking is Hajdúszoboszló (2012: 6th position), which is considered a highly visited bathing town even on a European scale (Erfurt-Cooper–Cooper, 2009). In addition, it hosts the headquarters of Tigáz Zrt. (owned by the Italian ENI), one of Hungary's largest gas suppliers, operating in nearly 1100 settlements and serving some 1.2 million consumers.

New economic centres

After the change in the political and economic system, the tendencies and territorial allocation of foreign direct investments (FDI) strongly influenced which settlements – how and when – could become the new nodes of the economy. At the beginning of the 1990s, foreign multinational corporations first appeared in Budapest, and then in the cities of the Western Transdanubian region (Barta 2005). Lux (2005, p. 85.) suggests that a west-to-east decline became prevalent due to the distance, but, even among the areas of the Transdanubian region, only those that had appropriate transportation facilities could attract capital. Consequently, it was the settlements of the Budapest Agglomeration (e.g. Budaörs, Gödöllő, and Törökbálint) that first closed the gap with the traditional economic centres. The large cities of the Western Transdanubian region (e.g. Győr, Székesfehérvár, and Szombathely) later strengthened their positions.

At the beginning of the 1990s, the target areas of FDI were generally small- or medium-sized cities and towns, where foreign multinational corporations established considerable manufacturing subsidiaries within the framework of greenfield investments. A general characteristic of the new economic centres is that a small number of (or frequently just one) multinational corporations determine their EP value, while previously they used to have very weak power. Immediately after the change in the political system, one of the largest investments was the establishment of General Motor's Opel factory in Szentgotthárd, which started to manufacture engines in 1992. In that year, based on its EP value, Szentgotthárd was still not among the first 1000 settlements. Although, in 1993, it

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

ranked 25th, by 1995 – ahead of large cities such as Szeged, Székesfehérvár, Pécs and Miskolc – it was the 5th most important economic centre. Naturally, Szentgotthárd could preserve that position only until other settlements also became scenes of investments of a similar scale. Therefore, it is not surprising that in 2012, it was only 65th in its EP ranking. Other new economic nodes, such as Bicske, Esztergom, Hatvan, Jászfényszaru, Komárom, Rácalmás, and Tab, witnessed very similar economic processes. These processes also contributed to the growth of the Budapest Agglomeration's new economic centres (e.g. Biatorbágy, Budaörs, Fót, and Törökbálint).

The course of growth taken by Szentgotthárd is characteristic of all the new economic centres. Table 3 shows that these settlements had much poorer positions in 1992 than in 2012; moreover, in terms of the rate of EP growth (more than 1000 %), Jászfényszaru, Rácalmás, Biatorbágy and Szentgotthárd were entirely special cases.

Table 3

Change of the position and Economic Power of the new economic centres, 1992/2012

City/town	Population, 2012	1992		2012		EP change, 1992=100%	Headquartered MNCs
		EP	Rank	EP	Rank		
Budaörs	27,306	0.24	45	1.42	5	491.76	BAT, Tchibo, Metro, Tesco, Auchan, Total
Bicske	11,813	0.14	72	0.69	13	382.53	Spar
Törökbálint	13,015	0.26	41	0.47	17	76.39	Telenor, dm-Drogerie Markt
Gödöllő	32,792	0.40	28	0.42	19	3.38	Teva, Avon Cosmetics
Komárom	19,200	0.13	79	0.40	22	214.40	Nokia, Foxconn
Biatorbágy	12,638	0.03	178	0.37	25	1021.19	Lindab, Scania, Atlas Copco, Ruukki
Jászfényszaru	5,664	0.01	562	0.36	26	6755.63	Samsung Electronics, Samsung C&T
Jászberény	26,809	0.26	42	0.36	27	36.37	Electrolux
Esztergom	28,550	0.10	92	0.35	28	245.88	Suzuki, Tyco Electronics
Fót	18,927	0.09	97	0.33	29	243.74	Philip Morris
Vecsés	20,164	0.04	165	0.30	31	715.32	Wizz Air (until 2011)
Hatvan	20,525	0.12	83	0.23	39	93.26	Robert Bosch
Rácalmás	4,479	0.01	442	0.22	40	2715.10	Hankook Tire
Szentgotthárd	8,787	0.00	–	0.11	65	–	General Motors

A particular correlation can be detected between the territorial allocation of the new economic centres and the types of settled multinational corporations. The settlements of the Budapest Agglomeration (primarily Biatorbágy, Budaörs, Fót, Gödöllő, and Törökbálint) mostly accommodate retail companies (e.g. Metro, Tesco, Auchan, and dm-Drogerie Markt) as well as the Hungarian or Central European regional commercial centres of manufacturing companies (e.g. BAT, Teva, Avon, and Philip Morris). On the other hand, the new economic centres lying farther from the capital (e.g. Esztergom, Jászfényszaru, Jászberény, Komárom, and Szentgotthárd) tended to attract manufacturing firms (e.g. Suzuki, Samsung, Electrolux, Nokia, Foxconn, and Opel).

Nevertheless, a critical factor is that the new economic centres – even the relatively populous Komárom and Esztergom – have fairly one-sided economic structures because they are determined by a single or a small group of manufacturing units. The inherent risks are most clearly observed in the example of Esztergom. The Japanese Suzuki Motor Corporation launched its Hungarian production in Esztergom in 1992; in the same year, the city ranked 92nd in the EP ranking. By 1994, manufacturing was expanded to two shifts, export production commenced, and the city found itself rising to the 42nd position. In 1996,

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

one of every five new cars commissioned in Hungary was released from Esztergom's Suzuki factory, resulting in the 31st position for the city. In 2007, Esztergom was the 16th largest economic centre, although this upward trend was broken by the economic crisis. The same view is reflected in Magyar Suzuki Zrt's announcement made in December 2008: "*As a consequence of the global financial and economic crisis as well as the deterioration of market conditions, Magyar Suzuki Zrt's sales volumes have been drastically dropping, and therefore it has become necessary to reduce the production output in adaptation to the actual demands. On 8 December 2008, manufacturing was reset to the two-shift work schedule*."[5] Therefore, it is not surprising that in 2009, Esztergom slid back to 63rd position in the ranking. It is true, however, that due to the improvement of the global economic environment and Magyar Suzuki Zrt's more massive export operations, today the city has re-established itself as one of the leading economic centres (28th position in the 2012 EP ranking).

The case of Szentgotthárd and Esztergom clearly reflects the fact that individual multinational corporations settling in the new economic centres have the potential to influence the positions of these cities radically. For similar reasons, it is expected that in the future, Komárom will lose its 22nd position of 2012; the parent company of the city's largest firm, Nokia, announced at the beginning of 2012 that the manufacturing of smartphones will be relocated from Komárom (similarly to the Mexican Reynosa and Finnish Salo) to Asian plants. By the end of 2012, Komárom's Nokia factory had dismissed half of the employees (2300 workers), and the production output was cut back; in 2014, the firm ceased operation. Obviously, similar external effects may be suffered by economic centres of the Budapest Agglomeration, although because of the proximity of the capital as well as the retail and non-producing character of these companies, it is not very likely.

Offshore cities/towns

Countries that impose relatively small tax burdens – if any – on multinational corporations have become highly important actors of the global economy (Hines 2004). A common characteristic of tax havens is that, by keeping their corporate income tax rates low, they offer appropriate business environs to multinational corporations (Dharmapala 2008). As Hines (2004) suggests, the pivotal point in this respect is the application of a tax rate that is much smaller – or even 0% – than the effective 35% corporate income tax rate in the United States. This encourages corporate giants of the world's largest economy to find a place for their profits in tax havens so that, consequently, income-related taxes paid will be reduced in the United States,[6] and instead paid – yet obviously to a much smaller extent – in tax havens. Although opportunities to achieve considerable savings encourage multinational corporations to take similar steps in all developed countries, it is companies from the United States that excel in trying to avoid domestic tax payment. Dharmapala and

5 Announcement of the Magyar Suzuki Zrt. on the reduction of the workforce (http://www.suzuki.hu/pages/display/magyar_suzuki_zrt./cikk/cikk:2008_december_19._-_a_magyar_suzuki_zrt._ kozlemenye_a_letszamcsokkentessel_kapcsolatban)

6 Bloomberg claims that Microsoft, Apple, and Google together keep a fortune of USD 134.5 billion outside the United States. The offshore list compiled by the news agencies is topped by the conglomerate of General Electric, with USD 108 billion followed by the pharmaceutical giant Pfizer, with USD 73 billion. (http://www.bloomberg.com/news/2013-03-08/offshore-cash-hoard-expands-by-183-billion-at-companies.html)

Hines (2006) identified nearly 40 tax havens, such as Bermuda, Hong Kong, the Cayman Islands, Liberia, Panama, and Singapore, as well as Cyprus, Ireland, Luxemburg and Switzerland in Europe. In the past, Hungary was not classified by the OECD (2000) or researchers (Hines 2004, Dharmapala–Hines 2006, Dharmapala 2008) among official tax havens, but according to Gravelle (2009, 2013), it would be justified to apply this category to Hungary, in the same company as Austria, the United States (Delaware, Nevada, and Wyoming), the United Kingdom and Canada.[7]

In Hungary, foreign companies were first allowed to establish offshore entities not involved in domestic economic operations in 1994. These companies did not pay local taxes or VAT, and until 2004 they were required to pay 3%, and then until 2005 4% corporate tax. It is not surprising that the number of offshore entities steadily rose, while after the initial few billions of Hungarian Forint (HUF), the aggregate amount of their registered capital in 2005 exceeded HUF 2000 billion. This growth of offshore companies came to an end with the country's accession to the European Union because the corporate income tax rate was increased to the standard 16% level (which is still much lower than the 35% rate applied in the United States); currently, it is 19%. Despite this, the tax regulations still permitted the entities registered in Hungary to transfer profits to the foreign parent companies without paying withholding taxes, and arrange funding within the individual groups with no need to pay the local trading tax. In spite of the increased corporate tax rate, these allowances proved to be extremely favourable conditions.

Table 1 reflects that in 2012, Újlengyel (ranked 9th) and Csomád (ranked 15th) were among Hungary's leading economic centres. The former settlement first hit the level of the EP 1000 ranking as early as 1994; the latter reached this category in 2004, yet their EP values drastically increased until 2012. Csomád's EP value showed a growth of 17.616%, whereas Újlengyel's EP value rose by 41.742%. The following question can be legitimately raised: which giant domestic companies or large foreign multinational firms have created huge manufacturing units in these settlements (similarly to e.g. Győr or Székesfehérvár), or which international retail chains or banks have relocated their Hungarian headquarters there (similarly to e.g. Budaörs or Törökbálint)? The answer is simple: there are no such companies or banks. Neither Újlengyel, with a population of 1672, nor Csomád with a population of 1546, has factories, banks, or retail chain headquarters. These small towns in Pest County are the most well known Hungarian offshore towns where companies state compelling financial parameters while failing to perform any actual domestic operation. Obviously, there are also large cities where offshore companies have settled – one of the best examples being Szombathely (HVG 2011). However, offshore towns, e.g., Újlengyel and Csomád, are not scenes of any considerable economic activity.

Based on the economic performance of registered offshore companies, Hungary's largest offshore town is Újlengyel in Pest County, as the combined assets of companies headquartered in the town correspond to the combined assets of Miskolc, population 163,000, or Kecskemét, population 112,000. In Újlengyel, a seemingly powerful corporate empire has been built up by Transocean Ltd, the world's largest offshore drilling contractor, under the name of Triton. Obviously, the offshore subsidiaries of foreign multinational

7 Others state that Hungary is not a tax haven, but rather a country of low tax regime. (OECD Tax Database: http://www.oecd.org/ctp/tax-policy/tax-database.htm#C_CorporateCaptial)

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

corporations have also established themselves in other settlements, such as Fibria Celulose[8] (Sao Paulo, Brazil) operating in the pulp and paper industry in Csomád or the Petrobras (Rio de Janeiro, Brazil) oil giant in Szombathely (HVG, 2011).

After Hungary acceded to the European Union, foreign multinational corporations still saw benefits in running offshore enterprises in Hungary despite the increased rates of incomes tax and the bureaucratic tax regime, as it proved to be a huge advantage that Hungary was not declared a tax haven. The tax authorities of the United States tend to sharply distinguish companies that have registered themselves in tax havens from those settling in countries of average tax rates[9], irrespective of whether they are actually involved in offshore operations. It is highly likely that the – otherwise not too substantial – correlation unveiled between the 2010 disaster of the Deepwater Horizon drilling rig and Újlengyel may contribute to Hungary's judgment as an offshore tax haven[10] (Gravelle 2013). In case this negative image does affect the offshore operations of multinational corporations, the positions of domestic offshore settlements are expected to weaken considerably in the future.

Summary

This study has attempted to use a method, applied in international practice, to identify and rank cities and towns as centres of the Hungarian economy. Therefore, a complex indicator (Economic Power – EP) has been introduced to integrate financial parameters describing economic entities that operate in the individual cities and towns. As data were available from the early 1990s, it was possible to demonstrate relatively long-term changes in the positions of the economic centres instead of their less-informative static conditions. The strong centralization of the Hungarian economy is clearly reflected in that, every year, the 100 largest economic nodes provided an average of 80% of the national EP value, whereas Budapest itself covered 45–50%. The other 900 settlements involved in the analysis made only a 15% contribution overall to Hungary's economic performance.

The EP values of the cities and towns were influenced by vastly different economic factors that varied over time; however, it was possible to establish a classification corresponding to the objectives of the analysis. In the analysis, we identified the following categories of leading economic centres: a metropolis with a complex industrial profile (Budapest); large and medium-sized cities with a complex industrial profile (the county seats and the quasi-regional centres); traditional industrial cities; new economic centres; tourism cities and towns; offshore cities/towns, and other cities, towns and villages. The

8 Fibria Celulose SA (http://www.fibria.com.br/rs2012/fibria-sustainability-report-2012.pdf)

9 The International Consortium of Investigative Journalists. Authorities Announce Tax Haven Investigation. (http://www.icij.org/blog/2013/05/authorities-announce-tax-haven-investigation)

10 Of course, Újlengyel's offshore function has always been known in Hungary (it is indicated by the numerous newspaper articles written about offshore activities in the settlement), yet international organizations have started to focus on the settlement just recently. The underlying reason is that in the Gulf of Mexico, 2010 witnessed the sinking of the Deepwater Horizon deep-sea drilling rig rented at that time by BP, the British oil multinational corporation, from its owner, Transocean Ltd. In the history of the United States, it has been the largest environmental disaster so far. In terms of the expanse of the oil contamination over the sea, it was the second largest such catastrophe of all time. The follow-up investigation and court proceedings revealed that the drilling rig was once owned by Triton Hungary Asset Management LLC, based in Újlengyel, although in 2009 – less than a year before the disaster – it was transferred to Triton Asset Leasing GmbH, headquartered in Zug, Switzerland.

economic structures of Budapest and the regional centres are highly complex; for them, no dominant economic factors can be observed, although there is no doubt that the economic performance of regional centres was perceptibly affected by changes in FDI. Traditional industrial centres featuring mostly weakening positions were gradually replaced by the new economic centres.

These generally small- or medium-sized cities and towns, which used to be negligible actors of the economy, have become significant economic nodes because of investments by foreign multinational companies. On the other hand, their achieved positions are very unsteady – probably, only the economic centres in the Budapest Agglomeration possess relatively stable positions – because, as a consequence of any negative investment decision, they could disappear as quickly as they emerged. Less-known actors of the Hungarian economy are offshore settlements that are usually not examined in similar analyses (primarily because they do not contribute at all to employment). In these settlements with a population of 1000-2000, offshore companies pursue activities only on a superficial level – yet they have real economic power (which is reflected by the huge amounts of corporate tax they pay). Offshore cities and towns have tried to exploit the particular characteristics of domestic tax regulations to become the target areas of foreign multinational companies pursuing offshore activities, while their positions – moreover their existence in this case – are primarily influenced by external factors.

In 1992, the leading economic centres were arranged along a relatively logical scheme: Budapest with its complex economic structure was followed by the regional centres and larger county seats, and just certain special industrial centres were included among them. By 2012, this structure had been transformed, becoming more complicated, with domestic and foreign effects equally influencing its changes.

* * *

Acknowledgement: This paper is supported by the **János Bolyai Research Scholarship** of the Hungarian Academy of Sciences.

REFERENCES

Antalóczy, K.–Sass, M. (2005): A külföldi működőtőke-befektetések regionális elhelyezkedése és gazdasági hatásai Magyarországon *Közgazdasági Szemle* 52 (5): 494–520.

Antalóczy, K.–Sass, M.–Szanyi, M. (2011): Policies for attracting foreign direct investments and enhancing its spillovers to indigenous firms: The case of Hungary In: Rugraff, E.–Hansen, M.W. (eds.) *Multinational Corporations and Local Firms in Emerging Economies* pp. 51–69. Amsterdam University Press, Amsterdam.

Bajmócy, P.–Hegedűs, G. (2008): Transformation of the settlement system in post-socialist Hungary In: Kertész, Á.–Kovács, Z. (eds.) *Dimensions and trends in Hungarian geography (Studies in Geography in Hungary, 33.)*. pp. 138–161., Hungarian Academy of Sciences, Geographical Research Institute, Budapest.

Baráth, G.–Molnár, B.–Szépvölgyi, Á. (2001): A külföldi működőtőke szerepe a Közép-Dunántúl átalakuló gazdaságában *Tér és Társadalom* 15 (2): 183–200.

Barta, Gy. (2001): A nagyvállalatok szervezeti – tulajdoni – térbeli változásai *Tér és Társadalom* 15 (1): 36–64.

Barta, Gy. (2005): The role of foreign direct investment in the spatial restructuring of Hungarian industry In: Barta, Gy. (ed.) *Hungarian Spaces and Places: Patterns of Transition* pp. 143–160., Centre for Regional Studies, Pécs.

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Barta, Gy.–Kukely, Gy.–Lengyel, B.–Ságvári, B. (2007): Magyarország a globális K+F térképén: Fejlődő országok a multinacionális vállalatok változó K+F stratégiájában *Tér és Társadalom* 21 (3): 31–50.

Beluszky, P. (2003): *Magyarország településföldrajza (Általános rész)* Dialóg Campus Kiadó, Budapest–Pécs.

Beluszky, P.–Győri, R. (2004): Fel is út, le is út... (városaink településhierarchiában elfoglalt pozícióinak változásai a 20. században) *Tér és Társadalom* 18 (1): 1–41.

Behringer, Zs.–Kiss, K. (2004): The role of foreign direct investment in the development of tourism in post-communist Hungary In: Hall D (ed) *Tourism and Transition: Governance, Transformation, and Development* pp. 73–81. CABI, Wallingford.

Boudeville, J.R. (1966): *Problems of Regional Economic Panning* Edinburgh University Press, Edinburgh.

Cséfalvay, Z. (2001): The new role of Budapest in the Central European City System In: Meusburger P, Jöns H (eds) *Transformations in Hungary: essays in economy and society* pp. 273–289., Physica-Verlag, Heidelberg.

Csomós, Gy. (2009a): A monocentrikus térszerkezet kialakulásának okai és a policentrikus területfejlesztés jövőképe: különbségek és hasonlóságok Franciaország és Magyarország területi tervezésében *Társadalomkutatás* 27 (2): 163–184.

Csomós, Gy. (2009b): A regionális centrumok súlyának meghatározása Magyarország településhálózatában *Területi Statisztika* 49 (2): 186–198.

Csomós, Gy. (2013): The command and control centers of the United States (2006/2012): An analysis of industry sectors influencing the position of cities *Geoforum* 50: 241–251.

Csomós, Gy.–Derudder, B. (2014): European cities as command and control centres, 2006-11. *European Urban and Regional Studies* 21 (3): 345–352.

Diczházi, B. (1997): A külföldi tőkebefektetések hatása a regionális gazdaságra *Tér és Társadalom*, 11 (2): 67–79.

Dharmapala, D.–Hines, J.R. Jr. (2006): *Which Countries Become Tax Havens?* Working Paper No. 12802, National Bureau of Economic Research, Cambridge, MA.

Dharmapala, D. (2008): What Problems and Opportunities are Created by Tax Havens? *Oxford Review of Economic Policy* 24 (4): 661–679.

Enyedi, Gy. (2000): Globalizáció és a magyar területi fejlődés *Tér és Társadalom* 14 (1): 1–10.

Erfurt-Cooper, P.–Cooper, M. (2009): *Health and wellness tourism: spas and hot springs* Channel View Publications, Bristol.

Gál, Z. (1998): A pénzintézeti szektor területfejlesztési kérdései Magyarországon *Tér és Társadalom* 12 (4): 43–67.

Glickman, N.J. (1987): Cities and the international division of labour In: Smith, M.P.–Feagin, J.R. (eds.) *The Capitalist City* pp. 66–86. Blackwell, Oxford.

Gravelle, J.G. (2009): Tax Havens: International Tax Avoidance and Evasion *National Tax Journal* 62 (4): 727–753.

Gravelle, J.G. (2013): *Tax Havens: International Tax Avoidance and Evasion* CRS Report for Congress Congressional Research Service, Washington.

Hines, J. R. Jr (2005): Do Tax Havens Flourish? In: Poterba JM (ed) *Tax Policy and the Economy, 19*. pp. 65–99., MIT Press, Cambridge, MA.

HVG (2011): Virtuális székhely: több száz cég fordult meg egy szombathelyi kertes házban. http://hvg.hu/kkv/20111110_szombathely_adooptimalizalas (downloaded: 01.08.2015)

Juhász, K.–Schottner, K. (2003): Németország és az Amerikai Egyesült Államok működőtőke-befektetései Magyarországon (1993-2000) *Tér és Társadalom* 17 (4): 85–99.

Koltai, Z. (2007): A magyarországi városok versenyképességének vállalati megítélése *Tér és Társadalom* 21 (2): 23–42.

Koós, B. (2004): Adalékok a gazdasági szuburbanizáció kérdésköréhez *Tér és Társadalom* 18 (1): 59–71.

Kukely, Gy. (2006): A nagyvárosok felértékelődése a külföldi működőtőke-beruházások telephelyválasztásában *Tér és Társadalom* 20 (4): 111–125.

Lausén, J.R. (1969): On growth poles *Urban Studies* 6 (2): 137–161.

Lengyel, I. (1999): Mérni a mérhetetlent? A megyei jogú városok vizsgálata többdimenziós skálázással *Tér és Társadalom* 13 (1–2): 53–74.

Lengyel, B. (2012): The Hungarian ICT Sector: A Comparative CEE Perspective with Special Emphasis on Structural Change In: Welfens PJJ (ed) *Clusters in Automotive and Information & Communication*

REGIONAL STATISTICS, 2015, VOL 5, No1: 66–85; DOI: 10.15196/RS05104

Technology. Innovation, Multinationalization and Networking Dynamics. pp. 59–85., Springer-Verlag, Berlin-Heidelberg.

Lux, G. (2005): A magyar fejlesztési politika térszemlélete és a második generációs programozás területi vetületei *Tér és Társadalom* 19 (3–4): 81–93.

Lux, G. (2013): A gazdaság szerepe a városi térségek fejlesztésében: A globális kihívásoktól a fejlesztéspolitikáig In: Somlyódyné Pfeil, E. (ed) *Az agglomerációk intézményesítésének sajátos kérdései: három magyar nagyvárosi térség az átalakuló térben* pp. 67–89., Publikon Kiadó, Pécs.

Pálné Kovács, I. (2001): *Regionális politika és közigazgatás* Dialóg Campus Kiadó, Budapest–Pécs.

Perczel, K. (1989): A magyarországi regionális tervezés történetéhez *Tér és Társadalom* 3 (3): 80–105.

Perroux, F. (1955): *Note sur la nation de pole de croissance* Économie Appliquée 8.

Rechnitzer, J. (1998): *Területi stratégiák* Dialóg Campus Kiadó, Budapest–Pécs.

Rechnitzer, J. (2004): A városhálózat és a régiók formálódása *Magyar Tudomány* 165 (9): 978–990.

Rechnitzer, J.–Csizmadia, Z.–Grosz, A. (2004): A magyar városhálózat tudásalapú megújító képessége az ezredfordulón *Tér és Társadalom* 18 (2): 117–156.

Rechnitzer, J.–Smahó, M. (2012): *Járműipar és regionális versenyképesség Nyugat- és Közép-Dunántúl a kelet-közép-európai térségben* Széchenyi University Press, Győr.

Sági, Zs. (2000): A külföldi tőke szerepe a budapesti agglomerációs övezet feldolgozó-ipari térszerkezetének kialakításában *Tér és Társadalom* 14 (2-3): 73–87.

Sassen, S. (2006): *Cities in a world economy. 3rd Edition* Pine Forge Press, Thousand Oaks, CA.

Sulyok, J.–Kiss, K. (2006): A magyarországi turizmus szezonalitása, 2000-2004 *Turizmus Bulletin*, 10 (1): 57–69.

Szemző, H.–Tosics, I. (2005): Hungary In: van Kempen, R.–Vermeulen, M.–Baan, A. (eds) *Urban Issues and Urban Policies in New EU Countries* pp. 37–60., Ashgate, Aldershot.

Tóth, G. (2009): Kísérlet a regionális turisztikai GDP becslésére Magyarországon *Statisztikai Szemle* 87 (10–11): 1038–1057.

Wágner, I. (2004): Magyar regionális bankközpontok felkutatása és azonosítása *Tér és Társadalom* 18 (2): 107–116.

ANDRÉ MAIA GOMES LAGES[a) – FABRÍCIO RIOS NASCIMENTO SANTOS[b)]
– HUMBERTO BARBOSA FERREIRA JUNIOR[c)]

Preference for Liquidity of Agents: An Analyse of Brasilian Case

Abstract

This work is meant to show the relevance of the role of money in explaining regional disparities. Points out that before the currency and the banks are incorporated theories of regional development, had two views on regional development, founded on a convergence of unequal growth and divergence in another, where the rates would become increasingly unequal. The literature on the regional economy have given little attention to financial variables and their role in regional development. In this context, the currency has received secondary treatment in the analysis of the regional economy, perhaps in the belief of some theorists in the neutrality of money in the long run, or others who have relied on the assumption of perfect interregional mobility of capital. However, this perspective has been changing in recent years, particularly in post-Keynesian agenda. Thus, we intend to examine the behavior of the public regarding the preference for liquidity in the face of regional characteristics and the financial instability and therefore demonstrate their relevance to explain the differences in regional economic development. To analyze the decision to demand money was used educational and behavioral aspects. The hypothesis that there is a financial concentration in regions with a lower liquidity preference was ratified. For this, the study was developed based on the analysis of units of the Brazilian federation. The database of the Central Bank, and Datasus allowed the use of the formula suggested by the literature pertinent to the theme

Keywords: uncertainty, regional development, banks, Brazilian economy, interest rates.

