MAGYAR TUDOMÁNYOS AKADÉMIA
Közgazdaság- és Regionális Tudományi Kutatóközpont

MTA KRTK KTI

Centre for Economic and Regional Studies
HUNGARIAN ACADEMY OF SCIENCES

MŰHELYTANULMÁNYOK DISCUSSION PAPERS

MT-DP – 2015/26

Grid and shake - Spatial aggregation and robustness of regionally estimated elasticities

GÁBOR BÉKÉS – PÉTER HARASZTOSI

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
HUNGARIAN ACADEMY OF SCIENCES BUDAPEST, 2015

Discussion papers
MT-DP – 2015/26

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences

KTI/IE Discussion Papers are circulated to promote discussion and provoque comments.
Any references to discussion papers should clearly state that the paper is preliminary.
Materials published in this series may subject to further publication.

Grid and shake - Spatial aggregation and robustness of regionally estimated elasticities

Authors:

Gábor Békés
senior research fellow
Institute of Economics - Centre for Economic and Regional Studies
Hungarian Academy of Sciences
and CEPR
e-mail: bekes.gabor@krtk.mta.hu

Péter Harasztosi
analyst
Magyar Nemzeti Bank
e-mail: harasztosip@mnb.hu

June 2015

ISBN 978-615-5447-88-4
ISSN 1785 377X

Grid and shake - Spatial aggregation and robustness of regionally estimated elasticities

Gábor Békés – Péter Harasztosi

Abstract

This paper proposes a simple method measuring spatial robustness of estimated coefficients and considers the role of administrative districts and regions' size. The procedure, dubbed "Grid and Shake", offers a solution for a practical empirical issue, when one compares a variables of interest across spatially aggregated units, such as regions. It may, for instance, be applied to investigate competition, agglomeration, spillover effects. The method offers to (i) have carry out estimations at various levels of aggregation and compare evidence, (ii) treat uneven and non-random distribution of administrative unit size, (iii) have the ability to compare results on administrative and artificial units, and (iv) be able to gouge statistical significance of differences. To illustrate the method, we use Hungarian data and compare estimates of agglomeration externalities at various levels of aggregation. We find that differences among estimated elasticities found at various levels of aggregation are broadly in the same range as those found in the literature employing various estimation method. Hence, the method of spatial aggregation seems to be of equal importance to modeling and econometric specification of the estimation.

Keywords: economic geography, firm productivity, agglomeration premium, spatial grid randomization

JEL classification: R12, R30, C15

Acknowledgement: The authors gratefully thank the Hungarian Academy of Sciences for its Firms, Strategy and Performance Lendület Grant. Békés thanks MTA Bolyai research grant. The views expressed are those of the authors and do not necessarily reflect the official view of the Magyar Nemzeti Bank.

Rács és keverés: területi aggregálás és a regionális szinten becsült együtthatók robusztussága

Békés Gábor – Harasztosi Péter

Összefoglaló

A tanulmány regressziókból becsült együtthatók térbeli robusztusságának egyszerű mérési módszerére tesz javaslatot, valamint a közigazgatási körzetek és régiók méretének szerepét vizsgálja. A „rács és keverésnek" nevezett eljárás megoldást kínál egy gyakorlati empirikus problémára, amely akkor merül fel, amikor a vizsgálni kívánt változókat területileg aggregát egységeken keresztül hasonlítjuk össze. Az eljárást alkalmazni lehet például verseny, agglomeráció vagy átterjedési hatások vizsgálatára is. A módszer képes (i) a különböző aggregálási szinteken becslések végzésére és az eredmények összevetésére, (ii) a közigazgatási egységek méretének egyenlőtlen és nem véletlenszerű eloszlásának kezelésére, (iii) közigazgatási és mesterségesen létrehozott egységeken alapuló eredmények összehasonlítására, (iv) az eltérések statisztikai szignifikanciájának mérése. A módszer szemléltetésére magyar adatok felhasználásával agglomerációs externáliák számításait hasonlítjuk össze különböző aggregátsági szinteken. Azt állapítjuk meg, hogy a számított rugalmasságok különböző aggregátsági szinteken mutatkozó eltérései körülbelül hasonló tartományban mozognak, mint amelyek a szakirodalomban a különböző identifikációs módszerek alkalmazásával előálltak. A térbeli aggregálás módszerének megválasztása tehát hasonló mértékben tűnik fontosnak, mint az ökonometriai modell specifikációja.

Tárgyszavak: földrajzi közgazdaságtan, vállalati termelékenység, agglomerációs prémium, területi rács randomizáció

JEL kódok: R12, R30, C15

Grid and shake - Spatial aggregation and robustness of regionally estimated elasticities[☆]

Gábor Békés[a,*], Péter Harasztosi[b]

[a] *MTA Center of Economics and Regional Studies, Central European University and CEPR*

[b] *Magyar Nemzeti Bank*

Abstract

This paper proposes a simple method measuring spatial robustness of estimated coefficients and considers the role of administrative districts and regions' size. The procedure, dubbed "Grid and Shake", offers a solution for a practical empirical issue, when one compares a variables of interest across spatially aggregated units, such as regions. It may, for instance, be applied to investigate competition, agglomeration, spillover effects. The method offers to (i) have carry out estimations at various levels of aggregation and compare evidence, (ii) treat uneven and non-random distribution of administrative unit size, (iii) have the ability to compare results on administrative and artificial units, and (iv) be able to gouge statistical significance of differences. To illustrate the method, we use Hungarian data and compare estimates of agglomeration externalities at various levels of aggregation. We find that differences among estimated elasticities found at various levels of aggregation are broadly in the same range as those found in the literature employing various estimation method. Hence, the method of spatial aggregation seems to be of equal importance to modeling and econometric specification of the estimation.

Keywords: economic geography, firm productivity, agglomeration premium, spatial grid randomization
JEL R12, R30, C15

[☆] The authors gratefully thank the Hungarian Academy of Sciences for its Firms, Strategy and Performance Lendület Grant. Békés thanks MTA Bolyai research grant. The views expressed are those of the authors and do not necessarily reflect the official view of the Magyar Nemzeti Bank.

[*] Corresponding author: Bekes: MTA KRTK Budaörsi ut 45, Budapest, 1112 Hungary; bekes.gabor@krtk.mta.hu

June 18, 2015

1. Introduction

Economic activity is neither evenly nor randomly spread in space. As a result, there are several areas of research where we are interested in how this uneven and non-random distribution may be related to performance or behavior of economic agents. One may look for evidence on the relationship between wages or firm performance and agglomeration externalities (Ciccone and Hall, 1996), entrepreneurship and the stock of firms (Ács and Armington, 2004), retail prices and local conditions (Iyer and Seetharaman, 2008) or the scope of local competition and vertical integration (Hortaçsu and Syverson, 2007; Csorba et al., 2011), or even study hospital performance in French regions (Gobillon and Milcent, 2013). In these cases, we have observations at a local level as independent variable and we are interested in how some feature of the neighborhood is related to the behavior of this observation.

Importantly, oftentimes, we need to understand how specific features of selecting that neighborhood may affect the outcome of the exercise. This offers a method to transparently investigate the impact of modeling choices of spatial aggregation.

These estimations have typically been carried out at some local level first (say, municipality) followed by additional estimation at some regional level. Regions were useful as they allowed for a larger area for spillovers to work or take into account competition for nearby customers. However, distance between economic agents may affects the ease of spreading of ideas, forming relationships, the likelihood of matching sellers and buyers and hence trade, competition of outlets, and other interactions. As a result, when we study how some outcome variable is affected by some feature of its neighborhood (such as population density), it is hard to know ex-ante what is the appropriate size of spatial aggregation we shall consider. Furthermore, the spatial scope of an activity may or may not coincide with the spatial scope of a polity. At the same time, policies may affect relationships wherever interacted. Hence, measurement at some regional level combines the role of size (inclusion of externalities within) as well as the possible impact of regional policies.

There is substantial evidence in the literature arguing that measurement matters (see Burger et al. (2010) for a review). The spatial scope matters as externalities are often found to be in operation but fading away on distance. For instance, Fotheringham and Wong (1991) argue that features of areal units affect regression results - often in an uncertain fashion, Dewhurst and McCann (2007) gave evidence that the relationship between specialization and urban structure is a function of the scale of analysis. Burger et al. (2010) found different effects of agglomeration forces across geographic levels in the Netherlands. In Andersson et al. (2014), Swedish employment data is used at a 1×1 km and 250×250 m sized grid. Furthermore, unit observation size and some of its features may be correlated and hence, estimates may be biased. For instance, large municipalities may be clustered in space because of first geography, specialization in agriculture, hard barriers to migration, distance to markets - many aspects of spatial economic structure we care about.

This paper offers a simple way to discuss robustness of estimates vis-a-vis spatial features of the data and allow us asking additional questions about spatial decay or administrative units. These features may include average size of administrative units, heterogeneity of unit size and the correlation of unit size with other variables of interest. Our method can offer to: (i) have easy scalability to compare evidence on differently sized units, (ii) treat uneven and non-random distribution of administrative unit size, (iii) be able to gouge statistical significance of differences, and (iv) have the ability to compare results on administrative and artificial units.

