_____________________ *Région et Développement n° 36-2012* _____________________

DOES THE CHOICE OF GEOGRAPHIC UNITS MATTER FOR THE VALIDATION OF GIBRAT'S LAW?

Boris A. PORTNOV *

Abstract - *As we hypothesize, the selection of observation units may become a critical consideration for an empirical investigation of Gibrat's Law, according to which the growth of cities is expected to be independent of their sizes. In order to verify this hypothesis, we investigate whether the validation of Gibrat's law may depend on the aerial units used for the analysis – i.e., individual localities vs. integrated urban areas. The present study examines that possibility, using 1990-2000 population growth data for two levels of geographic resolution – 4,667 local administrative units (i.e., municipalities) and 2189 contiguous urban areas in 40 European countries. According to our findings, the association between population size and growth tend to differ across subsets of localities, being positive across localities with fewer location advantages and negative elsewhere. As a result, the strength and direction of the 'growth-size' association may change, depending on the relative shares of location subsets used for aggregation.*

Key-words: POPULATION GROWTH; EUROPEAN URBAN SYSTEM; LOCATIONAL SUBSETS; LOCATIONAL ATTRIBUTES; GIBRAT'S LAW; URBAN AREAS; LOCAL ADMINISTRATIVE UNITS

JEL Classification: O18, R11, R23

Acknowledgements - *The paper is based on data sources used by Portnov et al. (2011) but contains results of a re-analysis of previously published data, with several newly added explanatory variables and interaction effects. The author would like to express his deep gratitude to Prof. Benjamin Reiser for his invaluable help with preparing Section 3 and his numerous helpful comments and suggestions on the earlier drafts of this paper.*

**Department of Natural Resources and Environmental Management, University of Haifa, Israel ; Portnov@nrem.haifa.ac.il*

INTRODUCTION

In empirical regional studies, different geographic units – municipalities, counties, regions, built contiguities, commuting zones and metropolitan areas – are used, often without particular justification apart from data availability. The situation is similar with empirical studies of Gibrat's Law for cities.[1] While in some studies, individual localities are used (e.g. Eeckhout, 2004; Ioannides and Overman, 2003), other studies investigate the applicability of this law to urban and metropolitan areas formed by several neighboring municipalities (see e.g. Pumain and Moriconi-Ebrard, 1997; Black and Henderson, 2003).

Several previous studies found support for the proportionate growth of places, stipulated by Gibrat's Law (see inter alia Clark and Stabler, 1991; Ioannides and Overman, 2003; Eeckhout 2004; Rose, 2005). However, in other studies, this law was not supported by empirical data (cf. e.g. Pumain and Moriconi-Ebrard, 1997; Black and Henderson, 2003; Portnov et al., 2011).

A possible explanation for these differences may reflect different levels of geographic resolution used in the analysis. Changes in the strength of correlation between variables in line with data aggregation into areal units of larger size are a well-known phenomenon. In an early study, Openshaw (1984) termed this phenomenon the 'modifiable areal unit problem' or MAUP.[2]

According to MAUP, different trends may emerge, if individual localities (e.g., local administrative areas (LAUs) or municipalities) are used in the analysis, as opposed to integrated territorial units (i.e., contiguous urban areas), combining several adjacent LAUs. The selection of observation units may thus become a critical consideration for a proper interpretation of urban phenomena, in general, and for empirical investigation of Gibrat's Law, in particular. In order to verify this hypothesis, in the present paper, we investigate whether the validation of Gibrat's law for cities may depend on the aerial units used for the analysis – i.e., individual localities vs. integrated urban areas. In particular, two main questions we attempt to answer are as follows:

- Can the direction and strength of association between population size and growth change as a result of areal aggregation of individual urban localities into geographic units of larger size?
- If such a change does tend to happen, under what circumstances can it occur and what factors may affect it?

---

1 According to Gibrat's Law, also called 'the law of proportional effect' (loi de l'effet proportionnel), the growth of an economic entity is independent of its size, as measured by e.g., the number of workers in the case of factories (Gibrat 1931, cited in Kalecki 1945). When applied to cities, growth rates, according to this law, are expected to be independent of the numbers of their residents, so that all cities, big and small, should expectedly grow at the same average rate (Eeckhout, 2004).

2 The MAUP is a part of a broader methodological issue known as 'change of units of support problem' of CUSP. According to this phenomenon, the outcome of the analysis often depends on the analytical units used for investigation, and may thus change as a result of merging, splitting of the analysis units or modification of their boundaries (Gotway and Young, 2002).

To answer these questions, we analyze two groups of areal units – 4,667 LAUs (i.e., municipalities) and 2,189 urban areas, formed by territorial contiguities of built-up areas in 40 European countries. Our main finding is that that the association between population size and growth appear to differ across subsets of localities, stratified by their location attributes (i.e., well-positioned localities vs. poorly positioned ones). As a result, the strength and direction of the 'growth-size' association may change, depending on the relative shares of location subsets used for aggregation.

1. PREVIOUS STUDIES OF GIBRAT'S LAW FOR CITIES

Portnov et al. (2011) provide a fairly detailed discussion of several previous studies investigating the applicability of Gibrat's law to cities and urban areas. In this section we shall focus on the most general strands of Gibrat's law research, pertinent to main topic of the present analysis, that is, difference in findings, potentially attributed to various units of supports used for investigation.

The simplest formulation of Gibrat's Law for cities (see, for example, Anderson and Ge, 2005) is:

$$\ln P_t = \mu + \ln P_{t-1} + u_t \qquad (1)$$

or alternatively:

$$\ln(P_t/P_{t-1}) = \Delta \ln P_t = \mu + u_t, \qquad (2)$$

where P_t is the population size of a locality at time t; μ represents the average rate of growth; u_t is a random error term, and $ln(P_t /P_{t-1})$ denotes the logarithm of the growth rate. [For a more formal examination of Gibrat's Law, see Appendix 1of Portnov et al., 2011)].

Clark and Stabler (1991) used a more complex specification for testing Gibrat's Law, which involved past growth rates, implying that the current one depends on them, an assumption not made in (2). Using historical records for seven largest Canadian cities, the authors of this study concluded that 'Gibrat's law cannot be rejected.

In another empirical study of some 25,000 localities, including 135 largest cities of the U.S.A., Eeckhout (2004) reached a similar conclusion. In particular, he found that the growth of cities appeared to follow Gibrat's Law perfectly, showing complete size-growth independence. In Eeckhout's view, the difference between his findings and those of previous studies, which failed to substantiate Gibrat's Law, was due to the fact that most previous studies used 'truncated' city-size distributions, whereas in his study he used "the entire population of American cities" (ibid.).

Characteristically, most studies which investigated Gibrat's law using integrated urban areas found little or no support for this law. Thus, in one of such studies, Black and Henderson (2003) investigated the relation between the population sizes of major metropolitan areas in the continental U.S., using population census data for 10-year periods between 1900 and 1990. The association between the population size and growth, in all historical periods studied, was found to be significant and negative, leading the authors to reject Gibrat's Law. It was also found that 'superior' sites (i.e., characterized by warmer climate, located on the coast and having a better market potential) grow faster than 'inferior' ones, thus indicating that urban growth patterns are not purely stochastic.

In their study of some 26,000 populated places worldwide, Pumain and Moriconi-Ebrard (1997) reached a similar conclusion, arguing that, over past decades, major metropolitan areas across the globe have grown 'systematically more rapidly' than the rest of their urban systems, thus effectively 'invalidating Gibrat's urban growth model.'

A similar conclusion was also reached by another recent analysis of urban growth in Europe by Portnov and al. (2011). In particular, the authors of this study did not find any strong evidence to support Gibrat's Law at any of the two levels of areal resolution they considered, that is, local administrative units (LAUs or municipalities) and urban areas. (The latter were formed by the aggregation of two or more adjacent LAUs). The authors considered Gibrat’s law for cities taking account of three location subsets – favorable, unfavorable, and others. As they found, the slope of the relationship of population growth rates versus population size (both in log units) was positive for unfavorable locations and negative for favorable locations when the regression models were computed for LAUs. However, when the computations were carried out for urban areas the associations between population size and growth were found to be positive. Portnov et al. (ibid.) discussed this point further although no detailed explanation for that change in the ‘size-growth’ relationship was suggested.

In this paper, we investigate under what circumstances, the areal aggregation of LAUs can result in a change in relationship between population size of localities and their growth, and, in addition, examine which factors may influence this change.

2. ANTICIPATED RELATIONSHIPS BETWEEN POPULATION SIZE AND GROWTH UPON AREAL AGGREGATION OF LOCATION SUBSETS

As discussed in Portnov et al. (2011), the simplest specification for the relationship between population growth and population size is:

$$G_{t,t-1} = \ln(P_t/P_{t-1}) = \ln P_t - \ln P_{t-1} = \alpha + \beta \ln P_{t-1} + u_t \qquad (3)$$

where u_t is a random error term, P_t is the population size of a locality or an urban area at time t, the end of the study period; P_{t-1} is population size of the locality at the beginning of the study period; $G_{t,t-1}$ represents the population growth rate(measured in *ln* units), and β is the slope of the regression line.

For simplicity’s sake, we can ignore the error term (u) and further assume that α is *zero*. There are two main reasons enabling this assumption: First, according to several empirical studies of Gibrat’s law, the regression intercept was found to be close to zero, even though estimates for β in these studies differed (Clark and Stabler, 1991; Eeckhout 2004; Pumain and Moriconi-Ebrard, 1997; Black and Henderson, 2003; Portnov et al., 2011). Second, α effectively means the growth rate of a locality with the population size of *one* resident (ln $P_{t-1} = 0$), which can safely be assumed to be close to *zero*.