Brazilian economy. Some historical fundaments

Brazil has five regions: North, Northeast, Central-West, Southeast and South. Historically, since its discovery in 1500, its economic evolution has developed mainly the Northeast, Southeast and South regions, facilitated by their proximity to the coast. The data indicates that the South and Southeast regions were the leaders in the development. The growth in the Northeast was, for a few centuries, attached to agricultural exporting activity, having

a) Faculdade de Economia, Administração e Contabilidade – Universidade Federal de Alagoas, A.C. Simões Av. Lourival de Melo Mota, Bloco 16, 1º andar Tabuleiro do Martins Maceió Alagoas, Brasil. E-mail:andre_lages@msn.com

b) Universidade Federal de Alagoas, Campus Sertão, Unidade Santana do Ipanema. Rua Prefeito Adeildo Naponuceno Marques, 472 - Monumento - 57500-000. Santana do Ipanema Alagoas, Brazil. E-mail:fabriciorioss@bol.com.br

c) Centro de Estudos Superiores de Maceió, Rua Cônego Machado, 918, Farol CEP: 57051-160, Maceió, Alagoas, Brazil. E-mail:hbfjr@ig.com.br

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

sugar as the main product. The sugar cycle growth was supported by slave labour (Africans) and to a lesser by natives (Indians). The Northeast's geography is dominated by called caatinga a common type of vegetation found in semi-arid regions. That condition is characterized by long dry periods with temporary rivers.

Other circumstances have leaded to Japanese as well as European immigration to Sao Paulo (the richest state in Brazil) and other states of the Southeast region of the country. Sao Paulo was a prime producer of coffee. Another flow of immigrants from Italy and Germany also headed to the South Region.

The end of the slavery in the 19th century, the Great Depression in the 1930s and two World Wars created the conditions for the South and Southeast regions to evolve economically faster, and São Paulo concentrating on the industrialization process up to the current day. In this context, it should be emphasized that the end of slavery monetized the Brazilian economy; the Know-how brought by immigrants and the strong expansion of the coffee culture created the seminal conditions for the industrial concentration in Sao Paulo. The other two regions, Central-West and North, with their logistical difficulties, were considered by that time an expanding agricultural frontier. By the end of the 1950s and beginning of the 1960s; it happened the construction of the new capital on Central-West region. The new Capital Brasilia induced a displacement of a considerable part of the Brazilian population to that region.

Table 1

Evolution of the Resident Population, 1991–2010

Brazil and Great Regions	1991	2000	2010
Brazil	146 825 475	169 799 170	190 755 799
North	10 030 556	12 900 704	15 864 454
Northeast	42 497 540	47 741 711	53 081 950
Southeast	62 740 401	72 412 411	80 364 410
South	22 129 377	25 107 616	27 386 891
Central-West	9 427 601	11 636 728	14 058 094

Source: IBGE – Censos Demográficos, 1991, 2000, 2010.

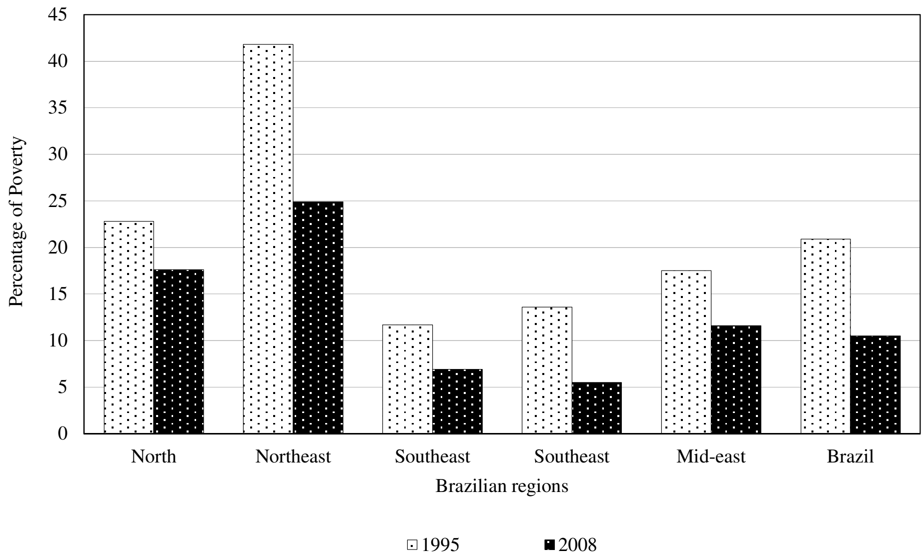
As Figure 1 shows below, despite development, up to 2010, the Northeast region suffers with the most social problems; here, the greatest percentage of the population earn one-quarter of the official minimum wage in comparison with others regions.

It is important to explain that the Brazilian population was distributed as indicated in Table 1. In the reality. The expansion of the less populated regions still implies a smaller number of banks in those regions; supporting observations from Table 1 and issues around liquidity (monetary).

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

Figure 1

*Extreme Poverty *Rate in the Brazilian Regions in 1995 and 2008*



* According to IBGE, extreme poverty rate are families with income around 1/4 of the offical minimum wage.
Source: IBGE – PNAD – (IPEA display elaboration).

To conclude, Figure 1 reflects the evolution of historical fundaments to the current day. It will become still more interesting with the Post-Keynesianism approach on sequence.

Uncertainty

Before the currency and banks were approached by regional development theories, Chick (2006) points out that there were two points of views on regional development -one founded on the convergence of uneven growth rates and another on divergence, where rates become increasingly unequal. The first is guided with the perspective that underdeveloped regions compared with developed regions, would offer higher returns on investments, since, with increasing investment in the given region, returns decrease; thusthose underperforming regions tend to receive greater incentives than the more prosperous areas. The second view holds that prosperous rather than the backward regions would attract investment because progress generates progress, which will increase the inequalities in growth rates.

The literature on the regional economy has given little attention to the financial variables and their role in regional development. In this context, the currency has received secondary treatment in the analysis of the regional economy; some theorists believing, perhaps, in the neutrality of money in the long run, or others who have relied on the assumption of perfect inter-regional mobility of capital, where money flows reproduce only regional differences in real terms (Dow–Fuentes 2006, Nogueira et al. 2010).

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

However, that perspective has been changing in recent years, mainly in the Post-Keynesian agenda.

Post-Keynesians, like Keynes, have the currency as a key element in the economic process and do not diminish its influence in the real side of the economic system. Therefore, it is not possible to completely distinguish the money side, and the real side of economy. The coverage given to the currency is based on the idea of a Monetary Production Economy defined by Keynes to explain certain peculiarities of the modern capitalist system. This approach points out that one of the keys caracterísitcas is the process of decision making in uncertain environment. So, the economic agents will define what to do with the money based on their expectations about the future. (Figueiredo–Crocco 2008, Crocco et al. 2011).

From the eighties, in his theoretical and analytical explaination of regional differences, the regional development issue received attention, especially as regards the use of money, because until then, the whole analysis was guided by real variables for this explanation. From Chick, and Dow's essays and contributions (1982, 1987), the currency has been considered fundamental in explaining such differences. That moment must be remembered by the construction of an autonomous post-Keynesian theoretical framework in terms of the relationship between the financial system and regional development (Figueiredo–Crocco 2008, Ribeiro–Marouvo 2011).

Unlike the neoclassical general equilibrium models and New Keynesians, Post-Keynesian theory, in its analysis, takes into account both the supply side as the demand of the regional credit market, using both the development stages of banking as the Keynesian principle of liquidity preference (Dow–Fuentes 2006).

Therefore, the supply and demand for credit are interdependent and affected by a preference for liquidity; these are related to the environment of uncertainty that pervades decision-making process (Nogueira et al. 2010).

Before discussing the concept of uncertainty, its variance from the concept of risk should be clarified; while one is identified by its characteristic of being non-measurable, the other can be measured in quantitative terms (Knight 1921).

According to Keynes (1991, p.274), "... the dismay and uncertainty as to the future which accompanies a collapse in the marginal efficiency of capital naturally precipitates a sharp increase in liquidity preference – and hence a rise in the rate of interest."

It is worth highlighting that this study will be basically used the demand for vision Keynesian currency exposed in his main book: The General Theory of Employment, Interest, and Money, (1936), and only for the purpose of organizing the argument. it t is recalled that Keynes (1991) defines the demand for money as based on the demand for transaction, precaution and speculation. He also emphasizes that the first reason for demanding currency can be subdivided into two reasons: income-motive and business-motive. In the first case, it is dependent on income and the normal period receipt and disbursement. For the second case, it makes it clear that the net proceeds are kept in terms of liquidity preference to ensure the time interval they start on the sales receipts. Thus, Keynes (1991) presents sets of reasons to retain liquid resources via the following formula:

$$M = M1\ (Y) + M2\ (R)$$

For analytical simplification, it is considered that M1 represents transaction and precautionary demand for money; where Y (income) represents a independent variable;

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

while M2 represents speculation motive, is dependent on and uniquely defines the behaviour of interest (R). Of course, one can point here to a certain reductionism in the reasoning of Keynes, but the purpose is purely didactic

Returning to the central argument, it is noted that the interest rate is a major component with respect to the liquidity preference and consequently, demand for money, since depending on the level, will lead the public to invest or not its cash balances in long-term assets. For Amado1998), the interest rate is the variable that reconciles the desires of savers and investors. According to Figueiredo and Crocco (2008), the interest rate is a monetary phenomenon and acts as a link between money and the real economy.

The behaviour of economic agents, concerning holding or not holding money will depend on its confidence regarding the current economic scenario as well as the expectations regarding the future, which Keynes coined in tha "Animal Spirits":

Keynes (1937 cited by Ferrari Filho–Conceição 2001, p.4), in response to critics of his famous obra4, writes an article entitled "The General Theory of Employment "which defines uncertainty, as follows:

> *"By 'uncertain' knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty … Or … the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence … About these matters, there is no scientific basis on which to form any calculable probability whatever. We simply do not know" (Keynes 1973 cited by Ferrari Filho–Conceição 2001, p.4).*

Post-Keynesians, as Keynes, state that in a changeable economy, agents prefer to hold money rather than make spending decisions. As Minsky stated (1975 cited by Ferrari Filho–Conceição 2001), to understand Keynes "it is necessary to understand his sophisticated view ofuncertainty, and the importance of uncertainty in his vision of the economic process. Keynes without uncertainty is something like Hamlet without the Prince".

In Post-Keynesian theory, all credit risk assessments are under the auspices of uncertainty, and the major uncertainty is the "gap" between the borrower and the lender, where the distance can be economic, social, spatial, cultural, etc. (Porteous 1995 cited by Dow–Fuentes 2006). Therefore, the greater the uncertainty, the greater the retention of currency, which in turn, results in a shortage of effective demand, making clear the influence of the currency on the side of real economy

On the demand side for credit, high preference for liquidity will affect the public and their respective portfolio decisions, making them demand more money and less active in the long-term, consequently, increasing liquidity (Crocco et al. 2006).

The hypothesis to be tested is that regions with higher liquidity preference presents a lower bank concentration, being clearly the most peripheral regions. It must be regarded in this case the theoretical and analytical perspective also drawn in Melissa et al. (2007).

[44] The General Theory of Employment, Interest and Money.

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

The reduced access to financial services, the lowest level of education, and the less degree of information help to form one negative institutional environment on less developed regions. On the other hand, it should not be forgotten that their business decisions and focus on resources are always more connected and in communion with the more developed regions, where they have usually installed the headquarters of banks

Following this, the paper stresses public behaviour in relation to the demand for money. The study will examine public behaviour in the 26 units of the Brazilian federation and the Federal District, observing signs relating to liquidity preference, and where there is a greater financial concentration, within the group of states and Distrito Federal. Concluding, where the public preference for liquidity is a key variable for distinguishing regional development; certain similarities of regional behaviour linked to the degree of regional development and/or its spatial location, according to a post-Keynesian perspective, can be perceived.

Methodology

The study was based on the secondary data tract. Collection sources were the sites of the Central Bank of Brazil (BCB) and DATASUS.

From the Central Bank site, information was collected on the number of bank agencies, demand deposits of the private sector and government, time deposits and savings in the period from 2004 to April 2011. DATASUS data on the population, aged 20 to 80 years or more in the 2004-2010 period, was also examined. The relationship is between population (within those limits) and number of bank branches. In Brazil, this is relevant because it should be noted States more developed or less developed is associated or not with such a relationship. That means when the ratio is smaller, implies certain state necessarily be developed? A simple algebraic division was used to assess in which states bank concentration predominates, and how this influences the regional dynamics.

To analyse public behaviour in relation to the allocation of its assets, the preference index for liquidity of the public (PLP) was used; this is presented by Crocco et al (2006) in his article entitled "regional polarization and the financial system", which captures how the public allocates decisions o its resources among assets or less liquidity, taking into account the degree of uncertainty and the amount of information available in the regions. Accounts used to generate the index were deposit accounts on demand, and savings and time deposits belonging to the liabilities of banks, which provide the public perception of the behaviour in general, and banks as intermediaries. The PLP formula is:

$$PLP = \frac{DV_P + DV_G}{DT} \quad (2)$$

where DV_P = deposits to private view, DV_G = government demand deposits, DT = total deposits (deposits include the private view and the government, term deposits and savings).

The higher the PLP index, the greater the liquidity preference of the public, signalling the public's distrust of the current economic scenario, i.e., they prefer to retain more liquid assets at the expense of less liquid investments in financial assets.

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

Results and Discussion

Looking at Table 1, it can be seen that in the Northeast and North regions, the greater preference for liquidity indices shows that in regions whose economic scenario does not transmit confidence. In this case, people don't invest their financial balances in long-term monetary assets, they will prefer to preserve liquidity. Then, they are on regions characterized by a reduced demand for credit. The uncertainty that permeates these regions is so strong .that even with the policy adopted in recent years by the federal government to increment such regions; the income transfer does not relax the situation. But the North has peculiarities such as: large investments in land belonging to businessmen residing in other parts of the country. In addition, lower population density and number of branches per inhabitant in some states.

Levino and Lages (2008), in analysing the public's attitude towards preference for liquidity in the years 1991, 1996 and 2000, found that the less developed areas showed a greater preference for more liquid assets while in dynamic and prosperous regions, the public would rather hold less liquid assets.

As can be seen, intraregional differences in the case of northeast Brazil, for the state of Ceará presented the lowest index compared to other states of Northeast region, allowing the conclusion that their population has shown reliable behaviour vis-a-vis future expectations. This in turn has led them to apply their resources to long-term assets and to demand more credit to realize investment opportunities. It is extremely important to note that Ceará achieved one of the lowest indexes in Brazil and not only in the Northeast, which allows it to be accredited it as a prosperous state.

On the other hand, it can be seen that the Southeast and South, and, in particular,the District Federal, the political centre, had a very small index; this indicates the regions with a more stable economic scenario, with a consolidated institutional apparatus and educational institutions as government presence. This in turn communicates greater public stability in relation to future expectations. It should be noted that the Federal District is a case in part because of its specific features (Melissa et. al. 2007). These regions are in the reality where are concentrated the most developed states of Brazil; as the numbers on tables show to exactly it.

It is noteworthy that the public had a higher preference for liquidity in 2007 (holding more liquid assets), a period that included international financial insecurity, and generating uncertainty as to the proper functioning of the economy.

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

Table 2

Preference for Public Liquidity (PLP), R$ 1000 (thousand)

State	2004	2005	2006	2007	2008	2009	2010	2011	2012
Rondônia	0.3839	0.4073	0.3676	0.3716	0.2776	0.2700	0.2843	0.2183	0.2130
Acre	0.4065	0.3631	0.4230	0.4164	0.2385	0.2688	0.2902	0.2354	0.3888
Amazonia	0.3181	0.3272	0.3518	0.3821	0.2533	0.2814	0.2622	0.2165	0.2188
Roraima	0.4541	0.3873	0.3702	0.3639	0.2614	0.2341	0.2705	0.1635	0.3164
Pará	0.3187	0.3052	0.3507	0.3744	0.3037	0.2905	0.3114	0.2587	0.2524
Amapá	0.3339	0.3797	0.3780	0.3704	0.2673	0.2668	0.2848	0.2419	0.4379
Tocantins	0.3347	0.3769	0.3646	0.3901	0.3395	0.3616	0.3392	0.3011	0.2806
Maranhão	0.3159	0.2818	0.2980	0.3341	0.2661	0.2624	0.2748	0.2413	0.2297
Piauí	0.2515	0.2453	0.2466	0.2541	0.2157	0.2377	0.2118	0.1806	0.1816
Ceará	0.1276	0.1277	0.1676	0.1787	0.1498	0.1401	0.1321	0.1113	0.1211
Rio G. do Norte	0.2389	0.2491	0.2361	0.2625	0.2108	0.2067	0.2086	0.1697	0.1753
Paraíba	0.2635	0.2626	0.2534	0.2467	0.2265	0.2350	0.2074	0.1720	0.1703
Pernambuco	0.2066	0.1976	0.2156	0.2263	0.1718	0.1798	0.1744	0.1491	0.1478
Alagoas	0.2275	0.2429	0.2404	0.2467	0.2065	0.2392	0.2122	0.1684	0.1624
Sergipe	0.1923	0.2080	0.1986	0.1949	0.1667	0.1694	0.1742	0.1537	0.1571
Bahia	0.2445	0.2411	0.2455	0.2694	0.1967	0.1770	0.1908	0.1567	0.1638
Minas Gerais	0.1814	0.1906	0.1970	0.2286	0.1630	0.1323	0.1335	0.1137	0.1119
Espírito Santo	0.2010	0.1913	0.1955	0.2007	0.1843	0.1809	0.1890	0.1524	0.1493
Rio de Janeiro	0.1856	0.1977	0.2201	0.2643	0.1617	0.1479	0.1461	0.1191	0.1292
São Paulo	0.1399	0.1395	0.1618	0.1970	0.1087	0.1261	0.1201	0.1055	0.1222
Paraná	0.1685	0.1571	0.1613	0.1787	0.1285	0.1264	0.1264	0.1083	0.1239
Saint Catarina	0.2552	0.2411	0.2473	0.2626	0.1794	0.1762	0.1822	0.1361	0.1446
Rio G. do Sul	0.1633	0.1506	0.1611	0.1857	0.1349	0.1347	0.1515	0.1102	0.1095
Mato G. Sul	0.3084	0.2888	0.2999	0.3054	0.2275	0.2187	0.2220	0.1939	0.1950
Mato Grosso	0.4124	0.3961	0.3718	0.3877	0.3208	0.3056	0.3238	0.2368	0.2409
Goiás	0.3191	0.3072	0.3083	0.3091	0.2502	0.2467	0.2439	0.1976	0.2083
Distrito. Federal	0.1431	0.1777	0.1357	0.1370	0.1546	0.1295	0.1134	0.1014	0.1233

Source: author's elaboration based on Central Bank data. Exchange rate € 1.0 = R$ 3.12.

In Table 2, it can be seen that there is more financial concentration exactly in regions with a lower preference for liquidity, depicting which regionsinspire greater confidence for the public. This encourages opportunities for investments, providing the banks in these regions the chance to expand their portfolio of loans, resulting in higher profits. It can be concluded that for the most developed states, the current value of the indicator suggested is lower. This is also explained by the higher level of education existing in these regions, mainly institutions directed to the financial market as well as including the headquarters of some important banks, and big companies. Thisallows a greater number of people over the age of twenty be more able to access financial services, compared to the number of bank branches. These results can be seen in Table 3.

Table 3

Population above 20 years / Number of bank branches

State	2004	2005	2006	2007	2008	2009	2010	2011	2012
Rondônia	9.6	9.9	10.1	10.3	10.1	9.1	9.0	7.6	7.2
Acre	10.6	9.7	10.1	10.8	10.5	9.2	9.5	7.9	7.1
Amazon	12.1	12.1	12.0	11.9	12.6	11.4	11.2	10.6	10.3
Roraima	11.1	10.8	11.0	11.8	12.5	10.0	10.2	9.0	7.4
Pará	13.2	13.0	13.1	13.5	14.1	12.9	13.0	10.6	10.6
Amapá	12.5	12.5	11.4	11.7	10.6	8.8	9.6	8.5	8.3
Tocantins	8.9	8.8	8.8	8.6	8.6	8.1	8.5	7.5	7.4
Maranhão	14.0	14.3	14.6	14.9	15.2	14.8	14.7	12.2	12.0
Piauí	15.1	15.3	15.0	15.5	15.8	15.6	15.9	13.2	12.2
Ceará	12.8	12.6	13.0	13.3	14.1	13.2	13.5	12.1	11.4
Rio Grande do Norte	13.2	12.3	12.4	12.5	13.2	12.1	12.6	11.1	10.6
Paraíba	12.6	12.6	12.5	12.7	13.2	12.4	12.6	10.8	10.5
Pernambuco	10.7	10.8	10.7	11.0	11.9	10.6	11.1	10.2	9.9
Alagoas	13.4	13.5	13.6	13.8	14.4	13.4	13.3	11.0	10.1
Sergipe	7.1	7.2	7.3	7.3	7.4	7.3	7.6	7.0	6.7
Bahia	10.8	11.1	11.0	11.2	12.2	11.2	10.8	9.7	9.0
M. Gerais	6.6	6.8	6.8	6.9	9.0	6.9	6.9	6.4	6.3
Espírito Santo	6.6	6.3	6.1	6.1	6.3	5.8	6.0	5.7	5.7
Rio de Janeiro	6.3	6.3	6.2	6.2	7.7	5.5	6.1	5.8	5.9
São Paulo	4.7	4.7	4.7	4.6	5.2	4.2	4.4	4.3	4.4
Paraná	5.2	5.2	5.2	5.4	7.1	5.3	5.3	4.8	5.1
Santa Catarina	4.4	4.5	4.6	4.6	4.9	4.7	4.9	4.7	4.9
Rio G. do Sul	5.0	5.0	5.1	5.1	5.3	4.8	4.9	4.7	4.6
Mato G. do Sul	6.1	6.2	6.3	6.4	6.6	6.1	6.3	5.9	6.5
Mato Grosso	7.0	7.0	7.1	7.1	7.5	6.7	6.8	6.6	6.8
Goiás	6.3	6.4	6.5	6.6	9.2	6.4	6.6	6.1	6.1
Distrito Federal	4.7	4.8	4.9	4.9	5.5	4.9	5.0	4.9	4.9

Source: author's elaboration based on Datasus and Central Bank data.

This data indicates that personal turnover is higher in prosperous regions. The spreading of social security policies, such as "bolsa família" does not interfere significantly in mitigating the results.

Conclusion

From the analysis of regional data, the regions where there is a lower public liquidity preferenceare exactly those with the highest financial concentration, as well as numbers of large companies. These usually require greater educational levels and qualifications. On the other hand, the presence of banksrests on public confidence; in turn, the banks see these great regions as business opportunities and for investment in financial assets due to the higher average purchasing power and greater financial education. For this to occur, there

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

is a need for the presence of a number of larg banks to extend credit, which in turn take advantage of the opportunity to increase their loan portfolio and thus generate higher profits.

From this, the hypothesis proposed by the study is ratified. It was dIt is demonstrated that there was an increase in the demand for money in 2007 due to the crisis, as can be seen in the relevant table. This research also state data on the liquidity preference, in short, allow observations and important signalizations about the degree of economic development of a state, region or country.

REFERENCES

Amado, A. M. (1998): Moeda, financiamento, sistema financeiro e trajetórias de desenvolvimento desigual: a perspectiva pós-keynesiana *Revista de Economia Política* 18 (1): 76–89.

Banco Central do Brasil (Bacen) (2011): *Disponível* http://www.bcb.gov.br (accesed: 05 jun. 2011)

Chick, V. (2006): Bancos e regiões: algumas lições da experiência européia In: Crocco, M. A.–Jayme JR, F. G. (Eds.): *Moeda e Território: Uma Interpretação da Dinâmica Regional Brasileira* pp. 79–92. Autêntica ,Belo Horizonte.

Crocco, M. A.–Cavalcante, A.–Barra, C.–Val, V. (2006): Polarização regional e sistema financeiro In: Crocco, M.–Jayme Jr, F. G. (Eds.): *Moeda e Território: Uma Interpretação da Dinâmica Regional Brasileira* pp. 231–269., Autêntica, Belo Horizonte.

Crocco, M.–Nogueira, M.–Andrade, C. M. C. de–Amaral, M. L. S. (2011): O Estudo do Gap Regional de Crédito e seus Determinantes, sob uma ótica Pós-Keynesiana *Revista Economia* 12 (2): 281–307.

Dow, S.–Fuentes, Rodrigues, C. J. (2006): Um "survey" da literatura de finanças regionais In: Crocco, M. A.–Jayme JR, F. G. (Eds.): *Moeda e Território: Uma Interpretação da Dinâmica Regional Brasileira* pp. 39–75., Autêntica, Belo Horizonte.

Ferrari Filho, F.–Conceição, O. A. C. (2001): A Noção de Incerteza nos Pós-Keynesiano e Institucionalistas: Uma Conciliação Possível? Nova Economia 11 (1): 99–122.

Figueiredo, A. T. L.–Crocco, M. (2008): The role of money in the locational theory: a post-keynesian approach *Revista Brasileira de Estudos Regionais e Urbanos* 2 (1): 33–54.

Keynes, J, M, (1991): *The General Thoery of Employment, Interest, and Money* Orlando, Harcourt Brace.

Knight, F.H. (1921): *Risk, Uncertainty and Profit* Houghton Mifflin, New York.

Levino, N. de A.–Lages, A. M. G. (2008): *Moeda e Comportamento Territorial dos Agentes Econômicos: Introdução a uma análise regional* Enaber, Aracaju.

Nogueira, M.–Crocco, M.–Figueiredo, A. T. L. (2010): Estratégias Bancárias Diferenciadas no Território: O Caso de Minas Gerais *Análise Econômic, Porto Alegre* 28 (54): 281–311.

Ribeiro, A. C.–Marouvo, C. C. (2011): *O Papel da Moeda no Processo de Desenvolvimento Econômico Regional* In: XVI Encontro Regional de Economia, Fortaleza.

REGIONAL STATISTICS, 2015, VOL 5, No1: 86–96; DOI: 10.15196/RS05105

ANNEX

Map of Brazil

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

ÉVA KOMLÓSI[a) – BALÁZS PÁGER[b)]

The impact of urban concentration on countries' competitiveness and entrepreneurial performance

Abstract

This paper aims to elaborate the role of Jacobs-type of agglomeration effects on countries' competitiveness and entrepreneurial performance. Our research allows for a better understanding of the relationship that exists between a country's urban system, characterized by spatial agglomeration (concentration) or deglomeration (deconcentration) processes, and its competitiveness and entrepreneurial performance.

Urbanization economies refer to considerable cost savings generated through the locating together of people, firms and organizations across different industries. It has recently become an axiom that the better performance of global cities (as they are important nodes of innovation and creativity) is derived from agglomeration effects. This general assumption follows that the more concentrated an urban system of a country, the more competitive and better its entrepreneurial performance. Even though this notion has gained quick and ardent acceptance from practitioners, the related literature shows contradictory results;this has induced a heated debate in academic circles, because it has raised serious doubts about the "bigger is better" theory. We hope to contribute to this debate with our detailed analysis.

To understand the impact of urban concentration, we selected 70 countries and calculated the so-called ROXY Index measuring the degree of agglomeration or deglomeration in their urban systems. To exemplify country-level competitiveness, we applied the Global Competitiveness Index (GCI) while the Global Entrepreneurship and Development Index (GEDI) was used to demonstrate country level entrepreneurial performance. Using these indexes correlation and cluster analysis were designed to obtain understanding of the relationship between them.

