The wiiw Balkan Observatory

Working Papers | 117 | November 2015

Roman Römisch

Estimating agglomeration in the EU and the Western Balkan regions

The wiiw Balkan Observatory

www.balkan-observatory.net

About

Shortly after the end of the Kosovo war, the last of the Yugoslav dissolution wars, the Balkan Reconstruction Observatory was set up jointly by the Hellenic Observatory, the Centre for the Study of Global Governance, both institutes at the London School of Economics (LSE), and the Vienna Institute for International Economic Studies (wiiw). A brainstorming meeting on Reconstruction and Regional Co-operation in the Balkans was held in Vouliagmeni on 8-10 July 1999, covering the issues of security, democratisation, economic reconstruction and the role of civil society. It was attended by academics and policy makers from all the countries in the region, from a number of EU countries, from the European Commission, the USA and Russia. Based on ideas and discussions generated at this meeting, a policy paper on Balkan Reconstruction and European Integration was the product of a collaborative effort by the two LSE institutes and the wiiw. The paper was presented at a follow-up meeting on Reconstruction and Integration in Southeast Europe in Vienna on 12-13 November 1999, which focused on the economic aspects of the process of reconstruction in the Balkans. It is this policy paper that became the very first Working Paper of the wiiw Balkan Observatory Working Papers series. The Working Papers are published online at www.balkan-observatory.net, the internet portal of the wiiw Balkan Observatory. It is a portal for research and communication in relation to economic developments in Southeast Europe maintained by the wiiw since 1999. Since 2000 it also serves as a forum for the Global Development Network Southeast Europe (GDN-SEE) project, which is based on an initiative by The World Bank with financial support from the Austrian Ministry of Finance and the Oesterreichische Nationalbank. The purpose of the GDN-SEE project is the creation of research networks throughout Southeast Europe in order to enhance the economic research capacity in Southeast Europe, to build new research capacities by mobilising young researchers, to promote knowledge transfer into the region, to facilitate networking between researchers within the region, and to assist in securing knowledge transfer from researchers to policy makers. The wiiw Balkan Observatory Working Papers series is one way to achieve these objectives.

The wiiw Balkan Observatory

Global Development Network Southeast Europe

This study has been developed in the framework of research networks initiated and monitored by wiiw under the premises of the GDN–SEE partnership.

The Global Development Network, initiated by The World Bank, is a global network of research and policy institutes working together to address the problems of national and regional development. It promotes the generation of local knowledge in developing and transition countries and aims at building research capacities in the different regions.

The Vienna Institute for International Economic Studies is a GDN Partner Institute and acts as a hub for Southeast Europe. The GDN–wiiw partnership aims to support the enhancement of economic research capacity in Southeast Europe, to promote knowledge transfer to SEE, to facilitate networking among researchers within SEE and to assist in securing knowledge transfer from researchers to policy makers.

The GDN–SEE programme is financed by the Global Development Network, the Austrian Ministry of Finance and the Jubiläumsfonds der Oesterreichischen Nationalbank.

For additional information see www.balkan-observatory.net, www.wiiw.ac.at and www.gdnet.org

Estimating agglomeration in the EU and the Western Balkan regions

Roman Römisch

The Vienna Institute for International Economic Studies

roemisch@wiiw.ac.at

Final version 30.11.2015

Abstract

The paper develops a model to consistently estimate agglomeration and agglomeration economies in European NUTS3 regions. It is based on the empirical observation that the size of population across regions as well as of other economic variables tend to follow a Zipf distribution. Furthermore, the model is extended to capture agglomeration effects in traditional regional convergence estimations. Agglomeration is analysed for 25 European countries, including Macedonia and Serbia, and the years 2000 to 2012. Results indicate significant agglomeration effects on the level and growth of regional economic development, with agglomeration and agglomeration economies generally declining in the Western European countries and increasing the Central East and South East European countries.

Keywords: Zipf's Law, agglomeration economies, Europe, Western Balkans

Introduction[1]

A prominent way to analyse the distribution of population is estimating Zipf's law for city size distribution[2]. The popularity of Zipf's law, postulating a log-linear relationship between the size of a city and its rank in the hierarchy of cities within a country, rests on the remarkable simplicity of testing it combined with the good approximation it delivers regarding empirical city size distributions[3]. Given this, the available literature is large, ranging from classic studies (Rosen and Resnick 1980; Alperovich 1984, 1988) over important methodological and theoretical improvements (Gabaix 1999a,b; Gabaix and Ibragimov, 2007), to a meta-analysis of 29 studies (Nitsch, 2005) and most recent publications (Giesen and Südekum 2011, González-Val et al. 2014). Despite being empirically highly intriguing, Zipf's law is much harder to catch from an economic theory point of view.

This does not mean that there are no theories or theoretical arguments why cities exist in the first place and why they are of a particular size. In fact there are plenty of theories for this. From the 'old' economic geography one may refer e.g. to the Von-Thünen model of concentric rings, to Marshallian arguments regarding the advantages of industrial districts (i.e. knowledge sharing, labour market pooling and the sharing of specialised inputs), to the Central Place theory or to Henderson type arguments, where the size of a city is determined by the tension of external economies and diseconomies connected with geographic concentration[4]. From the 'new' economic geography explanations for city sizes are provided by agglomeration externalities[5], increasing returns, transport costs and natural endowments[6].

Yet, Brakman et al. (2009) remark that these theories disregard the interdependencies between the cities and hence they explain only why one city is of a particular size but not why the cities follow a specific size distribution. Recent advances in this context have been made by Duranton (2006), Brakman et al. (ibid.) and most recently by Hsu (2012) who developed a Central Place model to explain the size distribution of cities within a country.

The literature on the distribution and agglomeration of economic activity across space is based on the same set of theories. Thus, they constitute the point where the two strands of research meet. One difference between the analysis of population and economic activity is that for the latter, with the

[1] Note to the discussant: The current version is indeed a very preliminary version. Due to contractual issues work on the paper has begun only late (about a month ago). It was quite difficult to come up with something reasonable, also given the time spent on collecting data. Though the basic idea seems not to bad, it is clear that still a lot has to be done to improve the paper. This is an ongoing process, so that for the conference it is likely that some new results not yet in the paper will be presented. Thanks a lot for reading the paper and please keep in mind that it just a first draft.

[2] and Gibrat's law for city size growth.

[3] Additionally, data are also easily available.

[4] See Fujita et al. for an overview (Fujita et al. 1999)

[5] Such externalities may be divided into Marshall (1890)-Arrow (1962)-Romer (1986) (MAR) externalities and Jacobs (1969) externalities. MAR externalities are associated with industrial specialisation, so that a concentration of a particular industry within a city facilitates knowledge spillovers across firms. In opposition, Jacobs regards inter-industry spillovers as the most important source of new knowledge creation. She argues that the agglomeration of different industries within an urban region fosters innovation due to the diversity of available local knowledge sources. (see also Greunz, 2004 on this)

[6] See Ottaviano and Puga (1998) for a survey of the 'New Economic Geography' and Fujita and Mori (2005) for the latest developments.

rising popularity and the academic appeal of the 'new economic geography' since the early 1990s[7], theoretical models are much more developed and diverse[8]. Correspondingly, there is also a wide range of empirical tools used to analyse the many aspects of economic agglomeration and the effects arising thereof.

Amongst these methods are a number of indices, such as straightforward measures of concentration and specialisation (Hallet, 2000), more sophisticated 'dartboard' indices of concentration (Ellison and Glaeser, 1997)[9] as well as general entropy (i.e. inequality) measures (Brülhart and Traeger, 2003). The vast majority of empirical studies are however econometric analysis largely concerned with estimating the size and causes of agglomeration economies. A review of 34 studies by Melo et al. (2009) shows that the most common approaches to estimate the impact of agglomeration externalities on economic output are either via a production framework or via estimating wages equations. At the same time Melo et al. report that the results of the studies are far from uniform and do highly depend on which countries, industries and estimation methods[10] are chosen. This and additional reviews by Redding (2010) and Puga (2010) suggest that there is still enough space for further and new analysis in this field, whereby "one promising area for further inquiry lies in integrating insights from urban economics" (Redding, 2010 p. 308).