The core idea of our procedure is to set up an artificial grid over the map of the country, assign municipalities and then randomize to get a distribution of results. To do this, first, we create a grid of squares of a given size, L. Second, we take this grid of $L \times L$ squares and randomize its place on a map and hence, create a set of random realizations of artificial spatial units. By repeating this procedure (in this paper, a 1000 times) we can get smooth kernel distributions. Third, we correct for the border effect by merging very small areas into bigger ones at the border of the country. Finally, we run our regression on each of these realizations and gather coefficients that yield a distribution of our parameter of interest.

After this procedure, we can look at moments of this distribution, relate distribution of coefficients to a single measure, say, at the administrative regional unit, or compare distributions from two randomizations.

To illustrate our procedure, we consider two examples. First, we take a classic issue in geographical economics, how is firm performance and hence, wage may be related to agglomeration, measured as population density. We estimate an agglomeration externalities or the elasticity of wage on population size controlling for the size of region. Second, we consider a simple case in entrepreneurship: the birth of new firms as the function of existing stock of firms. Here we estimate how the quantity of new firms in a region is related to the stock of firms and education. We compare elasticities at two administrative units, municipal (NUTS5) and micro-region (NUTS4) levels, and three artificial units. For both examples, we are interested in the relationship between an estimated coefficient at an administrative regional level and a comparable estimate at an artificial region of similar size as well as how the average size of area affects the estimated coefficient.

With over three thousand municipalities for a country of 10 million, there are plenty of small administrative units. For the similarly populous (but much larger) Sweden, we find 290 municipalities, Portugal has 309, while Austria has 2300 - the largest for medium-sized countries. Hence, using Hungarian data allows us a great deal of experimentation.

Treatment of the location issue will typically include spatial regression methods (Anselin, 2010), and spatial lags in various forms have been widely used to gauge uncertainty about the right spatial scope. A close exercise to ours was proposed by Briant et al. (2010) who would randomly place starting points on the map, and add up equal number of municipalities until the country is fully covered. We will argue that our methodology works similarly in some aspects to existing methods but is better able to capture size heterogeneity and study the impact of administrative borders. Once a digital map of the space in question is prepared our method is very easy to use and can be adjusted flexibly. Furthermore, it provides solutions for spatial measurement problems caused by systemic size differences of administrative units. Finally, this approach is also related to Abadie et al. (2010), who uses synthetic geographical control methods to evaluate a tobacco control program.

In what follows, we first describe four issues that shed light on why measurement methodology matters. Following these points we introduce the methodology we developed and offer some comparison with existing methods. Then, we present the data to illustrate the method and describe the examples. Finally, we show empirical results and discuss applications and conclude.

2. Measurement issues

Before discussing the method we are proposing, let us present some key considerations that motivated the procedure. First, we argue that unit size matters and there is a need to compare results across spatial units of different sizes. Second, uneven and non-random distribution of unit size matters as well. Third, in the presence of policy aspects, one shall be able to understand the influence of administrative borders when investigating polity effects. Finally, instead of a randomly chosen robustness test, we should be able to perform statistical inferences and hypothesis testing.

2.1. Unit size matters, scalability is key

There is substantial evidence suggesting that the impact of spatial externalities decays rapidly with distance, i.e. measured elasticities will decline based on unit size of measurement. Hence the presence of externalities may be detected at some level of spatial unit, but not at another. This issue is partly an econometric problem: A mismatch between the spatial unit of observation and the spatial extent of the phenomena under consideration will result in spatial measurement errors and spatial autocorrelation between these errors in adjoining locations. (Anselin and Bera, 1998). At the same time, this may be important for policy as well as simply comparing evidence across countries or regions.

The first issue the size of areas or the bandwidth of this filter. There may be several problems related to measurement at various aggregation levels of space. First to fully capture spillovers, size of areas of observation may matter, and different level of aggregation may give rise to different estimates. For instance, the impact of an externality may weaken over distance for a wide range of activities such as trade of intermediate inputs, productivity and wages, knowledge spillovers.[1] Similarly, the scope of spatial competition may be affected by infrastructure or regulation. As a result, understanding relationships over different levels of aggregation may shed light on the channels and mechanisms.

Consider the example of EU regions to show how the level of aggregation also changes correlation inferences. Figure 1 shows the population density distribution of Western European countries at NUTS1 and NUTS3 level. One can see that even within the same administrative level the size and shape of the basic units differ a lot.

For example, the correlation between population density and GDP per capita across European regions in Figure 1 at NUTS1 level is 0.74. The same relationship at the more disaggregated level of NUTS3 administrative units is much weaker, 0.54.[2] Consequently, simple ordinary least squares regressions on the elasticity of density on GDP also yield different estimates. At NUTS1 level we would find 17 percent, while at NUTS3 level only 13 percent.

Hence, the first important requirement is for a methodology that can be scaled by size of units of observation (ie. compare distributions from two randomizations, one at $L_1 \times L_1$ sized grid and another one at an $L_2 \times L_2$ grid), allowing for a direct comparison.

2.2. Uneven and non-random distribution of unit size matters

The second issue is the uneven distribution of size. More precisely the non-random uneven distribution of size. Maybe first geography (lakes or mountains) prompts people organize themselves in small units, or dense urban areas set up small but populous spatial units. By missing this, an exercise may yield a biased result. In some countries and some levels of aggregation this is a very minor issue. In France, for example, departments are fairly equal sized. [3] However, this is not the case in most countries and regions.

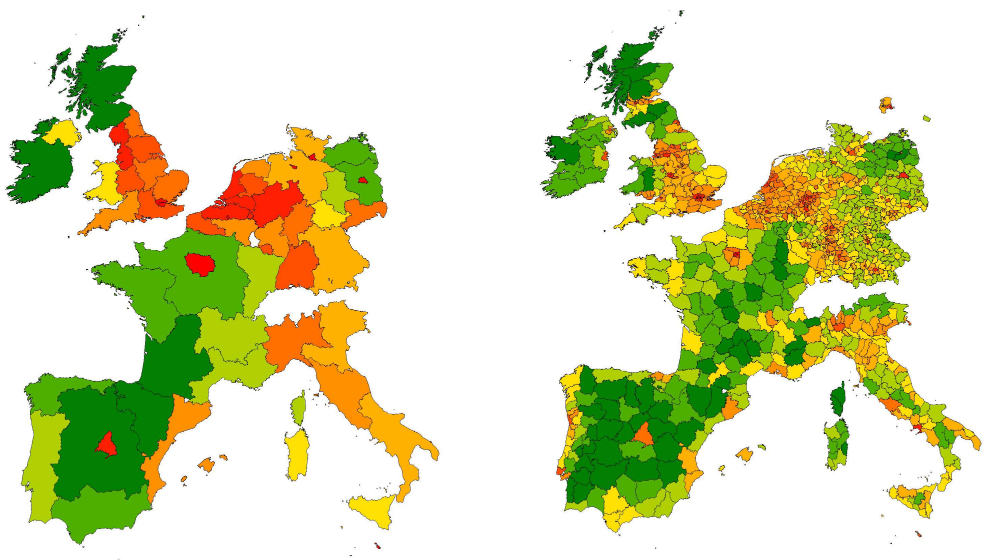
[1]For instance, trade proximity matters for localization for about 50km suggested by Duranton and Overman (2008) for manufacturing industries in Britain. Rosenthal and Strange (2008) estimate a wage increasing effect of being close to educated people falls to just 25% as the distance rises from 5 to 15 miles in the US. Lychagin et al. (2010) finds that the crucial range for proximity to labs is limited to 100-200 kms. Andersson et al. (2014) documents attenuation of wage density elasticity even within cities.

[2]We use 2007 euro per capita measure of GDP and 2006-2008 average population density from the Eurostat regional databases. The countries that constitute our sample: United Kingdom, Germany, Belgium, France, Spain, Portugal, The Netherlands and Italy.

[3]Note that, shape of units may be of interest.Briant et al. (2010) shows it does not matter. And we do not care here, but our method unifies shape as well.

Going back to EU NUTS3 regions, and looking at the representation of population density distribution, it is obvious that the dense regions are much smaller.[4]

Figure 1: Population density in European regions



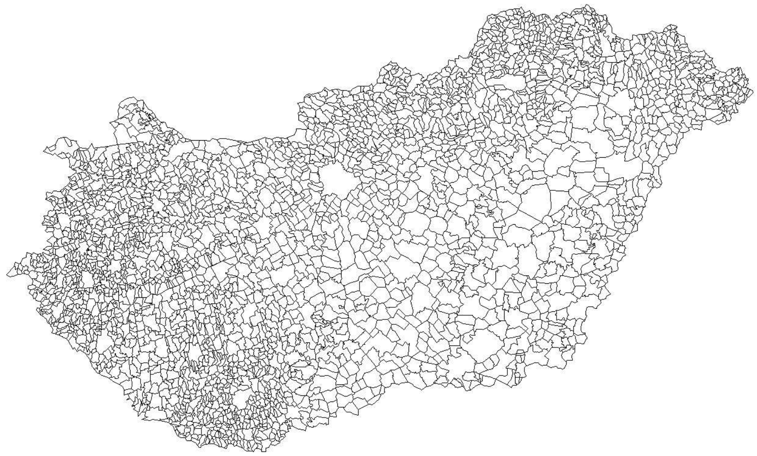
Importantly, not only do we see that small and large regions are clustered (i.e. North European regions are small, Spanish regions are large) but that this is correlated with first geography (Spanish mountains, access to Atlantic trade routes) and urbanization (abundance of mid-sized North-German cities, possibly related to system of Länder). Therefore, the more disaggregated data one uses, more the number of dense regions increase in the sample. This may affect inference about concentration and also give less weight to less populated areas.