Our analysis indicates that as urban concentration initially increases competitiveness, entrepreneurial performance also increases, but at a decreasing rate. Both of them eventually reaches a maximum, and then after a certain point decrease with further concentration. Therefore, the curve for this relationship is non-linear and folds back. This indicates that over- or under-concentration of the population within an urban system does not necessarily result in a better outcome. However, we should consider that a high concentration of population is only one important factor for competitiveness and entrepreneurial performance while other effects may exist.

a) Research fellow, MTA-PTE Innovation and Economic Growth Research Group, Faculty of Business and Economics, University of Pécs, H-7622 Pécs, Rákóczi út 80., Hungary. E-mail: komlosieva@ktk.pte.hu

b) Graduate research assistant, Centre for Economic and Regional Studies, Hungarian Academy of Sciences H-7621 Pécs, Papnövelde utca 22., Hungary. E-mail: pagerb@rkk.hu

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Keywords: urbanization economies, entrepreneurship, competitiveness, spatial cycles, ROXY Index.

Introduction

It was the classical economist, Alfred Marshall (1920), who first identified the main characteristics and sources of location-specific economies of scale labelled in economic literature as *agglomeration* or *external economies*. Since that time, other researchers have actively been contributing to the further development of his theory. Today it is widely accepted that agglomeration economies refer to an increasing return of scale derived from the clustering of different or, on the contrary, similar and specific economic activities. Co-locating near other firms, organizations and people can result in an increased benefit from greater diversity and/or specialization. Some often mentioned reasons for these advantages are the following: (1) market size effect (demand side), (2) local skilled labour pool effect (supply side), (3) local non-traded inputs (e.g. special business services), as well as (4) tacit knowledge and spillover effects. Clustering of firms together at a location implies frequent personal contacts and social ties among local actors that facilitate the circulation of information among them. Hence, intense knowledge spillovers are fostered by the proximity of firms, thereby promoting their innovation activity, and ultimately resulting in higher profit. Local non-traded inputs refer to peculiar products or services whose provision can be very expensive. However, if there are many firms located together, the cost of such special products or services can be dispersed among them. Co-locating of many firms also goes hand in hand with the accumulation of a skilled workforce. In many sectors, the cost of searching and (re)training people can be extremely high and time-consuming, particularly in a rapidly changing environment. Therefore, easy access to a pool of skilled labour also benefits firms (Duranton–Puga 2004, Rosenthal–Strange 2004, Puga 2010, McCann 2013). Even though agglomeration economies can be classified in several ways, typically three major categories are distinguished: (1) benefits of *localization economies (Marshall – Arrow – Romer externalities)* are derived from the agglomeration of specialized firms across the same industrial sector, (2) *urbanization economies* (*Jacobs' externalities*) refer to cost savings generated through the locating together of people and firms across different industries, and (3) *internal economies of scale* results in a significant return because of the size of the firms (Parr 2002, McCann 2013). In this paper, we solely study the impact of urbanization economies on countries' competitiveness and entrepreneurial performance.

Nowadays, there is a widely held view that global cities and large urban areas perform better and grow faster than the others. First of all, the appearance of globalization, new information and communication technologies have given rise to the recognition that metropolis are the '*space of flows*' as they are important nodes of innovation and creativity (Castells 1996). According to a large extent of the literature, the better performance of big cities can be explained by the advantages of agglomeration economies. Several theoretical and empirical studies have affirmed that agglomerations associated with a high density of economic activity does matter for productivity and economic growth (e.g., Chinitz 1961, Glaeser et al. 1992, Ellison and Glaeser 1997, Ciccone and Hall 1996, Ciccone 2002, Henderson 2003a, Rosenthal–Strange 2003). Studies dealing with urbanization economies

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

ascertain that a doubling of city size results in a productivity gain of between 2 and 8 percent (see overview from Vreeker et al. 2009).

This general assumption follows that the more concentrated an urban system of a country, the more competitive and better its entrepreneurial performance. For instance, in his well-known theoretical general equilibrium model, Krugman explained the initial spatial concentration processes of economic activities by identifying the main reasons, and introduced a new stream in spatial economic science, the New Economic Geography (NEG). NEG interprets agglomeration economies as the outcome of three reasons: (1) increasing returns, (2) trade costs and (3) the demand for manufacturing products. Krugman's theory implies that production is prone to concentrate in a few regions, which will become populous and competitive (Krugman 1991, 2009). Despite the remarkable novelty of Krugman's theory, it does not give any satisfying answer about divergent regional growth (Acs–Varga 2002). Also, findings from empirical studies indicate the positive effect of urban concentration on entrepreneurial performance mainly measured by new firm formation rates. Highly populated urban areas offer more opportunity for entrepreneurial success, because they can provide firms with a large consumer base, relatively cheap physical infrastructure, tacit knowledge, special services or a skilled workforce., therefore new firms prefer highly urbanized areas (Reynolds et al. 1994, Acs–Armington 2004, Van Stel–Suddle 2008, Knoben et al. 2011).

The 'bigger is better' notion has gained quick and ardent acceptance from practitioners, particularly from policymakers. Consequently, many EU territorial strategic concepts support the same view, viz. metropolitan areas are the most important drivers of European competitiveness, even if only a few studies *unambiguously* prove a positive contribution from agglomeration effects on economic growth. In fact, the majority of studies have highlighted contradictory results. *De facto*; there are some papers that support the view that urbanization economies tend to increase with the size of the city and they have a positive impact on economic growth; on the other hand, other studies have found no clear evidence that urbanization economies would generate growth (David et al. 2013). Then, there are some other studies that may explain these contradictory results. Findings of these papers refer to the negative effects of urban concentration, such as higher costs of skilled labour or higher rent for land, environmental contamination and severe congestion. Recent studies suggest that spatial competition (for qualified labour and other inputs), as a centrifugal force, can restrain the above-mentioned positive effect of urban concentration, possibly leading to a decrease in start-up rates and productivity growth (Rizov et al. 2012). Henderson (2003b) has estimated the impact of urbanization and urban concentration on productivity growth at the country level for the period between 1960 and 1990. According to his results, productivity growth is not strongly affected by urbanization, because *"urbanization is not a growth stimulus per se, is it a by-product"*, but there is a "*best degree of urban concentration in terms of maximizing productivity growth that varies with the level of development and country size*" (Henderson 2003b, p. 50). Consequently, both over- and under-concentration have negative effects on growth: "*City size affects positively the degree of local information spillovers, which interactively affects local knowledge accumulation, promoting productivity growth. However, cities of extensive size draw resources away from investment and innovation in productive activity to try to maintain the quality of life in a congested local environment.*" (Henderson 2003b, p. 67). According

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

to Duranton and Puga (2001, p. 1454) cities can be conceived as areas "*facilitating search and experimentation in innovation*". They also found evidence that under-sized cities have too little experimentation affecting productivity nationally, while over-sized cities waste excessive amounts of resources on other activities, which, drawing resources away from experimental activities, accordingly also inhibiting growth (Henderson 2003b). *David et al.* (2013) have systematically tested the relationship between city size (urban concentration) and the economic performance of cities in the European context. Their analysis also confirmed that the comparative advantage of cities also depends on the country in which they are located: *"In highly developed and densely urbanized areas, congestion effects might counteract the advantages of agglomeration. ... Hence, it may be that, in the dense Western part of Europe more than in the rest of Europe, the performances of the cities are more linked to their economic structures, their heritage, and the quality of their governance than to their size and centrality"* (David et al. 2013, p. 249). Castells-Quintana and Royuela (2014), in their study, explain that agglomeration (as urban concentration) fosters growth particularly in low-income developing countries, while urbanization has a positive effect on high-income developed countries. Since the large cities in Europe are highly urbanized areas, the positive link between urbanization and economic growth has already vanished. At the same time, dis-economies as congestion, pollution or high housing prices may have a negative effect. Consequently, economic growth in developed countries has been observed in small- and medium-sized cities, because of their intense urbanization: *"among the rich countries, twelve out of fifteen most entrepreneurial cities are small to medium-sized cities…"* (McCann–Acs 2012, p. 23).

In our study, we focus on the understanding of the relationship between a country's urban system characterized by spatial agglomeration (concentration) or deglomeration (deconcentration) processes and its competitiveness, as well as its entrepreneurial performance. The innovative component of our paper is that we demonstrate the impact of urbanization economies classified by the four stages of the spatial-cycle path on economic performance; and using a large sample of countries, a long time span and a method that has never been used for such purpose. Our results are consistent with other studies' findings related to the emerging literature on the limits of agglomeration.

We selected 70 countries and calculated the "ROXY Index", measuring the degree of agglomeration or deglomeration in their urban system. To exemplify country-level competitiveness, we applied the Global Competitiveness Index (GCI) while the Global Entrepreneurship and Development Index (GEDI) was used to demonstrate country level entrepreneurial performance.

The remainder of this paper is organized in the following fashion. First, the descriptions of the employed indexes (GEDI, GCI and ROXY) are reviewed (Section 2.1). We then describe the data and analysis methodology in Section 2.2. Our results are reported in Section 3. Finally, a concluding summary and discussion is provided in Section 4.

Data and methodology

In this section, we summarize the applied indexes and methodology. We employed three indexes for the analyses: (1) the *Global Competitiveness Index* (GCI) as a comprehensive tool to characterize country-level competitiveness, (2) the *Global Entrepreneurship and*

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Development Index (GEDI) as a composite indicator of entrepreneurship performance, and (3) as a third index the *ROXY Index,* which indicates the direction and size of population changes (concentration or deconcentration processes) within an urban system. We conducted correlation and cluster analysis to understand the relationship between urban concentration/deconcentration trends and economic performance.

Measuring country-level competitiveness and entrepreneurship

Since 2004, the yearly published Global Competitiveness Report – developed by the World Economic Forum (WEF) – ranks countries according to their competitiveness based on a composite indicator, the "*Global Competitiveness Index"* (GCI). According to the WEF, competitiveness can be defined *"as the set of institutions, policies, and factors that determine the level of productivity of a country"* (Schwab 2013, p. 4). Therefore, GCI builds up from many different indicators that characterize the institutions, productivity or policies of countries. Altogether 12 pillars are created from the identified set of indicators, which can be divided into three sub-indexes[1]: "basic requirements"(4 pillars), "efficiency enhancers" (6 pillars) and "innovation and sophistication factors" (2 pillars). The three sub-indexes are calculated by using weights that express the development level of a country's economy. Three development categories are used by WEF: factor-driven, efficiency-driven and innovation-driven economies. The involved countries are grouped into five groups, which are determined by the three development levels and two transition stages. Finally, the GCI Index is composed of the weighted average of the three sub-indexes. In our research, we used data derived from several GCI reports over the period 2006–2014.

The Global Entrepreneurship Development Institute lead by Zoltan J. Acs and László Szerb developed the *Global Entrepreneurship Index* (GEDI). The GEDI Index is a composite index that measures productive entrepreneurship in a multidimensional way. It examines the connection between entrepreneurship and economic development and provides policy recommendations regarding economic policies (Szerb et al. 2013). The basic idea of the GEDI Index is based on the theory of the National System of Entrepreneurship that "(…) *is the dynamic, institutionally embedded interaction between entrepreneurial attitudes, ability, and aspirations, by individuals, which drives the allocation of resources through the creation and operation of new ventures*" (Acs et al. 2014, p. 479). The index builds on individual data derived from the Global Entrepreneurship Monitor (GEM) Adult Population Survey. It focuses not only on the process of business creation but also captures the qualitative aspects, the 'institutional context' of the country. The index consists of three sub-indexes (attitudes, abilities, aspirations), and each sub-index has four or five pillars. The GEDI pillars are determined by a complex method and indicate the combined effect of individual and institutional data.[2]

1 The whole descriptions of GCI sub-indexes and pillars are available in the 2013–14 edition of the Global Competitiveness Report: http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2013-14.pdf

2 The whole description of GEDI sub-indexes and pillars are available in the latest GEDI book: Zoltan J. Acs et al. (2013): *Global Entrepreneurship and Development Index*. Edward Elgar Publishing.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

To express urbanization economies, we calculated the ROXY Index[3], which is *"an indicative instrument to quantitatively identify the major stages of the spatial cycles. This index can be used in conducting both of the intra- and inter-city analysis to study the spatial agglomeration and deglomeration processes"* (Fukatsu – Kawashima et al. 1999, p. 395). The ROXY Index captures the effect of migration through the periodic change of the population. It measures the change in population by a weighted average growth ratio and by a simple average growth ratio (see the formula below).

$$ROXY\ Index = \left(\frac{WAGR_{t,t+1}}{SAGR_{t,t+1}} - 1{,}0\right) * 10^4 = \left\{\frac{\sum_{i+1}^{n}(x_i^t * r_i^{t,t+1})}{\sum_{i=1}^{n} x_i^t} * \frac{n}{\sum_{i=1}^{n} r_i^{t,t+1}} - 1{,}0\right\} * 10^4 \quad (1)$$

where: x^t_i = population of city "i" in year "t", $r^{t,\ t+1}_i$ = annual growth ratio of population in city "i" for the period between years "t" and "t+1", which is defined as the "k"th root of x_i^{t+k}/x_i^t, n = number of cities, $WAGR_{t,\ t+1}$ = weighted average of annual growth ratios of population "n" cities for the period between years "t" and "t+1", which is equal, in the case where population level of each city is used as a weighting factor, to $\sum(x_i^t * r_i^{t,t+1})/\sum x_i^t$, $SAGR_{t,\ t+1}$ = simple average of annual growth ratios of population in "n" cities for the period between years "t" and "t+1", which is equal to

$$\sum r_i^{t,t+1}/n$$

Marginal value of the ROXY Index (ΔROXY)

$$\Delta ROXY = \frac{ROXY\ Index_{(t+1,t)} - ROXY\ Index_{(t,t-1)}}{Cf} \quad (2)$$

where: $ROXY\ Index_{(t+1,t)}$ = the value of ROXY Index for period "t", $ROXY\ Index_{(t,\ t-1)}$ = the value of ROXY Index for period "t-1", Cf = the difference between the mid-point time for "t" period and the mid-point time for "t-1" period (*Source*: Kawashima et al. 1997, p. 221 and Fukastu – Kawashima 1999, p.407.).

The index is based on the spatial-cycle hypothesis originally conceptualized by *Klaassen* (1979, 1981). Klaassen differentiated four stages of the spatial-cycle path: Stage 1 – Accelerating concentration, Stage 2 – Decelerating concentration, Stage 3 – Accelerating deconcentration and Stage 4 – Decelerating deconcentration (Kawashima et al. 1997). The first version of the ROXY Index was published in an empirical study written by Kawashima (1978). Since then, the index has been developed further and used in numerous empirical studies to identify the spatial agglomeration – deglomeration phenomena associated with the changes in population or other social and economic variables (see Kawashima 1982, 1985, 1986, Hirvonen et al. 1997, Fukatsu–Kawashima 1999).

According to the size and direction of the ROXY Index, four hypothetical stages of the spatial-cycle process can be distinguished. The positive value of the ROXY Index indicates concentration while the negative value shows deconcentration. The direction of change depends on the value of ΔROXY: if it is positive, there is an accelerating concentration or decelerating deconcentration; if it is negative, it indicates an accelerating deconcentration or decelerating concentration (Table 1).

3 ROXY means "**R**atio **o**f *Weighted Average Growth Ratio* (abbreviated as **X**) to *Simple Average Growth Ratio* (abbreviated as **Y**)" (*Kawashima* 1985)

Table 1

The characteristics of different stages of urbanization

Stages of urbanization	The size and direction of change
AC (accelerating concentration)	positive ROXY value, ΔROXY positive
DC (decelerating concentration)	positive or negative ROXY value, ΔROXY negative
DD (decelerating deconcentration)	negative or positive ROXY value, ΔROXY positive
AD (accelerating deconcentration)	negative ROXY value, ΔROXY negative

Source: own compilation based on Kawashima et al. 1997.

There are two crucial points in the computation of the ROXY Index: (1) the length of the examined period and (2) the number of cities that are involved in the examination. Therefore, we tried to find a rule or concept that could help us identify the most important cities of each examined country. However, data like GDP, which could represent the most important cities in a country, are available for only a few countries at the city level. After reviewing some of the relevant literature (see Gabaix 1999; Eeckhout 2004; Tabuchi et al. 2005; Czaller 2012), we realized that although the problem of determining the adequate number of cities is known, no clear solution exists. Therefore, we decided to analyse three cases and conducted the analysis for the first 20, 30 and 40 most populated cities of the 70 countries. It was important to be aware that the sample contained very different countries with regard to their size. Thus, examining an urban system with less than 20 cities was considered too small; on the other hand, in the case of some countries, it was not possible to examine their urban systems with more than 40 cities, because data were not available. The three mentioned cases may also serve as a robustness check of the results.

The other important factor to calculate the ROXY Index is the time period in which the index indicates the agglomeration or deglomeration trends. Therefore, we used the three latest available data of city populations. Thus, we created two periods ($ROXY_t$ and $ROXY_{t-1}$) and calculated the ΔROXY that shows the direction and scale of the change. Our original idea was that the time periods used by the GEDI/GCI Index would be considered by the calculation of the $ROXY_t$ and $ROXY_{t-1}$ indexes. However, because of data availability, the first or last years, and also the lengths of the periods were not the same for the different countries (see Appendix Table A1).

Originally, we planned to carry out analyses for all the countries involved in the GEDI research during the examined period. It was altogether 76 countries, but we excluded some of them due to the lack of city population data. Thus, we could involve 70 countries. Our country set contains both developed and developing nations. Because of the lack of former city population data, it was not possible to calculate $ROXY_{t-1}$ for some countries (Jordan, Malaysia, Portugal and South Africa), hence we had to exclude them from the later examinations. In the case of the United Arab Emirates, city population data are available only for its nine biggest cities; therefore, only the ROXY20 was calculated for the available cities of the United Arab Emirates (UAE).

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Method

To examine the intensity and direction of the relationships among the indexes, we conducted correlation analysis. We carried out the correlation analyses separately for the two indexes – GCI and GEDI – using the three versions of the ROXY Index (ROXY20, ROXY30 and ROXY40). The analyses were not limited to the main indexes alone. We also analysed the relationships between the ROXY Index and the different sub-indexes of GCI and GEDI. Furthermore, those sub-indexes of GCI and GEDI that showed the highest correlation with the ROXY Index were also examined.

As a first step, we checked the characteristics of our descriptive statistics. GCI and GEDI did not require any data transformations, but a relatively high skewness was discovered considering the ROXY Index. We managed this problem with a transformation process. Many data transformation processes were checked that might solve the problem of skewness. The results of the correlation analysis with different transformation processes did not show significant differences. Hence, we decided to apply the Box-Cox transformation method, in the same way as Annoni–Kozovska (2010). Finally, the transformed ROXY Index data were rescaled to a scale from 0 to 10.

We endeavour to use not only the annual values of GEDI and GCI but also to represent the changes in their values during a given period as well. Therefore, we calculated the average value of both indexes for the whole period (GCI_AVE and GEDI_AVE). To catch the changes within the examined period, the changes from year to year were calculated and averaged for each country as well (GCI_CH and GEDI_CH). Finally, we multiplied the 'average values' with the 'change values' in the case of both indexes (GCI_AVG_CH and GEDI_AVG_CH). Then we rescaled both modified indexes to a scale from 0 to 10.

As a next step, K-means cluster analysis was conducted. First, the observed outliers were excluded from the analysis. The examination started with 66 countries in the case of ROXY20 (no available / t-1 / data for Jordan, Malaysia, Portugal and South Africa) and 65 countries in the case of ROXY30 and ROXY40 (no available / t-1 / data for Jordan, Malaysia, Portugal and South Africa and UAE). We used the original and transformed ROXY indexes in the cluster analysis as well. We tested different numbers of clusters (2, 3, 4, 5 and 6 groups). However, the results of the ANOVA test (the optimal F- and significance-values) indicated the need to create 3 clusters in the case of GCI with the original ROXY Index and 4 clusters with the transformed ROXY values, while countries were classified into 4 groups in the case of the GEDI Index using the transformed ROXY values. The tests proved that the groups are significantly different from each other at every significance level.

Results

Results of the examination: ROXY and GCI indexes

In this sub-section, we examine the intensity and direction of the relation between the GCI Index and the three version of ROXY Index. Table 2 contains both the original ROXY indexes for different pools of cities (ROXY20, ROXY30 and ROXY40) and also the three Box-Cox transformed and rescaled ROXY indexes (ROXY_BOXCOX_10).

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

According to the correlation analysis, there is a positive relationship between the ROXY and the GCI Index, but the intensity of this relationship is quite moderate, and it is significant only with the Box-Cox transformed ROXY indexes. The ROXY40_BOXCOC_10 variable and the GCI Index show the strongest correlation coefficient (r = 0.321). If we analyse the relationship between the ROXY Index and the three sub-indexes of GCI (BASIC – GCI Basic sub-index, EFF – Efficiency sub-index, INN – Innovation sub-index), the strongest correlation can be observed between the ROXY Index and the GCI Efficiency sub-index, but only a loose connection can be confirmed among them (r = 0.350). We can assume that concentration or deconcentration of the population within an urban system has a moderate effect on efficiency (Table 2). To investigate this presumption, we detach the relationship between the transformed ROXY Index and the different pillars of the GCI Efficiency sub-index (Table 3).

Besides the intensity of the connection, the direction is also very important. A positive correlation coefficient between the ROXY Index, the GCI Index and its sub-indexes means that *the more concentrated the population within a country's urban system, the higher the value of the GCI Index. Higher GCI value refers to the higher competitiveness of the country.*

Table 2

The correlation coefficients between the GCI Index, its sub-indexes and ROXY Index

ROXY INDEX *(original and transformed)*	GCI_(AVG_CH)	BASIC_(AVG_CH)	EFF_(AVG_CH)	INN_(AVG_CH)
ROXY20	0.187	0.187	0.220	0.139
ROXY30	0.216	0.211	0.242 a)	0.154
ROXY40	0.218	0.210	0.241 a)	0.153
ROXY20_BOXCOX_10	0.221	0.223	0.267 a)	0.183
ROXY30_BOXCOX_10	0.295 a)	0.279 a)	0.347 b)	0.261 a)
ROXY40_BOXCOX_10	0.321 b)	0.305 a)	**0.350** b)	0.292 a)

Note: BASIC = GCI "Basic" sub-index, EFF = GCI "Efficiency" sub-index, INN = GCI "Innovation" sub-index.

a) Correlation is significant at the 0.05 level (2-tailed). b) Correlation is significant at the 0.01 level (2-tailed).

Source: own calculations.

Table 3

The correlation coefficients between the pillars of GCI Efficiency sub-index and ROXY Index

Pillars of GCI Efficiency sub-index	ROXY20_BOXCOX_10	ROXY30_BOXCOX_10	ROXY40_BOXCOX_10
HT_(AVG_CH)	0.284 a)	**0.362** b)	**0.363** b)
MEFF_(AVG_CH)	0.148	0.255 a)	0.283 a)
LEFF_(AVG_CH)	0.298	**0.404** b)	**0.382** b)
FIN_(AVG_CH)	0.112	0.192	0.206
TECH_(AVG_CH)	0.295 a)	**0.385** b)	**0.396** b)
MSIZE_(AVG_CH)	0.143	0.108	0.095

Note: HT = Human capital pillar, MEFF = Market efficiency pillar, LEFF = Labour efficiency pillar, FIN = Financing pillar, TECH = Technological readiness pillar, MSIZE = Market size pillar, AVG = average, CH = change.

a) Correlation is significant at the 0.05 level (2-tailed). b) Correlation is significant at the 0.01 level (2-tailed).

Source: own calculations.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Efficiency sub-index group's pillars related to human capital, market efficiency, labour productivity, financing, technology readiness and market size. Three of the pillars – Human capital, Labour efficiency and Technology readiness – show positive medium-strength correlation coefficients with the transformed and rescaled ROXY Index (Table 3). The strongest connection is shown between the GCI Labour efficiency pillar and the transformed ROXY30 (r = 0.404). The positive coefficient means that *the higher the concentration in a country's urban system in a given period, the higher the labour efficiency, the technological readiness and the quantity of skilled human capital of the country.*

As a next step, we conducted cluster analysis for each county's urban system consisting of the first most populous 20, 30 and 40 cities, respectively. Determining the direction of change, we calculated the ROXY values for the previous period (t-1) to receive the ΔROXY. We carried out the cluster analysis with different cluster numbers (2, 3, 4, 5 and 6 groups). Using original ROXY values, the ANOVA analysis showed the best F-test values with 3 clusters (the interpretation of the clusters below or over 3 clusters was problematic). However, using the Box-Cox transformed ROXY values, the interpretation of 4 clusters seemed to be more appropriate (ANOVA analysis showed appropriate F-test values for both 3 and 4 clusters). Consequently, we present here the results of the cluster analysis for 3 and 4 clusters. According to the ANOVA test, the significance was lower than 0.05, proving that the clusters in terms of the variables differ from each other. Here, we only show the results of the cluster analysis conducted between the ROXY30 and GCI index (cluster analyses with ROXY20 and ROXY40 are available in the Appendix and serve as robustness checks).

As mentioned above, using the original ROXY Index, we could distinguish three clusters (Table 4). *Cluster 1* contains those countries whose urban system is characterized by a strong concentration trend represented with a high ROXY value (final cluster centre = 5.21). The results show that the competitiveness of those countries in which the urban system is highly concentrated is high (GCI final cluster centre = 5.33). It contains 11 countries characterized by the acceleration of concentration (AC), and 29 countries characterized by the deceleration of deconcentration (DD). *Cluster 2* consists of 12 countries, among them 9 characterized by the deceleration of deconcentration, but they are still in the deconcentration stage. This cluster generally contains countries in which the urban system is heading to a concentration from the deconcentration stage. This cluster is a transitional category. *Cluster 3* contains those countries in which the deconcentration of the urban system is accelerating (AD, ROXY final cluster centre = -168.86). According to the results of the analysis, *if deconcentration is strengthening, competitiveness will drop* (GCI final cluster centre = 2.63).

On the other hand, using the Box-Cox transformed ROXY Index, we could identify 4 clusters of countries (*Table 4*). An important difference compared to the 3 cluster case is that here, if a country's urban system is characterized *either by a strong concentration trend* (Cluster 1) *or by a strong deconcentration trend* (Cluster 4), *competitiveness will fall* (Cluster 1 – GCI final cluster centre = 2.77, Cluster 4 – GCI final cluster centre = 2.82). While if the county's urban system is in a transition stage – meaning it is not so concentrated, or not so deconcentrated – competitiveness will be more outstanding (Cluster 2 – GCI final cluster centre = 8.33, Cluster 3 – GCI final cluster centre = 4.57).