Under these assumptions, we can rewrite (3) as:

$$\ln(P_t/P_{t-1}) \approx \beta \ln P_{t-1} \tag{4}$$

Now, let us consider a simple case where is the whole set of localities, ***C*** is formed by the combination of *two* location subsets, denoted by ***A*** and ***B***, giving that:

$$P_{A,t} + P_{B,t} = P_{C,t}. \tag{5}$$

We further assume that (4) holds for ***A*** and ***B*** with corresponding regression slopes β_A and β_B, respectively. Then,

$$P_{A,t} = P_{A,t-1}^{1+\beta_A} \tag{6}$$

and similarly,

$$P_{B,t} = P_{B,t-1}^{1+\beta_B} \tag{7}$$

Substituting (6) and (7) into (5) gives:

$$P_{C,t} = P_{A,t-1}^{1+\beta_A} + P_{B,t-1}^{1+\beta_B} \tag{8}$$

Consequently:

$$G_{C,t-1} = ln\left(\frac{P_{C,t}}{P_{C,t-1}}\right) = ln\left(P_{A,t-1}^{1+\beta_A} + P_{B,t-1}^{1+\beta_B}\right) - ln\left(P_{A.t-1} + P_{B,t-1}\right) = ln\left(\frac{P_{A,t-1}^{1+\beta_A} + P_{B,t-1}^{1+\beta_B}}{P_{C,t-1}}\right) \tag{9}$$

where $G_{Ct,t-1}$ is the growth rate of the amalgamated set of localities, ***C***.

In order to compare the growth rate for ***C*** with its population size, $P_{c,t-1}$ (ln), we compute:

$$\frac{G_{C_{t,t-1}}}{ln(P_{C,t-1})} = ln\left(\frac{P_{A,t-1}^{1+\beta_A} + P_{B,t-1}^{1+\beta_B}}{P_{C,t-1}}\right) / ln(P_{C,t-1}) \tag{10}$$

As one can easily see, this ratio is not constant for any given β_A and β_B but varies with the population sizes of **A** and **B** subsets. This indicates that the relationship between G_C and $P_C(ln)$ can change depending on both the relative shares of location subsets, **A, B** used for aggregation, and on their βs.

To illustrate this observation, the function defined by Formula (10) will be studied numerically in the next section, where the connection between population size and growth will be examined by a series of empirical tests.

3. POPULATION SIZE-POPULATION GROWTH ASSOCIATION IN AN AMALGAMATED SET OF LOCALITIES: EMPIRICAL TESTS

Figure 1 illustrates changes in the regression slope of the integrated set of localities, C formed by the aggregation of two locational subsets *A* and *B* (providing that $P_A+P_B=P_C$). The X and Y axes on the diagrams feature population sizes of the original subsets, while the radiating lines define the values of the $G_c/P_c(ln)$ ratio for the integrated set, C [see (10)] for specific combinations of P_A and P_B and several *a priori* defined β_A and β_B values.

In particular, the following four separate β_A and β_B settings were used for calculation:

- Proportionate growth in both subsets: $\beta_A = \beta_B =0$ (Fig. 1a);
- Positive association between of the population size and growth in the first subset ($\beta_A=0.001$) and a negative association in the second subset ($\beta_B=-0.001$) (Fig. 1b);
- Positive population size-population growth associations in both subsets ($\beta_A =0.001$ and $\beta_B =0.002$) (Fig. 1c);
- Negative population size-population growth association in both subsets ($\beta_A =-0.001$ and $\beta_B =-0.002$) (Fig. 1d).

The above β values used for calculation are similar to those reported in Portnov et al. (2011) and several other studies of Gibrat's Law (see *inter alia* Eeckhout, 2004) and may thus be considered plausible. The total population sizes of the input sets, **A** and **B,** are also set to a plausible range – 1 through 1,000,000 – although that range can be extended, without any change in the output relationship, due to the assumed linearity of relationship between P_{t-1} and $G_{t,t-1}$.

The simulation tests are run using Formula (10) in the Surfer v.8.0^{TM} software.

As Fig. 1 shows, *if* both β_A and β_B are set to *zero* (meaning that, in both subsets, growth rates are strictly proportionate to the population sizes of localities forming these subsets), the $G_c /P_c(ln)$ ratio in the amalgamated set is also equal to *zero*, irrespectively to the population sizes of the 'input' subsets (Fig. 1a). However, the outcome differs if subsets with *positive and negative relationships* between population size (S) and growth (G) are merged. Thus, for instance, if the population size of ***B***-subset ($\beta_B<0$) is equal to 800,000 and that of the ***A***-subset ($\beta_A>0$) is e.g., 200,000, the regression slope for the unified set, β_c is calculated as -0.0006 (Fig. 1b). Concurrently, in the opposite situation (i.e., when Pop(B)=200,000 while Pop(A)=800,000), the sign of the regression slope coefficient for the amalgamated set reverses to a positive value: β_c =0.0006 (see Fig. 1b). However, if G-S associations are positive in both subsets ($\beta_A >0$ and $\beta_B >0$), such an association remains positive in the unified set of localities as well, albeit exhibiting different β_c values for different combinations of population sizes of the input sub-sets (see Fig. 1c).

By the same token, the G-S association remains *negative* in the unified set of localities if such associations are negative in both input sub-sets, that is, if $\beta_A < 0$ and $\beta_B < 0$ (see Fig. 1d). In other words, the G-S association in the unified set of localities appears to response to changes in *both* regression slopes of G-S associations in the original subsets and the population sizes of the input sub-sets. Furthermore, it *changes its sign if location subsets with different signs of G-S association are amalgamated.*

Using the European urban system as a case study, in the following sub-sections, we shall attempt to verify whether the above predicted reversion of the P-G association may actually occur upon areal aggregation of geographic units used for the analysis.

4. RESEARCH METHODS AND DATA SOURCES

4.1. Data sources

In the present analysis, we work with the dataset of approx. 4,700 European *municipalities*, previously used and described in detail by Portnov and Schwartz (2008; 2009a,b) and Portnov *et al.* (2011). The database covers LAUs spread over 40 European countries, ranging from 2,000 to 7,000,000 residents.[3] [The analysis is performed separately for the original observation units reported in the database (that is, municipalities or LAUs) and for integrated urban areas, as described in some detail later in this section].

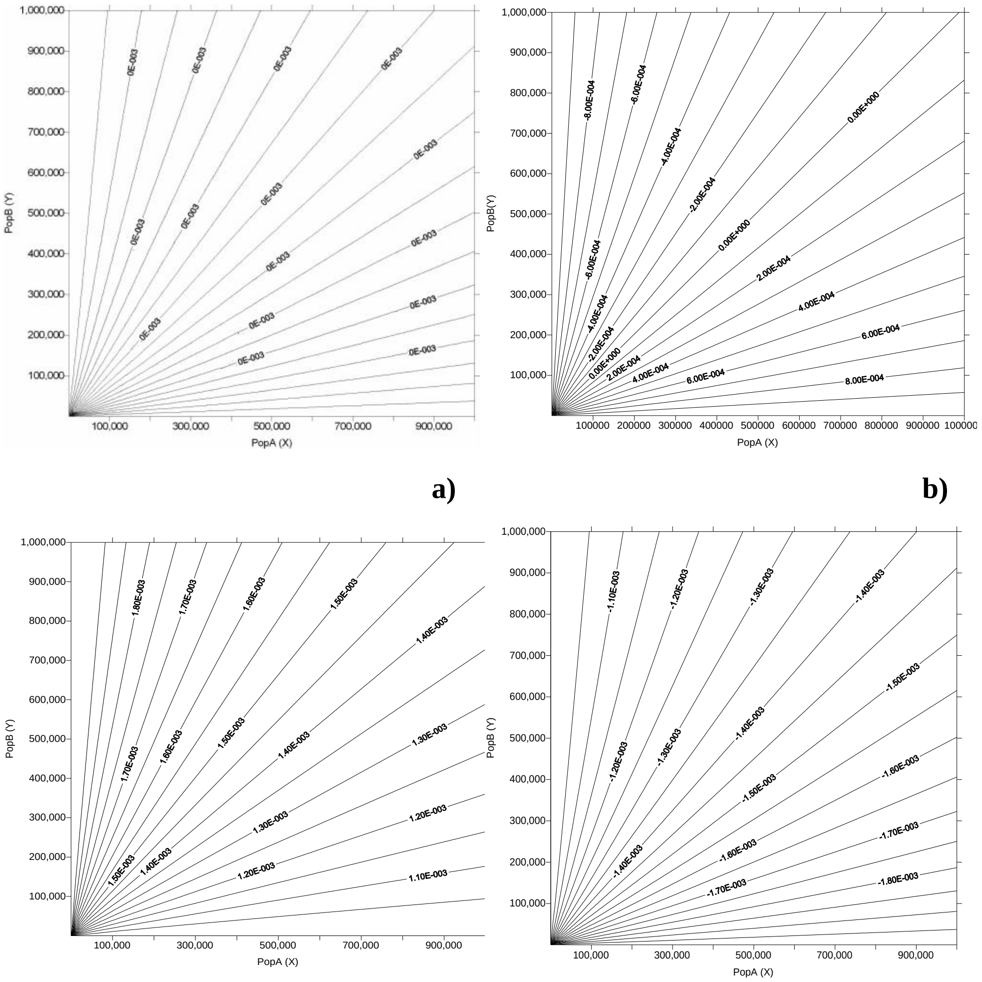
The data on the longitude and latitude of the settlements, and on their elevation above sea level, were obtained from the Geonames Database, which contains such data on urban and rural settlements worldwide (Geonames, 2007). Data on the population growth rates of localities were obtained from the City Population Database (Brinkhoff, 2007), whereas proximity of municipalities to location landmarks (the sea shore, and the closest city larger than 500,000 residents, etc.) was calculated in the ArcGIS9.x^{TM} software, using geographic layers obtained from the geo-coverage database maintained by ESRI (2000). The proximities were calculated as aerial distances between specific location features and the settlements' 'reference points' (which normally coincide with the location of city hall or some other local landmarks).