This paper contributes to existing literature firstly by combining the two different strands of literature to empirically analyse the role of cities and districts in a number of West, Central and Eastern and South-Eastern European economies. It uses Zipf's law to derive a new way to estimate the size and importance of agglomeration economies. Secondly, as the sample consists of countries at different stages of economic development, results may be indicative to which extent economic concentration tends to differ depending on the level of prosperity. This is the more important as Shepotylo (2011) claims that, because of the legacy of socialist urban policies restraining growth of the largest cities, the adjustment in Central and Eastern Europe and SEE to urban-rural patterns like in the more developed Western European countries is more difficult. Hence, significant East-West differences in the spatial distribution of economic activities may point towards potential problems of exploiting agglomeration economies fully in the East, and thus may demand some policy intervention.

The remainder of the paper is organised as follows: The first section develops a model to measure for agglomeration and agglomeration economies based on a Zipf dstribution and compares these measures to existing measures of concentration and agglomeration. Secondly the basic model is extended to include additional explanatory variables also outlining a convergence model that not only captures convergence but also agglomeration effects. The third part contains the empirical analysis and the final part concludes.

[7] Basically starting with Krugman (1991).
[8] In addition to above literature see also Redding (2009).
[9] For a variant see e.g. Maurel and Sedillot (1999).
[10] Melo et al. report that a) Japan, China and Sweden show lower urban agglomeration externalities than the US, France and Italy; b) services industries tend to benefit more from agglomerations than manufacturing industries, and c) panel data estimation methods tend to reduce the size of agglomeration externalities.

Measuring agglomeration and economies of agglomeration

The starting point of the analysis is the empirical observation that in many countries the distribution of population across of cities or regions follows a Zipf distribution. Formally it may be expressed by:

$$P_j = \frac{A}{r_j^b} \tag{1}$$

where P_j is the population size of region j, A is a constant, r_j is the rank of region j in the hierarchy of regions ranked by their population size (rank 1 for the largest region, rank 2 for the second largest region and so on), and b is the Pareto exponent determining the shape of the distribution. The special case in the Zipf distribution when b equals 1 is also known as 'Zipf's law' or the rank-sizerule. In this case A corresponds to the population size of the largest region, while the second largest region has exactly half of the population of the largest region. More generally, according to Zipf's law the largest region is exactly k times as large as the k-th region.

Testing whether population or other variables follow a Zipf distribution and whether Zipf's law holds is simple, as in logarithmic form it reduces to a linear equation that is easily tested econometrically by OLS:

$$\ln(P_j) = \ln(A) - \mathrm{b}\ln(r_j) \tag{2}$$

Empirically it shows[11] that overwhelmingly the estimated b is different from one, so that in the case b is smaller than 1, city size distribution is more even than suggested by Zipf's law and vice versa. Remarkably though, independent of the size of the estimated b the goodness-of-fit of the linear form is very high, with R^2 being in most cases higher than 0.9. That is, even though Zipf's law does not hold in many instances, the distribution of population across regions is well approximated by a Zipf distribution.

A less discussed aspect in the literature is that the estimated Pareto exponent b can be interpreted as an agglomeration measure. This follows directly from the definition of the Zipf distribution. The higher the Pareto exponent b is, the higher is the weight of the largest regions in the total population of a country (if population is the dependent variable), being equivalent to a higher population concentration. Vice versa, the lower b is, the more equal is the population distribution across regions, and if b equals 0, population is equally distributed over all regions (as equation 1 reduces to $P_j = A$).

As a conjecture, the estimated b may be a more consistent and reliable indicator of agglomeration than other agglomeration measures, like the Herfindahl index, the mean logarithmic deviation index and other entropy measures, or more sophisticated measures like the Ellison-Glaeser index or the Maurel and Sedillot index, the latter two using inter alia the Herfindahl index. The reason for this is that the estimated b is a distribution parameter, describing the shape of a distribution (in this case a Zipf distribution). Thus, b as an agglomeration measure is, unlike many other indicators, independent of the number of observations (regions). Hence, it does not change if the number of observations changes, which is of high empirical importance when comparing the agglomeration in countries with different numbers of regions. Additionally, as will be illustrated in the empirical analysis, b is as-

[11] See e.g. Brakman et.al (2009), p. 318-319

sumed to remove inconsistencies between the different agglomeration indices, as the more common indices tend to provide different results with respect to the size of agglomeration. Thus, in cross-country analysis of population agglomeration, the ranking of countries with highest and lowest population agglomeration may differ depending on the agglomeration index that is used.

To illustrate this conjecture it is assumed that the distribution of population in a country is exactly represented by a Zipf distribution given in equation 1. Based on this, it is feasible to express the mean logarithmic deviation (MLD) index and the Herfindahl index in terms of this distribution.

The MLD index is usually defined as:

$$MLD = \frac{1}{n}\sum_{j=1}^{n} \ln\left(\frac{\mu}{P_j}\right) \tag{3}$$

With n being the total number of regions, μ is the average population per regions and P_j is the population in region j.

Since it is assumed that the regions' population follows a Zipf distribution, P_j is given by $P_j = 1/r_j^b$ to use a normalised version of equation 1. Using this, equation 3 can be also expressed as:

$$MLD = \frac{1}{n}\sum_{j=1}^{n}\left[\ln(\mu) - \ln(1) + b\ln(r_j)\right] \tag{4}$$

For given n μ is constant and can be expressed as:

$$\mu = \frac{\sum_{j=1}^{n} 1/r_j^b}{n} \tag{5}$$

The numerator on the r.h.s. represents the sum of population over all regions as population is determined by a Zipf distribution. At the same time the numerator is a Riemann zeta function (if n goes to infinity) or can also be viewed as a general harmonic series (or hyperharmonic series). Thus, for finite n the numerator can be expressed a the n-th (general) Harmonic number $H_{n,b}$, so that $\mu = H_{n,b}/n$. Using this, reminding that ln(1) is zero and also considering that r represents the rank of the regions and thus goes from 1 to n, so that $\sum_{j=1}^{n} \ln(r_j) = \ln(n!)$ allows rewriting equation 3 as:

$$MLD = \ln\left(\frac{H_{n,b}}{n}\right) + \frac{b}{n}\ln(n!) \tag{6}$$

Under similar considerations the Herfindahl index (HI), defined as: $HI = \sum_{j=1}^{n} s_j^2$, with s_j being the share of region j in total country population can be expressed on the basis of a Zipf distribution as:

$$HI = \frac{H_{n,2b}}{H_{n,b}^2} \tag{7}$$

where H is again the (general) n-th Harmonic number (squared in the denominator), b the Pareto coefficient and 2b is b times 2. In both examples, the size of the respective index depends only on the number of observations and the distribution parameter b.

It follows that, when comparing two countries that have the same distribution of population across regions (i.e. b is the same in both countries), yet a different number of regions, both the MLD and Herfindahl index will differ between the countries, and thus may not be reliable measures for a cross-country comparison of agglomeration.

The empirical fact that the distribution of population and other variables across regions can be expressed via a Zipf distribution also allows developing a framework to estimate agglomeration economies. This framework is based on a simple idea:

Provided that the population of a country follows a Zipf distribution, then, if agglomeration economies do not exist, the distribution of economic activity across regions should correspond to the distribution of population. If, however, agglomeration economies exist, the distribution of economic activity should be different, so that the more populated regions exhibit higher agglomeration economies than the less populated regions. In its pure form this idea requires the restriction that except for population size, regions are identical in all other characteristics, including their natural and physical endowments, sectoral structure, their size etc.

Assuming for the moment that this restriction holds then this idea may be formulated as such:

Let E_j be some measure for economic activity in a region j of a country. Correspondingly, γ_j is defined as a variable measuring E_j per head of population.

$$\gamma_j = \frac{E_j}{P_j} \tag{8}$$

In the absence of agglomeration economies and provided that regions are identical except for their population size, γ can be expected to be equal across regions; hence $\gamma_j = \gamma_k$ for all j and k. However if there are agglomeration economies that vary with the size of regions[12], being highest for the largest regions, the relation should be $\gamma_1 > \gamma_2 > \gamma_3 > \cdots > \gamma_n$.

Assuming that E_j just as population follows a Zipf distribution it can be expressed as:

$$E_j = \frac{B}{r_j^c} \tag{9}$$

with B being a constant, approximately indicating economic activity in the largest region and c being the Pareto coefficient specific to the E. Using this and the power law for population size in equation (1) gives:

$$\gamma_j = \frac{D}{r_j^{c-b}} \tag{10}$$

with $D = B/A$.