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Table 4

Results of the cluster analysis: ROXY30 – GCI index

Final Cluster Centres	Cluster 1 *Concentration with generally high GCI values*	Cluster 2 *Deceleration of deconcentration with GCI values higher than the average*	Cluster 3 *Deconcentration with generally low GCI values*
Number of cases	45	12	5
ROXY 30	5.21	–55.51	–168.86
GCI	5.33	3.49	2.63
AC	*11*	-	-
DC	2 *(conc)*	-	1 *(deconc)*
DD	*29 (16 deconc, 13 conc)*	*9 (deconc)*	1 *(deconc)*
AD	3	3	*3*

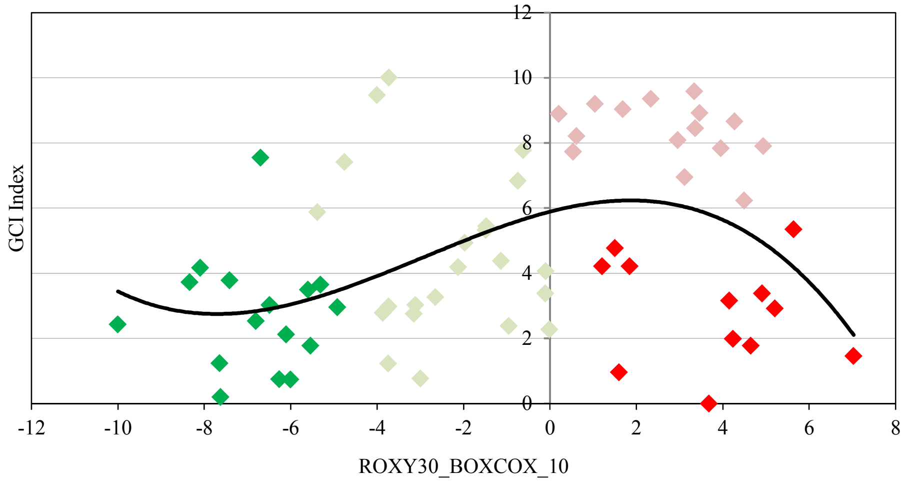
Final Cluster Centres	Cluster 1 *Concentration with generally low GCI value*	Cluster 2 *Acceleration of concentration with high GCI value*	Cluster 3 *Deceleration of deconcentration with high GCI value*	Cluster 4 *Deconcent-ration with generally low GCI value*
Number of cases	13	15	22	15
ROXY 30_BOXCOX	5.04	4.30	–2.02	–6.85
GCI	2.77	8.33	4.57	2.82
AC	*5*	*6*	–	–
DC	–	*2 (conc)*	–	1 *(deconc)*
DD	*7(conc)*	*7 (conc)*	*18 (deconc)*	*8 (deconc)*
AD	1	–	4	*6*

Notes: same as in Table 3.
Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Figure 1

Connection between GCI scores and ROXY30_BOXCOX



Note: green colour refers to countries where deconcentration is accelerating, light green colour refers to countries where deconcentration is decelerating while red and pink colours represent counties where concentration is accelerating or decelerating, respectively.

Source: own calculation.

The results of ROXY and GEDI examination

For the analysis, we used both the original index values (ROXY) and the Box-Cox transformed values (ROXY_BOXCOX_10). The original values represented a relatively high level of skewness; hence, the results of their examination should be taken into account with care. The analysis shows that there are positive relationships between the GEDI Index and the different ROXY indexes, but the intensity is moderate. We examined the relationships for the three sub-indexes of GEDI as well (ATT – Entrepreneurial Attitudes, ABT – Entrepreneurial Abilities, ASP – Entrepreneurial Aspirations). The Attitudes sub-index has the weakest while the Aspiration has the strongest relationship with ROXY indexes among the sub-indexes (Table 5).

Table 5

The correlation coefficients between the GEDI Index, its sub-indexes and ROXY Index

ROXY INDEX *(original and transformed)*	GEDI (AVGCH)	ATT (AVGCH)	ABT (AVGCH)	ASP (AVGCH)
ROXY20	0.309 b)	0.252 a)	0.278 a)	0.334 b)
ROXY20_BOXCOX_10	0.277 a)	0.198	0.264 a)	0.305 a)
ROXY30	0.328 b)	0.279 a)	0.297 a)	0.343 b)
ROXY30_BOXCOX_10	**0.355** b)	0.281 a)	**0.353** b)	0.358 b)
ROXY40	0.310 b)	0.246 a)	0.284 a)	0.335 b)
ROXY40_BOXCOX_10	0.335 b)	0.254 a)	0.316 b)	**0.360** b)

Notes: ATT = "attitudes", ABT = "abilities, ASP = "aspirations", AVG = average, CH = change
a) Correlation is significant at the 0.05 level (2-tailed). b) Correlation is significant at the 0.01 level (2-tailed).
Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

The analysis indicates the strongest correlation between the ROXY Index and GEDI in the case of ROXY30_BOXCOX_10 (r = 0.355). The lowest results were measured in the case of ROXY20 while the other two ROXY cases have almost the same scores. We can see almost the same results at the sub-indexes. The positive correlation coefficient between the ROXY indexes and GEDI means that *the more concentrated the population in a given country's urban system, the better the entrepreneurial performance there*. To discover more about the attributes of entrepreneurial aspirations, we analysed the pillars of this sub-index as well (Table 6).

Table 6

The correlation coefficients between the pillars of GEDI Aspiration sub-index and ROXY Index

Pillars of GEDI Aspirations sub-index	ROXY20_BOXCOX_10	ROXY30_BOXCOX_10	ROXY40_BOXCOX_10
ProdInnov_AVG_CH	0.191	0.230*	0.271*
ProcInnov_AVG_CH	0.257*	0.270*	0.317**
HGrowth_AVG_CH	0.307**	0.382**	0.346**
Internation_AVG_CH	0.268*	0.345**	0.326**
RiskCap_AVG_CH	0.246*	0.261*	0.261*

Notes: ProdInnov = "product innovation", ProcInnov = "process innovation", HGrowth = "high growth", Internation = "internationalization", RiskCap = "risk capital", AVG = average, CH = change

a) Correlation is significant at the 0.05 level (2-tailed). b) Correlation is significant at the 0.01 level (2-tailed).

Source: own calculation.

Almost all of the pillars of the Aspiration sub-index have a significant relationship with the different ROXY indexes, but two pillars are outstanding among them: High growth and Internationalization have the strongest correlation coefficients with the ROXY Index. It means that *the high growth (of firm size) and internationalization of firms are relatively dependent on the concentration or deconcentration of a country's population*.

The cluster analysis was carried out for each ROXY Index case (cluster analyses with ROXY20 and ROXY40 see the Appendix) with the GEDI Index. We applied the original ROXY Index for this examination. The ROXY30 cases showed the best correlation values with the GEDI Index and its sub-indexes. In this case, we used the data of 65 countries because it was not possible to involve the United Arab Emirates. The first results of cluster analysis helped us to filter the extreme values. There were altogether 7 countries that had extreme positive or negative values (Costa Rica, Dominican Republic, Ecuador, Guatemala, India, Panama and Zambia). Thus, 58 countries have been involved in this cluster analysis (Table 7).

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Table 7

Results of cluster analysis: ROXY30 and GEDI Index

Final Cluster Centres	Cluster 1 *Concentration with generally low GEDI value*	Cluster 2 *Acceleration of concentration with high GEDI value*	Cluster 3 *Deceleration of deconcentration with high GEDI value*	Cluster 4 *Deconcentration with generally low GEDI value*
Number of cases	10	18	21	9
ROXY 30_BOXCOX	27.76	4.24	–10.31	–62.66
GEDI	39.35	69.40	41.07	31.78
AC	*5*	*5*	2	–
DC	–	*2 (conc)*	–	–
DD	*5(conc)*	*6 (conc)* *3 (deconc)*	*16 (deconc)* *2 (conc)*	*6 (deconc)*
AD	–	2	1	*3*

Source: own calculation.

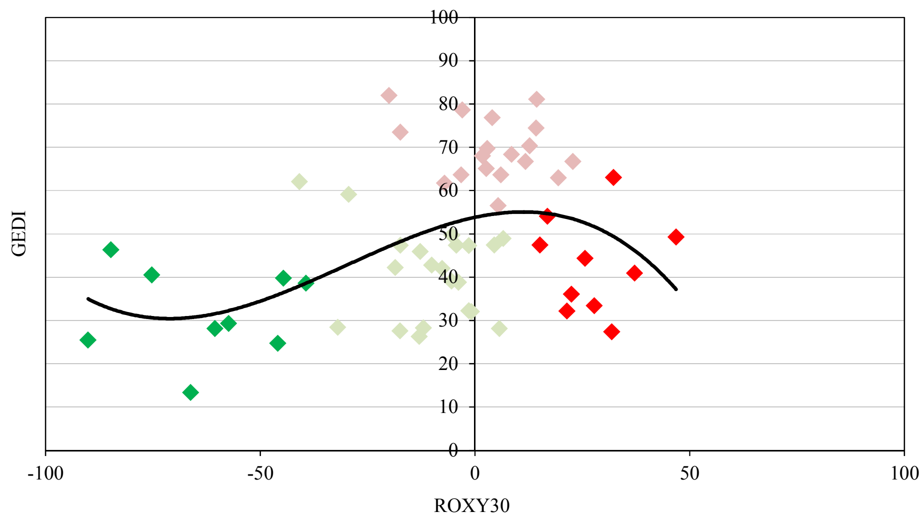
Cluster 4 contains those countries whose urban system is characterized by the acceleration of deconcentration or deceleration of deconcentration (but the cases typically remain in the deconcentration stage) represented with a relatively low ROXY30 and GEDI values (GEDI final cluster centre = 31.78). *Cluster 3* contains those countries where deconcentration is decelerating, but these countries are mostly still in the deconcentration stage. In this cluster, the GEDI Index is a bit higher than in Cluster 4 (GEDI final cluster centre = 41.07). Those countries have the highest GEDI values that belong to *Cluster 2* (GEDI final cluster centre = 69.4). This cluster can be characterized by the acceleration of concentration trends. This cluster contains countries in which the urban system is heading to a concentration from the deconcentration stage. This cluster is a transitional category: it shows the deceleration of deconcentration in counties that have already changed into the concentration stage (DD) or are still in the deconcentration stage (DC). The countries in *Cluster 1* have the highest ROXY30 values (final cluster centre = 27.76), but their GEDI values (final cluster centre = 39.35) are lower than countries in Cluster 2.

These results mean that those countries that show deconcentration trends have lower GEDI values than other countries characterized by concentration. However, this does not mean an obvious nexus between concentration and high entrepreneurial performance. It seems that concentration of the population has a positive effect on entrepreneurship, but on the other hand, *there is a threshold and after that, further concentration of the population may not improve the entrepreneurial performance, but has a negative effect on it* (Figure 2).

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Figure 2

The connection between the GEDI scores and ROXY30



Note: green colour refers to countries where deconcentration is accelerating, light green colour refers to countries where deconcentration is decelerating while red and pink colours refer to counties where concentration is accelerating or decelerating, respectively.

Source: own calculation.

Conclusions

The correlation analyses confirmed that *the more concentrated the population within a country's urban system, the higher its competitiveness and entrepreneurial performance*. This result seemingly supports the "bigger is better" concept. The correlation analysis has shown that the concentration or deconcentration of the population is *only one important factor* in the explanation of countries' entrepreneurial performance and competitiveness. This has been proven by the moderate correlation coefficients between the GCI/GEDI and ROXY indexes (both original and transferred). Consequently, we should consider that other effects may exist (e.g. differences in institutional settings, creativity and openness of human resources, culture).

However, in-depth analysis (conducting cluster analyses) confirmed that *relatively high-levels of concentration or deconcentration within an urban system are coupled with lower GCI/GEDI values*. Those countries have the highest GCI/GEDI values that have a ROXY Index value close to zero. It means that they have a moderate level of concentration (positive ROXY values) or moderate level of deconcentration (negative ROXY values). Our analysis indicates that, initially, as concentration increases (or deconcentration decreases) competitiveness and entrepreneurial performance also increase, but at a decreasing rate. Both of them eventually reaches a maximum and then after a certain point decrease with further concentration. Therefore, the curve that apprehends this relationship is non-linear and folding back. As follows, our results support the view that concentration is useful until a certain threshold, but excessive concentration could not help to improve

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

competitiveness or entrepreneurial performance. In other words, this indicates that *under- or over-concentration of the population within an urban system is not a useful phenomenon considering competitiveness or entrepreneurial performance*.

REFERENCES

Acs, Z. J.–Armington, C. (2004): Employment growth and entrepreneurial activity in cities *Regional Studies* 38 (8): 911–927.

Acs, Z.–Autio, E.–Szerb, L. (2014): National Systems of Entrepreneurship: Measurement issues and policy implications *Research Policy* 43 (3): 476–494.

Acs, Z.–Varga, A. (2002): Geography, Endogenous Growth, and Innovation *International Regional Science Review* 25 (1): 132–148.

Annoni, P.–Kozovska, K. (2010): *EU Regional Competitiveness Index 2010* JRC Scientific and Technical Reports, European Commission Joint Research Centre Institute for Security and Protection of the Citizens, Luxemburg.

Castells, M. (1996): *The rise of the network society* Blackwell, Oxford.

Castells-Quintana, D.–Royuela, V. (2014): Agglomeration, inequality and economic growth *Annual Regional Science* 52 (2): 343–366.

Chinitz, B. (1961): Contrasts in Agglomeration: New York and Pittsburgh. *American Economic Review* 51 (2): 279–289.

Ciccone, A. (2002): Agglomeration Effects in Europe. *European Economic Review* 46 (2): 213–227.

Ciccone, A. –Hall, R. (1996): Productivity and the density of economic activity *American Economic Review* 86 (1): 54–70.

Czaller, L. (2012): A Zipf-törvény érvényesülése a világ országaiban *Területi Statisztika* 52 (5): 461–478.

David, Q.–Peeters, D.–Van Hamme, G.–Vandermotten, C. (2013): Is bigger better? Economic performances of European cities, 1960-2009 *Cities* 35: 237–254.

Duranton, G.–Puga, D. (2004): Micro-foundations of urban agglomeration economies. In: Henderson, J. V. – Thisse, J. F. (eds.): *Handbook of Regional and Urban Economics*, Edition1, Chapter 48, pp. 2063–2117., Elsevier, Amsterdam.

Duranton, G.–Puga, D. (2001): Nursery Cities *American Economic Review* 91 (5): 1457–1477.

Eeckhout, J. (2004): Gibrat's Law for (All) Cities *The American Economic Review* 94 (5): 1429–1451.

Ellison, G –Glaeser, E. L. (1997): Geographic concentration in US manufacturing industries: A dartboard approach *Journal of Political Economy* 105 (5): 889–927.

Gabaix, X. (1999): Zipf's Law for Cities: An Explanation *The Quarterly Journal of Economics* 114 (3): 739–767.

Glaeser, E. L.–Kallal, H.–Scheinkman, J.–Shleiter, A. (1992): Growth in the Cities. *Journal of Political Economy* 100 (6): 1126–1152.

Henderson, J. V. (2003a): Marshall's Scale Economies *Journal of Urban Economies* 53 (1): 1–28.

Henderson, J. V. (2003b): The Urbanization Process and Economic Growth: The So-What Question. *Journal of Economic Growth* 8 (1): 47–71.

Hirvonen, M.–Hiraoka, N.–Kawashima, T. (1997): Long-term Urban Development of the Finnish Population: Application of the ROXY-index Analytical Method *Gakushuin Economic Papers* 36 (2): 243–263.

Kawashima, T. (1978): *Recent Urban Evolution Processes in Japan: Analysis of Functional Urban Regions*. Presented at the 25th North American Meeting of the Regional Science Association, Chicago, Illinois, USA.

Kawashima, T. (1982): Recent urban trends in Japan: analysis of functional urban regions. In: Kawashima, T. – Korcelli, P. (eds.): *Human settlement systems: spatial patterns and trends*, pp. 21–40., International Institute for Applied Systems Analysis, Laxenburg.

Kawashima, T. (1985): Roxy Index: An Indicative Instrument to Measure the Speed of Spatial Concentration and Deconcentration of Population *Gakushuin Economic Papers* 22 (2): 183–213.

Kawashima, T. (1986), People Follow Jobs in Japan?: Suburbanization of Labour and Job Markets *Gakushuin Economic Papers* 23 (1–2): 157–183.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Kawashima, T.–Azis, I.–Tane, M. (1997): *Comparative analysis of intercity agglomeration* http://www.econbiz.de/Record/comparative-analysis-of-intercity-agglomeration-kawashima-tatsuhiko/10001297129 [downloaded: 1st February 2014]

Knoben, J.–Ponds R.–van Oort, F. (2011): Employment from new firm formation in the Netherlands: Agglomeration economies and the Knowledge Spillover Theory of Entrepreneurship *Entrepreneurship & Regional Development* 23 (3-4): 135–157.

Fukatsu, A.–Kawashima, T. (1999): Urbanization, Suburbanization and Revived-urbanization: ROXY-index Analysis for the Chou-line Region of Tokyo *Gakushuin Economic Papers* 36 (3): 389–414.

Klaassen, L. H.–Paelinck, J. H. P. (1979): The Future of Large Towns. *Environment and Planning A* 11 (10): 1095–1104.

Klaassen, L. H.–Bourdrez, J. A. et al. (1981): *Transport and Reurbanization*, Gower, Aldershot.

Krugman, P. (1991): Increasing returns and economic geography *Journal of Political Economy* 99 (3): 483–499.

Krugman, P. (2009): The Increasing Returns Revolution in Trade and Geography *The American Economic Review* 99 (3): 561–571.

Marshall, A. (1920): *Principles of Economics*. Macmillan and Co., Ltd., London.

McCann, P.–Acs, J. Z. (2012): Globalization: Countries, cities and multinationals *Regional Studies* 45 (1): 17–32.

McCann, P. (2013): *Modern Urban and Regional Economics*. Oxford University Press, London.

Parr, J. B. (2002): Agglomeration economies: ambiguities and confusions *Environment and Planning A* 34 (4): 717–731.

Puga, D. (2010): The magnitude and causes of agglomeration economies *Journal of Regional Science* 50 (1): 203–219.

Reynolds, P.D.–Storey, D.J.–Westhead, P. (1994): Cross-national comparisons of the variation in new firm formation rates *Regional Studies* 28 (4): 443–456.

Rizov, M.–Oskamb, A.–Walshc, P. (2012): Is there a limit to agglomeration? Evidence from productivity of Dutch firms *Regional Science and Urban Economics* 42 (4): 595–606.

Rosenthal, S. S.–Strange, W. C. (2003): Geography, industrial organization and agglomeration. *Review of Economics and Statistics* 85 (2): 377–393.

Rosenthal, S. S.–Strange, W. C. (2004): Evidence on the nature and sources of agglomeration economies. In: Henderson, J. V., Thisse, J. F. (eds.): *Handbook of Regional and Urban Economics*, Edition 1, Chapter 49, pp. 2119–2171., Elsevier, Amsterdam.

Schwab, K. (ed.) (2013): *The Global Competitiveness Report*. World Economic Forum. http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2013-14.pdf [downloaded: 1st March 2014]

Szerb, L–Ács, Z.–Aidis, R. (2013): *A comparative analysis of Hungary's entrepreneurial performance in the 2006–2010 time period based in the GEM and the GEDI methodologies*. Pécsi Tudományegyetem Közgazdaságtudományi Kar, Pécs.

Tabuchi, T.–Thisse, J.–Zeng, D. (2005): On the number and size of cities *Journal of Economic Geography* 5 (4): 423–448.

Van Stel, A.–Suddle, K. (2008): The impact of new firm formation on regional development in the Netherlands *Small Business Economics* 30 (1): 31–47.

Vreeker, R.–Deakin, M.–Curwell, S. (eds.) (2009): *Sustainable Urban Development, Volume 3: The Toolkit for Assessment.* Routledge: New York.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Appendix

Table A1

The length of examined periods

Country	Period (t)	Period (t–1)	Country	Period (t)	Period (t–1)
Algeria	2008/1998	1998/1987	Jordan	2004/1994	no results
Argentina	2010/2001	2001/1991	Kazakhstan	2012/1999	1999/1989
Australia	2011/2006	2006/2001	Korea	2012/2002	2002/1997
Austria	2013/2001	2001/1991	Latvia	2013/2000	2000/1989
Bangladesh	2011/2001	2001/1991	Macedonia	2010/2002	2002/1994
Belgium	2013/2000	2000/1990	Malaysia	2000/1991	no results
Bolivia	2010/2001	2001/1992	Mexico	2010/2005	2005/2000
Bosnia Herzegovina	2013/1991	1991/1981	Montenegro	2011/2003	2003/1991
Brazil	2010/2000	2000/1991	Morocco	2004/1994	1994/1982
Canada	2011/2006	2001/1996	The Netherlands	2013/2000	2000/1990
Chile	2012/2002	2002/1992	Norway	2013/2000	2000/1990
China	2010/2000	2000/1990	Panama	2010/2000	2000/1990
Colombia	2010/2005	2005/1993	Peru	2007/1993	1993/1981
Costa Rica	2011/2000	2000/1984	Philippines	2010/2000	2000/1990
Croatia	2011/2001	2001/1991	Poland	2012/2002	2002/1992
Czech Rep	2011/2001	2001/1991	Portugal	2011/2001	no results
Denmark	2013/2000	2000/1990	Puerto Rico	2010/2000	2000/1990
Dominican Rep	2010/2002	2002/1993	Romania	2011/2002	2002/1992
Ecuador	2010/2001	2001/1990	Russia	2013/2002	2002/1989
Egypt	2006/1996	1996/1986	Saudi Arabia	2010/2004	2004/1992
Finland	2012/2000	2000/1990	Serbia	2011/2002	2002/1991
France	2011/2006	2006/1999	Slovakia	2012/2001	2001/1991
Germany	2012/2001	2001/1995	Slovenia	2013/2002	2002/1991
Ghana	2010/2000	2000/1996	South Africa	2011/2001	no results
Greece	2011/2001	2001/1991	Spain	2013/2001	2001/1991
Guatemala	2008/2002	2002/1994	Sweden	2012/2005	2005/2000
Hungary	2013/2001	2001/1990	Switzerland	2012/2000	2000/1990
Iceland	2013/2005	2005/2000	Taiwan	2012/2006	2006/2001
India	2011/2001	2001/1991	Uganda	2011/2002	2002/1991
Iran	2011/2006	2006/1996	UAE	2005/1995	1995/1985
Ireland	2011/2006	2006/2002	United Kingdom	2011/2001	2001/1991
Israel	2012/2008	2008/1995	United States	2012/2000	2000/1990
Italy	2012/2001	2001/1991	Uruguay	2011/2004	2004/1996
Jamaica	2011/2001	2001/1991	Venezuela	2011/2001	2001/1990
Japan	2010/2005	2005/2000	Zambia	2010/2000	2000/1990

Source: edited by the authors.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Table A2

Results of the cluster analysis: ROXY20 – GCI index

Final Cluster Centres	Cluster 1 *Concentration with generally high GCI value*	Cluster 2 *Deceleration of deconcentration, with GCI values higher than the average*	Cluster 3 *Deconcentration, with generally low GCI values*
Number of cases	33	20	10
ROXY 20	15.10	–17.43	–82.13
GCI	5.58	4.60	2.61
AC	*9*	–	–
DC	2 *(conc/deconc)*	1 *(deconc)*	2 *(deconc)*
DD	*22 (conc)*	*18 (deconc)*	3
AD	–	1	*5 (deconc)*

Final Cluster Centres	Cluster 1 *Concentration with generally low GCI value*	Cluster 2 *Acceleration of concentration with high GCI value*	Cluster 3 *Deceleration of deconcentration with high GCI value*	Cluster 4 *Deconcent-ration with generally low GCI value*
Number of cases	18	18	8	22
ROXY 20_BOXCOX	4.35	4.69	4.35	–5.81
GCI	2.78	7.62	7.94	2.49
AC	*6*	*4*	–	–
DC	*1 (deconc)*	*2 (conc)*	–	*2 (deconc)*
DD	*11(conc)*	*12 (conc)*	*7 (deconc)*	*13 (deconc)*
AD	–	–	1	*7*

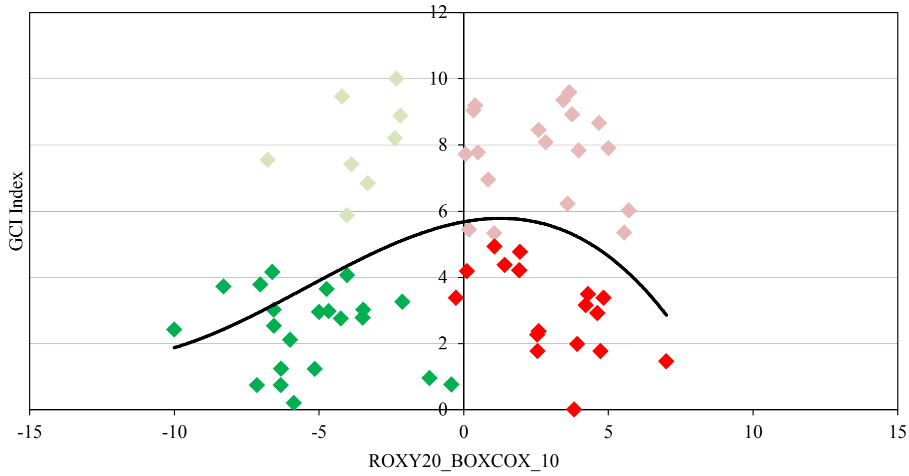
Notes: **AC** = Acceleration of concentration. **DC (deconc)** = Deceleration of concentration and it has already changed to deconcentration stage. **DC (conc)** = Deceleration of concentration, but it is still in concentration stage. **DD (deconc)** = Deceleration of deconcentration, but it is still in deconcentration stage. **DD (conc)** = Deceleration of deconcentration and it has already changed to concentration stage. **AD** = Acceleration of deconcentration.

Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Figure A1

The connection between GCI scores and ROXY20_BOXCOX



Note: green colour refers to countries where deconcentration is accelerating, light green colour refers to countries where deconcentration is decelerating while red and pink colours refer to counties where concentration is accelerating or decelerating, respectively.

Source: own calculation.

Table A3

Results of the cluster analysis: ROXY40 – GCI index

Final Cluster Centres	Cluster 1 *Concentration with generally high GCI value*	Cluster 2 *Deceleration of deconcentration, with GCI values higher than the average*	Cluster 3 *Deconcentration, with generally low GCI values*
Number of cases	26	28	8
ROXY 40	16.38	–21.28	–96.10
GCI	6.12	4.03	3.03
AC	*10*	–	–
DC	4 *(conc)*	–	1 *(deconc)*
DD	*12 (conc)*	*24 (deconc)*	*1 (deconc)*
AD	–	4	6

(Table continues on next page.)

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

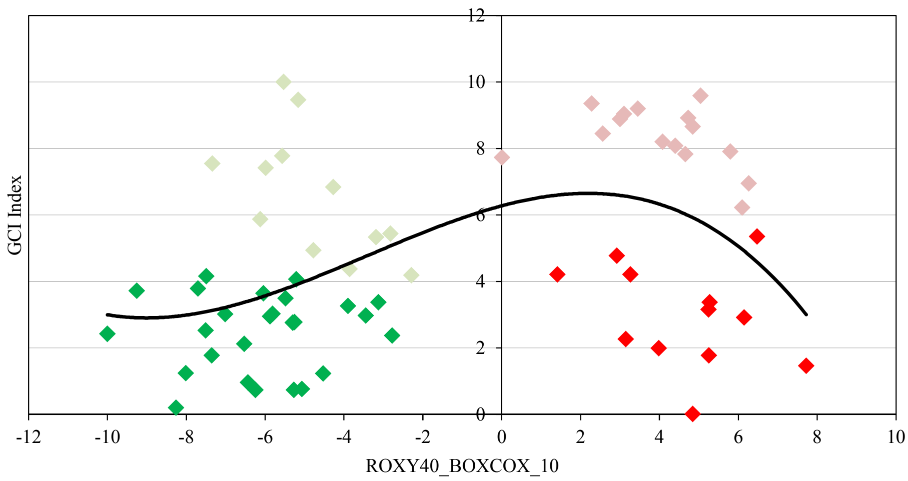
(Continued)

Final Cluster Centres	Cluster 1 *Concentration with generally low GCI value*	Cluster 2 *Acceleration of concentration with high GCI value*	Cluster 3 *Deceleration of deconcentration with high GCI value*	Cluster 4 *Deconcent-ration with generally low GCI value*
Number of cases	12	15	12	26
ROXY 40_BOXCOX	4.64	4.02	–4.74	–6.11
GCI	2.96	8.33	6.60	2.47
AC	*5*	*5*	–	–
DC	*1 (conc)*	*3 (conc)*	1 *(deconc)*	–
DD	*6 (conc)*	*7 (conc)*	*8 (deconc)*	*17 (deconc)*
AD	–	–	3	*9*

Notes: same as in Table 3 and Table 4.
Source: own calculation.