Although access time may seem to be the most accurate measure of inter-urban proximity, we opted for *aerial distances*, which are commonly used in urban and regional studies (see *inter alia* Henry et al., 1997; Partridge et al., 2007). Our decision was motivated by the shortcomings of travel time between any two given places, such as considerable variation by season of the year (especially in countries with rainy and snowy winters), and even by time of the day. If the infrastructure and quality of services are more or less uniform throughout the study area, aerial distance may be a fairly accurate measure of inter-urban proximity.

---

[3] *Nearly all cities and towns of Europe with a population of 20,000+ residents are covered by the study. Smaller localities are less fully represented, due to incomplete data on population growth. This limitation will be further discussed in the concluding section.*

Figure 1. Sensitivity tests of the changes in the regression slope of the 'population size – population growth' association in the integrated set of localities (C), resulting from the integration of 2 subsets of localities, characterized by varying population sizes (Pop) and regression slopes (β)



a) b)

c) d)

Notes: X and Y axes on the diagrams feature population sizes of the original sets of localities [Pop(A) and Pop(B), Pop(C)=Pop(A)+Pop(B)]. Sensitivity tests are run using Formula (10).

a) Proportionate growth in both subsets ($\beta_A = \beta_B = 0$).

b) Positive association between of the population size and growth in the first subset ($\beta_A = 0.001$) and a negative association in the second subset ($\beta_B = -0.001$).

c) Positive population size-population growth association in subsets ($\beta_A = 0.001; \beta_B = 0.002$).

d) Negative population size-population growth association in subsets ($\beta_A = -0.001; \beta_B = -0.002$).

Table 1. Descriptive statistics of selected research variables for European localities and urban areas datasets

Variable	N	Minimum	Maximum	Mean	Std. Dev.	Skewness	Kurtosis
				A. Localities			
Population growth (ln)	4667	-0.047	0.126	0.006	0.012	2.528	13.609
Population size (ln)	4667	6.915	15.820	10.461	1.003	0.409	2.438
Population density (ln)	4667	-2.528	6.608	4.428	1.286	-0.730	1.027
Elevation (m)	4667	-7.000	1941.000	175.759	209.988	2.585	10.524
Latitude	4667	27.917	69.967	48.842	5.898	-0.130	0.366
D_mcity	4667	0.000	20.234	1.525	1.717	5.336	43.633
D_mroad	4667	0.000	17.505	0.166	1.085	13.009	181.212
D_water	4667	0.000	19.650	0.514	0.995	7.520	90.458
D_shore	4667	0.000	17.134	1.780	2.046	2.339	8.935
January temperature (oC)	4667	-14.500	17.500	2.298	4.335	-0.339	1.627
Evaporation index (mm)	4667	10.200	21.800	18.955	1.669	-1.748	4.737
				B. Urban areas			
Population growth (ln)	2189	-0.023	0.126	0.006	0.012	3.000	18.496
Population size (ln)	2189	7.074	16.186	10.806	1.150	1.110	1.954
Population density (ln)	2189	2.512	10.795	6.331	0.945	0.129	2.353
Elevation (m)	2189	-7.000	1941.000	210.036	244.463	2.559	9.497
Latitude	2189	28.048	69.670	49.203	6.178	-0.014	0.085
D_mcity	2189	0.000	19.520	1.780	1.581	4.683	37.951
D_mroad	2189	0.000	16.740	0.185	0.952	13.173	200.084
D_water	2189	0.000	19.650	0.540	0.943	8.389	124.436
D_shore	2189	0.000	17.100	2.193	2.336	2.178	7.286
January temperature (oC)	2189	-14.500	17.500	1.125	4.460	-0.390	1.316
Evaporation index (mm)	2189	10.200	21.800	18.772	1.881	-1.976	4.962

Lastly, in order to integrate individual localities into contiguous urban areas and average their performance indicators, we used the geographic layer of *urban areas* worldwide generated in the framework of the Millennium Ecosystem Assessment Report (UNEP & UNESCO, 2004). The integration was performed using the 'spatial join' tool in the ArcGIS 9.x^{TM} software, which helps to join data from different geographic layers (maps), based on the relative location of features in the layers (Minami and ESRI 2000).

Descriptive statistics of the research variables in the analysis are reported in Table 1.

4.2. Population growth rates

Annual population growth rates were calculated as the natural logarithm of difference in population size at the beginning and the end of the study period (see Formula (3)), as commonly done in Gibrat's Law studies. For most countries covered by the present study, population data are available for 1990/91 and 2000/2001. However, for some countries, the analysis covers a slightly different time span. Thus, population data for Belorussia are only available for 1989 and 1998, whereas the data on French urban settlements can be obtained for 1990 and 1999, etc.

Another comment is important. Gibrat's Law may be interpreted as the convergence, in the long run, to a lognormal city size distribution. Several previous studies (see *inter alia* Robson, 1973; Vlora, 1979; Pumain, 1982; Guérin-Pace, 1993) have indeed investigated the size-growth relationship over relatively long time periods, that is, over several decades or even centuries. However, more recent studies of this law investigated its applicability to populated places, using relatively short-term data, mainly for past decades, using data for metropolitan areas (see *inter alia* Clark and Stabler, 1991; Pumain and Moriconi-Ebrard, 1997) or municipalities (cf. e.g., Eeckhout, 2004; Anderson and Ge, 2005) or both (Portnov et al., 2011). Although it would be desirable to run a test of Gibrat's Law using the distribution of growth rates for more than one time period (and not for 1990-2000 only), this was not feasible due to restrictions on data availability and comparability.

4.3. Explanatory variables

In addition to our main explanatory variable – *population size (ln)* – the following factors served in our study as additional predictors of population growth of LAUs and urban areas:

- *Density*: population density in a 75-km range from a locality, assumed to be a practicable commuting range (in the case of LAUs), and per km^2 of total area, in the case of contiguous urban areas;
- *D_shore*: distance to the sea shore (km);
- *D_mcity*: distance to the closest major city (km);
- *D_mroad*: distance to the nearest highway (km);
- *D_water*: distance to the nearest major water body, i.e., a major river or lake (km); *Latitude*: a place's latitude (decimal degrees);
- *Elevation*: elevation above sea level (meters), and

- Two climatic variables – *average daily temperature* in January (^{o}C) and *evaporation index* (mm).

In previous studies of urban growth (see *inter alia* Ades and Glaeser, 1995; Gallup et al, 1999; Duranton, 1999; Black and Henderson, 2003; Rappaport, 2006; Cheshire and Magrini, 2006) these factors were found to be associated with population growth of urban places and urban areas. [For a more detailed discussion of these factors and their potential contribution to the population growth patterns, see Portnov et al., 2011].

While these indicators do not include all possible predictors of urban development (e.g., industrial productivity or unemployment rates could also serve that purpose), they do cover essential aspects of urban development, such as population size, location and environmental conditions, etc. The inclusion of these variables in the analysis thus makes the variable set (though restricted, due to data availability, to a relatively small number of explanatory variables) fairly parsimonious. In addition, we represent individual countries in our analysis by country dummies, i.e., dichotomous variables taking on the values 1 if a locality is in a given country and 0 otherwise (for the sake of brevity, regression estimates for individual countries' dummies are not reported in the following analysis). These indicator variables help to 'fine-tune' our models to country-specific conditions, not captured by the above 'system-wide' variables, such as population density, population size, locational attributes, etc.[4]

4.4. Locational grouping

In order to test our hypothesis that the direction of 'population size-growth' relationship differs by settlement location, we split the set of localities covered by the study into *three subsets*, reflecting their 'package of location advantages.' According to Portnov and Schwartz (2009b), who suggested the 'location package' (LP) concept, there are several possible approaches to calculating the value of LP for a locality. First, LP may be estimated by adding up the number of positive location attributes a populated place has. Second, some weighting scheme may be applied, assuming that the relative weight of individual attributes in the LP may not be equal, with some attributes contributing to the 'package' more than others. Lastly the LP of a locality may be estimated using interaction terms of individual location attributes, assuming that some location attributes may be needed for the 'activation' of the others.

---

[4] *The effect of individual location attributes (e.g., topography, proximity to networks, etc.) may depend on how much they stand out in their regional or national contexts. In a region or country where a given advantage or disadvantage are commonplace, they are likely to have lesser effects than where they are uncommon (Polese and Shearmur, 2006; Portnov and Schwartz, 2008). To reflect this relativity of location attributes, location variables (proximity to the coast, proximity to major cities, and climatic harshness) were represented in the analysis by their 'relative' values, estimated by dividing the 'absolute' values by country-specific average values. While models with country-normalized location variables were also used in the initial stages of the analysis, the results were found to be similar to the models based on the 'absolute' variable set and are not reported in the following discussion for brevity's sake).*

In the present study, we use the first and simplest approach to calculating the LP of a locality, previously used in and Portnov et al. (2011). In particular, we calculated the LP values as the total number of favorable location attributes (i.e., without applying any weighting or testing interaction terms).

Five location factors (i.e., proximity to the coast, to major cities, to highways, to water bodies and elevation), estimated for each settlement individually (see the subsection on data sources) were re-coded as follows: each locality received integer values between 0 and 5, depending on the number of its location advantages. Thus, the locality received the maximum score of 5, if all its location attributes were defined as favorable, and a 0 score, if *none* was. Similarly, the locality's score was set to 4, if only four of the five attributes were defined as favorable, and so on.

The selection of parameters used to decide whether a location attribute should be classified as favorable or not, is somewhat arbitrary. However, the geographic distribution of urban settlements in Europe is extremely uneven, with some locations being particularly 'favored,' which reduces the risks of bias due to arbitrary classification. As several empirical studies indicated, most urban places on the continent are less than 150 m above sea level, close to major population centers, and in coastal areas (Gallup et al., 1999; Duranton, 1999; Cheshire and Magrini, 2006; McGranahan et al., 2007).