Taking logarithms results in:

$$\ln(\gamma_j) = \ln(D) - (c-b)\ln(r_j) \tag{11}$$

which provides a simple testing procedure for the existence, importance and size of agglomeration economies.

As a matter of fact, this simple model is not too far off from other model used to estimate agglomeration economies. For example, based on a regional model Ciccone (2002) agglomeration effects estimates for four European countries, regressing productivity, i.e. output per employed, on employment density (employment per area) and some controls. Thereby the coefficient for employment density represents the agglomeration variable. In the model based on the Zipf distribution employ-

[12] and implicitly the density, given that regions areas are of approximately the same.

ment is substituted by population, and more or less the same relation is estimated, given that the rank of a region is proxy for density (which should be the case if the areas of the districts are approximately equal).

Model extension

Lifting the restrictive assumption that, except for population size, regions are identical with respect to all other characteristics the model described in equation (10) may be augmented by a number of control variables X representing e.g. their endowments with skills, their sectoral structure, comparative advantages etc. Hence equation (10) could be rewritten as:

$$\gamma_j = \frac{D}{r_j^{c-b}} X^a \tag{12}$$

where 'a' represents the coefficient for the respective control variable X.

When estimating equation (12) (in logarithmic form) attention needs to be paid whether X itself is subject to agglomeration effects. For example, if X represents the sectoral structure of regions or, more specifically, the share of services in total employment or GVA, data shows that more developed and higher populated regions tend to have indeed a higher share of services than other regions. Hence, if it is indeed the case that X is subject to agglomeration effects, it can be modelled as:

$$X_j = \frac{Z}{r_j^x} \varepsilon_j \tag{13}$$

where x is the agglomeration coefficient of X, and ε_j is the variation over the data generating process. Using this in equation (12) results in

$$\gamma_j = \frac{D}{r_j^{c-b}} \left(\frac{X}{r_j^x} \varepsilon_j \right)^a = \frac{DX^a}{r_j^{c-b+ax}} \varepsilon_j^a \tag{14}$$

Thus, taking the logarithmic form of equation (12) and estimating:

$$\ln(\gamma_j) = \ln(\mathrm{D}) - (c-b)\ln(r_j) + a\ln(X_j) + u \tag{15}$$

would give an unbiased estimator for X_j, but a biased estimate for the agglomeration variable.

To avoid this and to take account of the agglomeration effects in X, an auxiliary regression like equation (16) may be used:

$$\ln(X_j) = \ln(Z) - x\ln(r_j) + \ln(\varepsilon_j) \tag{16}$$

where X is split into one part that is explained by agglomeration effects, i.e. $x\ln(r_j)$, and another part, $\ln(\varepsilon_j)$, identifying the region's characteristics clear of agglomeration effects. This allows re-estimating equation (12) using

$$\ln(\gamma_j) = \ln(\mathrm{D}) - (c-b)\ln(r_j) + \beta_x \ln(\varepsilon_{X,j}) + u \tag{17}$$

where $\varepsilon_{X,j}$ are the explanatory variables' residuals from the auxiliary regressions. Equation (17) provides both, an unbiased estimator for X_j and the agglomeration variable r_j.

A notable variant of this is the estimation with respect to the growth of agglomeration economies over time. Let the growth of the economic activity variable γ_j be defined as $\gamma_{j,t}/\gamma_{j,0}$, with t referring to the value of γ_j to the end of the observation period, and 0 referring to its begin. It can also be written as:

$$\frac{\gamma_{j,t}}{\gamma_{j,0}} = \frac{\frac{D_t}{r^{c_t - b_t}}}{\frac{D_0}{r^{c_0 - b_0}}} = \frac{D_t/D_0}{r^{(c_t - b_t)-(c_0 - b_0)}} \tag{18}$$

Equation (18) relates the growth of economic activity to the changes in agglomeration economies given by $(c_t - b_t) - (c_0 - b_0)$. Extending the model to a convergence type model requires adding $\gamma_{j,0}$ on the right hand side of the equation. However as $\gamma_{j,0}$ is a function of agglomeration economies, this would, as above, induce a bias to the estimation. Thus, as before using the residuals from a regression of $\ln(\gamma_{j,0})$ on $\ln(r_{j,0})$, they can be used in a convergence model including agglomeration economies of the form:

$$\ln\left(\frac{\gamma_{j,t}}{\gamma_{j,0}}\right) = \ln(\mathrm{D}) - [(c_t - b_t) - (c_0 - b_0)]\ln(r_j) + \beta_{\gamma_{j,0}} \ln\left(\varepsilon_{\gamma_{j,0}}\right) + u \tag{19}$$

where $D = D_t/D_0$. In contrast to the traditional convergence model: $\ln(\gamma_{j,t}/\gamma_{j,0}) = \alpha + \beta\ln(\gamma_{j,0}) + u$, the model in equation (19) splits the convergence parameter β into the effects of changes in agglomeration economies, given by $[(c_t - b_t) - (c_0 - b_0)]$, and an initial income effect $\beta_{\gamma_{j,0}}$.

Data

The paper estimates agglomeration effects and agglomeration economies at the NUTS3 level of regions for 25 countries. These include 23 countries from the EU-28 as well as Macedonia and Serbia. Five EU-28 countries are not covered because of data issues, these are: Cyprus, Denmark, Estonia, Luxembourg and Malta. For all countries, except Serbia, data for the analysis is taken from Eurostat regional and national accounts database and covers the years 2000 to 2012. Data for Serbia is taken from the annual publication 'Municipalities of Serbia', published by the Serbian Statistical Office. Serbian data was collected initially at the district level and aggregated to the NUTS3 level currently in use in Serbia. The main data collected were population data and data on employees and GVA as indicators of economic activity. The choice of these data was largely determined by data availability in Serbia, as there are no long time series on regional GDP and no time series on employment. Additionally, there is also no data on GVA at the regional level in Serbia, so that wage data has been used instead for this country. Overall, there are some gaps in the data. Serbian data has not been collected for the years 2004 and 2010, in France regional GVA data does not exist for the years 2000-2006, in Macedonia employees' data for 2000 to 2009 is missing, just as employees' data for Portugal in 2012. Additional data, e.g. on the regional sectoral structures have also been collected. All estimations are done at the country level.

Estimation

Agglomeration

The analysis starts with the estimation of population agglomeration using equation 2. The dependent variable is the natural log of population and the independent variable is the log rank of the NUTS3 regions. Estimation is done individually for each of the 25 countries and for each year in the period

2000 to 2012, resulting in 325 OLS regressions. The estimated agglomeration coefficients are shown in Figure 1.

The estimated population agglomeration coefficients are highly significant (at the 1% level or better) in each country and year: Moreover the R^2 in each regression is at least above 0.7 and in many countries above 0.9. Thus, the distribution of population is reasonably well approximated by a Zipf distribution throughout countries and years. Population agglomeration varies considerably across countries, being lowest in Slovakia, Poland and Romania and highest in Spain, Portugal and Slovenia.