Figure A2

The connection between GCI scores and ROXY40_BOXCOX



Note: green colour refers to countries where deconcentration is accelerating, light green colour refers to countries where deconcentration is decelerating while red and pink colours refer to counties where concentration is accelerating or decelerating, respectively.
Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Table A4

Results of cluster analysis: ROXY20 – GEDI Index

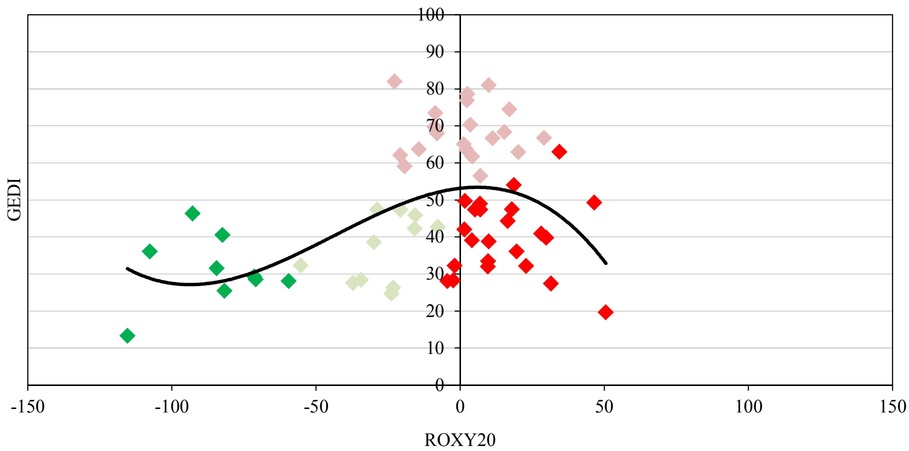
Final Cluster Centres	Cluster 1 *Concentration with generally lower GCI value*	Cluster 2 *Acceleration of concentration with high GCI value*	Cluster 3 *Deceleration of deconcentration, with high GCI value*	Cluster 4 *Deconcentration, with generally low GCI value*
Number of cases	23	20	11	9
ROXY 40_BOXCOX	15.8	1.16	–26.58	–85.12
GEDI	40.04	68.52	36.66	31.01
AC	*6*	*3*	–	–
DC	–	*2 (conc)*	1 *(deconc)*	2 *(deconc)*
DD	*14 (conc), 3 (deconc)*	*8 (conc), 7 (deconc)*	8 *(deconc)*	3 *(deconc)*
AD		–	2	*4*

Notes: **AC** = Acceleration of concentration. **DC (deconc)** = Deceleration of concentration and it has already changed to deconcentration stage. **DC (conc)** = Deceleration of concentration, but it is still in concentration stage. **DD (deconc)** = Deceleration of deconcentration, but it is still in deconcentration stage. **DD (conc)** = Deceleration of deconcentration and it has already changed to concentration stage. **AD** = Acceleration of deconcentration.

Source: own calculation.

Figure A3

The connection between the GEDI scores and ROXY20



Note: green colour refers to countries where deconcentration is accelerating, light green colour refers to countries where deconcentration is decelerating while red and pink colours refer to counties where concentration is accelerating or decelerating, respectively.

Source: own calculation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Table 12

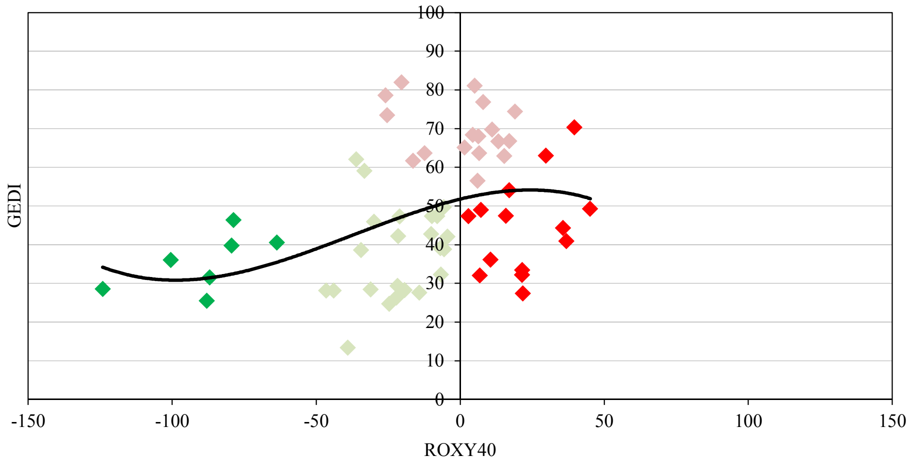
Results of cluster analysis: ROXY40 and GEDI Index

Final Cluster Centres	Cluster 1 *Concentration with generally lower GCI value*	Cluster 2 *Acceleration of concentration with high GCI value*	Cluster 3 *Deceleration of deconcentration, with high GCI value*	Cluster 4 *Deconcentration, with generally low GCI value*
Number of cases	14	18	22	7
ROXY 40_BOXCOX	22.31	0.43	–22.27	–88.71
GEDI	42.29	68.26	37.21	35.45
AC	6	*4*	–	-
DC	2 *(deconc)*	2 *(conc)*	–	1 *(conc)*
DD	6 *(conc)*	*6 (conc),* 3 *(deconc)*	*21 (deconc)*	1 *(conc)*
AD	–	3	–	*5*

Source: own calculation.

Figure 6

The connection between the GEDI scores and ROXY40



Source: own computation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 97–120; DOI: 10.15196/RS05106

Appendix 3

Table 13

The results of cluster analysis between ROXY Indexes and GCI Index

ANOVA	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
ROXY_20	36945,376	2	181,007	60	204,110	,000
GCI_AVE_CH_10	34,468	2	7,247	60	4,756	,012
ROXY_20_BOXCOX_10	523,973	3	3,277	62	159,916	,000
GCI_AVE_CH_10	136,622	3	1,772	62	77,086	,000
ROXY_30_BOXCOX_10	465,608	3	1,431	61	325,341	,000
GCI_AVE_CH_10	100,020	3	3,593	61	27,841	,000
ROXY_40	39659,711	2	220,599	59	179,782	,000
GCI_AVE_CH_10	43,303	2	7,044	59	6,147	,004
ROXY_40_BOXCOX_10	529,564	3	2,936	61	180,339	,000
GCI_AVE_CH_10	135,866	3	1,830	61	74,261	,000

Table 14

The results of cluster analysis between ROXY Indexes and GEDI Index

ANOVA	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
ROXY20	23860,079	3	223,237	59	106,882	,000
GEDI_AVGCH	4616,308	3	81,006	59	56,987	,000
ROXY30	14152,718	3	162,240	54	87,233	,000
GEDI_AVGCH	4123,310	3	96,702	54	42,639	,000
ROXY40	20903,343	3	205,731	57	101,605	,000
GEDI_AVGCH	3827,858	3	120,616	57	31,736	,000

GYÖRGY JÓNA[a)]

New trajectories of the Hungarian regional development: balanced and rush growth of territorial capital

Abstract

The basic assumption of the paper is that numerous *similarities* exist between the patterns of economic growth and territorial capital growth. The rush economic growth and rush growth of territorial capital are compared empirically at Hungarian micro-regional level from 2004 until 2010. After normalizing the dataset, a very novel spatial econometric method is applied, called a penalty for bottleneck. The results show that the constant rush growth of territorial capital is as harmful as economic recession. On the other hand, the decrease of infrastructural and social capital caused the rush growth of territorial capital in this period. Moreover, the key findings of two case studies suggest that the balanced growth of territorial capital will be created by the falling social inequalities and increasing infrastructural capital.

Keywords: territorial capital, rush growth, balanced development, endogenous assets.

Introduction

Year by year the territorial capital of certain Hungarian micro-regions were increasing significantly (Jóna–Hajnal 2014, Jóna 2015). On first sight, we might think that the rush growth of territorial capital (RGTC) may be desirable and benevolent but the rush growth prevents the harmonic and sustainable development in the long run. The constant RGTC is as harmful as economic recession.

János Kornai's rush growth theory has been adopted to analyse and to compare to the harmonic growth of territorial capital (HGTC) and RGTC. According to the basic assumption, *empirical similarity* exists between rush economic growth and RGTC; so many similarities can be founded between these mentioned two concepts.

The study focuses on the special changes of the Hungarian micro-regions' territorial capital. It is not concerned about demonstrating the theoretical background of territorial capital as a large number of scholars have already been carried this out; it rather demonstrates the results of empirical research. In this study the most important features of HGTC and RGTC are designed, moreover the paper answers the questions which Hungarian micro-regions emerged RGTC and what created it between 2004 and 2010. Finally, partial effects of the HGTC and RGTC are compared in the study.

a) University of Debrecen, Department of Social Science, H-4400, Nyíregyháza, Sóstói u. 2-4., Hungary. E-mail: jona.gyorgy@foh.unideb.hu

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Theoretical background

The concept of territorial capital enroots the paradigm of new regionalism, accepting the supply-oriented approach (Capello–Nijkamp 2009). It can gather and classify the endogenous and exogenous assets of territorial development and can be operationalized with special spatial econometric methods. The new paradigm needed to be introduced due to the early conjecture, that proved empirically, that endogenous and exogenous goods can significantly determine the regional economic growth (Aroca–Stimson–Stough 2014, Camagni–Capello 2014, Maillat–Kibir 2001, Stimson–Stough 2009).

In these days the politicians, decision-makers and territorial planners focus on issues of economic growth. The topic of speeding up this growth has become a public affair (sometimes almost a fetishism of growth can be found), thus, at the first sight, the notion of rush growth may be desirable and benevolent. Kornai, however, highlights that rush growth prevents the harmonic and sustainable development on the long run (Kornai–Dániel–Jónás–Martos 1971). Applying a radical but expressive simile: the man who grows 180 cm in 23 years is healthy but a man who grows 180 cm in 5 years (namely he increases extremely fast) is ill. Economic growth is similar to human development; it also has a special growth trajectory. It needs to move uniformly and predictably, to guarantee the sustainability of harmonic growth. If rush growth emerges over the long term, it can deform the developmental structure, finally leading to economic depression and crisis. In short term, the rush growth is acceptable and natural but in the long run it has harmful effects on the local economic system (Kornai–Simonovits 1977, Kourtit–Nijkamp–Stimson 2014).

Kornai identifies rush growth as a type of disequilibrium, meaning the regional growth rate accelerates, compared to its own and the countries growth rate (Kornai 1971). It can be explained in that the partial effects of the development factors are fluctuating consistently, excessively and disproportionately; the potential of a region is discordant. The other typical mark of rush growth (the opposite of the previous situation) is that the determinants of growth define, almost to the same extent, the development trajectory, and this static status exists over the long term. It rarely emerges in practice, it is only a theoretical opportunity and the first one can be found in the real economic process (Kornai 1972, Mihályi 2011).

The important feature of harmonic growth, as an opposite, is the rate of growth of certain regions at nearly the same rate as the nation's growth rate, occasionally, short-term outstanding increases can occur sometimes (Kornai 2006). Moreover, there are differences among the partial effects of development determinants but these are not significant and the standard deviation is also small. Later, the growth factors may shift to a greater or lesser extent; one time the first factor and then another factors overcomes but extreme changes cannot be revealed in the development structure; the growth potential of the region is harmonic. Simply put, balanced growth is a dynamic equilibrium state that is not static nor excessive. The rare and short-run spurts are not called rush growth only the permanent and static regional development can be defined as rush growth.

Hypothesis 1 is that rush economic growth is similar theoretically and empirically to RGTC.

According to hypothesis 2, significant differences can be revealed *among the partial effects* of the factors of HGTC and RGTC.

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

The point of hypothesis 3 is that essential differences can be found in *the rank of the partial effects* of HGTC and RGTC at micro-regional level.

Hypothesis 4 alleges that RGTC is a barrier of sustainable economic growth; after the RGTC the decrease will be sharply higher than the result of rush growth.

Two aspects of the theoretical and methodological background of research are highlighted in the paper. Firstly, the trends of territorial capital growth differ in a certain country, some regions have a dynamic development trajectory, in other regions stagnation can be revealed, on the other hand, in some regions decrease can be measured. It explains that the study scrutinizes the RGTC at micro-regional level (LAU-1 or NUTS-4 region) between 2004 and 2010; in this period Hungary had 174 micro-regions. Secondly, regional economic growth and development are also directly and significantly determined by the endogenous and exogenous assets; both have to be analysed together, this is the main reason why the concept of territorial capital is applied. Kornai (1971) accepts and highlights these approaches too.

Research model and applied methods

Before beginning of operationalization, some important details need to be clarified. To date, RGTC has not been scrutinized, empirically, therefore in this study inductive methods were applied in some rare cases (Vieira–Tsotras 2013).

When researching territorial capital, seven types of capital can be identified; these arithmetic averages show the territorial capital[1] (Capello 2007). The seven kinds of capital are the following: Economic, institutional, infrastructural, relational, human, social and cultural capital. Accepting the set theory, it may be more understandable, if the territorial capital of a micro-region is calculated as $\mathrm{tc}_{r,t} = \{\frac{\mathrm{HC}_{r,t}+\mathrm{InfC}_{r,t}+\mathrm{InsC}_{r,t}+\mathrm{SC}_{r,t}+\mathrm{CC}_{r,t}+\mathrm{EC}_{r,t}+\mathrm{RC}_{r,t}}{N_{\mathrm{kc}}}\}$ where tc is territorial capital, HC is human capital, InfC is infrastructural capital, InsC is institutional capital, SC is social capital, CC is cultural capital, EC is economic capital, RC is relational capital, NKC is the number of the kind of capital, r is region and t is time. The territorial capital of nation can be operationalized as logically $TC_t = \sum_{l=1}^{n}\{tc_t\}$ where TC is territorial capital at national level Perucca 2013).

In those micro-regions where the growth of territorial capital was harmonic, the territorial capital increased between half and double the average growth of territorial capital. So: $0{,}5\overline{G_t} \leq HG_t \leq 2\bar{G}$, where HG is a region where harmonic growth of territorial capital is measured, G is average growth of territorial capital.

The equation of RGTC is: $RG_{r,t} \geq 4(HG_{N,t})$ where RG is a region where rush growth of territorial capital is measured, N is national level namely the 174 micro-regions of Hungary.

In some years the average growth of territorial capital decreased, in this case, the RGTC is measured: $RG_{r,t} \geq \left|3(HG_{N,t})\right|$.

The paper uses the set theory for plain language. Basically, the first set contains micro-regions in which the growth of territorial capital was harmonic, this set is called as HG. According to the previous definition, the second set embraces micro-regions in which the

1 $G = \sqrt[n]{a_1 \bullet a_2 \bullet a_3 \dots a_n}$, $a_n \in \mathbf{R}_0^+, n \in \mathbf{Z}^+$ where G refers to the arithmetic average, hereinafter average.

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

growth of territorial capital was rush, this set is named as RG. The numbers of both sets changed year on year. A certain micro-region does not belong to both sets; HG and RG were emerging as disjoint sets, so $HG \cap RG$.

Actually, the members of these sets were scrutinized separately and compared over a seven year term. The study only deals with RG and HG sets, the micro-regions where other territorial capital occurred, are excluded from the paper.

Every micro-region has territorial capital and these can be ranked based on their territorial capital. Moreover, within a framework of longitudinal research, the trend of territorial capital growth can be operationalized thus the rush and harmonic growth of territorial capital can be investigated.

The rush and balanced growth of territorial capital can be defined by flow variables, meaning chain relatives can demonstrate how territorial capital of micro-regions were shifting year by year and which micro-regions belong to HG or RG.

The items of research were mustered from the database of the Land Information System (Hungarian: TEIR). By applying Q-type principal component analysis (PCA), the multicollinearity was reduced and the number of variables can be eliminated (Yuz–Goodwin 2014). After the PCA, 48 variables remained in the model and these were classified within seven sub-indices, the average of seven sub-indices show the territorial capital of a micro-region. Ultimately, a matrix was created which consists of 58.464 (48 variables x 174 micro-regions x 7 years) cells (see Table 1).

Table 1

Variables and sub-indices constructing the territorial capital

Aggregated index	Sub-index/ dimension	Variables
Territorial capital	Economic capital	Total domestic income per capita
		Net export sales revenue per 1000 people
		Issued capital for 1 firm
		Equity for 1 firm
		Total firms for 1000 residents
		High-tech business service per 1000 residents
		Output per 1 firm
		Entrepreneurship
	Infrastructural capital	Footpath and pavement per 1 km^2
		Cycle path per 1 km^2
		Length of the national road per 1 km^2
		The length of the gas pipe per 1 km^2
		Drinking water system for public utility per 1km^2
		The length of the sewer per 1 km^2
		The size of the reservation per 1 km^2
		The size of the total green area
	Institutional capital	The number of libraries per 1000 people
		The number of the institutions for public culture per 1000 people
		The number of museums per 1000 people
		The number of theatres per 1000 people
		The number of cinema seats per 1000 people
		The number of post offices per 1000 people
		The number of art communities per 1000 people

(Table continues next page.)

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

(Continued.)

Aggregated index	Sub-index/ dimension	Variables
Territorial capital	Human capital	The number of infant mortality per 1000 live-birth The number of General Practitioners per 1000 people The number of chemist's per 1000 people The number of people enrolled in libraries per 1000 people The number of students taking part in tertiary education per 1000 people The number of teachers working in tertiary education per 1000 people
	Social capital	The number of the registered unemployed per 1000 people in active age The daily average number of people in the communal kitchen The number of people paying taxes per 1000 people The number of crimes with prosecution per 1000 people The number of economic crimes per 1 company Domestic migration difference Hoover-index
	Relational capital	The number of Internet users per 1000 people The number of mobile phone subscriptions per 1000 people The number of non-profit organizations per 1000 people The number of clubs for old people and the number of the members in them per 1000 old people
	Cultural capital	The number of the members of the art communities per 1000 people The number of theatre-goers per 1000 people The numbers of the participants on a cultural event per 1000 people The number of the museum visitors per 1000 people The number of the people going to permanent theatres per 1000 people The number of cinema visits per 1000 residents The number of monuments per 1000 people

Source: own calculation.

After normalization and weighting, the figures were corrected with a very novel way that is referred to as Penalty for Bottleneck (Ács–Szerb 2012). Finally, the micro-regions with harmonic and rush growth were selected according to the above definitions. This calculation was carried out for each year.

As already indicated, this model includes seven explanatory variables. The economic capital shows the economic performance of the region; Camagni calls these rivalry and tangible goods. Many scholars used similar indices measuring the economic capital (Capello et al. 2009, Brasili 2010, Veneri 2011, Brasili et al. 2012). Furthermore, the infrastructural capital contains the aptness and size of the elements of the infrastructure; the same indicators were used in studies by Capello et al. 2009, Brasili 2010, Russo et al. 2010, Brasili et al. 2012, Russo–Servillo 2012.

The institutional capital represents a micro-region's public institutions and their services. It is typical that the cultural institutions appear (one exception is the post office), between the institutional and the cultural capital emerge a close theoretical and empirical correlation in this model. Caragliu and Nijkamp (2008) also applied a similar indicator system to these variables. Human capital expresses two aspects of local society: on the one hand, the health condition of the population, and, on the other hand, the region's knowledge level. The health status also determines the regional welfare; if the population is healthy, the economic output and performance may increase, in parallel the social transfers decrease. The health condition is operationalized with the traditionally accepted indicators

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

(e.g. infant mortality, life expectancy and so on), the knowledge degree is measured by the number of students and teachers participating in tertiary education and the number of people enrolled in the libraries (Kunzmann 2007, Caragliu–Nijkamp 2008, Camagni et.al. 2011, Brasili et.al. 2012, Russo–Servillo 2012).

The social capital refers to the level of social integration (Bourdieu 1983). It has two dimensions: (1) employment and (2) local social inequalities (Putman 2000). The first one is measured – among the others – with the employment rate of the sub-regions and the second one embraces the Hoover-index and so on; both effectively demonstrate the level of social integration. Additionally, homelessness causes social disintegration (it does not mean that the homeless people are harmful to society but the phenomenon itself can lead to disintegration in the local society). It is notable that the institution system for the homeless functions in bigger cities and metropolises (of course there are some exceptions), but the public kitchens (kitchens for the poor) are concerned about the groups living in social exclusion in small villages or towns. In general, the local trust relations can be measured through the observation of economic crime (Camagni–Capello 2015, Russo et al. 2010, Veneri 2011, Brasili et al. 2012).

The relational capital firstly includes the communication devices. The relational nets can come into being formally and informally. The first one usually occurs in civilian organizations, the latter evolves in formal and informal clubs (club-goods). In Hungary both are relatively new. In the third sector the relational capital can be accumulated, which can be converted to the economic capital and becomes the driving force of regional growth. The clubs for the old (here special club-goods are formed) belong to the relational capital on account of the local welfare is determined too by the quality of interactions of old people (Camagni et al. 2011). Moreover, the cultural capital embraces the number of the different cultural institutions and their capacity (Caragliu–Nijkamp 2008, Brasili et al. 2012). The more details can seen in Table 2.

Table 2

Main data of principal component analysis

Year	Number of principal components	Measure of Sampling Adequacy (MSA-test)	Kaiser–Meyer–Olkin (KMO-test)	Durbin–Watson-test	Total variance explained, %	Redundancy[a)]
2004	7	0.68	0.711	1.985	71.45	0.514
2005	7	0.71	0.763	2.249	82.58	0.532
2006	7	0.67	0.757	2.198	78.42	0.494
2007	7	0.78	0.801	1.995	84.62	0.501
2008	7	0.79	0.812	2.366	78.29	0.551
2009	7	0.77	0.809	2.341	79.32	0.519
2010	7	0.81	0.825	2.113	80.35	0.533

N=174; sig.: p<0,05.

a) $Red = \sqrt{\frac{\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m} r_{i,j}^{2}}{m(m-1)}}$.

Source: own calculation.

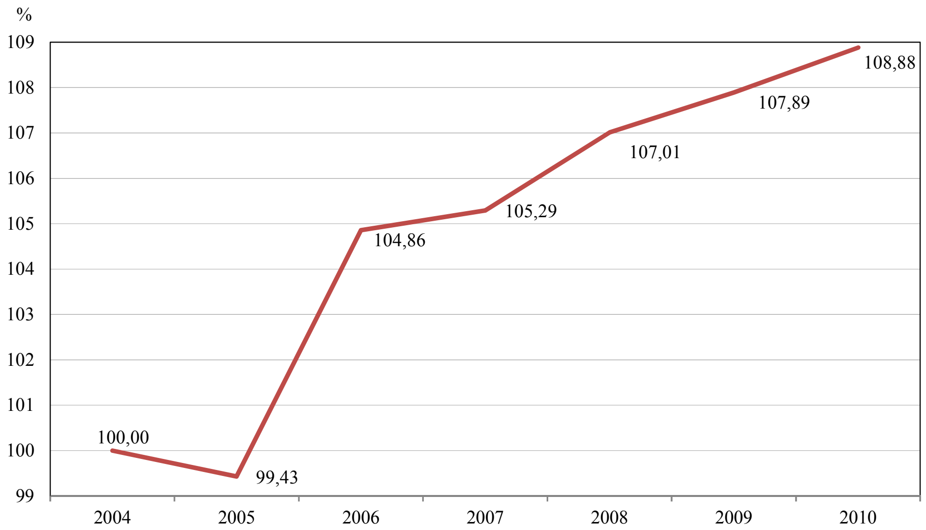
REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Results

Calculating the arithmetic average (average) between 2004 and 2010 the territorial capital as the national level was increasing by 9,66%, indicating that the annual average growth of territorial capital was 1,38%. Between 2005 and 2006, the highest accumulation of territorial capital emerged and then the trend slowed owing to the credit crisis of 2008, after that the trend of territorial capital was corrected (see Figure 1).

Figure 1

The change of the territorial capital at a national level between 2004 and 2010
2004=100%



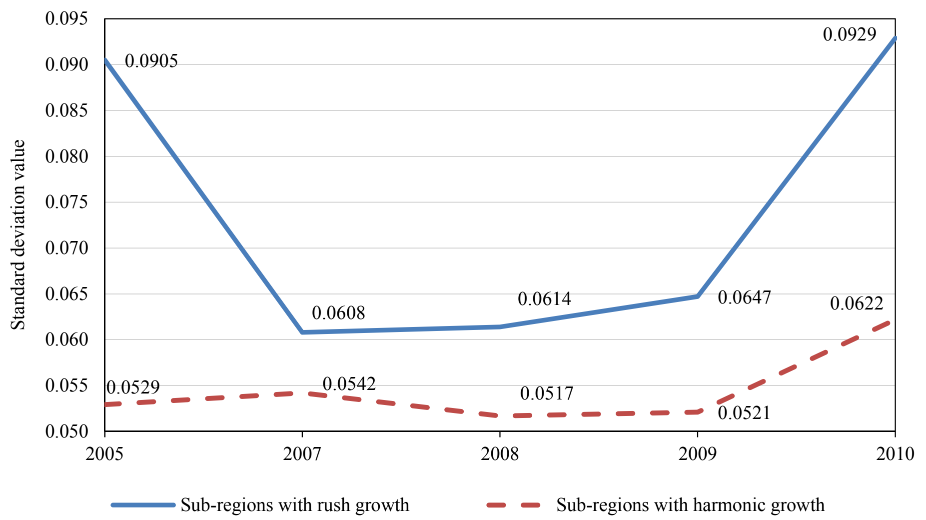
Moreover, the paper answers the question as to the extent of standard deviation among the partial effects of territorial capital. According to Kornai, the rush growth is determined significantly as well by the high standard deviation of partial effects. Accepting this qualitative thesis, the high standard deviation of the partial effects of territorial capital (one has high partial effect, and another has low partial effect specifically the standard deviation is high among them) causes RGTC. Basically, Hypothesis 2 will be tested empirically; it is confirmed statistically by the Figure 2.

In general, the Figure 2 compares the standard deviation of the partial effects of HG and RG. It is self-evident that the standard deviation of RG is higher than HG, strengthening that rush economic growth is similar to RGTC so Hypothesis 2 can be accepted. The standard deviation of RG was fluctuating hectically in the first period; from 2007 it was decreasing but did not level off to the standard deviation of HG. The standard deviation of the partial effects of RG started increasing again due to the effect of the debt crisis of 2008. The HG factors were determined as well by the financial depression but its standard deviation increased slightly.

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Figure 2

The change of the deviation between the capital types determining the HG and RG



The paper then scrutinises what types of capital define RG and HG sets; it is calculated by multiple linear regression analysis. Simply put, it can be inspected what types of capitals determine the shift of territorial capital.

First, multiple correlation analysis is applied to reveal the statistical connection among the variables of RG and HG. In this dynamic model the connection of the strength and direction of the explanatory and dependent variables were calculated as a set. So: $\Delta HG_{r,t} = HG\{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_7 X_7 + \varepsilon\}$, and $\Delta RG_{r,t} = RG\{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_7 X_7 + \varepsilon\}$. Finally, the results of regression can explain difference among the factors of RG and HG.