Take another example. In this paper, we will work with data from Hungary, the municipal and micro regional structure is rather unbalanced (Figure 2) -unlike in France, where large administrative regions called départements are about equal size. We see an average municipal size of 15-20 km^2 in South-West and 70-75 km^2 in North-East, and theres is a great deal of variation in the size of micro regions, as shown in Table 1. The small size of typical administrative units in the South-West is correlated with rugged terrain, history, and agricultural traditions. These are all issues that may bias estimated coefficients.[5]

[4]This is of course the direct consequence of the political use of spatial units. Districts tend to have an equalizing role across population.

[5]When regressing the (log) number of municipalities within a district on (log) wages, we find a negative correlation. This is primarily due to the fact that there are plenty of small municipalities, close to Vienna with decent market access and higher wages. Using a cross section of 2002, we have: $log(No)$=11.8+0.8*$log(Wage)$, significant at 5%; $log(No)$=1.41+

Figure 2: Map of Hungary at municipal level



As a result, the second expectation is the ability to treat clustering of large or small units, and thus, the ability to compare aggregated units of the same size rather than the same sub-units.

2.3. The role of administrative borders may be directly investigated

The last issue is related to administrative unit borders. There is yet another consequence of the size variation administrative units. For instance, municipality size contains also the agricultural land that surround the cities, hence those with more agricultural land will appear less dense. In Hungary eastern municipalities have larger plots of land at their disposal which implies that western municipalities will be calculated as denser.

Administrative units will also determine policy. They may be official seat of policy making bodies such as employment agencies, set their taxes (such as Bacher and Brülhart (2013)) or offer subsidies.

Regions are often polity areas, where local bureaucracies set policies on labor market (often by regional employment agencies), firm subsidies (by regional development agencies) or regulation (e.g. by regional environment protection agencies). It may be an explicit research question to understand the effectiveness of such policies. Hence the ability to compare administrative regions with a set of spatial units artificially defined matching average regional size, may offer a useful insight.

0.05*$log(Wage)$+0.45*$log(Distance_{Budapest})$ -0.24* $log(Distance_{Vienna})$, both distance variables are significant at 1%

The third requirement is hence the capacity to compare administrative regional units with artificial units where size of units are kept to match (on average) the administrative regional size. With our model, we can compare this distribution of coefficients gathered from the randomized $L \times L$ grid to the single measure, say, at the administrative regional unit. This allows us understanding to what extent do we have a special realization of results, can we claim, for instance, that 95% of grids yield a smaller or larger coefficient than the administrative one.

2.4. Statistically differentiating results across aggregation

The core idea of MAUP is that location and specificities of spatial borders may determine correlations. As a result, any particular aggregation may suffer from a set of idiosyncratic shocks, particularities of regional border. To see robustness of a particular relationship, one may need to go beyond comparisons and understand the statistical significance of differences of elasticities measured at various levels of aggregation. For instance, we cannot determine if results comparing elasticities at district and regional levels (e.g., Burger 2008) differ because of imprecise estimation or because of a significant difference.

To allow for comparisons across different sized units as well as artificial versus administrative units, the method will have to able to generate randomized aggregation outcome allowing to create standard errors for regression outcome at various levels of aggregation.

3. The "Grid and shake" method

The method in this paper is simple and easy to replicate. First, we digitize the map by evenly placing markers on the country and hence pinning down municipalities. Second, we define a grid made up of $L \times L$ squares and place it over the map. Our markers will help allocating each municipality to a particular unit of the grid. Third, we define grid unit size based on some economic consideration, such as matching the average size of administrative units. Fourth, we shake this grid and thus, randomly create possible realization of spatial units. Finally, make cosmetic adjustments around the country borders to avoid having too small units. When running regressions, we can repeat any given regression on all realization and thus, get a distribution of elasticities. We dubbed this method "grid and shake".

3.1. Setting up the grid

To rearrange municipalities into spatial groups of approximately equal size, we use an idea analogous to offset printing technology. First, we place markers arranged in a raster-like manner on the municipality-level map of Hungary. Markers are defined on the positive quadrant of a co-ordinate system[6] with the origin is defined as the (0,0) km. They are placed rather close to each other (800 meters) such that the smallest municipalities have at least 3 points within the boundary of the polygon that define them. The markers (we have 153,113 marker points for the 93,030km^2 of Hungary) in turn uniquely identify the spatial territory of a municipality. Markers are matched to the polygon of the municipality[7] In Figure 3, the markers are illustrated using a small part of Hungary on the Croatian border.

As markers are assigned to municipalities they make it possible to group municipalities into larger spatial units by aggregating them. If we place a grid consisting of size L squares over the map, then we can identify the markers that belong in the same square. The square of the grid is defined by a number pair (g_x,g_y). We assign each marker to a grid-square with: $g_x = int(x/L) + 1$ and $g_y = int(y/L) + 1$, where x any y are the coordinates of the markers, L is the length of side of the square and *int* is a function that takes the integer part of a number. The coordinates of the markers being given in km's from the origin (0,0), g_x will then give the number of grids-squares we need to put on the map starting from zero to arrive at the longitude of marker (x,y). That is, if the marker coordinates are, say, 23 and 42 then using a grid size 10, the marker will be in the 3rd grid-square from the left and 5th from the bottom.

As a next step, we can form a new spatial unit by joining municipalities whose points are in the same grid. If the points that define a municipality belongs to more grids, then the municipality joins the spatial unit that holds most of its points. This method gives one alternative realization of larger spatial units anchored by the starting point (0,0). However the size-L grid could be placed on the map at random locations.

3.2. Defining grid sizes

To analyze the role of size and administrative units, we'll create three types of grids: small, medium and large. Small grids are 15km×15km, and it takes typically 440 of such squares to cover the country. This is the size that fits the largest of municipalities (basically the handful of large cities). That is, a realization with grid-size 15 produces a spatial division as if municipalities were of the same size - by aggregating small ones and keeping large cities as they are.

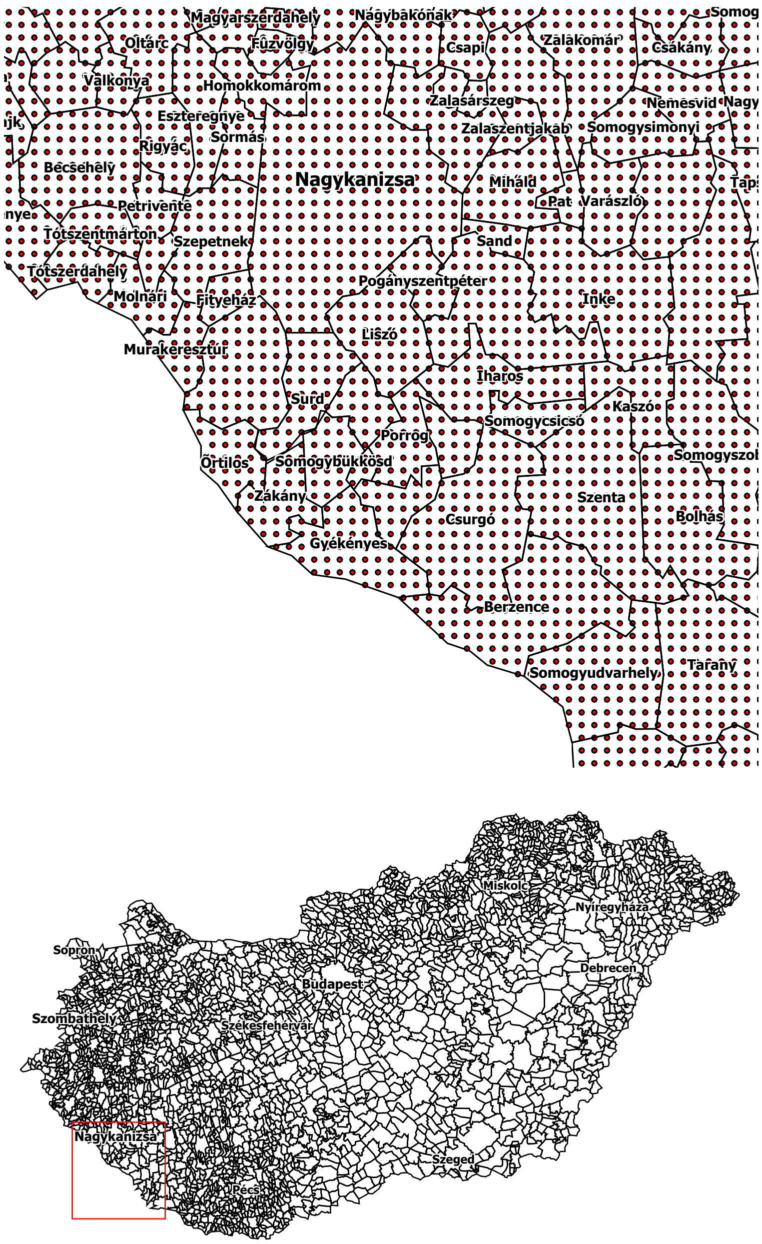
The medium sized grid is made up of 26×26 kilometer squares. We picked this size to match the number of micro-regions, 150. This is apparent from Table 1 which summarizes area and count data for the different spatial aggregations.

Figure 4 shows the distribution of number of municipalities per grid. As we argued

[6] Mercator WGS84 EPSG:41001 projection

[7] Polygon matching is done in a GIS mapping software such as Mapinfo. Otherwise, estimations may be done in a statistical software.