Figure 1: Population agglomeration coefficients, 2012

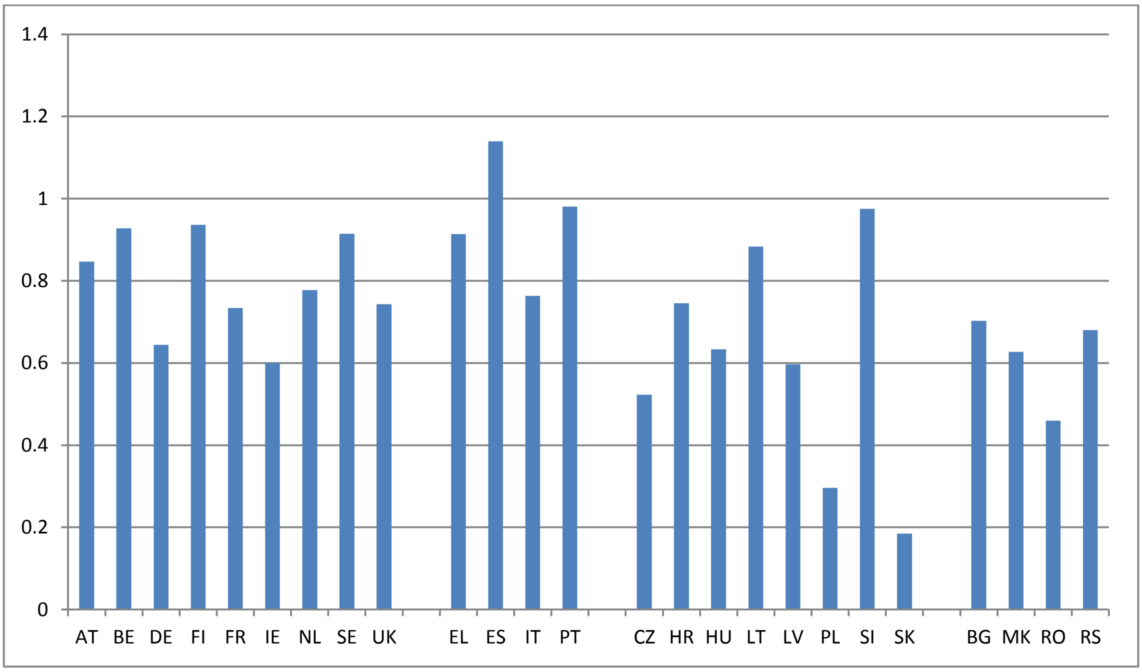
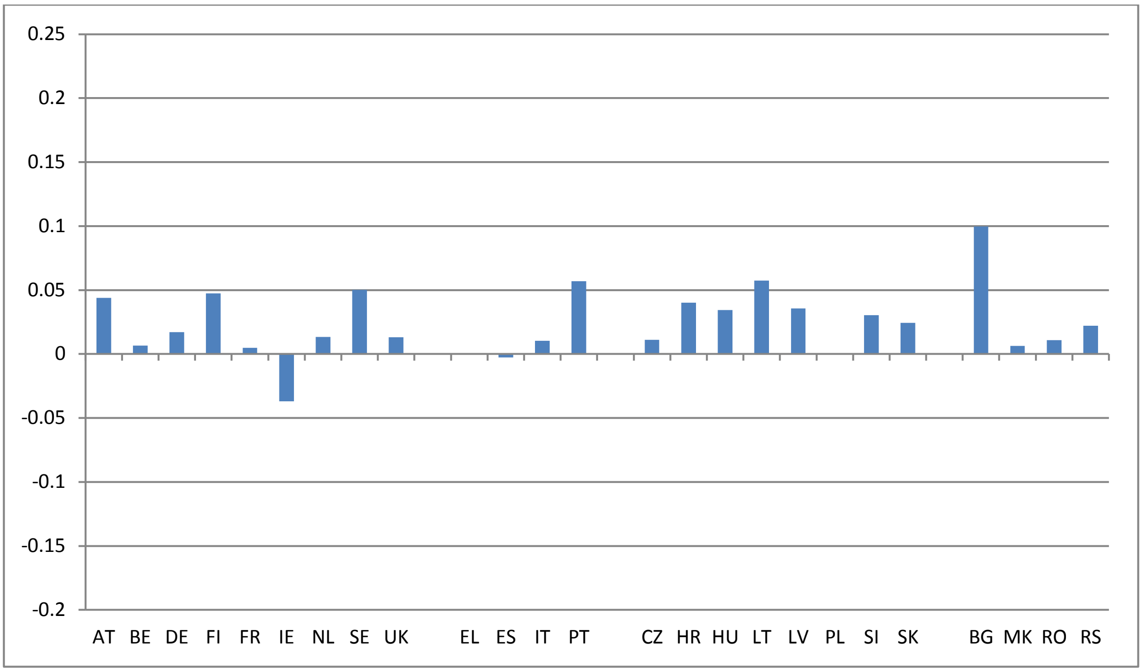


Figure 2 shows the changes in population agglomeration from 2000 to 2012. In most countries of the sample population became increasingly agglomerated, though these increases were mostly moderate, the exception being Bulgaria. Only in two countries, Spain and Ireland population became more dispersed over time.

Figure 2: Change in population agglomeration coefficients 2000-2012



In the light of the conjecture raised above it is of interest to compare the estimated agglomeration coefficient to other measures of agglomeration, notably the Herfindahl and the Mean logarithmic deviation (MLD) index. This is shown in Table 1, where the countries are ranked according to the size of the respective index. The table indicates a considerable inconsistency in the assessment of agglomeration between the various indicators, as the ranking of countries is different for each indicator. Thus, the Herfindahl index generally ranks the countries with the highest number of NUTS3 lowest, while countries with only few NUTS3 regions show a high agglomeration. The MLD index appears to be a bit more consistent and its ranking is relatively close to the ranking of the estimated agglomeration coefficient, though its accuracy still may suffer from the differences in number of regions per country. For comparison reasons also a hypothetical Herfindahl and MLD index are estimated with equations (6) and (7), using the estimated agglomeration coefficient b and the assumption of an equal number of regions across countries (n=20). As both estimated indices only depend on b and the number of regions they provide, by definition, the same ranking as the agglomeration coefficient itself.

Table 1: Comparison of agglomeration measures, NUTS3 population agglomeration 2012

Herfindahl index		Mean logarithmic deviation index		Agglomeration coefficient		Estimated Herfindahl index		Estimated Mean logarithmic deviation index	
DE	0.6	SK	0.8	SK	0.184	SK	5.1	SK	1.1
UK	1.2	PL	4.7	PL	0.296	PL	5.4	PL	3.0
FR	1.6	LV	7.4	RO	0.460	RO	6.0	RO	7.5
PL	1.7	CZ	9.1	CZ	0.523	CZ	6.4	CZ	9.9
IT	2.0	IE	9.3	LV	0.596	LV	6.9	LV	13.1
RO	3.1	RO	10.0	IE	0.600	IE	6.9	IE	13.3
NL	4.0	MK	10.6	MK	0.627	MK	7.2	MK	14.6
BE	4.3	HU	15.9	HU	0.633	HU	7.2	HU	14.9
ES	5.0	HR	21.3	DE	0.644	DE	7.3	DE	15.5
BG	6.5	LT	21.4	RS	0.680	RS	7.7	RS	17.4
AT	6.7	BG	21.9	BG	0.702	BG	7.9	BG	18.7
HU	7.6	RS	23.4	FR	0.734	FR	8.3	FR	20.5
PT	7.6	NL	25.8	UK	0.743	UK	8.4	UK	21.1
HR	7.8	DE	26.1	HR	0.745	HR	8.4	HR	21.2
RS	8.3	FR	26.7	IT	0.764	IT	8.6	IT	22.4
CZ	8.5	SI	27.2	NL	0.777	NL	8.8	NL	23.2
SE	10.8	UK	29.0	AT	0.847	AT	9.7	AT	28.0
FI	12.0	IT	31.7	LT	0.883	LT	10.3	LT	30.7
SK	12.7	AT	34.5	EL	0.914	EL	10.8	EL	33.0
SI	13.7	FI	35.6	SE	0.915	SE	10.8	SE	33.1
EL	14.5	BE	37.0	BE	0.928	BE	11.0	BE	34.1
LT	15.2	SE	38.1	FI	0.937	FI	11.2	FI	34.8
IE	15.5	PT	40.4	SI	0.976	SI	11.9	SI	38.0
MK	16.2	ES	60.3	PT	0.981	PT	12.0	PT	38.5
LV	19.8	EL	63.3	ES	1.140	ES	15.3	ES	53.3