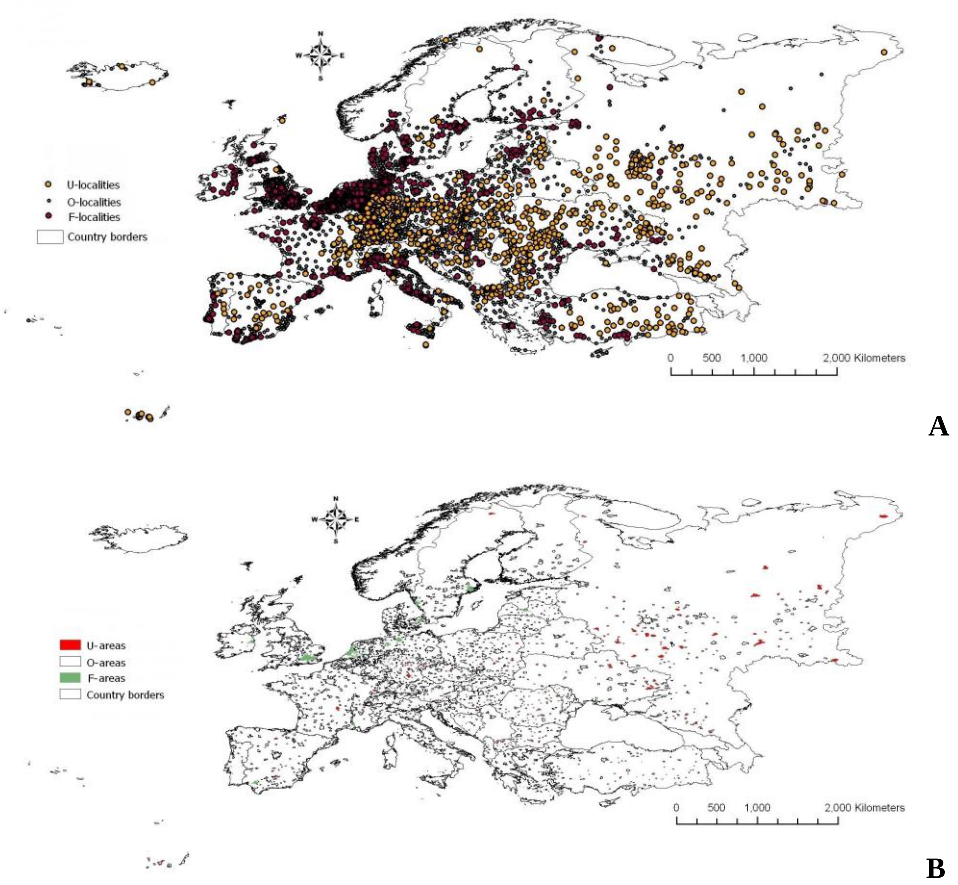
In calculating the 'Location Package' (LP) variable, the following location values were thus conditionally defined as favorable: elevation – 0-150 m; proximity to the sea shore and the nearest major population center <75 km (1 dd), generally considered practicable for daily commuting (Strutzer and Frey, 2004). In addition, 'fresh water' and 'highway' proximities were defined as favorable if distances were less than 0.1 dd (or ~7.5 km), thus giving the locality a LP score of 1, and 0 otherwise.

Using the LP scores of individual localities we split the sample (approximately 4,700 settlements) into three groups: 'favorable localities' (LP = 4 or 5), 'least favorable localities' (LP = 0 or 1) and the rest (LP = 2 or 3) (see Appendix 1). Similar criteria were also used for urban areas: 'favorable locations' ($LP \geq 4$), 'least favorable locations' ($LP \leq 1$) and the rest ($1 < LP < 4$) (see Appendix 2). In line with our initial assumption, we expected that among populated places with a favorable 'location package' (i.e., the first location group), the importance of population size would be smaller than for localities or urban areas for which this factor may be the only advantage (i.e., in 'least favorable' locations).

4.5. Statistical analysis

The analysis was performed in several steps. First, we examined the population growth rates of the whole set of localities available to us in order to check whether their population sizes and growth rates are independent. To this end, we first used scatter-plots and simple bi-variate OLS regressions. In the next phase, we used multiple regression analysis (MRA) to investigate how population size affects growth rates, while controlling for several location attributes.

Figure 2. Location subsets covered by the analysis – Individual localities (A), and Urban areas (B)



Note: U[nfavorable]-locations (LP = 0.1); B – F[avorable]-locations (LP = 4,5); C – O[ther]-locations (LP = 2,3).

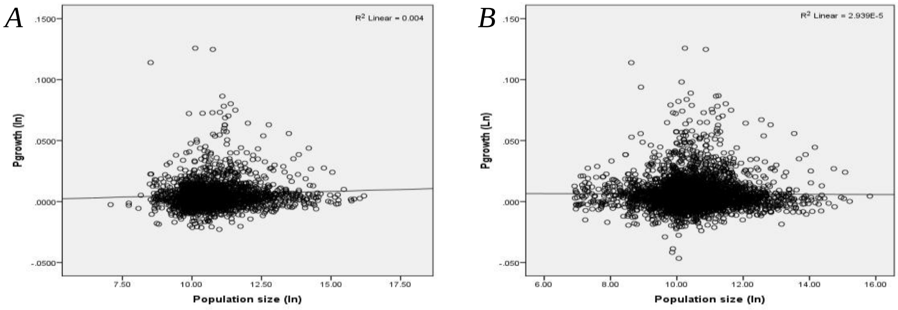
As discussed in Section 5.1, the analysis was performed separately for two levels of geographic resolution: first, individual localities (or LAUs), and, then, for urban contiguities, which are formed by adjacent individual LAUs (see Fig. 2 and the sub-section on data sources). Then we performed separate analyses for location subgroups (favorable locations, unfavorable ones and the rest of the sample), to determine whether the size-growth relationship differ across them, and whether the observed relationships change upon aggregation of individual LAUs into contiguous urban areas. We finally considered combined models containing dummy variables for location subgroups along with ‘population size-location group’ interaction terms.

5. RESULTS

5.1. General trends

The scatter-plots of G-S association and bi-variate regressions reported in Table 2 (Models 1-L and 1-U) point out that there appears to be no significant relationship between population sizes and growth rates of *localities (*t = -0.370; P>0.7; Model 1-L), with the trend line being nearly parallel to the population axis (see Fig. 3a). This outcome is fully consistent with Gibrat's Law. In contrast, for *urban areas*, population size and growth rate are positively associated (t = 2.808; P < 0.01; Model 1-U), with the trend line showing a slight upward trend (Fig. 3b). This indicates that, on the average, larger urban areas tend to grow faster than smaller ones. This reversal of G-S growth relationships upon aggregation of localities into contiguous urban areas is thus fully in line with a possibility predicted by our numerical tests (see Fig. 1b).

Figure 3. Annualized population growth rates of localities (A) and urban areas (B) as a function of their population size (ln)



5.2. Multivariate analysis

The resulting models for the whole sample of localities and urban areas are reported separately in Table 2. As previously mentioned, Models 1-L and 1-U are simple bi-variate regressions, in which population growth of individual localities (Model 1-L) and urban areas (Model 1-U) are regressed on their population size. Models 2-L and 2-U add population density (ln) to the regression models. Lastly, Models 3-L and 3-U include location attributes of localities (Model 3-L) and urban areas (Model 3-U), but *exclude* fixed effects (i.e., countries' dummies). Models 4-L and 4-U include these fixed effects. In addition, Figure 4 shows changes in the values of *t*-statistic for the population size variable in different regression models, thus helping to compare at glance, *t*-statistics and their significance levels.

Table 2. Factors affecting the rates of population growth across urban localities and urban areas (All locations; method – multivariate regression; dependent variable – population growth rate (ln))

Variable	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.^{c}
	A. Localities											
	Model 1-L			**Model 2-L**			**Model 3-L**			**Model 4-L**		
(Constant)	6.81E-03	3.673	0.000	7.09E-03	3.739	0.000	5.84E-02	19.461	0.000	4.88E-02	9.383	0.000
Pop1990_ln	-6.53E-05	-0.370	0.711	-5.10E-05	-0.287	0.774	-5.61E-04	-3.380	0.001	-1.17E-03	-7.400	0.000
Density_ln	-	-	-	-9.84E-05	-0.710	0.477	1.31E-03	7.386	0.000	1.77E-03	9.214	0.000
D_shore	-	-	-	-	-	-	-3.79E-05	-0.083	0.934	-3.51E-04	-0.714	0.475
D_mcity	-	-	-	-	-	-	3.59E-04	2.175	0.030	4.12E-04	2.558	0.011
D_mroad	-	-	-	-	-	-	-9.04E-04	-4.307	0.000	-6.89E-05	-0.209	0.834
Elevation	-	-	-	-	-	-	1.70E-03	4.452	0.000	1.05E-03	2.868	0.004
Latitude	-	-	-	-	-	-	3.42E-04	4.904	0.000	-1.38E-04	-0.996	0.319
D_water	-	-	-	-	-	-	4.40E-04	2.228	0.026	8.45E-05	0.383	0.702
Tmp_january	-	-	-	-	-	-	4.16E-04	4.942	0.000	5.13E-04	3.377	0.001
Evaporation index	-	-	-	-	-	-	-3.77E-03	-21.683	0.000	-1.51E-03	-4.987	0.000
Country dummies	No			No			No			Yes		
No of obs.	4667			4667			4667			4667		
R^2	0.000			0.000			0.172			0.366		
R^2-adjusted	0.000			0.000			0.170			0.359		
SEE	0.012			0.012			0.011			0.010		
F	0.137			0.321			96.77			54.377		
	B. Urban areas											
	Model 1-U			**Model 2-U**			**Model 3-U**			**Model 4-U**		
(Constant)	-1.08E-03	-0.448	0.654	-7.93E-03	-3.040	0.002	0.040	8.450	0.000	0.036	4.846	0.000
Pop1990_ln	6.25E-04	2.808	0.005	1.63E-04	0.705	0.481	-3.94E-04	-1.938	0.053	-3.77E-04	-2.031	0.042
Density_ln	-	-	-	1.87E-03	6.651	0.000	2.42E-03	9.040	0.000	1.36E-03	4.921	0.000
D_shore	-	-	-	-	-	-	2.83E-04	1.907	0.057	4.15E-04	2.661	0.008
D_mcity	-	-	-	-	-	-	1.14E-04	0.574	0.566	5.02E-04	2.625	0.009
D_mroad	-	-	-	-	-	-	-7.67E-04	-2.664	0.008	-3.27E-04	-0.977	0.329
Elevation	-	-	-	-	-	-	-9.96E-07	-0.855	0.392	-1.30E-06	-1.131	0.258
Latitude	-	-	-	-	-	-	8.13E-04	8.481	0.000	-6.22E-05	-0.339	0.735
D_water	-	-	-	-	-	-	4.61E-04	1.710	0.087	-1.34E-04	-0.463	0.644
Tmp_january	-	-	-	-	-	-	7.26E-04	6.037	0.000	5.46E-04	2.669	0.008
Evaporation index	-	-	-	-	-	-	-4.62E-03	-20.011	0.000	-1.70E-03	-4.458	0.000
Country dummies	No			No			No			Yes		
No of obs.	2189			2189			2189			2189		
R^2	0.004			0.023			0.283			0.500		
R^2-adjusted	0.003			0.022			0.280			0.488		
SEE	0.012			0.012			0.010			0.009		
F	7.884			26.138			86.104			43.572		