Figure 3: Markers serve as basic elements to be shocked



earlier, Hungarian municipality sizes vary a great deal, non-randomly, hence the deviation. About 30% of units house only few number of units (no more than 10) while, about 25% include many (31+). Obviously there is a great deal of variation. Note that as the graph suggests, the distribution of number of municipalities per unit (both when considering the first realization or the average 1000 permutations) matched fairly closely that of NUTS4

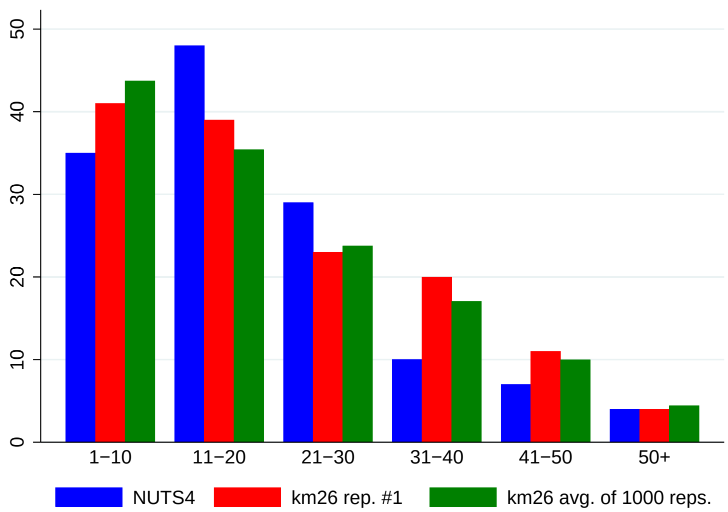
Table 1: Summary of Hungarian administrative spatial units and grids

	number of units	area km sq. mean	min	max
Municipality	3125	29.55	0.71	483.22
km15 grid	440	210.37	1.52	607.21
Micro-region	150	620.23	103.08	1573.10
km26 grid	150	605.35	67.61	1146.01
km39 grid	75	1235.50	152.20	2060.82

Hungarian municipalities are NUTS5 level units, micro-regions are NUTS4 level units.
Figures for grids are averages over 1000 replications

regions.

Figure 4: Distribution of units that include few and many municipalities



The largest size of grid squares in 39×39 to match the largest micro-region, 1,541km^2. This will allow us simulate what would happen if all micro-regions were equally sized.

Having these three artificial grids of sizes, with 15×15, 26×26, 39×39 squares, will allow simply comparing the spatial features of relationships, such as the spatial magnitude of agglomeration spillovers.

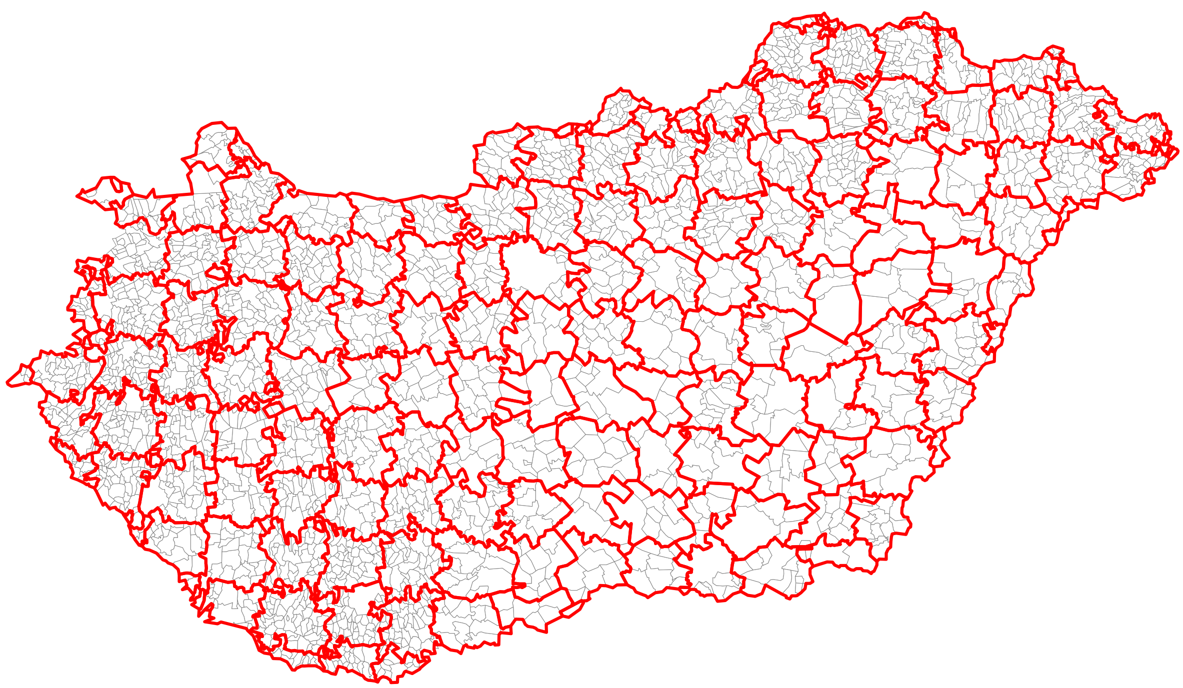
3.3. Randomization

The most problematic aspect of both the administrative setting and the artificial grid is based on the arbitrary location of units - one of the basic MAUP problems. When considering robustness or validity of results at a certain level aggregation, the choice of the actual location of units is rather arbitrary. As we will show later, two randomly placed grids may often yield rather different outcomes. Hence, we will draw not one but a thousand grids to make sure we would not face this MAUP problem. To treat this arbitrary nature, we randomly generate 1000 possible grids - by taking the grid of LxL and "shaking it" over the map of Hungary.

The procedure works as follows. We select a random staring point by adding independently drawn random numbers ξ and ζ to the (0,0) starting point, which are uniformly distributed over the 0 to L interval. Then using the algorithm above, we assign markers to L×L sided grid-squares. The grid thus defined will be independent from the starting point (0,0). It is worth noting that picking random numbers from (0,L) is enough as ξ+L and ζ+L would yield the same realization as ξ and ζ.

Having assigned a grid (a pair of numbers) to each point, we define a municipality to belong to a grid if most of the points that constitute a municipality is coupled with the same number-pair. Figure 5 illustrates a realization of random grid based spatial units. Based on the data defined by this realization we can run regressions and save coefficients.

Figure 5: A realization of the 26×26 km spatial units



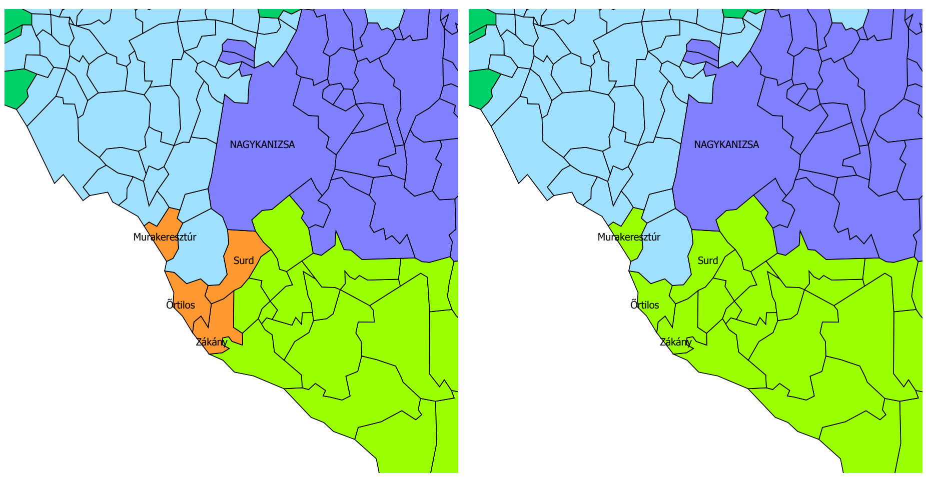
This is repeated a thousand times, yielding a thousand coefficients. Finally, we take these saved coefficients and consider the distribution at various grid sizes, with L=15km, 26km, 39km.

3.4. Border correction

A disadvantage of the method is that small areas are likely to get drawn by the border. This is illustrated in the left panel of Figure 6, where the random grid joins only three municipalities at the border, including Őrtilos and Zákány.

To partially offset this, we considered all areas (created by the grid) that are smaller than the a pre-determined threshold (set as 10% of $L \times L$) and added it to a randomly chosen neighboring region. The right panel of Figure 6 shows the adjusted spatial units. In this case, the three municipalities are added to the larger region to the east, hence increasing its size.