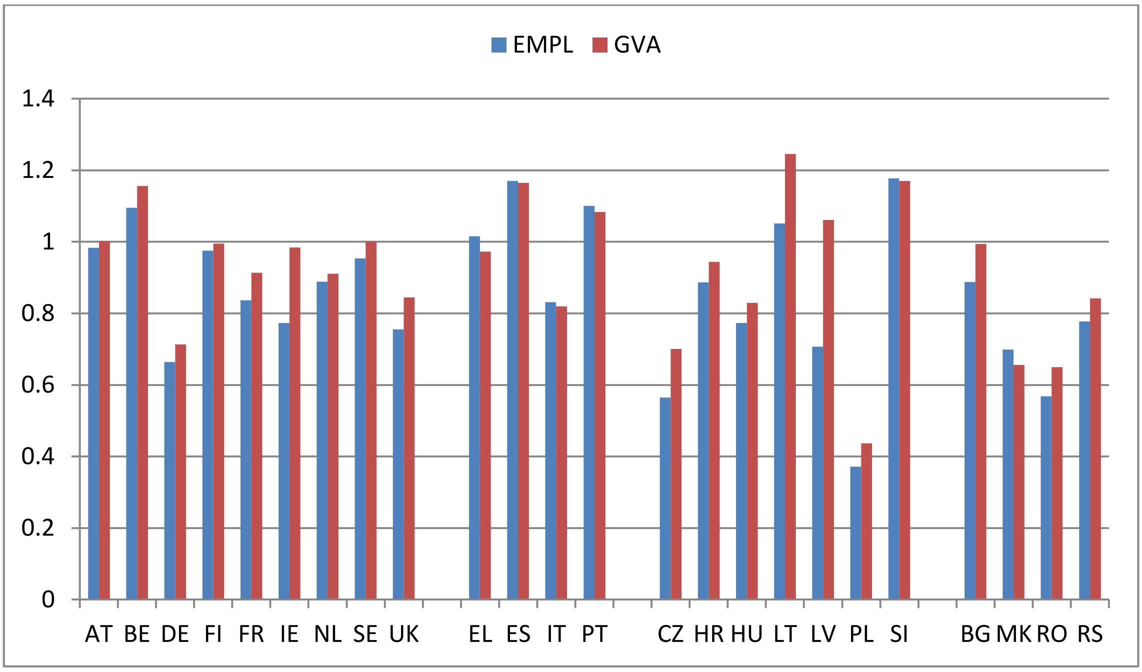
The size of agglomeration is also estimated for employment and GVA for each country and year. Again, equation 2 is used with the dependent variable either being log employment or log GVA. The independent variable is the log rank of population[13]. The estimated agglomeration coefficients for employment and GVA in the year 2012 are reported in Figure 3. In the case of Slovakia, for both employment and GVA, the estimated agglomeration coefficients were not significantly different from zero, indicating an almost equal distribution of economic activity over the Slovak regions. Because of the insignificant coefficients Slovakia is also excluded in Figure 3. For the remaining countries, the estimated employment and GVA coefficients are highly significant throughout the years.

Both, employment and GVA agglomeration tend to vary widely across countries without an intuitively emerging pattern across country groups. High agglomeration of both employment and GVA are found e.g. in Belgium, Spain, Portugal, Lithuania and Slovenia. Low levels of agglomeration are present in Germany, the Czech Republic, Poland and Romania. Still, in each country agglomeration of employment and GVA is higher than population agglomeration, with GVA agglomeration in most countries being higher than employment agglomeration, especially in Ireland and the two Baltic

[13] Alternatively, the log rank of employment or GVA could be used, though for consistency and comparison reasons this is not done in this paper.

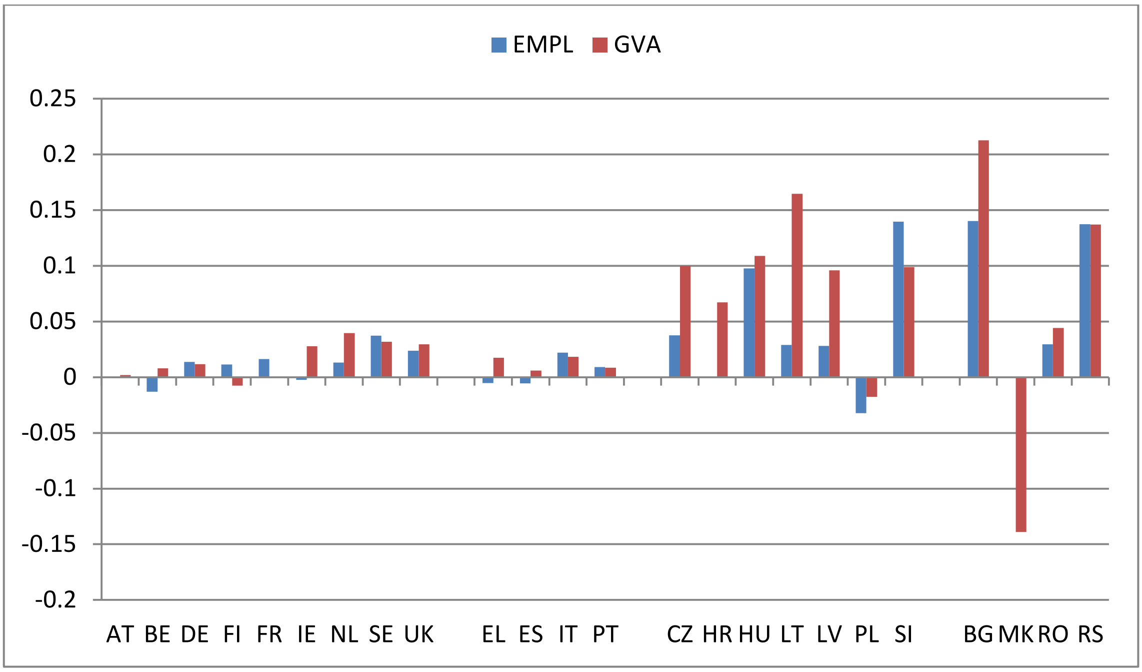
states. Overall the reliability of results is decently high, given that the regression R^2 is in most cases above 0.73 (as a rule, the R^2 in the employment regression tend to be higher than the R^2 in the GVA regressions). The exceptions to this are Poland and Macedonia, where the R^2 are in the range of 0.44 to 0.6 in the employment and GVA regressions.

Figure 3: Employment and GVA agglomeration coefficients, 2012



The changes in employment and GVA agglomeration from 2000 to 2012 are shown in Figure 4. In the more developed Western European as well as Southern EU countries both employment and GVA agglomeration changed by little. Contrastingly, agglomeration tended to increase strongly in the CEE and Balkan countries (with the exception of Macedonia). In most of these countries agglomeration in GVA increased by far more than employment agglomeration, e.g. in the Czech Republic, the Baltic states and Bulgaria. In Hungary and Serbia employment and GVA agglomeration increased approximately the same, while in Slovenia employment agglomeration increased more than GVA agglomeration.

Figure 4: Change in employment and GVA agglomeration coefficients 2000-2012



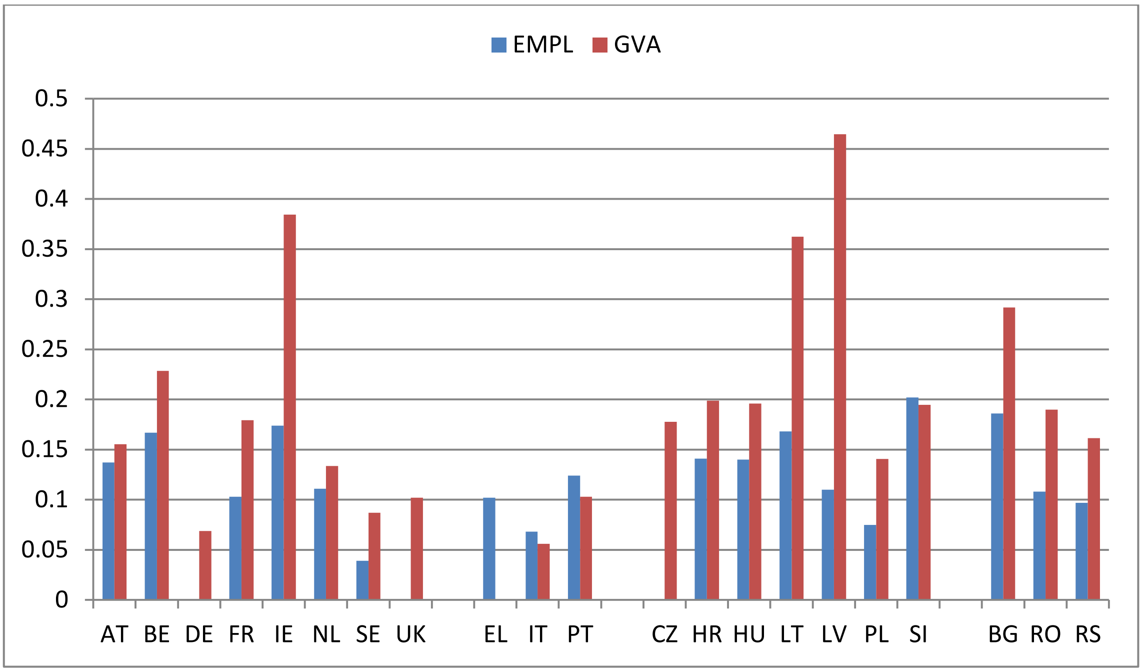
Note: no data for changes in GVA agglomeration in France and for employment agglomeration in Croatia and Macedonia.