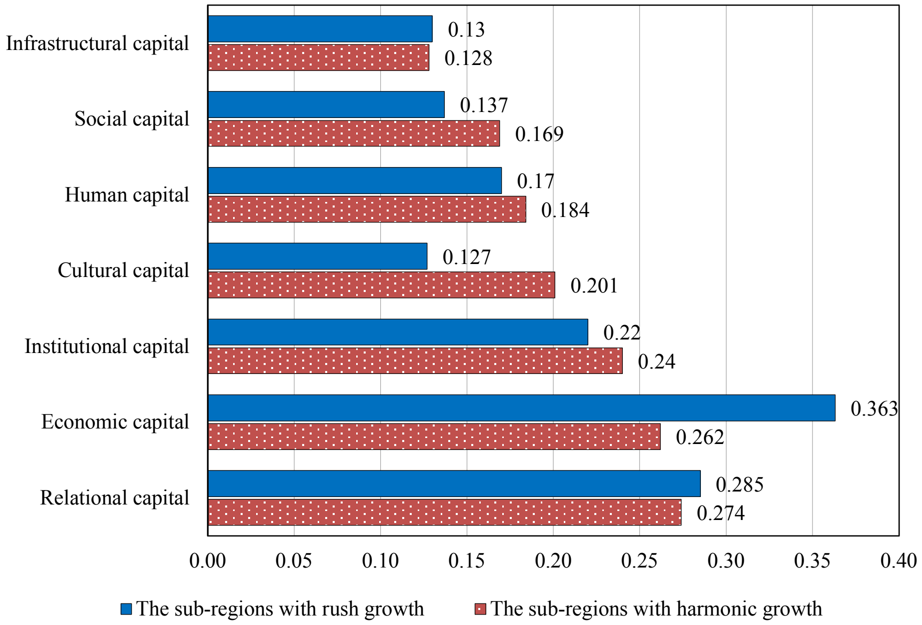
The partial effect of independent variables can be measured by standardized regression coefficient; it is the so-called beta value. The beta value demonstrates the effect of explanatory power of independent variable on depend variable. It can only be taken into account, if the significance level of t-probe is suitable ($p<0,05$).

The RGTC can be operationalised in 20 micro-regions from 2004 to 2005. Figure 3 shows that among the territorial capital structures of RG and HG, significant differences can be revealed; for example the partial effect of economic capital was the highest and the partial effect of cultural capital was the lowest in RG. Moreover, within the ranking of the partial effect of types of capital a large number of differences can be found between two sets: the territorial capital growth of HG was defined by the relational-, economic- and institutional capital.

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Figure 3

The capital types determining the territorial capital of the sub-regions with rush and harmonic growth in 2005 (beta-value)



Nevertheless, the partial effects of three types of capital were the highest in the RG as well, but the ranking of types of capital is confused: the economic capital is the strongest than relational- and institutional capital. Interestingly, in RG the partial effect of social capital was significantly lower than in HG. The growth of RG was determined by the exogenous assets, the HG was defined in particular by endogenous goods. More broadly, the endogenous development could cause HGTC, the hegemony of exogenous goods could impose RGTC.

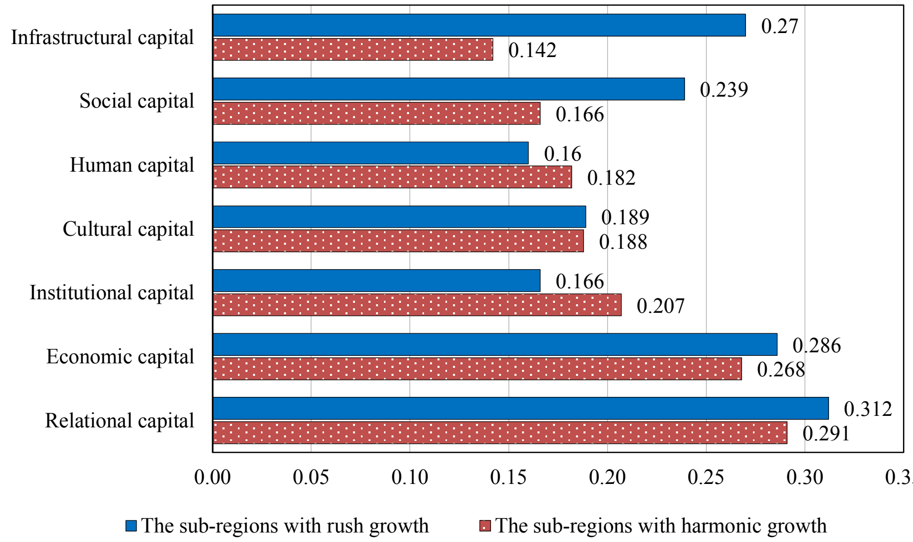
In 2006, the growth of territorial capital peaked; the territorial capital growth was 6.05 percentage points. Compared to the average annual growth of territorial capital, it is relatively high but not extreme. In such growth conditions, if the RGTC had been defined by the average growth of territorial capital of one and a half, only one micro-region (Ibrány-Nagyhalászi micro-region) would have belonged to the RG. In this year RG became empty, RG=Ø. Put another way, the 6.05 percentage point growth of territorial capital can create a stable equilibrium of territorial capital growth at national and micro-regional level. This growth trend may impede the RGTC.

32 micro-regions belonged to RG in 2007. The HG was determined consistently by relational-, economic- and institutional capital (see Figure 4). In contrast, the HG was defined by relational-, economic- and infrastructural capital; the partial effect of last one is significant. In a nutshell, in RG, the partial effects of social- and infrastructural capital were stronger, and the partial effect of institutional capital was weaker than in HG. These three differences caused RGTC in 2007.

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Figure 4

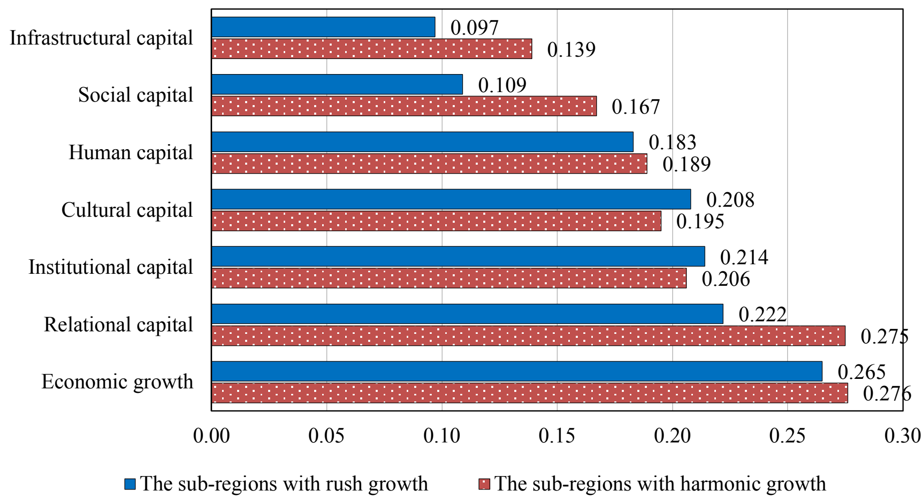
The capital types determining the territorial capital of the sub-regions with rush and harmonic growth in 2007 (beta-value)



In 2008 RG had 39 objects and their territorial capital structure changed basically compared to previous years: the partial effect of types of capital decreased in RG (see Figure 5). The partial effect of infrastructural-, relational- and social capital decreased significantly, and this circumstance caused RGTC in 2008.

Figure 5

The capital types determining the territorial capital of the sub-regions with rush and harmonic growth in 2008 (beta-value)

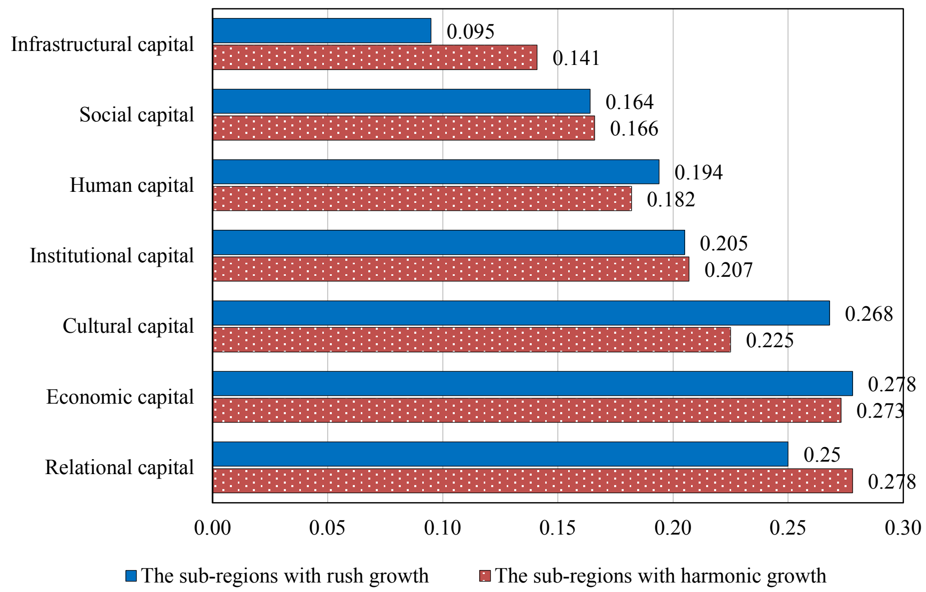


REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

In 2009, the following year, 53 micro-regions could be found in RG. The effect of the economic crisis on territorial capital structure and the following trend was directly very harmful. Compared to previously, the growth factors of RG and HG also reshaped. Most evidently, the partial effect of economic and cultural capital increased in RG micro-regions; in parallel, the partial effect of relational capital decreased considerably (see Figure 6). The spatial equilibrium overturned because of the shifted situation.

Figure 6

The types determining the territorial capital of the sub-regions with harmonic and rush growth in 2009 (beta-value)



In RG, there were 52 objects in 2010, see the Figure 7. It is important that only two differences can be found between RG and HG: the partial effects of infrastructural and social capital are lower in RG than in HG. The territorial capital structure of RG was almost similar to HR except for the above mentioned two capitals. In this year, on the one hand, RGTC was founded owing to both partial effects of kind of capitals decreased, and the other, the standard deviation shot up significantly among the types of capital. More precisely, after the economic crisis, RGTC was formed by the sudden rise of social inequalities and the gradual fall of the partial effect of infrastructural capital.

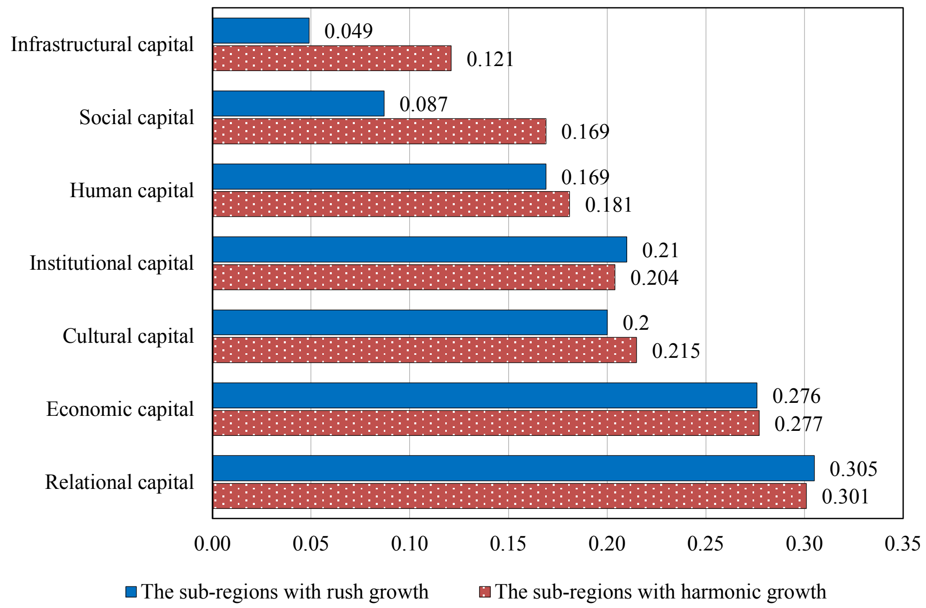
Finally, in RG the effect of the kind of capital emerged disproportionately compared to HG. In the territorial capital structure, the hegemony of exogenous assets can hamper the steady and harmonic development, meaning the endogenous goods have to be exploited more to reach a balanced growth of territorial capital. Those micro-regions were able to stand on HGTC that could utilize endogenous process, mostly their relational capital and complex economic networks. In every year, the territorial capital structure of RG was changing but in general, the partial effects of infrastructural and social capital were falling sharply. Overall, the infrastructural conditions were decreasing (other sectors were

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

developed instead of infrastructure) in this period, in contrast, the local social inequalities were rising and this complex status caused RGTC.

Figure 7

The capital types the territorial capital of the sub-regions with harmonic and rush growth in 2010 (beta-value)



Basically, these figures show that Hypothesis 3 can be accepted as well.

Table 3 presents how many times RGTC formed in certain micro-regions. The number of objects of RG doubled almost because of the debt crisis of 2008; it then decreased slightly.

Table 3

The change of the number of objects RG and HG between 2005 and 2010

Denomination	2005	2006	2007	2008	2009	2010
The number of objects of RG	20	0	32	39	53	52
The number of objects of HG	84	102	71	69	64	61

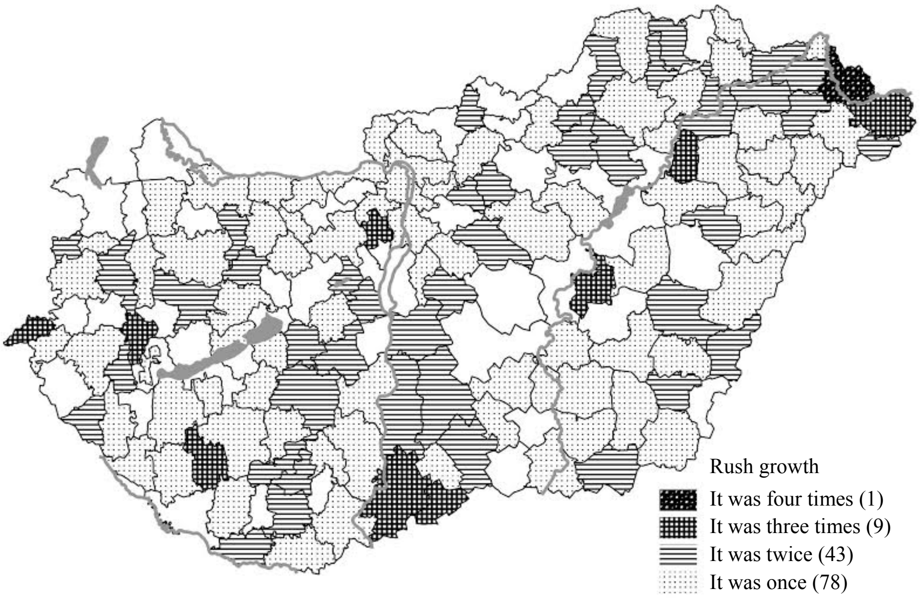
Source: own calculation.

The RGTC has no geographical patterns in Hungary. However, the results suggest that the RGTC created regularly in the micro-regions with low competitiveness but it rarely occurred in developed micro-regions as well (see Figure 8).

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Figure 8

Change of territorial capital of two micro-regions



A short case study regarding two micro-regions

In this section, two case studies are outlined to illustrate empirically the characters, patterns and consequences of RGTC in practice. It is important to recognize empirically the effects of RGTC on regional economic development and the accumulation of territorial capital at local level. It may also provide a good example and systematic evidence for other regions where RGTC has a negative effect on the balanced regional growth and development.

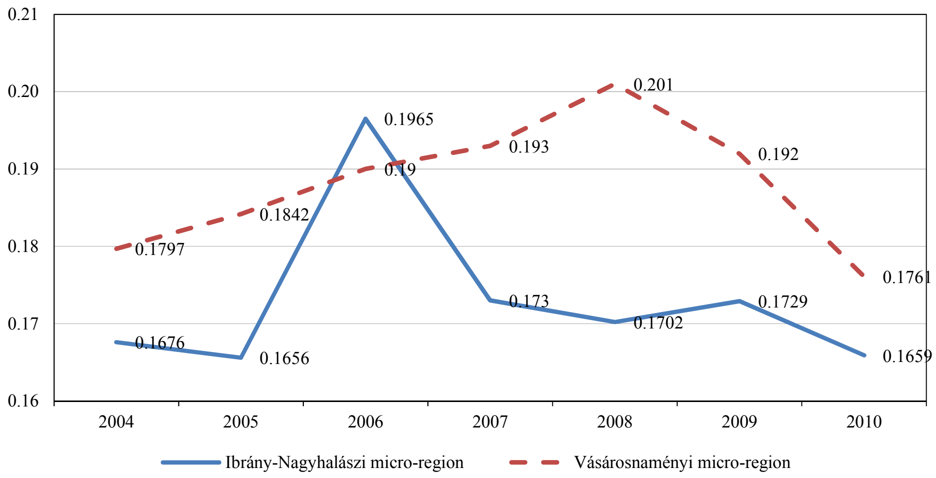
First, the territorial capital trends of Vásárosnamény and Ibrány-Nagyhalász have to be scrutinized; the results are shown in Figure 8. RGTC appeared in Vásárosnamény four times and only once in Ibrány-Nagyhalász. In Vásárosnamény the RGTC was growing rapidly for four years, following which, the trajectory of territorial capital fell dramatically. It simply means that the RGTC prevents harmonic and equilibrium growth, Kornai's qualitative theses and Hypothesis 4 of the paper can be fortified empirically too. Moreover, between the regional economic development and territorial capital growth so many similarities can be found.

The development of the Ibrány-Nagyhalász micro-region was similar to Vásárosnamény. The difference that emerged between them is that in 2006 the RGTC lasted only one year in Ibrány-Nagyhalász and after that its territorial capital was decreasing sharply.

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Figure 9

Change of territorial capital of two micro-regions



Finally, if the average growth of territorial capital had been higher, the number of objects of RG would have been lower. Empirical evidence demonstrates that the growth of annual 6 percent point of territorial capital was able to reduce the number of RG, see 2006. Moreover, the economic crisis of 2008 owing to the number of RG shot up significantly, meaning that the territorial capital structure of the most micro-regions was deformed. Furthermore, between the partial effects and the rank of kind of capitals of RG and HG emerged substantially. To conclude, the partial effects of social- and infrastructural capital decreased mostly caused the RGTC; the RGTC was determined by exogenous assets, as opposed to the HGTC, which was determined by endogenous goods.

Conclusions

Kornai's theses have actual messages for regional development policy. Those territorial assets should be mobilized which guarantee slow but sustainable economic development. Contrary to this, mainly in Hungary, those territorial goods have been exploited which impose RGTC. It is because the growth-oriented world needs rapid, significant and spectacular economic development that it contributes to the rapid but unstable growth of territorial capital.

The HGTC causes strong regional economic development trajectory on the long run; notwithstanding, RGTC was preferred by regional decision makers, territorial experts, planners and stakeholders owing to bringing political capital and legitimacy. It is because the local society believes that rush growth is a benevolent phenomenon thus it has been established by political leaders and territorial experts. If rush growth is conducted in practice by stakeholders or/and territorial planners, the local society legitimatise the leaders acts. Fundamentally, this is the dilemma: the RGTC brings political legitimacy in the short run but can upset local regional development; on the contrary, the HGTC may achieve lower political legitimacy but can contribute sustainable growth in the long run. The

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

empirical study recommends that for local leadership HGTC is preferable to RGTC. Moreover, it is relevant to understand that harmonic growth does not mean regional efficiency reduction.

To conclude, the HGTC will be created by falling social inequalities and increasing infrastructural background. The partial effects of RGTC ought to be harmonized to the partial effects of HGTC thus the micro-regions with RGTC will be able to maintain harmonic growth trajectory.

Acknowledgement

The research was financed by Pallas Athéne Geoploitical Foundation.

REFERENCES

Ács, J. Z.–Szerb, L. (2012): *Global entrepreneurship and development index 2012* Edward Elgar, Cheltenham.

Aroca, P.–Stimson, R.–Stough, R. (2014): Using a structural equation model to analyze potential determinants to spatial variations in endogenous regional growth performance In.: Kourtit, K.–Nijkamp, P.–Stimson, R. (eds.): *Applied regional growth and innovation models* pp. 335–360., Springer, Berlin–Heidelberg.

Bourdieu, P. (1983): The Forms of Capital In: Richardson, J. G.: *Handbook of Theory and Research for Sociology of Education* pp. 241-258., Greenwood, New York.

Brasili, C.–Saguatti, A.–Benni, F. (et.al.) (2012): *The impact of the economic crisis on the territorial capital Italian regions* http://www-sre.wu.ac.at/ersa/ersaconfs/ersa12/e120821aFinal00646.pdf (downloaded: 1. 08. 2015)

Brasili, C. (2010): *Local and urban development in the European Union* http://www-sre.wu.ac.at/ersa/ersaconfs/ersa11/e110830aFinal01382.pdf (downloaded: 1. 08. 2015)

Camagni, R.–Capello, R. (2015): Agglomeration economies and urban location benefits In.: Michelangeli, A. (ed.): *Quality of life in cities* pp. 143–162., Routledge, New York.

Camagni, R. (2011). Creativity, culture and urban milieux In.: Girard, L. F. (ed.): *Sustainable city and creativity* pp.183–199., Ashgate, Burlington.

Camagni, R.–Capello, R. (2015): Second-rank cities dynamics: theoretical interpretations behind their growth potentials *European Planning Studies* 23 (6): 1041–1053.

Capello, R.–Nijkamp, P. (eds.) (2009): *Regional growth and development theories in the XXI. century* Edwar Elgar, Cheltenham.

Capello, R. (2007): A forecasting territorial model of regional growth: the MASST model *The Annals of Regional Science* 41 (4): 753–787.

Coleman, J. S. (1988): Social capital in the creation of human capital *American Journal of Sociology* 94: S95–S120.

Jóna, Gy.–Hajnal, B. (2014): A magyarországi kistérségek területi tőkéjének alakulása *Területi Statisztika* 54 (2): 99–118.

Jóna, Gy. (2015): Determinants of Hungarian sub-regions' territorial capital European Spatial Research and Policy 22 (1): 101–119.

Kornai, J. (1971): *Anti-equilibrium* North-Holland Publishing Company, Amsterdam.

Kornai, J.–Dániel, Zs.–Jónás, A.–Martos, B. (1971): Plan-sounding *Economics of Planning* 11 (1–2): 31–58.

Kornai, J.–Simonovits, A. (1977): Decentralized control problems in Neumann-economies *Journal of Economic Theory* 14 (1): 44–67.

Kornai, J. (1972): *Rush versus harmonic growth* North-Holland Publishing Company, Amsterdam–London.

Kornai, J. (2006): *By Force of Thought. Irregular Memoirs of an Intellectual Journey* MIT Press, London–Cambridge.

Kourtit, K.–Nijkamp, P.–Stimson, R. (eds.) (2014): *Applied regional growth and innovation models* Springer, New York–London.

REGIONAL STATISTICS, 2015, VOL 5, No1: 121–136; DOI: 10.15196/RS05107

Kunzmann, K. R. (2007): *Medium-sized cities and creative industries. Cultural Identity, Cultural Mapping and Planning in the Oresund Region* http://www.kulturplan-oresund.dk/pdf/ aktivitetsplan/071017_ Potsdam_Creativity2.pdf (downloaded: 1. 08. 2015)

Maillat, D.–Kibir, L. (2001): The learning region and territorial production systems In.: Johansson, B.–Karlsson, C.–Stough, R. R. (eds.): *Theories of endogenous regional growth* pp. 255–277., Springer, Heidelberg.

Mihályi, P. (2011): A gazdasági növekedés gyorsításának esélyei Magyarországon 2030-ig *Competitio* 10 (2): 5–27.

Perucca, G. (2013): The role of territorial capital in local economic growth: evidence from Italy *European Planning Studies* 21 (4): 502–521.

Putman, R. D. (2000): *Bowling alone* Simon and Schuster, New York.

Russo, A. P.–Servillo, L. (2012): The attractiveness of regions and cities for residents and visitors www.espon.eu/export/sites/.../31_A_Russo.ppt (downloaded: 1. 10. 2012)

Stimson, R. J.–Stough, R. R. (2009): *Leadership, institutions and regional endogenous development* Edward Elgar, Cheltenham.

Tranos, E.–Nijkamp, P. (2013): A távolság halálának új vizsgálata *Tér és Társadalom* 27 (3): 3–27.

Veneri, P. (2011). *Territorial identity in Italian NUTS-3 regions* Universitá Politecnica delle Marche, Ancona. http://www.grupposervizioambiente.it/aisre/pendrive2011/pendrive/Paper/paper_vert_AN_june_2011.pdf (downloaded: 01. 08. 2015)

Vieira, M. R.–Tsotras, V. J. (2013): *Spatio-temporal databses complex motion pattern queries* Springer, New York–London.

Yuz, J. I.–Goodwin, G. C. (2014): *Sampled-data models for linear and nonlinear systems* Springer, London–Heidelberg.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

GÉZA TÓTH[a] – ÁRON KINCSES[b]

Accessibility Models Based On the Gravity Analogy: In Theory and Practice

Abstract

The most commonly applied types of accessibility models are based on the gravity analogy. In these models, researchers use different types of resistance factors, but they rarely give any elaborate explanation for their choice of a specific type of factor in their research. Another problem with this kind of analysis is that in many cases, the authors do not describe precisely how they determine the constants for a line of calculations in a given model. Thus, the results cannot be fully accepted since they cannot be reproduced by the reader. Finally, we consider it to be yet another huge problem that the results of the models are rarely compared to the real (for example traffic) parameters, therefore, it is also impossible to detect what would happen if the researcher used a different model. In this study, we tried to line up the most commonly used models, and by enlightening the resistance factors, examine their possible usage and their boundaries through exact Hungarian examples.

Keywords: accessibility models, gravity analogy, resistance factors.

Introduction

The models based on the gravity analogy are certainly the most commonly used type of models in the professional literature. (See for example: Stewart 1947, Hansen 1959, Ingram 1971, Vickerman 1974, Harris 1954, Huff 1963, Keeble et al. 1988, Dalvi–Martin 1976, Linneker–Spence 1991, Spence–Linneker 1994, Geertman–Ritsema van Eck 1995, Bruinsma–Rietveld 1998, Brunton–Richardson 1998, Kwan 1998, Levinson 1998, Smith–Gibb 1993, Gutiérrez 2001, Scheurer–Curtis 2007, etc.). These models attempt to take into consideration the impacts affecting the behaviour of the passenger. (Therefore, no matter which destination the passenger chooses, the possibility of any destination will be included in the model). One of the most common types of gravity analogy-based models is the potential model. The potential models – after forming disjoint, totally covered areal divisions – estimate the accessibility potentials of the area i., regarding the other areas, from which those with the smaller mass or those who fall further have less of an impact (Rich 1980, Geertman–van Eck 1995).

The goal of this study is to present the background of the calculation of the models' constants. With the resulting constants, other accessibility models can be calculated, and

a) Hungarian Central Statistical Office, H-1024 Budapest, Keleti Károly utca 5-7., Hungary. E-mail: geza.toth@ksh.hu
b) Hungarian Central Statistical Office, H-1024 Budapest, Keleti Károly utca 5-7., Hungary. E-mail: aron.kincses@ksh.hu

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

then the results can be compared to the real flow data. We would like to demonstrate what kind of methods are used when selecting a given model, and how much it can affect the results derived from them. In this study, depending on their conditions, we offer suggestions on the applicability of the resistance factors.