Figure 6: Adjustment for small areas near the border: 26 $\times$ 26 km



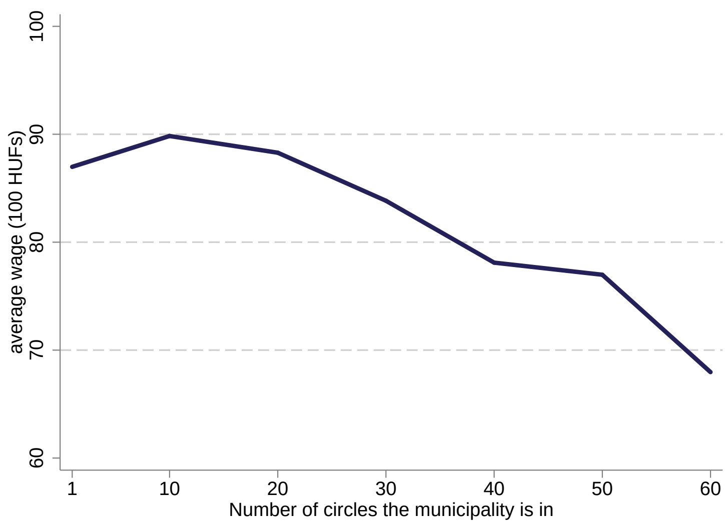
(a) Before adjustment (b) After adjustment

3.5. Comparison with other methods

To evaluate the performance of our "Grid and Shake" method, let us compare our method to existing methods. The most basic method is aggregating units (markets) within a circle with a given radius around an observation. For instance Martin-Barroso et al. (2010) use this method to correct for agricultural area of municipality size when estimating agglomeration elasticities. This simple spatial aggregator function method assumes that we are uncertain about the market size or the extension of some externality and consider access to be symmetric around the observation. In this process, we set up circles around each (left hand side) observation and aggregate units. The advantage of this method is simplicity and flexibility regarding size.

The disadvantage is a bias introduced by oversampling of areas composed of many small municipalities. To better see this, we look at how the Circles method would work in Hungary. There are 3125 municipalities, so we create 3125 circles with a center located in each of these municipalities. The circles have a radius of 14.5km so that its area would match the average size of NUTS4 regions as well as the average area of the size of our grid of 26×26km. For each municipality, we calculate the average wage, as well as count the number of circles the municipality is included in. We find the following correlation: richer municipalities are counted less frequently than poorer ones, thus biasing estimates, see Figure 7. This does not happen when estimates are based on disjunct areas, where all municipalities are counted once.

Figure 7: Frequency a municipality is included in a circle and it's mean wage



Number of circles is created by first calculating for each municipality, the number of circles this municipality is included, and then grouping them by this count. Average wage is calculated for groups of municipalities by this count as well.

Spatial dynamic or distributed lags models had also been used to capture the potential correlation of the independent variable across space. For instance Cainelli et al. (2014) analyses firm exits as a function of agglomeration and uses neighboring Italian regions in a dynamic panel model estimated by GMM. Similarly to our model, this allows testing of decay over space. However, it will include equal number of neighbors independently of size, whether an observation is surrounded by small units or large ones.

Our approach of creating spatial district is closest to Briant et al. (2010), who investigate the effects of MAUP on inference using French administrative units and use two approaches. First, they place a grid over France which creates the same number (N) of new spatial units as the one they wish to compare to. This creates square shaped units with only one realization. Second , to match the numbers of existing administrative units, they randomly pick basic units of the same number (N) and merge them with neighboring units. The merging is repeated until all shall units are out. Repeating this exercise many time yields various representation of France[8]. Note, that the first method is without repetition can be influenced by how the grid is placed. In the case of the second method, the size distribution of the basic units will influence the size of the created spatial unites. In France, due to the roughly equal size basic spatial units, this is not a problem. However would not be entirely satisfactory in the case of Hungary. Considering the large amount of small municipalities in the west of Hungary, and large municipalities in the east part of the country, a replication of the Briant et al. (2010) method would result in very large units the east and tiny ones in the east. Our main goal is however to use basic units of approximately equal size. Consequently, our method is an alternative to the Briant et al. (2010) that can be applicable to countries with basic spatial units that are rather heterogeneous in size.

We believe that our method improves on existing ones (spatial lags (SL), dots to boxes (DtB), added area under fixed radius circles (Circles), or in other words, unifies some advantages from these approaches.

First, it offers a more transparent way to compare estimates for differently sized areas. Both SL and DtB add units rather than distance while distance weighted SL and circles are similar to our approach.

Second, our approach ends up comparing equally sized units thus avoiding bias from endogeneity of spatial structure (i.e., concentration of small units in particular areas). This is true for the Circles model, but both SL and DtB adds units rather than distance. For Hungary this is rather important given large variation in size - as suggested in Graph 7. Aggregating equal number of units (ie 21 in our case) would obviously yield a different structure. Importantly, our method counts all units only once and hence, does not oversample areas with many small units (such as Circles). While it improves on some SL models as well, it is less certainly flexible in modeling non-linear relationships within units.

[8]Their goal is to create artificial spatial units in the same number as the number of an existing districting scheme (N) such as French départements. To do that, they draw N of the smallest spatial unit they observe - called seeds, joined repeatedly with randomly chosen, neighboring units. The procedure is repeated, for each seed, until there no more small neighboring units to join, all of the basic units are joined to one of the new spatial unit. This procedure results in differently shaped and sized N spatial units depending on which seeds are chosen and on the sequence neighboring units are added. The size also depends on the spatial distribution of the size of the basic units. If within the country, largely sized and small basic units are clustered, the procedure may result in large size differences across the newly created artificial units.

Third, our approach allows for analyzing the impact of MAUP by creating repeated random artificial units, and generates - in a transparent fashion - a set of coefficients. Hence, we can calculate a mean of coefficient estimates as well as standard deviation of those estimates over replications. As a result, it allows for a statistical comparison of results measured at differently sized units. To our knowledge, this is not available for SL or circles. DtB may also generate a distribution[9]. This method is not only transparent, it is also simple. Once the map is digitized, it is fast and may be modified easily.

Fourth, the method can compare administrative units to a set of artificial ones. Thus is possible in DtB but not in other models as their outcome is not disjunct spatial units.

4. Empirical applications - two examples

In this section, we present details of the Hungarian data we use for illustration purposes, introduce two topics that use regionally aggregated data. We start by looking at how wages are correlated with density followed by looking at how agglomeration may be correlated with bolder entrepreneurship measured as new firm formation. For both these cases, we present results at various administrative levels and compare them with artificial grid results.

4.1. Data

For the empirical work, we use Hungarian firm level balance sheet data from the Central Statistics Office in Hungary. We use the 1999 information for the cross section OLS, and the 2002-1999 difference when using first differences. Firms include all double accounting manufacturing companies with at least 5 employees. We have information on earlier years, allowing us creating variables for new firm formation.

Firm level financial data is matched with location information at municipal level. Unfortunately, we only know about headquarters. However, this is not a major problem for manufacturing firms in Hungary.[10] Firm level data are then matched with municipality level information on municipality size and population from the T-STAR database.

Wages are calculated from balance sheet data, using additional information on full time employees and total wage cost. The proxy for human capital is obtained from the Hungarian Labour Force Survey (HLFS). The survey includes a sample of workers in the manufacturing industries and documents various employee characteristics. This information includes their highest attained year of schooling, which is proposed here as a proxy

[9]The features of such distribution was not evident from the paper.

[10]In their appendix, Békés and Harasztosi (2013) show that only 7% of firms have multiple sites, and that on average, these firms have 1.15 plants. this suggest a small potential bias.

for human capital. The data connects workers to municipalities, however their actual employers are not identified, which prevents joining the firm level information with worker data. Using the HLFS sample, municipality level average highest year of schooling variable can be calculated to serve as proxy for labor quality. The human capital proxy shows a considerable variation from 7 to 14 years of education attained.

4.2. Example I. Agglomeration elasticity of labor productivity

One key feature of economic geography is that more densely populated areas show higher productivity and also higher wages. The positive correlation between density and productivity stems from two sources. First, natural advantages, the so-called first nature geography foster economic activity. Second, the fact that economic agents are relatively close to each other creates external economies that boost productivity and was first described by Marshall (1920). The Marshallian externalities play important role in the micro-founded explanations of growth, innovation and trade theory and the agglomeration and dispersion forces of new economic geography. See, e.g., Krugman (1991). Understanding their strength is important to evaluate economic policies that for example aim at cluster formation or at helping backward EU regions to close gap with more developed regions.

The identification of agglomeration effect is not straightforward.[11] To address spatial quality differences of industries and workers Ciccone (2002) suggests the use of human capital measures. Combes et al. (2010), who favor using individual wage panel to control for spatial skill differences, argue that controlling for the endogeneity of quality is empirically more important than controlling for that of quantity. Typically the literature finds that on average, wages or productivity of firms located in a region with twice as many people, will be 5-10 percent higher, and this higher productivity will be then translated into higher local wages. For a thorough meta-analysis on results, see Melo et al. (2009).

In this paper, we opt for considering wage as the dependent variable, as this is the most widely investigated relationship and hence, allows for international comparisons. We estimated agglomeration effect on average municipality wage with this model:

$$Wage_i = \beta_1 POP_i + \beta_2 AREA_i + \beta_3 EDUC_i + \nu_i \tag{1}$$

[11] Ciccone and Hall (1996) apply instrumental variables approach to handle simultaneity and to identify the productivity advantage stemming from spatial density. Combes et al. (2012) and Combes et al. (2008) offer some mechanism to control for spatial sorting, when more productive firms and high-skilled labor is more likely to be attracted by more dense location - is treated by . Sectors that benefit more from increasing returns in their technology may find it more profitable to locate in denser areas.