Agglomeration economies

Turning from estimating agglomeration to the estimation of agglomeration economies, the paper estimates equation 6, using two different definitions of the dependent variable γ. Firstly, γ is defined as the employment to population ratio (i.e. employment rate) and secondly as the GVA to population ratio, which approximately corresponds to GDP per capita. The independent variable is again the log rank of population.

In contrast to the estimation of agglomeration coefficients, the estimation of agglomeration economies produces a lesser number of significant results, indicating that not in all countries agglomeration economies might be present. Moreover, for some countries the significance of the estimated agglomeration economies coefficients tends to change over years (as regressions are run over all years). Thus, Greece, Spain and Macedonia no significant agglomeration economies are found over the whole period, while in Finland significant agglomeration economies were found at the start of the period, but those economies declined over time and became not significantly different from zero in the 2011 and 2012 regressions. The Czech Republic is the opposite case, as for many years no significant agglomeration externalities were found and only in 2012 the estimated coefficient became significant (only in the case of GVA agglomeration economies).

Figure 5: Employment and GVA agglomeration economies coefficient 2012

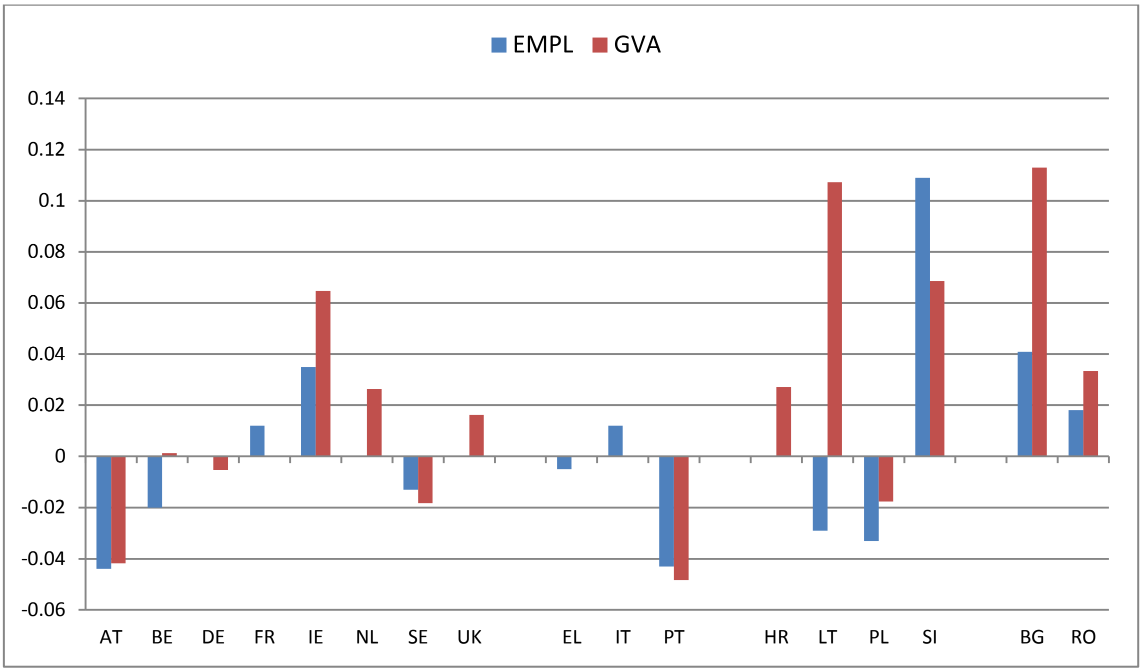


Note: Only significant values reported

The estimated and significant coefficients for employment and GVA agglomeration economies in 2012 are shown in Figure 5. Agglomeration economies tend to be stronger in terms of GVA than in employment terms, especially in the Baltic States, Bulgaria and Ireland. Exceptions to this are Italy, Portugal and Slovenia. Additionally, with the exception of Ireland and Belgium, agglomeration economies tend to be stronger in the CEE and SEE countries than in the Western and Southern EU countries.

The changes the employment and GVA agglomeration economies from the year 2000 to 2012 are shown in Figure 6. Like in the case of levels, agglomeration economies mostly increased by more in the CEE and SEE countries than in the Western EU countries, except for Poland where agglomeration economies declined. In the Western EU countries agglomeration hardly increased (except for Ireland), and in Austria, Sweden and Portugal the declined from 2000 and 2012.

Figure 6: Change in employment and GVA agglomeration economies coefficient 2000-2012



Note: Portugal change 2000-2011; only significant values reported

As in Figure 5 and Figure 6 countries are grouped according to their level of economic development, both figures suggest that the level of agglomeration economies as well as the changes therein are to some extent correlated with GDP per capita levels and GDP growth. This is analysed in bit more detail in Figure 7 showing the correlation between GVA agglomeration economies and GDP per capita levels in 2012 and Figure 8 showing the correlation between the changes in GVA agglomeration economies and average real GDP growth rates from 2000-2012.

The first of these two figures indicates a weak negative correlation between agglomeration economies and GDP per capita levels, though this correlation would become stronger if Ireland was considered an outlier and would be removed from the sample. Still, Figure 7 suggests that higher GDP per capita levels tend to be associated with lower agglomeration economies or a more even spatial distribution of economic activity. Contrastingly, Figure 8 indicates a strong correlation between high GDP growth rates and increasing agglomeration economies. Taken together both figures describe a contradicting or conflicting situation. High GDP per capita levels are associated with low agglomeration economies, while high GDP growth rates that are necessary (especially for the less developed countries) to reach higher levels of GDP are associated with increasing agglomeration economies. This conflict can only be solved assuming an inverse U-shaped pattern of agglomeration economies over the stages of economic development, with agglomeration economies increasing (causing an increase in disparities) in early stages of economic development and decreasing in more advanced stages.

Figure 7: Correlation of GVA agglomeration economies and GDP per capital levels, 2012

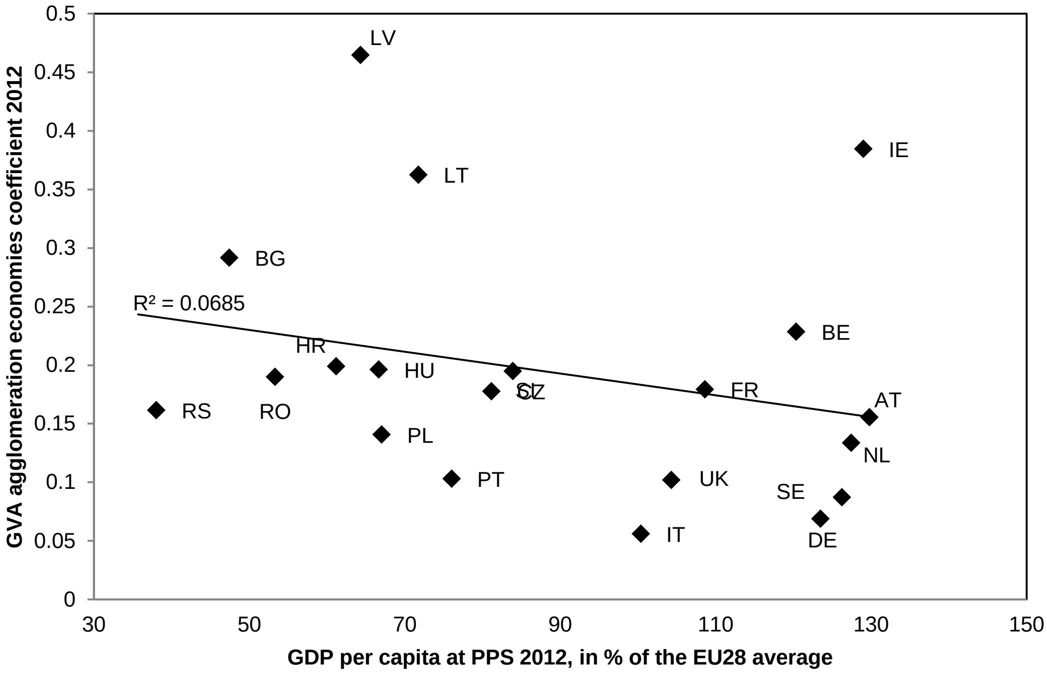
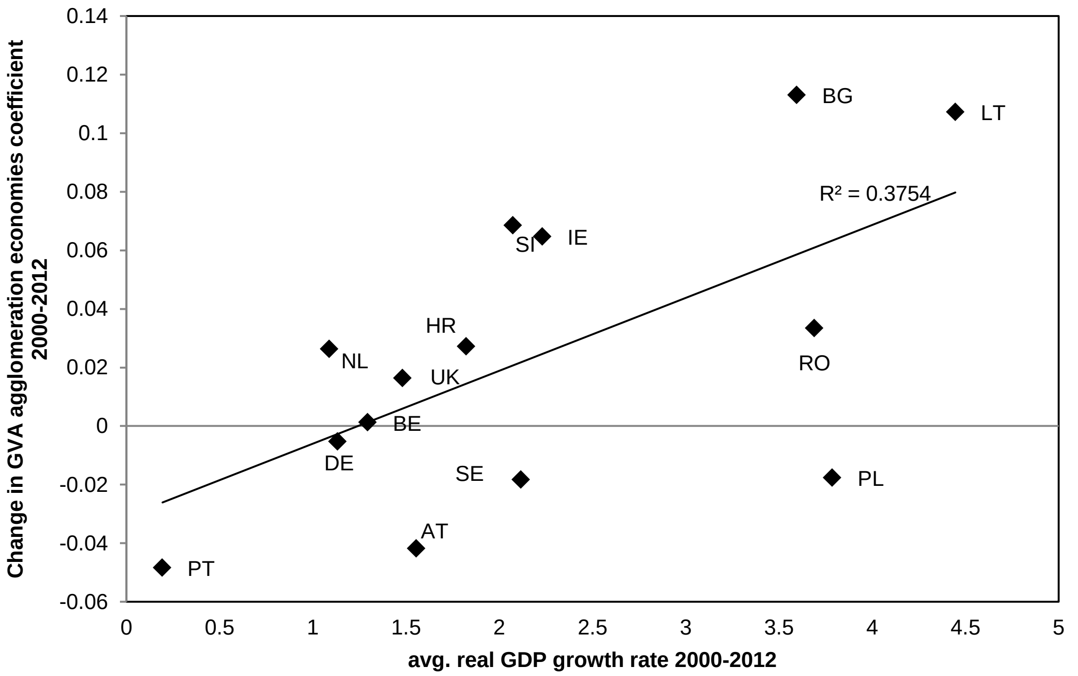