Material and methods

In this study, the definition of accessibility is: "Accessibility must refer to the role taken by the usage of an area and traffic system in a society that allows individuals and groups of individuals to participate in activities taking place in different locations." (Geurs–van Wee 2004, p128). In this case – according to the definition – accessibility should be measured as a data numericizing the social space, which shows the density of the spatial connections of traffic systems and usage of lands in some areas. For the calculations, the actual areas must be simplified and modelled. This paper tries to describe the real spatial interfering impacts using the data resulting from the models.

The general potential formula of the potential models is the following:

Equation #1:

$$A_i = \sum_j D_j \cdot F(c_{ij}), \qquad (1)$$

where A_i is the accessibility of the area; D_j is the mass of the area derived from i; c_{ij} is the general cost of transportation between i and j; $F(c_{ij})$ is the resistance factor (function).

The earliest antecedent of this study is the gravity model by Hansen (1959). Hansen (1959, p78) claimed that accessibility is "the generalization of the connections of the population, reaching across distances". The accessibility potentials of accessible destinations are strongly connected to the interaction of masses based on the gravity models.

The general formula of the Hansen-model:

Equation #2:

$$A_i = \sum_j \frac{W_j}{f(c_{ij};\beta)}, \qquad (2)$$

where A_1 is the accessibility of the area; W_1 is the mass accessible by the passengers, regardless of whether they actually want to access it or not; f (c_{ij}; ß) is the resistance factor (the general marking of the resistance factor suggests that ß can be part of an exponentiation as well as a multiplication, depending on the choice of the researcher); c_{ij} is a variable defining the cost of transportation between points i and j; and ß is a chosen constant.

The Hansen-model very much relied on the equations of gravity used in physics, which is also reflected in the fact that the constant in the equation, according to the physical deduction of the model, is bound to be a square number: ß=2. (See: Calvo–Pueyo Campos–Jover Yuste, 1992).

Equation #3:

$$A_i = \sum_j \frac{W_j}{c_{ij}^{\beta}}, \qquad (3)$$

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

The gravity and potential models are connected in many ways. What the two models have in common is that the size of a possible interaction between two settlements or locations is inversely proportional to the distance between them. Another similarity is that in the given settlement, any person (or other item) generates the same amount of interaction. Therefore, the size of the interaction between the two settlements is directly proportional to the size of the settlement, which is equivalent to chosen "mass". Thus with the growth of the mass of the examined items, the extent of the interaction grows as well.

The problem with choosing the constant (ß) occurred many times in previous scientific studies, since in sociological analogies, it is not essential to stick with the squared powers used in the laws of gravity in physics. If the constant is larger than one, that means the researcher emphasizes the distances more. There are some types of analysis though where they model such infrastructural methods (for example, intercity trains), which can operate at their best in medium-length distances, in which case, the scale of the constant is one. (Martín–Gutiérrez–Roman 1999, Capineri 1996). These types of models have the advantage of being easy to comprehend and calculate. The measurements that consider each potential destination reflect the impacts affecting the behaviour of the passengers. Another great advantage is that with these measurements, the accessible locations can be distinguished from each other, meaning that their different values can be taken into account for the calculations. However, these models have a disadvantage in that is they are unable to manage the passengers who have multiple traveling preferences. The results and the differences between the calculated numbers are also hard to interpret.

Choosing the Resistance Factor

The usage of the distance dependence in social geography is mainly induced by the fact that the insularity in the area impedes the cooperation of different areas; hence, it should be numericized. The easiest cases are naturally the ones with geographical distances. Regarding the data on accessibility, only the travel time, distance or cost of travelling by some means of transportation should be taken into account. The distance to be covered between two given points is called the resistance factor.

The most notable difference between the adaptation of the potential model of accessibility and potential model of physics is that, unlike in physics, in sociology space is not continuous, but "discrete". The social-economic forms (such as settlements or towns) tend to be concentrated in one distinctive point in space, and their "masses" are also concentrated in that spot. Since these "points of masses" do not fill the space, the potential value of a point – which obviously is altered by the impact of other points – in a limited part in the space (for example a country) could hardly be measured. (Tagai 2007). The mass points are concentrated in space to different extents; this induces different types of potential areas that might result in the distances between the points – and therefore the resistance factors – being represented by different functions. Very different resistance factors are required for dominance of small settlement structures versus large settlements, which affects the index exponent. Therefore, the equations of the resistance factor differ in the examinations of different areas, levels, or examinations of the same level, but not the same number of mass points.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Different forms of the resistance factor appear in the accessibility studies. In the case of models using boundaries, linear resistance factors are applied, or only destinations falling within a certain distance, time or cost are taken into account. Among models examining all accessible destinations and routes, the choice of the resistance varies considerably. The models distinguish the distances between the certain "masses" in their calculations. In the various approaches, the reciprocal or the exponentiated distance is applied. (See for example Hansen 1959, Davidson 1977, Fotheringham et. al. 2000, El-Geneidy–Levinson 2006.) The most commonly used types of models are the ones using a linear resistance factor (in the equation of the potential: in the denominator, the distance is raised to the first power). In these cases, no mathematical modifications are applied to the access time or the cost. However, this is not an unbreakable law, so in the models based on the gravity analogy, other exponents may occur. They numericize the probability of reaching the accessible destinations in the model. Essentially, researchers apply the exponential resistance factor for this purpose (Wilson 1971, Dalvi–Martin 1976, Martin–Dalvi 1976, Song 1996, Simma–Vritic–Axhausen 2001, Schürmann–Spiekermann–Wegener 1997, ESPON 2007, Papa–Coppola 2012). Other known models using resistance factors are the Gaussian ((Ingram 1971, Guy 1983) and the logistical ones (Bewley–Fiebig 1988, Hilbers–Veroen 1993). There are examples of other amendments to certain basic types, but this paper does not deal with them in detail (Reggiani, Bucci, Russo 2011).

Researchers divided the units of the access matrix (the units of the cost or duration of the passage between any examined parts of the area) into intervals (Simma–Axhausen 2003). What they found, is that the relationship between the frequencies and the average travel time/cost could best be described with an exponential regression function. These models are based on the presumption that within the examined area, with growth of the distance/travel time/cost, the probability of the frequency decreases exponentially, which is likely to affect the number of the potential destinations. According to this, the most subservient equation is:

Equation #4:

$$e^{-\beta c_{ij}}, \qquad (4)$$

where c_{ij} is the travelling expenses (time) between the points, ß is the constant. ß is the constant of the examined spatial layout, which should be determined separately for each spatial structure. The reason why this is necessary, is that the connection between the different levels and in the case of examining different targets, the frequencies, and the average travel time/costs can be represented by various types of functions. The role of this constant is to make the connection between the contribution of the potential of the specific parts of space and the whole space per se. (Later, the problem of choosing the constant in detail will be discussed.)

When studying certain spatial structures, in exponential regressional research, it is preferable to obtain an even more precise fit so we can determine the possibility of reaching certain destinations. To achieve this, it is vital to apply a Box-Cox (1964) transformation, which melds the residuals of the regression (making them homoscedastic), directing them to the normal distribution. Regarding the □$_i$ mistakes, there is the presumption that their expected value is 0 and their deviations in divided groups are equal. This is the so-called *homoscedastic* case. If the mistakes in the measurement change in line with the variable X

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

(homoscedastic case), the occurring huge deviations (squared) give a disproportionately inaccurate range, thus inaccurate parameters as well; therefore, the results of regressional or other models are not consistent with reality. When the homoscedastic condition is met, all the points of the regressional line are straight, or the hyperplane residuals with the same deviation are located.

Equation #5:

$$\mathrm{var}(\varepsilon_i) = \sigma^2 \quad \forall \quad i \in N^+ \text{-re} \qquad (5)$$

The Box-Cox transformation changes the values but not the order. For the usage of the Box-Cox resistance factor, the study by Willigers, Floor and van Wee (2007) gives an excellent example.

Ingram (1971) found that compared to the actual data, the values of some of the transformed resistance factors, distancing the origin, decrease too fast. He proposed the modified Gaussian resistance factor, which shows a slow decrease near the origin, and the extent of the decrease is smaller than that of the exponential and squared resistance factors. The smoothing nature of the Gaussian resistance function (it is convexo-concave) makes it apt to explore the spatial phenomena and – for instance – the migration of the population. (Grasland–Mathian–Vincent 2000) Equation #6:

$$f(d_{ij}) = 100 * e^{\frac{-d^2}{u}} \qquad (6)$$

In probability theory and statistics, the log-logistic distribution (in economics, "Fisk-distribution") shows the probability of a continuous, not negative variable. It is used in areas where the probability of the variable is high initially, but then gradually decreases. The logistic distribution is a probability distribution of a random variable, the logarithm of which has a logistic distribution. The log-logistic models are based on the logistic distribution. The log-logistic distribution near the expected value is symmetric, although it can be described with a larger deviation than the log-normal one (since its expected value is calculated).

Equation #7:

$$f(d_{ij}) = 1 + e^{a + b * \ln d} \qquad (7)$$

The Self-potential

The professional publications on potential models have been studying the notion of "self-potential" (see among others Frost–Spence 1995, Bruinsma–Rietveld 1998). The significance of this is that, within examined space, the extent of the place-dependent potential is not only defined by the sizes and distances of the masses around it, but also depends on how big of a force field it can generate around itself. In the studies of potential, it is vital to distinguish inner from outer potential (Nemes Nagy 1998, 2005). This severance comes from the distinction of the specifically examined area and the force of the space, outside that area, affecting it. So the potential of a given point is calculated from the sum, the self-, inner and outer potentials.

Self-potential should also be taken into consideration in the cases of accessibility studies. When calculating the self-potential of an area, it is presumed that the transportation

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

between two areas is not the only condition increasing accessibility, but also the transportation within areas or settlement. Thus, we conclude that a certain product or service does not necessarily have to be transported to another area if it can be sold within the given area. Neglecting the role of the self-potential might be misleading – for example – when examining the level of settlements. Naturally, in these cases, the accessibility of central settlements of the agglomeration or group of settlements would be much lower than the other settlements of the agglomeration.

When defining the self-potential – like when examining other types of potential – the calculations are derived from the size of the given area (preferably, not the administrative size, but the "clear" one). According to the generally used methods, the area is regarded as a circle, and then the radius for a given unit that is considered proportionate to the distances on roads within certain settlements is calculated; thus, it is also called "self-distance". This type of distance is used in models calculating geographical distances, whereas with those applying distances based on the road network, the distance can be converted by an average velocity/cost formula, which can then be placed in the equation. The difference between the various methods used for calculating the self-potential is that to what extent and in what way the radius is weighted (Tagai 2007). In this way, by examining it from different aspects, the role of the centre point can be further highlighted, or it can be better fitted to the surface of potential.

Tagai's study claims that there are some researchers who use distances equivalent to the radius since its length is close to the average distance within the area.

The most commonly used method is calculating a third of the radius as the self-potential. This is also connected to the probability of the mass distribution. Others, however, came to different conclusions after similar calculations; they believe two-thirds of the radius is the most suitable as the self-derived approximate value of distance.

Modifying the radius in any way is arbitrary, and consequently, hard to back up objectively. Therefore, in this study, it is always the whole radius that is considered when calculating the self-potential. Naturally, if actual data about the settlement is available, the various assumptions can be compared to the facts; although, in this case, it is also true that some models fit some settlements better.

The Total Potential

In the accessibility studies, usually the examined space is chosen so that it is larger than the narrow examination area, and in this way, the impact of the outer potential can be ignored. In this case, the space-dependent accessibility potential is calculated from the sum of the inner and self-potential:

Equation #8:

$$\sum A = SA + BA \qquad (8)$$

where ΣA_i is the total potential of area i, SA_i is the own potential of the spatial unit, and BA_i is the inner potential within the area analysed. There are some theories that take into account the potential outside the area, the "outer potential", however, within this study, on the level of smaller areas, the outer potential cannot be calculated due to methodological reasons, and hence they are not included. Although this affects the result of the examination

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

to some extent, it does not prevent us from drawing up basic connections. The effect can mostly be noticed on the peripheries; the work did not deal with the models that take the competition into account (Weibull 1976, Knox 1978, Van Wee Annema Hagoort, 2001, Joseph Bantock 1982, Fotheringham 1982).

Calculation

In the professional literature on accessibility, the method for calculating the constant is not often discussed. Thus, readers can only interpret the results or recreate the research with considerable difficulty. The next segment covers the theoretical background determining these constants in detail and the way they are calculated.

By definition, accessibility numericizes the possibility of approaching a location from the aspects of the passengers, regarding their (household, business) options and goals; furthermore, the services available at the desired destination, and relationships connected to the space. The models based on the gravity analogy – in our opinion – conclude from some elements of the definition that the number of the potential journeys from point A to point B is determined by four factors: the desired destination's mass and distance; the examined area's spatial structure and the role of random probability. Three out of these four factors can be modelled, whereas conclusions can be drawn from the forth, but they are hard to predict.

It would not be true to state that the number of journeys between two points half as far away from each other is twice as many; or – in line with that – a point three times the mass is visited three times as many times. Yet the modelling can be done. The goal is to numericize the probability of the potential journey. In this case, the potential is the function of three factors: the desired destination's mass, distance and spatial structure. The first two factors are specific to each potential model; however, in the case of models using a constant, the latter should be described with a factor. With the functions used (exponential, exponential modified by the Box-Cox transformation, Gauss-function), the role of the spatial structure has different impacts on the value of the potential; they infiltrate in different ways depending on the geographical distance. The constants of the mathematical functions in the resistance factors are defined by the frequency of the accessibility distances appearing in the examined space (Simma–Axhausen 2003 p. 184).

The paper will now focus on the exact calculations. The first unit in the potential's formula is the self-potential while the rest is the inner potential's impact on the self-potential. The formula can be applied to any optional area – for example a settlement, a county, or a region. The following equation determines the value of the constant ß.

The value of the place-dependent potential in point j of the space:

Equation #9:

$$A_i = W_i \cdot e^{-\beta \cdot c_{ii}} + \sum_{i \neq j} W_j \cdot e^{-\beta \cdot c_{ij}} = \frac{W_i}{e^{\beta \cdot c_{ii}}} + \sum_{i \neq j} \frac{W_j}{e^{\beta \cdot c_{ij}}} \tag{9}$$

where A_1 is the accessibility of the area; W_i and W_j are the masses – in this case, population – belonging to the appropriate level; c_{ij} is the time needed to get to area i from area j, on the

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

road network, measured in minutes. Lastly, ß is the constant of the examined spatial structure, which should be determined separately in each spatial structure.

It should be noted that the current definition of the potential implies there is a linear superposition between the units, thus there is no interaction between the impacts; they do not enhance or lessen each other, they simply add up. Looking for analogies, the gravitational, magnetic or electric fields are much similar. The definition does not take into account the interference described in string theory as in quantum physics.

The potential journeys on the road network between all possible pairs of all the 175 districts are investigated as follows.

The data has been laid out in a 175*175 matrix, aligned with the travel times (in minutes). The values are divided into intervals, and avoid placing too few occurrences in one interval, yet forming a sufficient number of intervals, because that would hinder the examination of the distribution of the values. In the study, the access times were divided into 50 equal intervals.

Take the set containing intervals that contain the durations of the journeys between the districts in minutes:

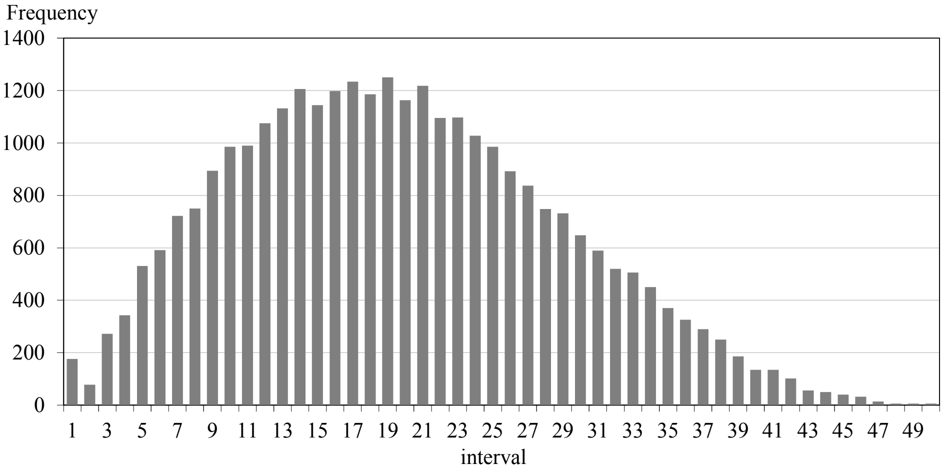
$$\forall \chi \in \Phi : \chi \in (0;429,17).$$

The 429.17 minutes result is the largest unit of the 175*175 matrix; thus, this is the longest distance between points of the district on the road network, measured in minutes. In this examination, theoretical access times were used for the calculations, meaning that the length of time was only affected by the speed limits and not any other conditions.

Divide the interval into 50 equal pieces. Interval number "i" contains the times (i*8.58; (i+1)*8.58)); where i = 1, 2, …50, thus dividing the maximum mid-district distances of 429 minutes into 50 equal pieces results in 8.58-minute-long intervals.

Figure 1

The frequency of the accessibility time



Source: own compilation.

A function-like connection should be found between the frequency of the intervals and the average travel times. The function is needed for modelling the frequency of journeys with the increase in the distances, which technically translates to the probability of the journey. Theoretically, the goal is to find the function most suitable for the frequency since that would result in the most accurate potential model. However, human behaviour (in this case, travelling) can barely be modelled – as can be seen in many cases of models based on a physical analogy. It is not certain at all that the most suitable model will give the most accurate results. Moreover, it may occur that the frequency of the intervals fit the given model very poorly, yet the model using the constant calculated from it gives the closest to the actual results.

At first glance, the polynomial approach may look like the obvious choice, but it would make the results difficult to interpret. Applying it would be necessary due to the fluctuation shown by the frequency. Naturally, this could vary in each sample. Further problems may occur when trying to interpret the rate of the frequency together with the increase of the distances. Probably the most obvious case is that of the exponential function since here the frequency decreases exponentially as the distances grow. As a result, the version that finds the exponential connection between the travel times and the mean values of the intervals is reviewed first.

The Exponential Resistance Factor

Equation #10:

$$\nu_l \approx e^{-\bar{c}_l} \quad l=1,2,\ldots,50, \qquad (10)$$

We assume that the duration of the journeys among the districts and the mean values of the intervals in minutes are proportional to each other. The constant ß makes the exact connection between the average length of time and the frequency.

Equation #11:

$$\nu_l = e^{-\beta \bar{c}_l} \quad l=1,2,\ldots,50, \qquad (11)$$

where v is the frequencies and stands for the average duration. So the presence of the exponential connection mentioned above should be checked for each examination; the value of the constant should also be calculated each time.

The equation (11) implies a regressional relationship, and we are looking for the constant ß that best fits the equivalences overall. After rearranging the equation, we get:

Equation #12:

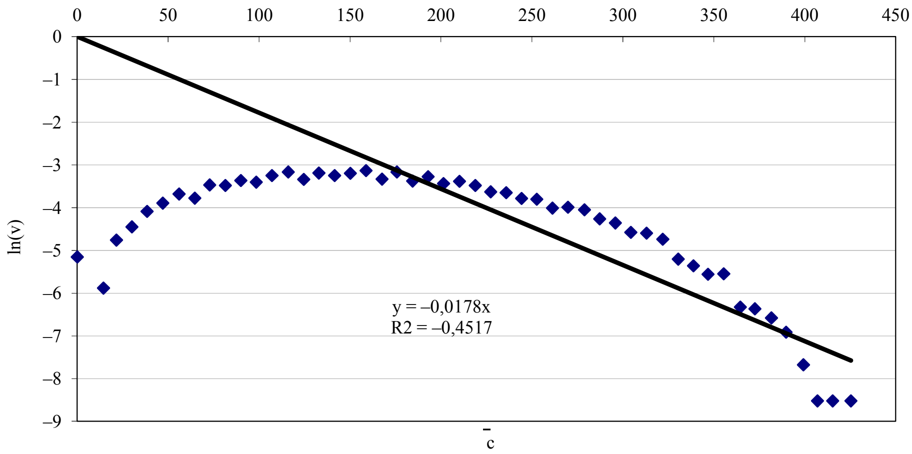
$$\ln \nu_l = -\beta c_l \qquad (12)$$

Depicting the e (a natural number) based logarithm of the frequency in the function of the average travel times, the value of ß can be determined by linear regression. In the study, the normalizing criterion of the joint line crossing the origin is required. Thus, the normalized frequency belonging to the average time of 0 is 1. The calculated value of ß is 0.0178, with 45.17% reliability.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Figure 2

The n based logarithm of frequency with the function of average durations



Source: own compilation.

The method described above is just the first, approximate solution. This schema ignores the statistical errors occurring during the division-based access times, since, during this process, only the outer multiplications and subtractions among the groups were taken into account. See the following:

Take $c_{ij,}$ (in this case ij does not mean the duration of travel between i and j!) as the value number "j" from the interval number "i". (i=1, 2, 3, …, 50; j=1, 2, …, pi). Now any optional element of the set Φ can be described as such:

Equation #13:

$$c_{ij} = \bar{c}_{i.} + e_{ij} \qquad (13)$$

where $\bar{c}_{i.}$ is the average of the interval i;

$\bar{c}_{i.} = \frac{1}{p_i}\sum_{j=1}^{p_i} c_{ij}$, and e_{ij} stands for the errors and residues of the groups.

Take a fully squared value:

Equation #14–15:

$$Q^2_{all} = \sum_{i=1}^{l}\sum_{j=1}^{p_i}(c_{ij} - \bar{c}_{..})^2 = \sum_{i=1}^{l}\sum_{j=1}^{p_i}(c_{ij} - \bar{c}_{i.} + \bar{c}_{i.} - \bar{c}_{..})^2 =$$

$$\sum_{i=1}^{l}\sum_{j=1}^{p_i}(c_{ij} - \bar{c}_{i.})^2 + \sum_{i=1}^{l}\sum_{j=1}^{p_i}(\bar{c}_{i.} - \bar{c}_{..})^2 + \qquad (14)$$

$$2\sum_{i=1}^{l}\sum_{j=1}^{p_i}(c_{ij} - \bar{c}_{i.})(\bar{c}_{i.} - \bar{c}_{..})$$

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Applying $2\sum_{i=1}^{l}\sum_{j=1}^{p_i}(c_{ij}-\bar{c}_{i.})(\bar{c}_{i.}-\bar{c}_{..})=0$, since $\sum_{j=1}^{p_i}(c_{ij}-\bar{c}_{i.})=\sum_{j=1}^{p_i}c_{ij}-p_i\bar{c}_{i.}=0$; (15)

Therefore $Q^2_{all} = Q^2_{inner} + Q^2_{outer}$.

And where

Equation #16-18:

$$Q^2_{inner} = \sum_{i=1}^{l}\sum_{j=1}^{p_i}(c_{ij}-\bar{c}_{i.})^2; \quad (16)$$

$$Q^2_{outer} = \sum_{i=1}^{l}\sum_{j=1}^{p_i}(\bar{c}_{i.}-\bar{c}_{..})^2; \quad (17)$$

$$\bar{c}_{..} = \frac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{p_i}c_{ij} \quad (18)$$

This applies to its deviation as well:

$$\sigma^2_{all} = \sigma^2_{inner} + \sigma^2_{outer} \quad (19)$$

where σ^2_{inner} is the deviation within the intervals squared; σ^2_{outer} is the deviation in between the intervals squared.

Therefore, the phenomenon described the way it was defined in the first part of the study is only accurate if e_{ij} of the equation $c_{ij} = \bar{c}_{i.} + e_{ij}$ has the following features:

1. Their expected value in every interval is zero.
2. Within the intervals, they have the same deviation. In other words, they are homoscedastic.

If the conditions are not met, the results will differ depending on the classification, thus will not be reliable enough to base realistic predictions on them. In the case regarding the districts – described above – we find that the residues are usually large, and the two requirements are only approximately fulfilled. Using the first method would not be a huge mistake by any means, but there is a method in statistics to eliminate even that little flaw. Namely, it is the Box-Cox transformation, which secures the random occurrence of residues. With this method, the data that best meets the two conditions above and fits a given spatial layout can be calculated.

The Box-Cox Resistance Factor

Two forms of the Box-Cox transformation are known.

Equation #20-21:

1. $$c_{ij}^{transformed} = \begin{cases} \frac{c_{ij}^{\lambda}-1}{\lambda}, \lambda \neq 0 \\ \ln(c_{ij}), \lambda = 0 \end{cases} \quad (20)$$

2. $$c_{ij}^{transformed} = c_{ij}^{\lambda} \quad (21)$$

Because of the rearrangement, C_{ij} must be larger than 0; this condition has to be met by every access time measured in minutes.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

With the restriction of λ>0, it can be achieved in both cases that the transformation is relation invariant, meaning that this transformation might change the values, but not the order. Both definitions lead to the same conclusion. We chose the latter. Therefore, we are looking for the value of λ that best suits the random distribution of differences. For the investigations, we used the software SAS 8.2, more specifically, the proc transreg. This program calculates the transformed duration times with all possible values of λ and very little class differences, and the log-likelihood functions belonging to these intervals. The estimate of the maximum log-likelihood function is technically an estimate of a point in which we consider those parameters that have the maximum combined partition function of the exanimation vector. The log-likelihood function:

Equation #22:

$$\log L(c_{ij}, f_k, \lambda) = F(c_{ij}, \lambda) \cdot F(f_k, \lambda) \qquad (22)$$

Estimating the maximum log-likelihood function has the advantage of being asymptotically effective, furthermore, if it cannot be given in closed-form, it can be examined by numeric maximization (this applies to this case). In other words, the population parameters can be estimated by the value where the log-likelihood function reaches its maximum; thus, the most probable outcome is a realized sample when taking a sample.

In the district example, λ = 1.07085 was calculated for λ belonging to the maximum output of the function. This is how the transformation constant was arrived at, which, by transforming the access times, makes the residues of the groups the most independent. Overall, with this method, the information resulting from the group formations with transformed access times, is more accurate.

In the case of a transformed variable, if the boundaries of the 50 intervals are transformed in the same way, we get an identical exponential curve. In this way, we can redefine the average travel times of the groups, and determine the value of ß by regression, based on the previous, exponential calculation. In our example, regarding already transformed pairs of travel times, β = 0.0119. Therefore, we can conclude that constant ß can be determined. This has to be done separately for each examination or spatial distribution. For the first estimate, we find that β=0.0178, while after further investigation, it was proven that the Box-Cox transformation carried out with the power of λ = 1.0785, allows a more accurate analysis. In this case, β = 0.0119.