Table 2: Wage regressions

units	municipal	micro region		km15	km26	km39
Population	0.052	0.127	Population	0.077	0.118	0.134
SE	0.007	0.020	(mean) SE	0.011	0.018	0.029
			SD coefficient	0.005	0.007	0.014
Area size	-0.025	-0.125	Area size	-0.046	-0.125	-0.131
SE	0.010	0.027	(mean) SE	0.021	0.026	0.049
			SD coefficient	0.013	0.016	0.023
Years of schooling	0.950	0.818	Years of schooling	1.092	1.103	1.009
SE	0.041	0.367	(mean) SE	0.172	0.410	0.396
			SD coefficient	0.126	0.206	0.178
			Replication	1000	1000	1000
Observations	2,873	150	Observations	437	150	75
R-squared	0.248	0.384	R-squared (mean)	0.25	0.364	0.396

All variables are in logs. Population is size of local working age population. Area is in km^2. Years of schooling is average of local population. Robust standard errors are displayed. In grid models we report two statistics for standard errors: (mean) SE is the mean of robust standard error of the coefficient over 1000 replications, while SD coefficient provides the statistic for the standard deviation of the coefficient itself calculated over 1000 replications.

where i is the index of the spatial unit, $Wage$ is the average wage of employees, POP is the number of working age population, $AREA$ is the size of municipality, $EDUC$ is a proxy for education levels. This is estimated with OLS on a cross section of 1999 data, with all variables used in logs.

Results, presented in Table 2 suggest that at all estimation level, there is a positive correlation between the average wage in an area and its density (size of population, holding area size constant). The estimated coefficient is lowest when using municipal level data (5.5%) , followed by small artificial units (7.7%) , while the estimate for micro-region and medium sized artificial units (11.8%, 12.7%) are about double the size of the municipal levele estimate. Having artificial units will allows us comparing these figures attaching statistical signficance to any gap we may find.

4.3. Example II. Entrepreneurship and knowledge spillovers

This example will consider new firm formation as an example for a process with great spatial variation, where policy interventions abound at local as well as regional levels. This topic is often approached from a point of view of entrepreneurship, looking at how people start companies. Our focus, instead, is the study of spatial variation in new formation of firms and the influence of local and close-by factors and policies. This is another question with policy importance, as it can answer questions about the usefulness of regional policies, grants offered by regional authorities. We may test whether regional boundaries matter by investigating if policy impact is bounded by administrative units lines.

Regional variation in new firm formation has amply been studied. Key explanatory variables include measures of unemployment, population density, industrial restructuring, and availability of financing (Armington and Acs, 2002), income as well as market dynamics such as exits and change in industrial structure (Sutaria and Hicks, 2004). Infrastructure, such as motorways, may also affect firm entry and location (Holl, 2004).

Regarding empirical approaches, least squares and count data models have both been often used.[12] Here we estimate a Poisson model, where we regress the number of new firms on existing number of firms or the population given the size of the location and labor market conditions proxied by wages.

$$\begin{aligned} Pr(y_i = N) &= f(\lambda_i) \\ \lambda_i &= \beta_1 POP_i + \beta_2 AREA_i + \beta_3 Wage_i \end{aligned} \tag{2}$$

Results are presented in Table 3. Estimated coefficients of the population coefficient are greater than unity, suggesting a net positive relationship between population and firm formation.

5. Discussion of differences

To illustrate the functioning of the *Grid and Shake* method, we estimated agglomeration elasticities on two types of administrative units (municipalities, micro-regions) and three artificial grids, made up of 15×15km, 26×26km and 39×39km units. For all artificial grids, we generated 1000 replications, and hence, regressions were repeated a 1000 times, yielding a distribution of elasticities.

In what follows, we discuss three applications: (i) the role of fragmentation, comparing municipalities with a grid simulating average city sized units (15×15km), (ii) looking the impact of size of units, comparing three grid sizes; and (iii) the role of administration, comparing micro regions with artificial units of 26×26km sized unit, matching the average

[12] Least squares models have the advantage of being simple and allow for various weighting structures. At the same time, the fact that there may great many zeroes - especially when area is narrowly defined makes use problematic given advantages of log-log specifications. One way of treating this would be to use the ratio of new firms to existing firms on the left hand side and estimate a Tobit regression with censoring at 0 and 1. Poisson models are used to estimate incidence in banking sector (Gobbi and Lotti, 2004) or firm exit (Pe'er and Vertinsky, 2008).

Table 3: Entry regressions

units	municipal	micro region		km15	km26	km39
Population	1.212	1.193	Population	1.240	1.230	1.163
SE	0.029	0.072	(mean) SE	0.044	0.067	0.081
			SD coefficient	0.010	0.017	0.029
Area	-0.108	-0.098	Area	-0.062	-0.124	-0.144
SE	0.044	0.089	(mean) SE	0.103	0.134	0.125
			SD coefficient	0.047	0.060	0.053
Wage	0.412	0.645	Wage	0.551	0.635	1.113
SE	0.123	0.319	(mean) SE	0.210	0.296	0.400
			SD coefficient	0.065	0.088	0.157
			Replication	1000	1000	1000
Observations	2,873	150	Observations	438	151	74
Pseudo R squared	0.703	0.673	Pseudo R squared (mean)	0.680	0.689	0.673

All variables are in logs. Population is size of local working age population. Area is in km^2. Years of schooling is average of local population. Robust standard errors are displayed. In grid models we report two statistics for standard errors: (mean) SE is the mean of robust standard error of the coefficient over 1000 replications, while SD coefficient provides the statistic for the standard deviation of the coefficient itself calculated over 1000 replications.

size of micro-regions. For all this, we use the examples we introduced earlier: the wage as well as firm formation elasticity of agglomeration.

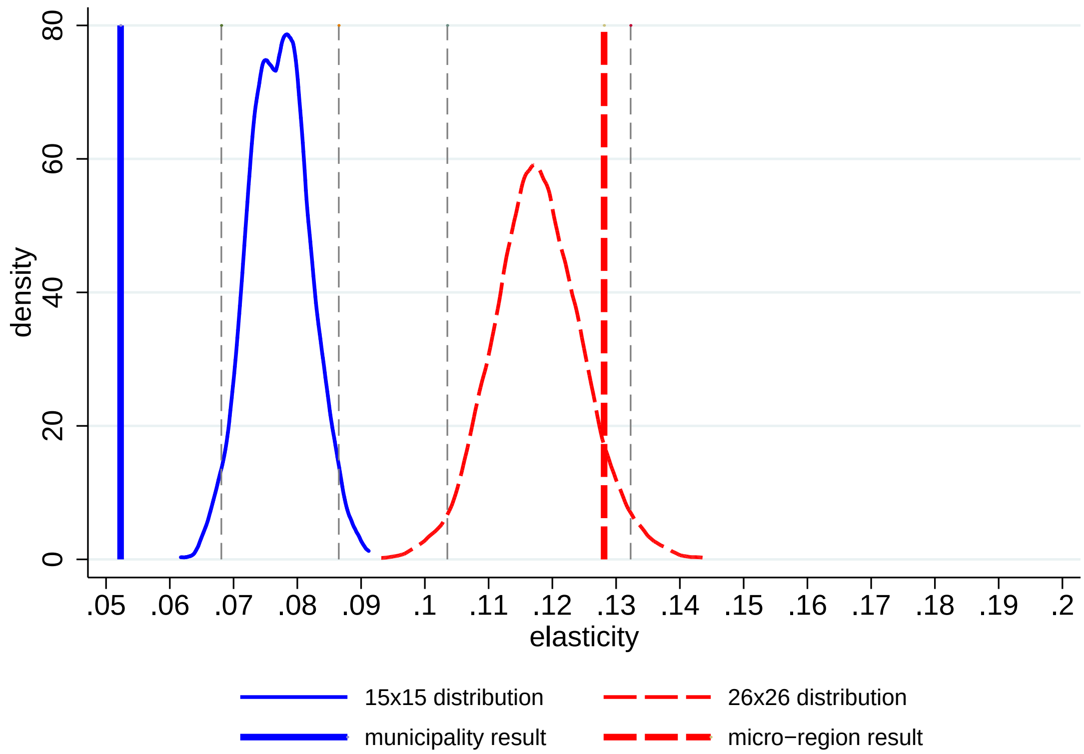
5.1. Fragmentation of municipalities

The first question is related to a lower level of public administration, the municipal system with 3125 local units in Hungary. We can investigate how this fragmentation affects the relationship between agglomeration and wage as well as new firm formation. To see this we can now compare regression results at the municipal level with mean elasticities from regression on the 15×15km grid. For instance, such result would allow comparisons with model estimates used other countries with less fragmented systems, such as Sweden and Portugal.

For both the wage OLS and the firm entry Poisson regressions, we find that the estimated elasticity for the municipal level lie outside 95% of the distribution of artificial unit elasticities. See blue (solid) lines in both Figures 8 and 9.