Note: ^{a} unstandardized regression coefficient; ^{b} t-statistic; ^{c} actual significance of t-statistic (two-tailed).

Models 1-L &1-U: *Bi-variate models for population size vs. population growth (density, location attributes and country-specific fixed effects are excluded);*

Models 2-L &2-U: *Multivariate models with density variable added, and location attributes and country-specific fixed effects excluded;*

Model 3-L &3-U: *Multivariate models with density variable and location attributes added while country dummies excluded;*

Models 4-L &4-U*: Multivariate models with all predictors added.*

Table 3. Factors affecting the rates of population growth across urban localities and urban areas (Unfavorable locations; method – multivariate regression; dependent variable – population growth rate (ln))

Variable	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.^{c}
	A. Localities											
	Model 5-L			**Model 6-L**			**Model 7-L**			**Model 8-L**		
(Constant)	-0.007	-1.482	0.139	-0.005	-0.923	0.356	0.082	10.026	<0.001	0.088	5.228	<0.001
Pop1990_ln	1.27E-03	2.610	0.009	1.26E-03	2.594	0.010	-3.27E-04	-0.762	0.446	-5.48E-04	-1.282	0.200
Density_ln				-6.69E-04	-1.488	0.137	1.40E-03	2.780	0.006	1.45E-03	2.600	0.009
D_shore							-4.56E-03	-2.191	0.029	-1.97E-03	-0.857	0.392
D_mcity							1.42E-03	3.639	<0.001	1.42E-03	3.628	<0.001
D_mroad							-1.17E-03	-2.394	0.017	-7.76E-04	-1.335	0.182
Elevation							-1.31E-03	-1.001	0.317	-9.29E-04	-0.730	0.466
Latitude							-2.47E-04	-1.059	0.290	-8.35E-04	-1.837	0.067
D_water							2.27E-04	0.488	0.626	8.66E-05	0.172	0.863
Tmp_january							-1.42E-04	-0.661	0.509	-5.13E-04	-1.232	0.218
Evaporation index							-3.73E-03	-8.129	<0.001	-1.87E-03	-2.079	0.038
Country dummies	No			No			No			Yes		
No of obs.	904			904			904			904		
R^2	0.007			0.010			0.344			0.495		
R^2-adjusted	0.006			0.008			0.337			0.469		
SEE	0.015			0.015			0.012			0.011		
F	6.814			4.519			46.833			19.163		
	B. Urban areas											
	Model 5-U			**Model 6-U**			**Model 7-U**			**Model 8-U**		
(Constant)	-0.020	-2.961	0.003	-0.032	-4.635	<0.001	0.064	5.657	<0.001	0.072	4.208	<0.001
Pop1990_ln	2.47E-03	3.871	<0.001	9.39E-04	1.412	0.158	-8.55E-04	-1.496	0.135	-3.02E-04	-0.540	0.589
Density_ln				4.39E-03	6.186	<0.001	3.24E-03	5.139	<0.001	2.50E-03	3.644	<0.001
D_shore							5.46E-04	2.149	0.032	5.63E-04	2.017	0.044
D_mcity							1.25E-03	2.939	0.003	1.02E-03	2.471	0.014
D_mroad							-1.14E-03	-1.863	0.063	-8.01E-04	-1.376	0.169
Elevation							-1.37E-06	-0.613	0.540	-4.33E-06	-1.878	0.061
Latitude							-8.80E-06	-0.032	0.974	-2.68E-04	-0.510	0.610
D_water							-4.73E-04	-0.841	0.401	7.64E-05	0.128	0.898
Tmp_january							1.02E-04	0.343	0.732	-1.04E-04	-0.196	0.845
Evaporation index							-3.95E-03	-6.914	<0.001	-3.85E-03	-3.724	<0.001
Country dummies	No			No			No			Yes		
No of obs.	584			584			584			584		
R^2	0.025			0.085			0.453			0.591		
R^2-adjusted	0.023			0.082			0.444			0.562		
SEE	0.014			0.014			0.011			0.009		
F	14.988			27.108			47.484			20.686		

Note: [a] unstandardized regression coefficient; [b] t-statistic; [c] actual significance of t-statistic (two-tailed).

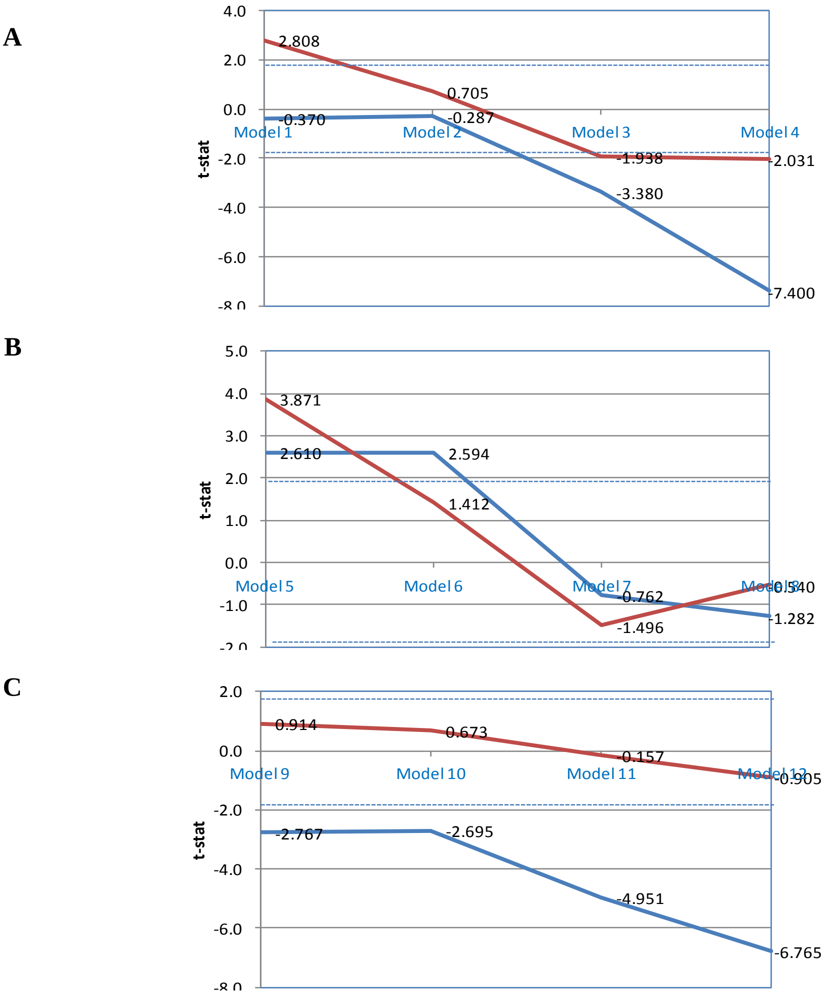
Models 5-L &5-U: *Bi-variate models for population size vs. population growth (density, location attributes and country-specific fixed effects are excluded);*

Models 6-L &6-U: *Multivariate models with density variable added, and location attributes and country-specific fixed effects excluded;*

Model 7-L &7-U: *Multivariate models with density variable and location attributes added while country dummies excluded;*

Models 8-L &8-U*: Multivariate models with all predictors added.*

Figure 4. Statistical significance of the association (t-statistic) between population growth and size for urban localities and urban areas stratified according to their location attributes



A - All locations; B - U[nfavorable]-locations (LP=<1); C – F[avorable]- locations (LP>=4).
Note: Doted lines indicate a two-tailed 0.05 significance interval.
Models 1, 5 & 9: *Bi-variate models for Population size vs. population growth (density, location attributes and country-specific fixed effects are excluded);*
Models 2, 6 & 10: *Multivariate models with density variable added, and location attributes and country-specific fixed effects excluded;*
Model 3, 7 & 11: *Multivariate models with density variable and location attributes added, and country dummies excluded;*
Models 4, 8 & 12: *Multivariate models with all the predictors added.*

As Table 3 shows, the *population size* variable emerges as statistically *insignificant* in the two first models for localities, in which *only* the population size and population density variables are included (Models 1-L and 2-L). However, it becomes highly statistically significant in all the models, in which locational attributes and countries' fixed effects are controlled ($|t|>3.3$; $P<0.01$), steadily exhibiting a negative sign (Models 3-L and 4-L). This indicates that, when locational differences are taken into account, *bigger cities grow slower than smaller ones*, thus contradicting Gibrat's Law. Possible reasons for this trend are more or less clear: emergence of disagglomeration economies in big cities, such as high housings costs, traffic congestions, air pollution, etc.