The Gaussian Resistance Factor

According to figure 3, when analysing the distribution of frequencies, the following relationship between the frequencies and average access times can be presumed:

Equation #23:

$$\nu \approx w * e^{\frac{-c_i^2}{u}} \Rightarrow$$

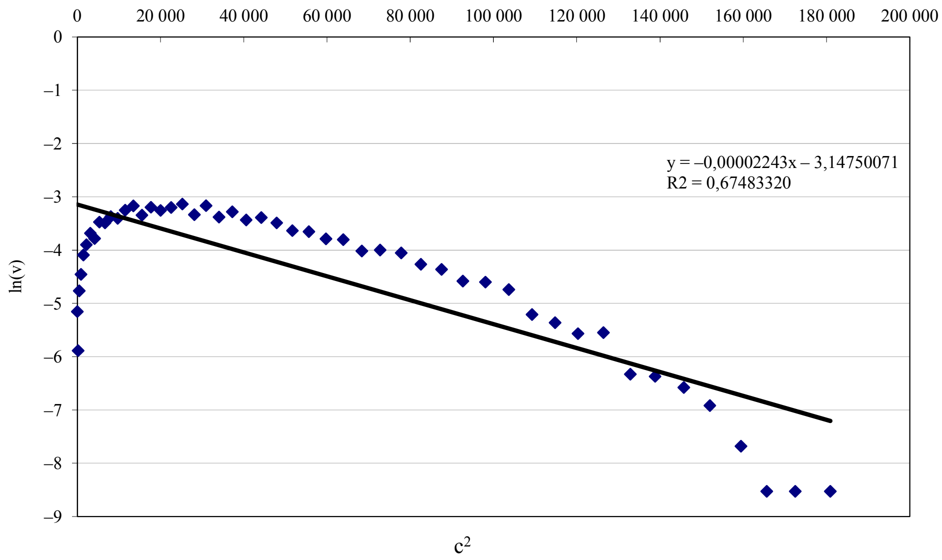
$$\ln(\nu) \approx \ln(w) + \frac{1}{u}(-c_i^2) \qquad (23)$$

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

where v is the frequencies, c is the average access times, and w and u are the constants. The goal is to calculate the constants that best fit the existing data. For this purpose, we represented the logarithm of the frequency of the elements belonging to the 50 intervals formed based on the access times between the capitals of the districts in the function of the average durations of the squared intervals.

Figure 3

The logarithm of frequencies and the average durations of the intervals squared



Source: own compilation.

From which w = 0.04295936, u = 44583.1476; with 0.67 accuracy.

The Gauss model better fits that in an exponential function.

The Log-logistic Resistance Factor

In this case, we predict the following relationship between the frequency and the average duration:

Equation #24:

$$\nu \approx 1 + e^{a + b \ln c_i} \Rightarrow \qquad (24)$$

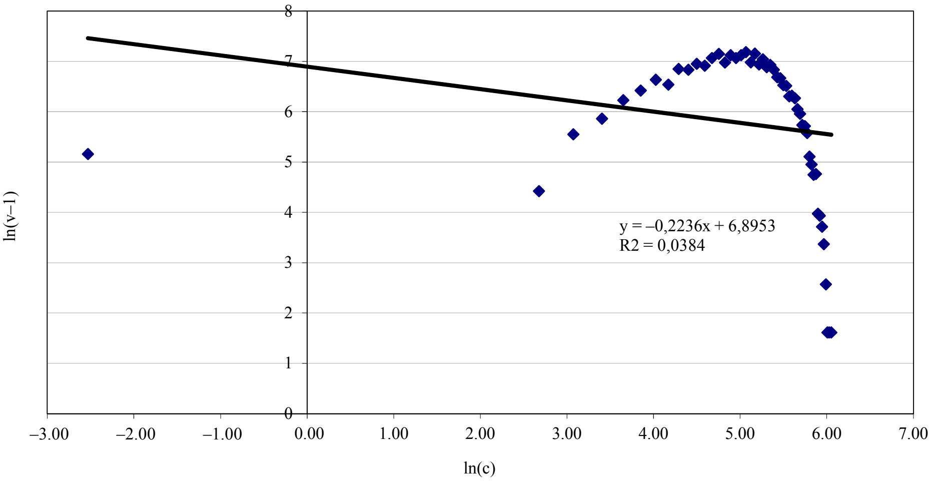
$$\ln(\nu - 1) \approx a + b \ln c_i$$

where v stands for the frequencies, c stands for the average durations; a and b are constants.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Figure 4

The relationship between the frequency and the average duration in a log-logistic case



Source: own compilation.

From which: b = –0.2236, a = 6.8953.

It should be noted that R^2 is smallest in this example, yet this method – as can be seen later – may lead to accurate results.

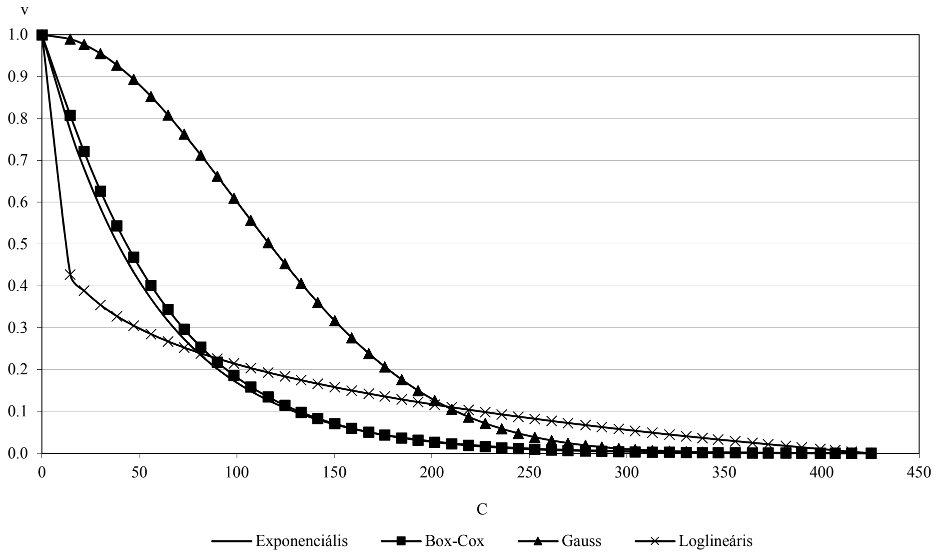
For comparing the different types of resistance factors, we redid these calculations. The recalculation of the Gaussian and log-logistic model was needed to fit the others, meaning that in case of c = 0, the frequency is 1 (or 100%). (According to these calculations, R^2 is decreasing; therefore, the data resulting from them cannot be used further, and are only represented in figure 5.)

By comparing the various resistance factors, the differences between them can be studied. The access times are measured in minutes with the distances constructed according to the specific models. Noticeably, the Gaussian model best works for the medium distances, the log-logistic approach for the very large or very small while the exponential method for the smaller distances. Theoretically, according to this, for the studies within settlements or districts, the log-logistic approach should used; for country-level studies the exponential; for a European study the Gaussian, and for a global examination once again the log-logistic model. Yet – as will be shown later – it is not certain that these approaches provide the most accurate results in the mentioned areas.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Figure 5

The relationship between the resistance factors of the models based on the gravity analogy



Source: own compilation.

Results

The study has so far focused on the structures of the different models, ignoring the question of how realistically the potential structure formed by basic elements and relationships describes the space (based on geometric interpretations). We have not yet dealt with how comparable the volume of traffic on the public roads is with the data derived from the models. So can the models′ conclusions be applied to the real social space? During the research, we only found studies in the professional literature that compare the results of different models to each other (see De Montis–Caschili–Chessa 2011); we could not find one that contrasts them with the traffic. Thus, this essay is unique in the professional literature.

The traffic data measured by the Hungarian Public Road Not-for-profit Inc. keeps track of the annual average daily traffic (AADT) for the cross-section of certain roads. (The latest nationwide traffic measuring included the cross sections of over 4500 roads.) These measurements are carried out by taking samples. This method allows researchers to calculate the average daily traffic – if they are aware of the time-specific fluctuation of traffic – from relatively little data (smaller sample, briefer counting) with sufficient accuracy and reliability.

The nationwide cross-section measurements take place at a large number of junctions and deviations throughout the year, on five different occasions, from 6 am to 6 pm.

Most of the annual measurements last for three days. A rotating method is used so that annually, measurements are only carried out in about 20% of the country, and then the data from the previous years are modified to fit the new results.

From the measured data, by simple averaging, and then multiplying by factors bearing the time-specific characteristics of the traffic at a given time of the day (a_x), a given day (b_i) or month (c_i), the annual average daily traffic can be calculated.

Equation #25:

$$ÁNF = \frac{1}{n} * \sum_{i=1}^{n} g_x * a_x * b_i * c_i \qquad (25)$$

where n is the number of the days counted; g_x is the traffic measured within the x hours; a_x is the time-of-day factor (the volume of traffic at a specific time of day in proportion to the whole 24 hours); b_i is the daily factor (a multiplier for the given day of the week that modifies the daily traffic to the monthly average); c_i is the monthly factor (a multiplier for a given month of the year to give the yearly average).

The results of the cross-section traffic count are expanded to the "validity sections" according to a convention of the experts. The district-level AADT data was provided by the Hungarian Public Road Not-for-profit Inc.

Table 1

The dimensions of the study

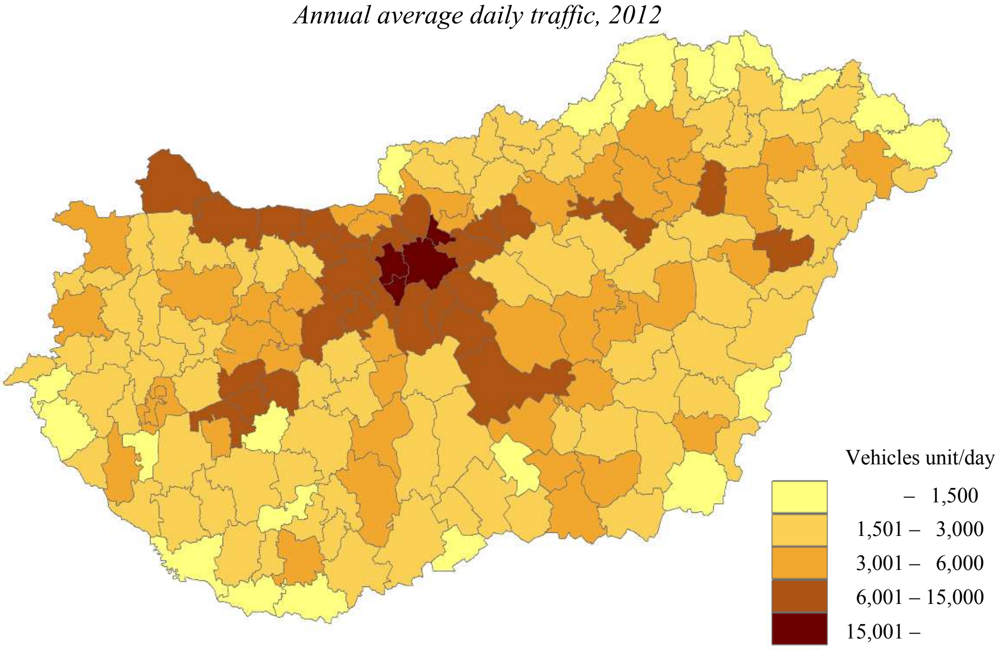
Dimension	Notes
Source	In the study, the accessibility is calculated from the aspect of each individual; also, we interpret but do not distinguish certain social groups or the varying destinations of the different passengers.
Goal	We numericize the aspired destinations by its population and income.
Resistance	In this case, the areal resistance factor is the theoretical access time between given district capitals on the public road network, measured in minutes. The resistance factor applied can be linear, squared, exponential, Box-Coy, Gaussian, or log-logistic.
Restrictions	When travelling on the roads between two districts, the maximum achievable speed on that specific type of road is the speed limit.
Boundaries	Hungary′s borders mean the boundaries of the examined area. Although it is an undeniable fact that the Hungarian potentials are affected by destinations outside the country since there was no sufficiently detailed map of the road network of other countries available, we had to neglect their impact.
Means of transportation	We did not distinguish personal or freight traffic for this study.
Modus	The calculations regarding public roads were based on unimodal accessibility.
Areal level	The study mainly focuses on the district level, LAU 1.
Diversity	The study′s main goal is to model the differences in accessibility in Hungary.
Dynamics	In the study, the population, income, and public network as of January 1st: 2004, 2008 and 2012, is included.

Source: own compilation.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Figure 6

Annual average daily traffic, 2012



Source: own compilation.

We compared 2004, 2008 and 2012 data with various potential models. As a mass factor, both the income and population was included in the calculations. (For details of these models see: Tóth–Kincses 2007). The dimensions of the study can be seen in Table #1.

Models Based on the Gravity Analogy

General gravity analogy

The examined models are the following:

Equation #26-31:

$$c_1 = \frac{W_i}{c_{ii}} + \sum_j \frac{W_j}{c_{ij}} \qquad (26)$$

$$c_2 = \frac{W_i}{c_{ii}^2} + \sum_j \frac{W_j}{c_{ij}^2} \qquad (27)$$

$$c_3 = \frac{W_i}{e^{\beta c_{ii}}} + \sum_j \frac{W_j}{e^{\beta c_{ij}}} \qquad (28)$$

$$c_4 = \frac{W_i}{e^{\beta\left(\frac{c_{ii}^{\lambda}-1}{\lambda}\right)}} + \sum_j \frac{W_j}{e^{\beta\left(\frac{c_{ij}^{\lambda}-1}{\lambda}\right)}} \qquad (29)$$

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

$$c_5 = \frac{W_i}{p^* e^{\frac{-c_{ii}^2}{u}}} + \sum_j \frac{W_j}{p^* e^{\frac{-c_{ij}^2}{u}}} \quad (30)$$

$$c_6 = \frac{W_i}{1+e^{a+b \ln c_{ii}}} + \sum_j \frac{W_j}{1+e^{a+b \ln c_{ij}}} \quad (31)$$

where C_{1-6} is the accessibility of area i, W_i is the mass of the district, W_j is the "mass" of the desired destination, c_{ij} and c_{ii} are the access times, while β, λ, p, u and a, b are constants. (Equation 26 is using the linear model, 27 is using the squared, 28 the exponential, 29 the Box-Cox, 30 the Gaussian and 31 the log-logistic model.)

As Tables 2 and 3 show us, the traffic data can be better fitted together with the income numbers rather than with the population numbers, although the difference is not significant. The experiments on a district level suggest that the most suitable potential models are those using the log-logistic factor; it should be noted that if other types of division had been applied, the results would have been the same. While calculating with the log-logistic resistance factor, the parity of the duration frequencies was quite weak. Despite this, the approach provided the most accurate model. This demonstrates that it is appropriate to use more than one type of model for the calculations; we cannot draw conclusions only from parity.

The reason the log-logistic function is favourable here is that it is the one that can estimate the differences of the areal traffic, with the help of the explanatory variable. In case of the areal differences of traffic in Hungary, the most important aspect is to find the function that can not only take the basic areal differences into account, but is able to estimate, with the smallest residue possible, the traffic value of the capital city, Budapest. If Budapest′s impact were not so outstanding or if there was a district with a similarly large volume of traffic in Hungary, the calculations carried out with log-logistic function might not give the most accurate results.

Table 2

The parity of the models based on the gravity analogy using population masses, compared to the district AADT statistics (R^2)

Years	c_1	c_2	c_3	c_4	c_5	c_6
2004	0.43	0.26	0.55	0.52	0.19	0.63
2008	0.45	0.45	0.56	0.52	0.13	0.69
2012	0.58	0.44	0.61	0.58	0.21	0.69

Table 3

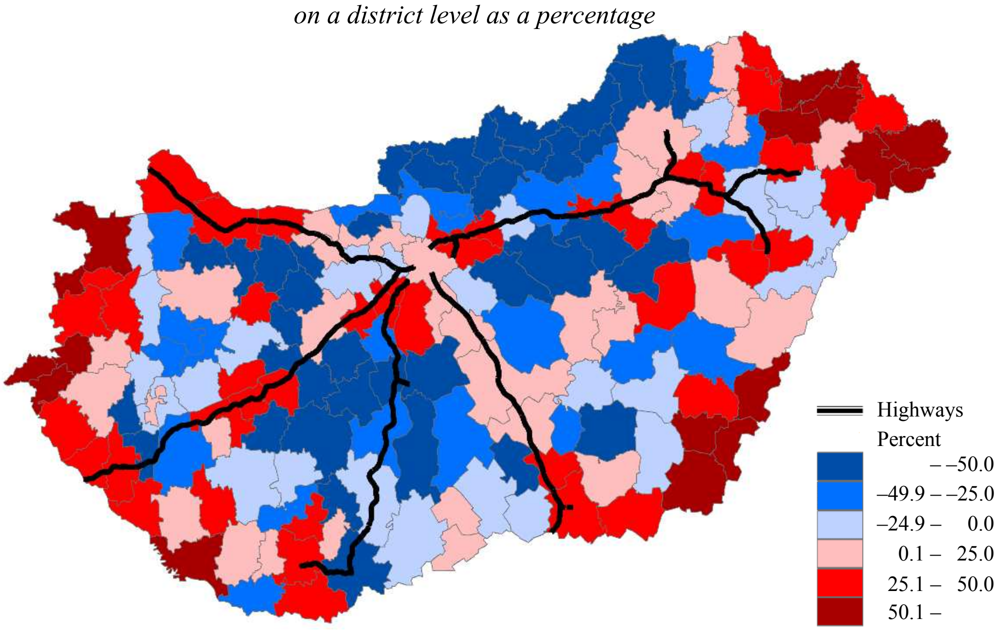
The parity of the models based on the gravity analogy using income masses, compared to the district AADT statistics (R^2)

Years	c_1	c_2	c_3	c_4	c_5	c_6
2004	0.42	0.24	0.56	0.53	0.18	0.73
2008	0.46	0.45	0.58	0.55	0.11	0.72
2012	0.58	0.45	0.60	0.57	0.22	0.69

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Figure 7

The 2012 average daily traffic; the potential model with the largest parity (c_6 – with log-logistic resistance factor); the difference of the 2012 estimated traffic on a district level as a percentage



Source: own compilation.

In one-half of the districts, mainly the busiest ones that are near a highway, the traffic estimated by the models is less than the actual (these are marked red on the map). In the rest of the districts, located in the middle of the country, the volume is usually underestimated while districts near the border are once again overestimated.

Results and Discussion

The methodological base for applying different types of resistance factor is to numericize the role of the spatial structure in the model. The possibility of the journeys – the potential – is determined by the mass of the desired destination, the distance and random chance. In this case, "spatial structure" refers to the size of the road network in the analysis, by the disruption of which, the distances/travel costs are modified with various functions.

The goal of choosing one resistance factor from the several available (squared, Box-Cox, Gaussian, exponential, log-logistic) is to differentiate the various relationships by the probability to be chosen. Out of the four factors of travel potential, the travel relationship is only one; the remaining three factors have further impacts on the probability of the choice. The spatial relationship can hardly be described mathematically. The choice of the resistance factor cannot be generalized according to the area or the extent of the analysis. It is essential – if it is all possible – to compare the calculations carried out with different resistance factors to find the most suitable model for the analysis.

* * *

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Acknowledgement

This paper is supported by the **János Bolyai Research Scholarship** of the Hungarian Academy of Sciences.

REFERENCES

Bewley, R.–Fiebig, D. G. (1988): A flexible logistic growth model with applications to telecommunications *International Journal of Forecasting* 4 (2): 177–192.

Box, G. E. P.–Cox, D. R. (1964): An analysis of transformations *Journal of Royal Statistical Society, Series B* 26 (2): 211–246.

Bruinsma, F.R.–Rietveld, P. (1998): The Accessibility of European Cities: Theoretical Framework and Comparison of Approaches *Environment and Planning A* 30 (3): 499–521.

Brunton, P.J.–Richardson, A.J. (1998): A Cautionary Note on Zonal Aggregation and Accessibility. *Paper presented at the 77th Annual Meeting of Transportation Research Board*, Washington, DC.

Calvo, P.–Pueyo Campos, A.–Jover Yuste, J. M. (1992): *Potenciales demográficos de España* Atlas Nacional de España. Tomo 14-b. Istituto Geográfico Nacional, Madrid.

Capineri, C. (1996): From Networks to Regional Development: Representations of Italian Regional Disparities. *Paper Presented at the Nectar Euroconference*, 24–28 September 1996., Mons, Belgium.

Dalvi, M. Q.–Martin, K. M. (1976): The measurement of accessibility: some preliminary result. *Transportation* 5 (1): 17–42.

Davidson, K. B. (1977): Accessibility in transport/land-use modelling and assessment. *Environment and Planning A*. 9 (12): 1401–1416.

De Montis, A.–Caschili, S.–Chessa, A. (2011): Spatial Complex Network Analysis and Accessibility Indicators: the Case of Municipal Commuting in Sardinia, Italy *European Journal of Transport and Infrastructure Research* 11 (4): 405–419.

El-Geneidy, A. M.–Levinson, D. M. (2006): *Access to Destinations: Development of Accessibility Measures*. Department of Civil Engineering University of Minnesota, Minneapolis.

ESPON (2007): *Update of Selected Potential Accessibility Indicators*. Final Report. Luxembourg.

Fotheringham, A. S.–Brunsdon, C.–Charlton, M. (2000): *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage, London.

Fotheringham, A. S. (1982): A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A*. 15 (1): 15–36.

Frost, M.E.–Spence, N.A. (1995): The Rediscovery of Accessibility and Economic Potential: The Critical Issue of Self-potential. *Environment and Planning A*. 27 (11): 1833–1848.

Geertman, S.C.M.–van Eck, J.R.R. (1995): GIS and models of accessibility potential: an application in planning. International *Journal of Geographical Information Systems* 9 (1): 67–80.

Geurs, KT.–van Wee, B. (2004): Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography* 12 (2): 127–140.

Grasland, C.–Mathian, H.–Vincent, J. (2000): Multiscalar analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences. *Statistical Journal of the United Nations Economic Commission for Europe* 17 (2): 157–188.

Guy, C. M. (1983): The assessment of access to local shopping opportunities: a comparison of accessibility measures. *Environment and Planning B: Planning and Design*. 10 (2): 219–238.

Hansen, W.G. (1959): How Accessibility Shapes Land-Use. *Journal of the American Institute of Planners*. 25 (2): 73–76.

Harris, C. D. (1954): The market as a factor in the localisation of industry in the United States. *Annals of the Association of American Geographers* 44 (4): 315–348.

Hilbers, H. D.–Veroen, E. J. (1993): *Het beoordelen van de bereikbaarheid van lokaties. Definiëring, maatstaven, toepassingen beleidsimplicaties*. INRO-VVG 1993–09, TNO Inro, Delft.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

Ingram, D. R. (1971): The Concept of Accessibility: A Search for an Operational Form. *Regional Studies* 5 (2): 101–105.

Joseph, A.E.–Bantock, P.R. (1982): Measuring potential physical accessibility to general practitioners in rural areas: a method and case study. *Social Science and Medicine* 16 (1): 85–90.

Keeble, D.–Offord, J.–Walker, S. (1988): *Peripheral Regions in a Community of Twelve Member States*, Commission of the European Community, Luxembourg.

Knox, P.L. (1978): The intraurban ecology of primary medical care: patterns of accessibility and their policy implications. *Environment and Planning A* 10 (4): 415–435.

Kwan, M. P. (1998): Space–time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical Analysis* 30 (3): 191–216.

Levinson, D.M. (1998): Accessibility and the Journey to Work. *Journal of Transport Geography* 6 (1): 11–21.

Linneker, B.J.–Spence, N.A. (1991): An Accessibility Analysis of the Impact of the M25 London Orbital Motorway on Britain. *Regional Studies* 26 (1): 31–47.

Martín, J.C.–Gutiérrez, J.–Román, C. (2004): Data Envelopment Análisis (DEA) Index to measure the accessibility impacts of new infrastructure investments: the case of the High-Speed Train Corridor Madrid–Barcelona–French border. *Regional Studies* 38 (6): 697–712.

Martin, K. M.–Dalvi, M. Q. (1976): The comparison of accessibility by public and private transport. *Traffic Engineering and Control* 17 (12): 509–513.

Nemes Nagy, J. (1998): A fekvés szerepe a regionális tagoltságban. In: *Munkaerőpiac és regionalitás*, pp. 147–165., MTA KK KI, Budapest.

Nemes Nagy, J. (ed) (2005): *Regionális elemzési módszerek*. Regionális Tudományi Tanulmányok 11. ELTE Regionális Földrajzi Tanszék–MTA–ELTE Regionális Tudományi Kutatócsoport, Budapest.

Papa, E.–Coppola, P. (2012): Gravity-Based Accessibility measures for Integrated Transport-land Use Planning (GraBAM) In: Hull,A.–Silva, C.–Bertolini, L. (Eds.) *Accessibility Instruments for Planning Practice*, pp. 117–124. COST Office.

Reggiani, A.–Bucci, P.–Russo, G. (2011): Accessibility and Impedance Forms: Empirical Applications to the German Commuting Network. *International Regional Science Review* 34 (2): 230–252.

Rich, D. (1980): *Potential Models in Human Geography*. Concepts and Techniques in Modern Geography. 26. Geo Abstracts, Norwich.

Scheurer, J.–Curtis, C. (2007): *Accessibility Measures: Overview and Practical Applications*. Department of Urban and Regional Planning, Curtin University, Perth.

Schürmann, C.–Spiekermann, K.–Wegener, M. (1997): *Accessibility Indicators*. Berichte aus dem Institüt für Raumplanung 39, IRPUD, Dortmund.

Simma, A.–Axhausen, K.W. (2003): Interactions between travel behaviour, accessibility and personal characteristics: The case of the Upper Austria Region *European Journal of Transport and Infrastructure Research* 3 (2): 179–197.

Simma, A.–Vrtic, M.–Axhausen, K. W. (2001): *Interactions of travel behaviour, accessibility and personal characteristics: The Case of Upper Austria, presentation*, European Transport Conference, Cambridge, September 2001.

Smith, D.M.–Gibb, R.A. (1993): The Regional Impact of the Channel Tunnel. A Return to Potential Analysis. *Geoforum* 24 (2): 183–192.

Song, S. (1996): Some Tests of Alternative Accessibility Measures: A Population Density Approach. *Land Economics* 72 (4): 474–482.

Spence, N.–Linneker, B. (1994): Evolution of the motorway network and changing levels of accessibility in Great Britain. *Journal of Transport Geography* 2 (4): 247–264.

Stewart, J.Q. (1947): Empirical mathematical rules concerning the distribution and equilibrium of population. *Geography Review* 37 (3): 461–485.

Tagai, G. (2007): A potenciálmodell erényei és korlátai a társadalomkutatásban *Tér és Társadalom* 21 (1): 145–158.

Tóth, G.–Kincses, Á. (2007): Elérhetőségi modellek. *Tér és Társadalom* 21 (3): 51–87.

REGIONAL STATISTICS, 2015, VOL 5, No1: 137–158; DOI: 10.15196/RS05108

van Wee, B.–Hagoort, M.–Annema, J.A. (2001): Accessibility measures with competition. *Journal of Transport Geography* 9 (3): 199–208.

Vickerman, R.W. (1974): Accessibility, attraction, and potential: a review of some concepts and their use in determining mobility. *Environment and Planning A* 6 (6): 675–691.

Weibull, J.W. (1976): An axiomatic approach to the measurement of accessibility. *Regional Science and Urban Economics* 6 (4): 357–379.

Willigers, J.–Floor, H.–van Wee, B. (2007): Accessibility indicators for locational choice of offices: An application to the intraregional distributive effects of high-speed rail in the Netherlands. *Environment Planning. A* 39 (9): 2086–2098.

Wilson, A. G. (1971): A family of spatial interaction models, and associated developments. *Environment and Planning A* 3 (1): 1–32.