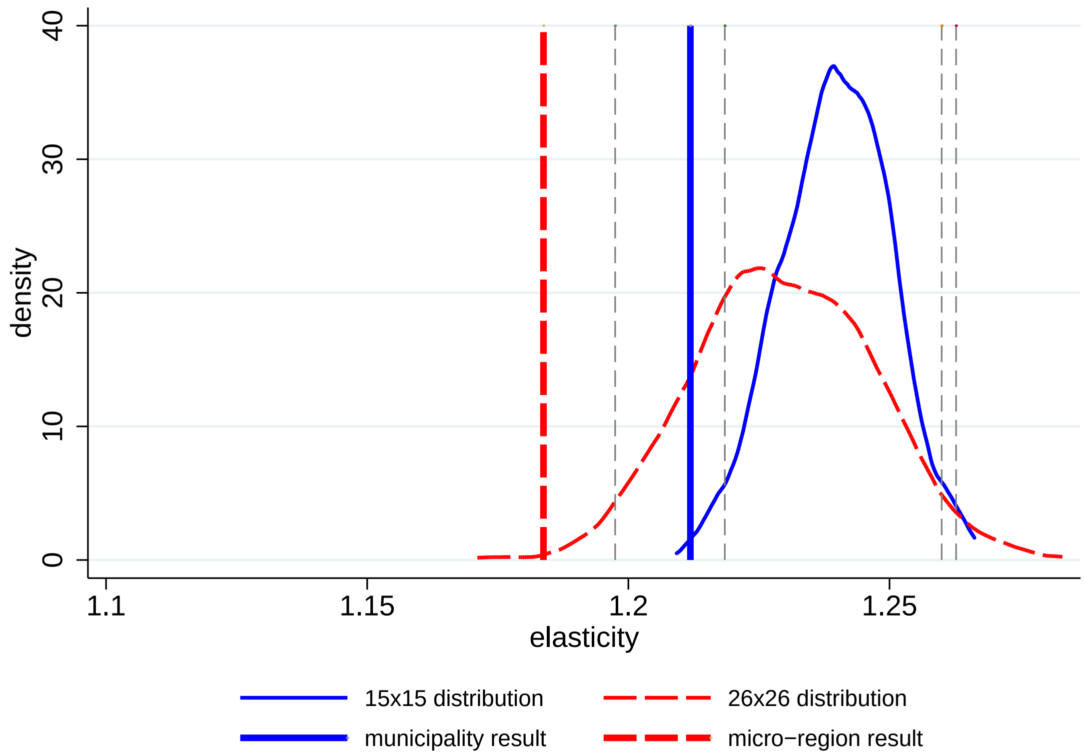
In this data, fragmentation of the local polity implies that locally estimated agglomeration elasticities are different to what it would in the case of artificially aggregating small units - thus, creating 437 artificial units in lieu of actual municipalities. The difference is particularly strong for the wage regressions: the estimated elasticity is 5.2% at the municipal level and 7.7% at the 15×15km sized artificial unit level. One possible explanation for this that by disregard spillovers beyond the boundary of (small) municipalities, we would underestimate these elasticities.

Figure 8: Agglomeration externality - administrative vs. artificial units



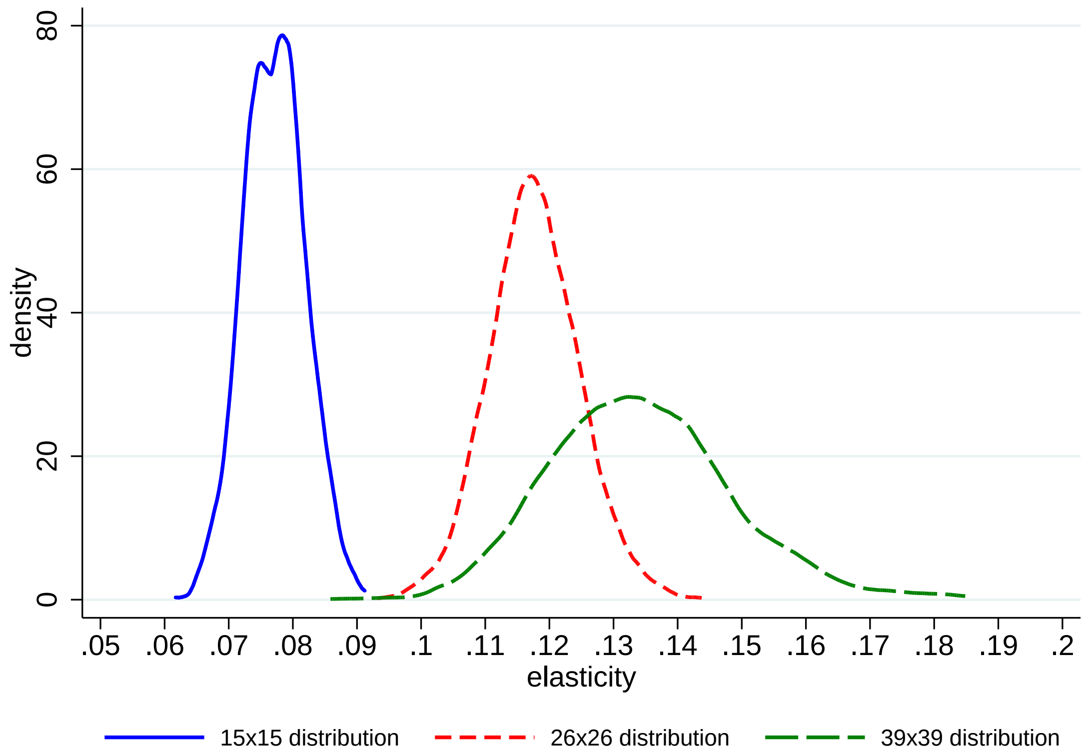
The Figure shows the coefficient estimate on population in Model 1 (agglomeration elasticity of wage). The thick blue (solid) line gives the point estimate for the municipality estimate. The thin blue (solid) line gives the distribution of point estimates over 1000 point estimates on the 15×15 grid. The thick red (solid) line gives the point estimate for the micro-region estimate. The thin red (solid) line gives the distribution of point estimates over 1000 point estimates on the 26×26 grid. The very thin dashed lines represent the 5% confidentiality thresholds.

Figure 9: Agglomeration externality - administrative vs artificial units



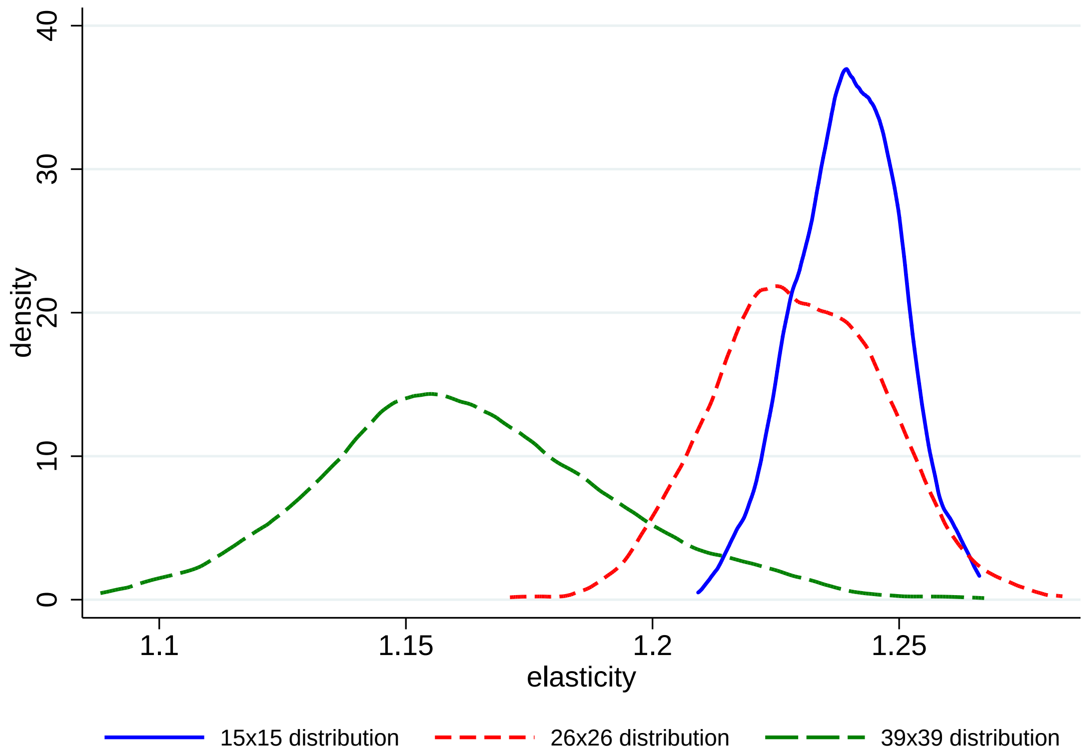
The Figure shows the coefficient estimate on population in Model 2 (agglomeration elasticity of firm formation). The thick blue (solid) line gives the point estimate for the municipality estimate. The thin blue (solid) line gives the distribution of point estimates over 1000 point estimates on the 15×15 grid. The thick red (solid) line gives the point estimate for the micro-region estimate. The thin red (solid) line gives the distribution of point estimates over 1000 point estimates on the 26×26 grid. The very thin dashed lines represent the 5% confidentiality thresholds.

Figure 10: Agglomeration elasticities at various grid size



The Figure shows the coefficient estimate on population in Model 1 (agglomeration elasticity of wage) over various grid sizes. The blue (solid) line gives the distribution of coefficient over 1000 estimations on the grid size 15×15. The red (short-dash) line gives the distribution of coefficient over 1000 estimations on the grid size 26×26 km. The green (long-dash) line gives the distribution of coefficient over 1000 estimations on the grid size 39×39 km.

Figure 11: Agglomeration elasticities at various grid size



The Figure shows the coefficient estimate on population in Model 2 (agglomeration elasticity of firm formation) over various grid sizes. The Figure shows the coefficient estimate on population in Model 1 (agglomeration elasticity of wage) over various grid sizes. The blue (solid) line gives the distribution of coefficient over 1000 estimations on the grid size 15×15. The red (short-dash) line gives the distribution of coefficient over 1000 estimations on the grid size 26×26 km. The green (long-dash) line gives the distribution of coefficient over 1000 estimations on the grid size 39×39 km.