Multivariate models with locational attributes, estimated for *urban areas* and also reported in Table 3 (Models 3,4-U), essentially lead to the same conclusion, with the regression coefficients for the population size variable being significant ($t>1.9$; $P<0.1$), albeit *positive, and not negative*, as in the models estimated for individual localities (see Table 3: Models 3,4-L). Moreover, the population density variable (estimated, as mentioned, as the average density of population residing in a given urban area) emerged as positive and highly statistically significant in all the models estimated for urban areas ($t>4.9$; $P<0.01$; see Models 2-U–4-U; Table 3). The importance of these results will be discussed in the concluding section of the paper.

5.3. Location sub-groups

The analysis of the G-S association in LAUs and urban areas which differ by their location attributes may help to shed some light on the reversal of the regression coefficients for the population size variable noted in the previous sub-section. As we hypothesized from the outset of the analysis, the G-S relationship should expectedly depend on a populated place's locational settings (see the sub-section on explanatory variables). To verify this hypothesis, we split, as previously mentioned, the entire set of localities into three location groups – localities with fewest location advantages (Group 1); well positioned localities (Group 2) and the rest of the sample (Group 3). To this end, the notion of 'location package' was used, as detailed in the statistical analysis section. Out of 4,667 individual localities covered by the study, 904 localities were in Group 1, 1061 in Group 2, and 2702 localities in Group 3. Some 2189 urban areas covered by the analysis were also classified into the above three locational groups using the same ‘location package’ criterion (see Appendix 2).

Figures 5-6 feature population size-growth relationship in each location subgroup covered by the study – either *LAUs* (Fig. 5) or *urban areas* (Fig. 6). Although regression fit (measured by R^2) in any of these groups is not especially high, the G-S association appears to be *positive* for 'unfavorable' localities (see Fig. 5A) and *negative* – for ‘favorable’ ones, that is, for LAUs with more location advantages (Fig. 5B). Concurrently, the rest of the sample (Fig. 5C) shows the 'no-trend' relationship, like that detected for the entire sample of settlements (see Fig. 3A). This suggests that in line with our initial hypothesis, *urban places with different location attributes do show different relationships between size and growth, but these contrasting trends appear to cancel each other out if the whole set of localities is considered.*

Table 4. Factors affecting the rates of population growth across urban localities and urban areas (Favorable locations; method – multivariate regression; dependent variable – population growth rate (ln))

Variable	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.^{c}	B^a	t^b	Sig.^{c}
	A. Localities											
	Model 9-L			**Model 10-L**			**Model 11-L**			**Model 12-L**		
(Constant)	0.016	4.685	<0.001	0.019	5.095	<0.001	0.064	9.317	<0.001	0.056	4.271	<0.001
Pop1990_ln	-8.70E-04	-2.767	0.006	-8.46E-04	-2.695	0.007	-1.53E-03	-4.951	<0.001	-1.99E-03	-6.765	<0.001
Density_ln				-6.59E-04	-1.999	0.046	7.73E-04	1.823	0.069	1.19E-03	2.067	0.039
D_shore							5.97E-04	0.662	0.508	-1.19E-04	-0.109	0.913
D_mcity							-1.61E-04	-0.344	0.731	2.82E-05	0.060	0.952
D_mroad							-1.20E-02	-0.842	0.400	-9.14E-03	-0.674	0.501
Elevation							-2.53E-03	-1.404	0.161	-3.16E-03	-1.855	0.064
Latitude							2.58E-04	1.611	0.107	7.85E-06	0.025	0.980
D_water							1.71E-03	2.044	0.041	1.01E-03	1.032	0.302
Tmp_january							3.31E-04	1.558	0.119	1.20E-03	2.950	0.003
Evaporation index							-2.93E-03	-6.818	<0.001	-1.18E-03	-1.555	0.120
Country dummies	No			No			No			Yes		
No of obs.	1061			1061			1061			1061		
R^2	0.007			0.011			0.135			0.317		
R^2-adjusted	0.006			0.009			0.126			0.290		
SEE	0.010			0.010			0.010			0.009		
F	7.654			5,836			16.326			11.548		
	B. Urban areas											
	Model 9-U			**Model 10-U**			**Model 11-L**			**Model 12-L**		
(Constant)	8.87E-04	0.145	0.885	-3.43E-04	-0.052	0.959	2.31E-02	1.840	0.067	1.77E-02	0.830	0.407
Pop1990_ln	5.10E-04	0.914	0.362	4.03E-04	0.673	0.502	-8.49E-05	-0.157	0.876	-4.22E-04	-0.905	0.366
Density_ln				3.69E-04	0.500	0.617	1.45E-03	1.877	0.062	1.34E-03	1.835	0.068
D_shore							1.95E-03	2.109	0.036	1.81E-03	2.036	0.043
D_mcity							2.98E-04	0.377	0.707	-1.42E-03	-1.901	0.059
D_mroad							-4.19E-03	-0.158	0.875	1.78E-02	0.796	0.427
Elevation							-2.16E-05	-2.470	0.014	-1.37E-05	-2.005	0.046
Latitude							1.36E-03	4.656	<0.001	1.59E-03	2.986	0.003
D_water							4.84E-03	2.617	0.009	2.63E-03	1.412	0.159
Tmp_january							1.34E-03	3.504	0.001	2.09E-03	3.186	0.002
Evaporation index							-5.08E-03	-7.281	<0.001	-3.49E-03	-3.127	0.002
Country dummies	No			No			No			Yes		
No of obs.	266			266			266			266		
R^2	0.003			0.004			0.265			0.637		
R^2-adjusted	-0.001			-0.003			0.236			0.575		
SEE	0.011			0.011			0.010			0.007		
F	0.835			0.541			9.195			12.631		

Note: ^{a} unstandardized regression coefficient; ^{b} t-statistic; ^{c} actual significance of t-statistic (two-tailed).

Models 9-L &9-U: *Bi-variate models for population size vs. population growth (density, location attributes and country-specific fixed effects are excluded);*

Models 10-L &10-U: *Multivariate models with density variable added, and location attributes and country-specific fixed effects excluded;*

Model 11-L &11-U: *Multivariate models with density variable and location attributes added while country dummies excluded;*

Models 12-L &12-U: *Multivariate models with all predictors added.*

Table 5. Factors affecting the rates of population growth in localities and urban areas (method – multivariate regression; dependent variable – population growth rate (ln); location-population size interaction terms and fixed effects added)

Variable	B[a]	t[b]	Sig.[c]	B[a]	t[b]	Sig.[c]	B[a]	t[b]	Sig.[c]	B[a]	t[b]	Sig.[c]
	Model 13-L			**Model 14-L**			**Model 13-U**			**Model 14-U**		
Constant	5.60E-02	15.017	<0.001	4.65E-02	7.911	<0.001	4.21E-02	8.405	<0.001	3.84E-02	5.034	<0.001
Pop1990_ln	-5.28E-04	-2.409	0.016	-1.23E-03	-5.982	<0.001	-5.68E-04	-2.196	0.028	-4.49E-04	-1.941	0.052
Density_ln	8.50E-04	4.783	<0.001	1.64E-03	8.381	<0.001	2.37E-03	8.846	<0.001	1.31E-03	4.724	<0.001
D_shore	3.71E-04	2.885	0.004	4.20E-04	2.978	0.003	3.80E-04	2.515	0.012	5.17E-04	3.267	0.001
D_mcity	-2.13E-05	-0.160	0.873	5.38E-04	3.298	0.001	2.76E-04	1.347	0.178	6.52E-04	3.315	0.001
D_mroad	1.06E-03	2.006	0.045	8.75E-04	1.771	0.077	-8.64E-04	-2.987	0.003	-4.01E-04	-1.199	0.230
Elevation	1.96E-07	0.182	0.855	-6.46E-07	-0.621	0.535	3.52E-07	0.288	0.773	-6.49E-08	-0.055	0.956
Latitude	3.89E-04	5.210	<0.001	-4.74E-05	-0.318	0.751	7.38E-04	7.57	<0.001	-1.33E-04	-0.723	0.470
D_water	5.62E-04	2.798	0.005	3.28E-05	0.149	0.881	5.36E-04	1.974	0.049	-8.35E-05	-0.289	0.773
Tmp_january	5.66E-04	5.664	<0.001	6.68E-04	3.936	<0.001	6.47E-04	5.295	<0.001	4.93E-04	2.408	0.016
Evaporation index	-3.72E-03	-19.203	<0.001	-1.60E-03	-5.003	<0.001	-4.46E-03	-19.136	<0.001	-1.62E-03	-4.243	<0.001
F-areas	1.31E-02	3.046	0.002	8.53E-03	2.224	0.026	5.63E-03	1.062	0.288	6.28E-03	1.389	0.165
U-areas	-1.09E-02	-2.471	0.014	-8.91E-03	-2.279	0.023	-7.74E-03	-1.422	0.155	-6.66E-03	-1.424	0.155
F-areas*Pop	-1.05E-03	-2.610	0.009	-7.19E-04	-2.007	0.045	-3.09E-04	-0.65	0.516	-4.55E-04	-1.122	0.262
U-areas*Pop	9.89E-04	2.326	0.020	7.27E-04	1.931	0.054	6.05E-04	1.192	0.233	4.67E-04	1.068	0.286
Country dummies	No			Yes			No			Yes		
No of obs.	4667			4667			2189			2189		
R^2	0.174			0.370			0.289			0.504		
R^2-adjusted	0.172			0.363			0.284			0.492		
SEE	0.011			0.010			0.010			0.009		
F	70.164			51.206			63.125			40.917		

Note: [a] unstandardized regression coefficient; [b] t-statistic; [c] actual significance of t-statistic (two-tailed).