Figure 8: Correlation of changes in GVA agglomeration economies and average real GDP growth rates 2000-2012.



Convergence

Finally, the analysis focuses on the estimation of the growth of agglomeration economies. For this, a convergence model outlined in equation (19) is estimated for each country using a cross-sectional spatial autoregressive model with spatial autoregressive disturbances. The following model is estimated:

$$y = \alpha + \beta_{r_{j,0}} \ln(r_{j,0}) + \beta_{\gamma_{j,0}} \ln(\gamma_{j,0}) + \lambda Wy + \theta X + u \qquad (20)$$

with $u = \rho Wu + \varepsilon$, and $y = \ln\left(\frac{\gamma_{j,t}}{\gamma_{j,0}}\right)$, i.e. the growth of the agglomeration economies variable. In this model W is a row-normalised spatial weighting matrix, $r_{j,0}$ refers to the rank of region j at the beginning of period, $\gamma_{j,0}$ is the size of agglomeration economies and X is a matrix of additional explanatory variables measuring the regions' sectoral structure and changes therein as well as population density. The model is estimated for GVA agglomeration economies, and since these are defined as the GVA to population ratio the model resembles a GDP per capital convergence model. In detail, the model estimates the growth of GVA per capita from 2000 to 2012 on the rank of regions, capturing the effects of changes in agglomeration economies, the initial GVA per capita, capturing convergence, as well on the regions' employment share in agriculture, industry and services (excluding construction) and the growth therein to capture the regions' sectoral structure as well as on population density as additional control variable.

A negative sign of $\beta_{r_{j,0}}$ indicates, according to equation (19) an increase of agglomeration economies over time, while a negative sign of $\beta_{\gamma_{j,0}}$ would indicate a convergence of GVA per capita. To illustrate the importance of controlling for agglomeration effects in the explanatory variables (according to equations (12)-(19)) the model is estimated twice, one time using the explanatory variables as they are and the second time using the residuals of auxiliary regressions of the explanatory variables on the rank of regions. The results of these regressions are shown in Table 2, whereby only those countries with 20 or more NUTS3 regions are covered.

Comparing the results of the regressions using the original explanatory variables with the results of the regression using the residuals it shows that the main difference is, indeed, the sign, size and significance of the agglomeration economies variable. Thus, the size and/or significance of this variable changes for almost all countries in the sample, with the number of significant agglomeration economies variables being much higher in the regressions using the residuals. At the same time the coefficients and signs of the other explanatory variables do not change by much, thus confirming the validity of the model using residuals in equation (15).

As far as the regression results of the model using residuals are concerned, they indicate that agglomeration economies in the more developed Western European countries tended to decline, i.e. $\beta_{r_{j,0}} > 0$, while they generally increased in the CEE and SEE countries, i.e. $\beta_{r_{j,0}} < 0$ (except for Poland). At the same time convergence, i.e. $\beta_{\gamma_{j,0}} < 0$, in GVA per capita was found in many, but not all countries. Thereby, this convergence process went pari passu with the decrease in agglomeration economies in Western Europe, e.g. in Belgium, Germany and Sweden, indicating a strong catching up process of less developed regions. Oppositely in the CEE and SEE countries this catching up process was less clear, as the convergence of less developed regions was offset by an increase in agglomeration economies that favoured the more developed regions in those countries.

As far as the other explanatory variables are concerned, their size, sign and significance tend to be highly country specific and there is no general pattern with respect to the extent the sectoral structure affects the level of economic development of the regions.

Table 2: Convergence estimation results; dependent variable: growth of GVA per capita

Regressions using original explanatory variables

	AT	BE	DE	NL	SE	UK	ES	IT	PT	HU	PL	BG	RO	RS
Rank - $\beta_{r_{j,0}}$	0.004	-0.048***	-0.002	-0.043*	0.027	-0.019**	0.003	0.004	0.099***	-0.042**	0.019*	-0.039	-0.035	-0.065***
Initial GVA/POP - $\beta_{\gamma_{j,0}}$	-0.011	-0.089**	-0.080***	0.096	-0.287	0.028	-0.146*	0.036	-0.256***	-0.240***	0.004	-0.149	-0.240*	-0.297***
Initial share Agriculture	-0.003	0.014*	-0.009	0.061**	-0.135***	0.000	-0.016	-0.023**	-0.080***	0.188*	-0.014	0.037	-0.151*	-0.114***
Initial share Industry	0.005	0.054	-0.049*	0.284*	0.059	0.010	0.094*	-0.023	-0.074	0.753*	-0.345***	-0.304	-0.333*	-0.952***
Initial share Services	-0.126	0.329	-0.448***	0.898	0.784	-0.068	0.284	0.139	-0.317*	0.691	-0.767***	-0.654	-0.371	-1.780***
Change in share Agriculture	-0.071**	-0.026	0.006	-0.141	-0.148**	-0.002	-0.057*	-0.019	-0.118	-0.314***	0.040	-0.125	-0.096	0.013
Change in share Industry	0.186	0.224*	0.005	-0.276	-0.006	0.003	-0.004	0.023	-0.242*	0.022	0.051	-0.019	0.294	-0.331
Change in share Services	0.189	0.628	-0.549***	-0.402	0.200	-0.135	-0.142	-0.099	-1.524***	0.130	-0.327	-1.118*	-0.568	-1.225*
Population density	-0.018	-0.051**	0.006	-0.061**	-0.083***	-0.010	-0.019*	-0.008	0.024	0.192**	0.021	0.311***	-0.017	0.045
λ	-0.411	5.297***	1.059***	-1.879	-3.701***	2.493***	1.883***	2.702***	-0.488	-4.226***	1.274***	1.861	4.799***	-5.316***
ρ	-1.284	-5.603***	-1.871***	-1.128	-4.965**	-2.804**	-0.840	-1.388	-2.085	-1.416	-3.594*	-2.885**	-4.092	1.851
Constant	0.518	-1.077***	-0.626***	2.672**	0.500	0.133	-0.630	-0.216	-1.539***	2.622***	-1.191***	-3.624***	-6.863***	-0.757