5.2. Size of externality impacts

The second issue is related to the size of units of observations, comparing results from grids of 15×15km, 26×26km and the 39×39km sized units. Figure 10 shows the population elasticities at various levels of grid size for our example 1 and Figure 11 plots results obtained from example 2.

As for wages, considering OLS models, we find that the population elasticity is different at 15×15 and 26×26 size, but not beyond. At the same time the relationship with years of schooling varies between 26×26 and 39×39km unit sizes. For the first difference model, there is no significant difference between grids regarding density but there is regarding the effect of schooling.

Regarding firm, entry, point estimates suggest that elasticities are different for the 39×39km grid from the other two, but given larger variation of estimates, these are not statistically different. As for the elasticity of wage level on firm entry, estimates on the large grids are statistically different indeed. These results broadly conform with estimates found earlier.[13]

5.3. Administrative units

Third, we consider the role of administrative units, comparing elasticities measured at the micro region level with a grid with 26×26km large units. The agglomeration elasticity of wage is presented in Table 2, while the estimated coefficients from the Poisson regression of number of new firms on population is shown in Table 3.

Table 2 shows that for model 1 (wage regressed on population), we find an estimated coefficient for the micro regions of 12.7%, and this may be compared with the *mean* estimated coefficient of 11.8% - averaged out on the 1000 replication of the 26 km grid . We also calculated the standard error of these estimates, 0.007. This suggests, that while there may be a small difference, the estimated coefficient of administrative unit is not statistically significant from the artificial set of results. Hence, we cannot reject the null of zero effect of administrative units of micro-regions compared to artificial units of equal average size (26×26km).

The story is somewhat different when we consider firm entry. For both cases, the estimated coefficients are significantly greater than unity. However, equality of estimated coefficients is here rejected at 5% confidence level, suggesting that there is a small difference between elasticities at artificial units compared to administrative regions. Note that estimates at different artificial units vary more than earlier; and tails are longer, too.

These results are better visible on Figures 8 and 9. Figure 8 reports regression results as well as the distribution of results on the artificial grids for model 1 (wage regressed on population), Figure 9 for model 2 (count of new firms regressed on population). For both panels, we compare the elasticity we get for micro regions (0.077) and the distribution of point estimates on the set of artificial grids with 26x26km sized units. We also displayed the 5% confidence intervals.

[13] Comparing it estimates found by Briant et al. (2010) using a different aggregation methodology, we find that here, the order of elasticities is different. In this paper, we find coefficients rising with size, but a significant difference found only between small and medium sized units.Briant et al. (2010) find no linear relations, and difference only between their medium and large sized units.

6. Conclusions

In this paper, we propose and novel and simple method to test robustness of empirical investigations at various levels of spatially aggregated data. We look at two frequently researched topics: local wages and creation of new firms. In particular, we investigate the role of spatial externalities such as knowledge spillovers.

Our methodology has a few attractive features compared to existing procedures. First, it is very flexible, may be easily automated and used. It is flexible to cover any size of the grid and match the size of the administrative regions. Second, it compares evenly sized areas rather than even sum of municipalities (like Briant et al. (2010)), this is useful when the average size of basic units is correlated with some unobserved feature this is the case in south-west versus north-east of Hungary or Benelux versus Spanish regions. Third, while being as simple as drawing circles of radius L around observations (or municipalities) it does not oversample dense regions (i.e introduce a bias of having the regional sample size being dependent on the number of observations from that area). Finally it can be considered as a generalization of regression discontinuity design when borders are numerous.

We illustrated the method on looking at relationship between wage, firm birth and agglomeration, and found that agglomeration elasticities found at various levels of aggregation are all statistically different from zero (wages) and unity (firm creation). We first found that administrative borders of the NUTS4 level of Hungarian micro-regions (*"kistérség"*) do not matter. Second, investigating at more disaggregated level, we posited that if small municipalities were merged to form typical city-sized unit, estimated elasticities would be larger than when measured as is. Third, the size of artificial grid units matters, mostly between small and medium size.

Importantly, differences among elasticities (such as 0.05 vs 0.13 for agglomeration elasticity of wage) found at various levels of aggregation, are broadly in the range of those found in the literature employing various estimation method (instrumental variables, selection equations, fixed effects, etc.). Hence, the method of spatial aggregation seems to be of equal importance to modeling and econometric specification of the estimation.

We believe this method it may be used for several other types of studies. Possible areas include investigating the role of local competition (in retail, the number of local firms) for pricing, maybe applied to competition authority studies on mergers. Our approach may also be used to study the role of administrative measures that vary across regions, such as employment policies or tax breaks. Furthermore, it may be used as an alternative to regression discontinuity design, in cases when discontinuity is complicated, driven by several borders.

References

Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. Journal of the American Statistical Association 105 (490), 493–505.

Ács, Z. J., Armington, C., 2004. The impact of geographic differences in human capital on service firm formation rates. Journal of Urban Economics 56 (2), 244–278.

Andersson, M., Klaesson, J., Larsson, J. P., 2014. The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? Papers in Regional Science 93 (4), 727–747.

Anselin, L., 2010. Thirty years of spatial econometrics. Papers in Regional Science 89 (1), 3–25.

Anselin, L., Bera, A. K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. statistics textbooks and monographs 155, 237–290.

Armington, C., Acs, Z. J., 2002. The determinants of regional variation in new firm formation. Regional studies 36 (1), 33–45.

Bacher, H., Brülhart, M., 2013. Progressive taxes and firm births. International Tax and Public Finance 20 (1), 129–168.

Békés, G., Harasztosi, P., 2013. Agglomeration premium and trading activity of firms. Regional Science and Urban Economics 43 (1), 51–64.

Briant, A., Combes, P.-P., Lafourcade, M., 2010. Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? Journal of Urban Economics 67 (3), 287–302.

Burger, M. J., van Oort, F. G., van der Knaap, B., 2010. A Treatise on the Geographical Scale of Agglomeration Externalities and the MAUP. Scienze Regionali 9 (1), 19–39.

Cainelli, G., Montresor, S., Vittucci Marzetti, G., 2014. Spatial agglomeration and firm exit: a spatial dynamic analysis for italian provinces. Small Business Economics 43 (1), 213–228.

Ciccone, A., 2002. Agglomeration effects in Europe. European Economic Review 46 (2), 213–227.

Ciccone, A., Hall, R. E., 1996. Productivity and the Density of Economic Activity. American Economic Review 86 (1), 54–70.

Combes, P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2012. The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. Econometrica 80 (6), 2543–2594.

Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters! Journal of Urban Economics 63 (2), 723–742.

Combes, P.-P., Duranton, G., Gobillon, L., Roux, S., 2010. Estimating Agglomeration Economies with History, Geology, and Worker Effects. In: Agglomeration Economics. NBER Chapters. National Bureau of Economic Research, Inc, pp. 15–66.

Csorba, G., Koltay, G., Farkas, D., 2011. Separating the ex post effects of mergers: an analysis of structural changes on the Hungarian retail gasoline market. IEHAS Discussion Papers 1118, Institute of Economics, Hungarian Academy of Sciences.

Dewhurst, J., McCann, P., 2007. Specialisation and regional size. Edward Elgar, pp. 204–220.

Duranton, G., Overman, H. G., 2008. Exploring The Detailed Location Patterns Of U.K. Manufacturing Industries Using Microgeographic Data. Journal of Regional Science 48 (1), 213–243.

Fotheringham, A. S., Wong, D. S., 1991. The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning 23 (7), 1025–1044.

Gobbi, G., Lotti, F., 2004. Entry decisions and adverse selection: an empirical analysis of local credit markets. Journal of Financial services Research 26 (3), 225–244.

Gobillon, L., Milcent, C., November 2013. Spatial disparities in hospital performance. Journal of Economic Geography 13 (6), 1013–1040.

Holl, A., 2004. Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal. Journal of Regional Science 44 (4), 693–712.

Hortaçsu, A., Syverson, C., 2007. Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices. Journal of Political Economy 115, 250–301.

Iyer, G., Seetharaman, P., 2008. Too close to be similar: Product and price competition in retail gasoline markets. Quantitative Marketing and Economics 6 (3), 205–234.

Krugman, P., 1991. Increasing Returns and Economic Geography. Journal of Political Economy 99 (3), 483–499.

Lychagin, S., Pinkse, J., Slade, M. E., Reenen, J. V., 2010. Spillovers in Space: Does Geography Matter? NBER Working Papers 16188, National Bureau of Economic Research, Inc.

Marshall, A., 1920. Principal of Economics: An introductory volume, 8th Edition. Procupine Press, Philadelphia.

Martin-Barroso, D., Nunez Serrano, J. A., Velazquez, F. J., 2010. A different look at agglomeration effects in Spain. MPRA Paper 33601, University Library of Munich, Germany.

Melo, P. C., Graham, D. J., Noland, R. B., 2009. A meta-analysis of estimates of urban agglomeration economies. Regional Science and Urban Economics 39 (3), 332–342.

Pe'er, A., Vertinsky, I., 2008. Firm exits as a determinant of new entry: Is there evidence of local creative destruction? Journal of Business Venturing 23 (3), 280–306.

Rosenthal, S. S., Strange, W. C., 2008. The attenuation of human capital spillovers. Journal of Urban Economics 64 (2), 373–389.

Sutaria, V., Hicks, D. A., 2004. New firm formation: Dynamics and determinants. The Annals of Regional Science 38 (2), 241–262.