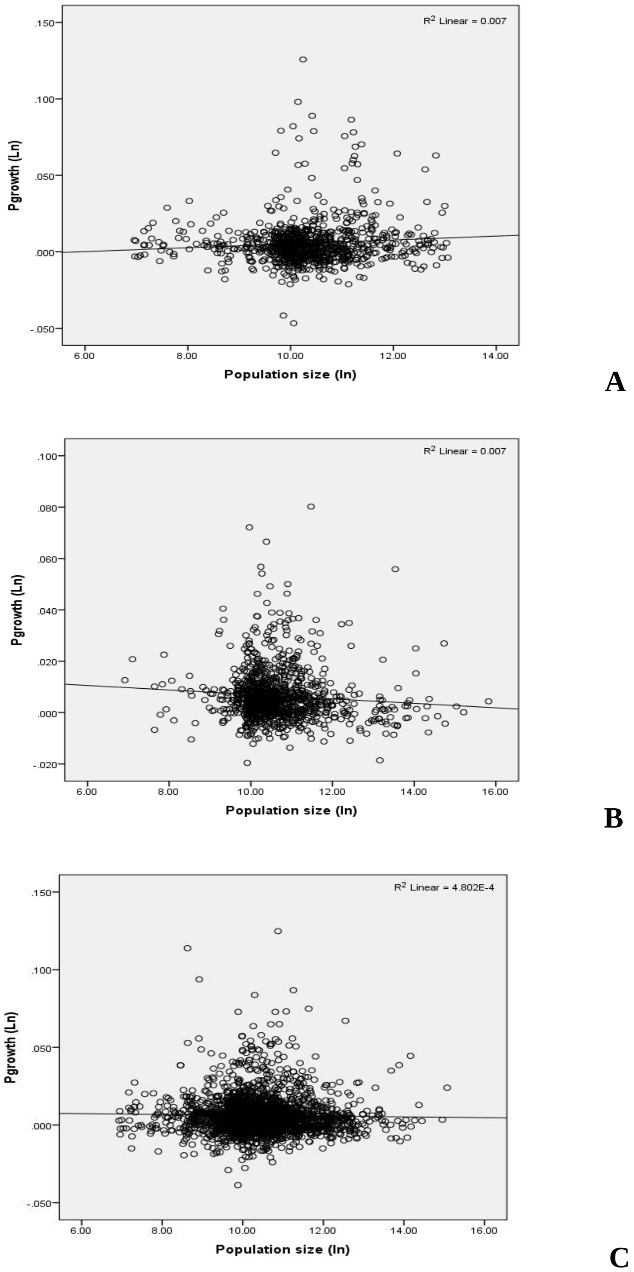
***Models 13-L**: Units – LAUs; all development predictors, location package fixed effects (F[avourable]-locations, U[nfourable]-locations) and interaction terms added; countries' fixed effects excluded;*

***Models 14-L**: Units – LAUs; all development predictors, location package fixed effects (F[avourable]-locations, U[nfourable]-locations) and interaction terms added; countries' fixed effects added.*

***Models 13-U**: Units – urban areas; all development predictors, location package fixed effects (F[avourable]-locations, U[nfourable]-locations) and interaction terms added; countries' fixed effects excluded;*

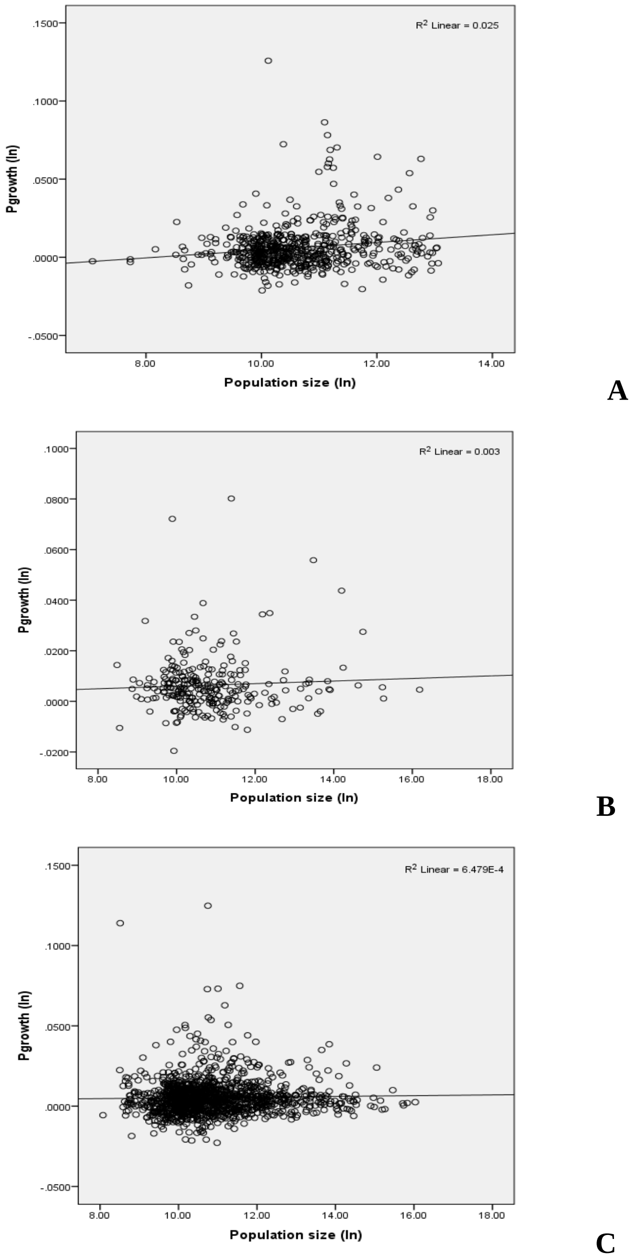
***Models 14-U**: Units – urban areas; all development predictors, location package fixed effects (F[avourable]-locations, U[nfourable]-locations) and interaction terms added; countries' fixed effects added.*

Figure 5. Annualized population growth rates vs. population size (ln) of localities stratified by location subsets



A –U[nfavorable]-locations (LP ≤ 1); C – F[avorable]-locations (LP ≥ 4); C – O[ther]- locations (1 < LP < 4)

Figure 6. Annualized population growth rates vs. population size (ln) of urban areas stratified by location subsets



A –U[nfavorable]-locations (LP=<1); B – F[avorable]- locations (LP>=4); C – O[ther] urban areas v

Simple bi-variate regression models, reported in Tables 3-4 (Models 5-L and 9-L), also confirms that the slopes of 'size-growth' regression lines for 'favorably' and 'unfavorably' located towns are different, being positive in the 'unfavorable locations' group and negative in the 'favorable locations' group ($P < 0.01$; Tables 3-4). This conclusion is further strengthened by two other, somewhat more sophisticated regression models in which ‘population size-location group’ interaction terms are included (F-areas*Pop: $t < -2.2$; $P < 0.05$; U-areas*Pop: $t > 1.9$; $P < 0.1$; see Models 13-L and 14-L in Table 5).

The *positive* association between population growth and city sizes observed in *unfavorable* locations, may have a simple explanation: Favorably located settlements may have enough advantages to retain current residents and become sufficiently attractive for potential newcomers even without 'size benefits'; however, for unfavorably located settlements, with fewer location advantages, larger population sizes may be a necessary precondition for an increase in the population growth rate, compensating, at least in part, for their location drawbacks.

Do the above relationships between size and growth persist if controlled for confounders, such as the country of a town's location, proximity to major city, sea shore, regional population density, etc.?

As additional models reported in Tables 3&4 (Models 6-L–8-L (Table 3) and Models 10-L–12-L (Table 4)) show, the *negative* association between population size and growth remains highly statistically significant in the ‘favorable’ subset of LAUs ($P < 0.01$; see Models 10-L–12-L in Table 4 and Fig. 4C) but loses its statistical significance in the ‘unfavorable’ subset ($P > 0.2$; Models 6-L–8-L in Table 3 and Fig. 4B).

Characteristically, when *urban areas* are used as observation units, instead of individual localities, the G-S association emerges as *negative*, albeit statistically weak ($P<0.05$), in most of the models considered (see Tables 3-5 and Fig. 4). The explanation for this trend may be fairly straightforward: the merging of groups of localities with different G-S associations (for most of which this association is negative) results in a weak negative G-S association for urban areas, when those are used as analysis units, which is fully in line with our numerical tests (see Section 4 and Fig. 1).

6. CONCLUSION

Upon the desegregation of data, several, often opposite, trends emerge. Thus, a *zero* net migration balance in a locality may 'hide' an outflow of wealthy residents to suburban areas and an inflow of poor families into central cities (such as in North America), or the opposite trend, observed in most European countries, with the two opposite currents being of similar strength (Portnov et al., 2011). A similar process may be at work when population growth and population size of localities are mutually compared. In particular, opposite G-S association, potentially observed in different subgroups of localities (stratified e.g., by their locational attributes or population composition) may cancel each other out, leading to the emergence of the proportionate effect of growth-size relationship expected under Gibrat's Law.

In this study, we considered a possibility that growth rates (G) *may depend* on size (S), while the direction of relationship differs, i.e., a *monotonic increase in unfavorable loci vs. a monotonic decrease elsewhere.* Over the entire settlement system, these opposite trends can expectedly cancel each other out, with the 'no-trend' relationship, expected under Gibrat's Law, emerging. As we further hypothesized, due to the varying G-S relationships observed in different groups of localities, the strength and direction of the G-S association may also change, depending on the relative shares of location subsets used for aggregation.