Regressions using residuals of explanatory variables

	AT	BE	DE	NL	SE	UK	ES	IT	PT	HU	PL	BG	RO	RS
Rank - $\beta_{r_{j,0}}$	0.043***	0.010**	0.013***	-0.021	0.025**	-0.014**	-0.002	-0.007	0.056***	-0.098***	-0.016	-0.141***	-0.042*	-0.089***
Initial GVA/POP - $\beta_{\gamma_{j,0}}$	-0.013	-0.090**	-0.082***	0.083	-0.479**	0.028	-0.137	0.037	-0.243***	-0.218***	-0.003	-0.120	-0.254**	-0.309***
Initial share Agriculture	-0.005	0.014*	-0.009	0.063***	-0.149***	0.000	-0.013	-0.022**	-0.083***	0.214*	-0.015	0.030	-0.157**	-0.116***
Initial share Industry	0.005	0.053	-0.047	0.292**	0.025	0.009	0.100*	-0.024	-0.091*	0.682	-0.300***	-0.385	-0.346*	-1.002***
Initial share Services	-0.121	0.326	-0.439***	0.922	0.883	-0.070	0.340	0.137	-0.381**	0.703	-0.687***	-0.734	-0.389	-1.876***
Change in share Agriculture	-0.075**	-0.026	0.004	-0.159*	-0.122**	-0.002	-0.059**	-0.021	-0.168**	-0.179*	0.034	-0.081	-0.098	0.014
Change in share Industry	0.174	0.222*	0.011	-0.362	0.111	0.001	0.004	0.022	-0.274**	0.432	0.065	0.120	0.318	-0.424
Change in share Services	0.262	0.622	-0.500**	-0.509	0.649	-0.137	-0.083	-0.099	-1.609***	0.709	-0.243	-0.912	-0.525	-1.322*
Population density	-0.020	-0.052**	0.005	-0.062**	-0.072***	-0.011	-0.019*	-0.008	0.020	0.265***	0.023	0.304***	-0.017	0.036
λ	-0.532	5.348***	1.054***	-1.830	-3.917***	2.487***	1.855***	2.696***	-0.430	-4.290***	1.335***	1.719	4.844***	-5.299***
ρ	-1.092	-5.629***	-1.850***	-1.326	-4.613**	-2.775**	-0.957	-1.378	-2.921	-2.095	-1.807*	-2.899**	-4.062	2.111*
Constant	0.467**	-1.484***	-0.079	0.832**	1.597***	-0.060	-0.308*	-0.324*	0.286	3.362***	-0.189	-0.332	-4.552***	2.883***
Number of observations	35	44	412	40	21	139	52	110	30	20	66	28	42	25

Note: Stars *, **, *** indicating significance at the 10%, 5% and 1% level respectively.

Conclusions

The paper has developed a model to consistently estimate agglomeration and agglomeration economies in a single framework. The estimated agglomeration (economies) coefficients are likely to be superior to other measures of agglomeration, as the former are based on the (statistical) distribution of economic activity across regions and are thus independent of the number of regions, while the latter are not and thus are prone to inconsistency, especially in cross country comparisons. Furthermore the paper also outlined a model of how to include agglomeration effects in regional convergence models, which is considered to be a useful extension of existing model and thus deserves some further attention.

From an economic policy point of view, the results suggest that high growth of less developed countries is connected to increasing agglomeration and agglomeration economies, while the high income levels in the more developed countries are usually associated with lower agglomeration economies. For economic policy this is a difficult gap to close, as it may mean that in the stage of economic catching up, an increase in regional disparities between low and high income regions may be inevitable, but over the longer run these disparities need to be reduced. Hence the choice policy has to make is whether to support agglomeration trends and thus to strengthen short run growth, or whether to target longer run economic development via supporting less developed regions and thus reducing potential gains from agglomeration.

References

ALPEROVICH G., 1984, The size distribution of cities: On the empirical validity of the rank-size rule, *Journal of Urban Economics*, 16.

- 1988, A new testing procedure of the rank size distribution, *Journal of Urban Economics*, Volume 23, Issue 2, March.

ARROW K. 1962, The economic implications of learning by doing, *Review of Economic Studies*, Vol. 29.

AXTELL R., 2001, Zipf Distribution of U.S. Firm Sizes, *Science*, Vol. 293 no. 5536.

BRAKMAN S., GARRETSEN H. and VAN MARREWIJK C., 2009, *The new introduction to geographical economics*, Cambridge Universtiy Press, 2nd edition.

BRÜLHART M., TRAEGER, R., 2003, An Account of Geographic Concentration Patterns in Europe, *IIIS Discussion Paper,* No. 2; *GEP Research Paper,* No. 03/11.

CHEN Y, 2011, Unraveling the Rank-Size Rule with Self-Similar Hierarchies, arXiv:1104.4381 [physics.soc-ph], http://arxiv.org/abs/1104.4381

CICCONE A., 2002, Agglomeration effects in Europe, *European Economic Review,* No. 46.

DURANTON G., 2006, Some foundations for Zipf's law: Product proliferation and local spillovers, *Regional Science and Urban Economics*, Vol. 36.

ELLISON, G., GLAESER, E. L. 1997, Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, Vol. 105, 5.

FUJITA M., KRUGMAN P. and VENABLES A.J., 1999, *The spatial economy: cities, regions and international trade*, MIT Press, Cambridge MA.

FUJITA M., MORI T., 2005, Frontiers of the New Economic Geography, *Papers in Regional Science*, Volume 84, Issue 3, August 2005.

GABAIX X., 1999a, Zipf's law for cities: an explanation, Quarterly Journal of Economics, 114.

- Zipf's law and the Growth of Cities, The American Economic Review, Vol.89 No.2 Papers and proceedings.

GABAIX X., IBRAGIMOV R., 2007, Rank-1/2: a simple way to improve the OLS estimation of tail exponents, NBER Technical Working Paper, No. 342.

GIESEN K., SÜDEKUM J., 2011, Zipf's law for cities in the regions and the country, Journal of Economic Geography, Vol. 11.

GONZÁLEZ-VAL R., LANASPA L. and SANZ-GRACIA F., 2014, New Evidence on Gibrat's Law for Cities Urban Studies, Vol.51(1).

GREUNZ L., 2004, Industrial structure and innovation – evidence from European regions, *Journal of Evolutionary Economics*, Vol. 14.

HALLET M., 2000, Regional specialisation and concentration in the EU, *DG ECFIN Economic Papers*, No. 141.

HINLOOPEN J., VAN MARREWIJK C., 2006, Comparative Advantage, the Rank-size Rule, and Zipf's Law, *Tinbergen Institute Discussion Paper* TI 2006-100/1.

HSU W., 2012, Central Place Theory and City Size Distribution, *The Economic Journal*, Vol. 122 (September).

JACOBS J., 1969, *The economy of cities*, Vintage, NewYork.

KRUGMAN P., 1991, Increasing Returns and Economic Geography, *Journal of Political Economy*, Vol. 99, No. 3.

MARSHALL A., 1890, *Principles of economics*, Macmillan, London.

MAUREL, F., SEDILLOT, B., 1999, A measure of the geographic concentration in french manufacturing industries, *Regional Science and Urban Economics*, Vol. 29.

MELO P.C., GRAHAM D.J. and NOLAND R.B., 2009, A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics*, Vol. 39.

NITSCH V., 2005, Zipf zipped, Journal of Urban Economics, Vol. 57.

OTTAVIANO G., PUGA D., 1998, Agglomeration in the Global Economy: A Survey of the 'New Economic Geography', *The World Economy*, Volume 21, Issue 6, August 1998.

PUGA D., 2010, The Magnitude and Causes of Agglomeration Economies, Journal OF Regional Science, Vol. 50, No. 1.

REDDING S.J., 2009, Economic Geography: A Review of the Theoretical and Empirical Literature, *Centre for Economic Performance Discussion Paper*, No 904, January 2009.

- 2010, The Empirics of New Economic Geography, Journal of Regional Science, Vol. 50, No. 1.

ROMER P., 1986, Increasing returns and long-run growth, *Journal of Political Economy*, Vol. 94: 1002–1037.

ROSEN K.T., RESNICK M., 1980, The size distribution of cities: An examination of the Pareto law and primacy, *Journal of Urban Economics*, Volume 8, Issue 2, September 1980.

SHEPOTYLO O., 2011, Cities in Transition, *Kyiv School of Economics and Kyiv Economics Institute Discussion Paper*, No. 42.