We performed our analysis at two levels of geographic resolution – 4,667 local administrative units (i.e., municipalities) and 2189 urban areas, formed by territorial contiguities of municipal built areas, using data available for 40 European countries. The choice of spatial aggregation units is a critical consideration for the analysis because alternative units (e.g., urban area or municipalities) may expectedly have different mechanisms behind their population growth. Thus, selecting *urban areas* as units of the analysis reflects the fact that urban areas are likely to function as a whole and may thus be considered as economically integrated units. However, development disparities between local administrative units (i.e., municipalities) may also have a profound effect on population growth patterns as favorably located and attractive municipalities may provide better services and facilities, thus appealing to more migrants and businesses.

Since Gibrat's growth model explicitly assumes the independence of individual observations, for its proper testing, local interdependencies that constitute regions should be minimized. A prominent type of such local interdependencies is agglomeration economies. For instance, the specific growth characteristics of a suburb of a city are inherently shaped by the present of a large urban center nearby. The characteristics of the same suburb would be different, if there was no urban center in its vicinity, and its role in the intra-metropolitan division of labor (including population density and dynamics) may thus be shaped more by the urban center or the metropolitan area as a whole, than by the suburb itself (Portnov et al., 2011). From this perspective, the detection of a positive association between population sizes and growth at the urban area level of spatial resolution, both before and after controlling for location attributes, is especially important, as it supports our initial research hypothesis that *growth rates do depend on population size, contrary to what is expected under Gibrat's Law.*

This conclusion is generally in line with Kalecki's (1945) hypothesis of a correlation between growth and size. Anderson and Ge (2005) have also argued, albeit from a different perspective, that Kalecki's assumption fitted Chinese city data better than the independence of growth on size which is the corollary of Gibrat's Law.

Using the ‘location package’ concepts, introduced by Portnov and Schwarts (2009a), we also split the whole set of European urban localities and urban areas into three location subsets – favorably located, unfavorably located and the rest. One clarification is required: hence urban growth is highly variable and fluctuant, as there are no *permanently ‘favorable’ or ‘unfavorable’ loci*, since relative importance of location attributes may change over time. The importance of particular location attributes may diminish (or increase overtime, such as with, e.g., sea-freight and railroad transportation) thus *changing the ‘anchoring’* of main *foci* of urban growth in geographic space, gradually responding to changes in development needs and in the appeal of particular location attributes at a given point of time. However, the spatial "anchoring" of favorable and unfavorable loci may conditionally be considered constant over relatively short periods of time, such as the ten-year period, between 1990 and 2000, covered by the present study.

As our analysis points out, the *negative* effect of *population size* on the population growth of individual towns appears to be positive (albeit statistically weak) in the 'unfavorable localities' group and negative elsewhere. A possible explanation for these differences may be as follows: the adverse effects of agglomeration may be less profound in disadvantaged areas, in which the absence of other favorable location attributes may boost the effect of size on urban growth, by offering a 'safety net' for residents in terms of employment and cultural opportunities and larger market opportunities for local businesses. Concurrently, in areas with more location advantages, the *importance of size* for the development of individual towns may be less profound (see *inter alia* Portnov and Erell, 2001).

Hence the G-S relationships appears to differ both in strength and direction across different locational subgroups of localities, the merging of groups of localities with different G-S associations (for most of which this association is negative) into contiguous urban areas thus resulted in a weak negative G-S association for urban areas, when those are used as analysis units. This was demonstrated by our empirical analysis and sensitivity tests. The difference in the results obtained at different levels of geographic resolution (that is, individual localities vs. integrated urban areas) is, in fact, a well-known phenomenon of changing relationships between variables in line with data aggregation into areal units of larger size, which Openshaw (1984) termed the ‘modifiable areal unit problem’ or MAUP.

There are several limitations in our study, which should be taken into consideration. According to a popular interpretation of Gibrat's growth model (see *inter alia* Robson, 1973; Guérin-Pace, 1993), it is conceived to explain the proportionate growth of localities over *long periods* of time. Yet, according to our findings, when individual localities are considered, 'proportionate' growth (expected under Gibrat's Law) does emerge at the aggregate (system-wide) level even for a relatively *short time-span* covered by the analysis (that is, one decade between 1990 and 2000). This implies that the observed relationships *do not* necessarily require a long time span to emerge. However, this size-growth rate independence 'dissipates' when the settlement system is disaggregated into two urban sub-systems, formed by ‘well-positioned’ localities and ‘poorly positioned’ ones, which is fully in line with our initial research hypothesis.

Lastly, while the present study covers most European cities and towns with populations of 20,000+ residents, localities of smaller size are less fully represented, due to incomplete data on population growth rates. Our findings are thus primarily pertinent to the larger settlements on the continent. Moreover, our classification of localities into 'favorable' and 'unfavorable,' based of the 'package' of location attributes, is somewhat arbitrary and may be improved on by more detailed classifications (e.g., based on additional location criteria and their interaction terms). Furthermore, different size-growth relationships may emerge along additional (i.e., non-locational) 'seam-lines', such as established vs. transitional economies, local towns vs. metropolitan areas, etc. Such possibilities may deserve investigation in future studies.

REFERENCES

Ades A. F. and Glaeser E.L. (1995) 'Trade and Circuses: Explaining Urban Giants,' *The Quarterly Journal of Economics*, 110(1), 195-227.

Black D. and Henderson V. (2003) Urban evolution in the USA, *Journal of Economic Geography*, 3: 343-372.

Brinkhoff T. (2007) City Population Database (http://www.citypopulation.de).

Cheshire P.C. and Magrini S. (2006) Population Growth in European Cities: Weather Matters – But only Nationally, *Regional Studies*, Vol. 40(1): 23-37.

Clark S.J., and Stabler J.C. (1991) Gibrat's Law and the Growth of Canadian Cities, *Urban Studies*, 28(4): 635-639.

Duranton G. (1999) Distance, land, and proximity, economic analysis and the evolution of cities. *Research Papers in Environmental and Spatial Analysis No. 53*, Department of Geography and Environment, London School of Economics.

Eeckhout I. (2004) Gibrat's Law for (all) cities, *The American Economic Review*, 94(5): 1429-1451.

ESRI (2000) ESRI Data and Maps, Redlands, CA: Environmental Science Research Institute.

Gallup J.L., Sachs J., Mellinger A. (1999) Geography and Economic Development, *International Regional Science Review*, 22: 179-232.

Geonames (2007) Geonames Project Database (http://www.geonames.org).

Gibrat, R. (1931) *Les inégalités économiques : applications aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel. Paris, Recueil Sirey*. (Cited in: Kalecki M. (1945) On the Gibrat Distribution, *Econometrica*, 13(2): 161-170 and Moeckel R. (2009) Simulation of firms as a planning support system to limit urban sprawl of jobs, *Planning and Design*, 36(5): 883-905).

Gotway C. and Young L.J. (2002) Combining incompatible spatial data, *Journal of the American Statistical Association*, 97 (48): 632-647.

Guérin-Pace, F. (1993) *Deux siècles de croissance urbaine*. Paris, Anthropos (Cited in: Pumain D. and Moriconi-Ebrard F. (1997), City size distributions and metropolization, *GeoJournal* 43(3): 307-314).

Ioannides Y.M. and Overman H.G. (2003) Zipf's Law for cities: an empirical examination, *Regional Science and Urban Economics*, 33: 127-137.

Kalecki M. (1945) On the Gibrat Distribution, *Econometrica*, 13(2): 161-170.

McGranahan G., Balk D. and Anderson B. (2007) The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones, *Environment and Urbanization*, 19(1): 17-37.

Minami, M. and ESRI (2000). *Using ArcMap: GIS.* Redlands, California, ESRI.

Openshaw S. (1984) The modifiable areal unit problem, In: *Concepts and Techniques in Modern Geography.* Monograph Series. London: Geo Books; 38, 41 pp.

Polese M. and Shearmur R. (2006) Growth and Location of Economic Activity: The Spatial Dynamics of Industries in Canada 1971-2001, *Growth and Change,* 37 (3), 362-395.

Portnov B. A., and Erell E. (2001) *Urban Clustering: The Benefits and Drawbacks of Location*, Aldershot: Ashgate.

Portnov B.A. and Schwartz M. (2008) On the Relativity of Urban Location, *Regional Studies*, 42(4): 605-615.

Portnov B.A. and Schwartz M. (2009a) Urban Clusters as Growth Foci, *Journal of Regional Science*, 49(2): 287-310.

Portnov B.A. and Schwartz M. (2009b) On the Importance of 'Location Package' for Urban Growth, *Urban Studies*, 46(8): 1665-79.

Portnov B.A., B. Reiser and Schwartz M. (2011) Does Gibrat's Law for Cities Hold when Location Counts? *Annals of Regional Science* (in press; DOI: 10.1007/s00168-010-0402-8).

Pumain D. (1982) La dynamique des villes. *Paris, Economica* 231 p. (cited in Suarez-Villa L. (1988) Metropolitan Evolution, Sectoral Economic Change, and the City Size Distribution, *Urban Studies,* 25: 11-20).

Pumain D. and Moriconi-Ebrard F. (1997), City size distributions and metropolization, *GeoJournal* 43(3): 307-314.

Rappaport J. 2006. 'Moving to Nice Weather,' *Regional Science and Urban Economics*, 37, 375-398.

Robson B.T. (1973) Urban Growth. An Approach, *London: Methuen*, 268 pp.

Rose A.K. (2005) Cities and countries, *NBER Working Paper Series*, Working Paper 11762, National Bureau of Economic Research, Canbridge, MA.

Strutzer A. and Frey B.S. (2004) Stress that doesn't pay: the Commuting Paradox, *Discussion Paper Series IZA* DP No. 1278.

UNEP & UNESCO (2004) *Millennium Ecosystem Assessment Report,* NY: UNESCO.

Vlora, N.R. (1979) *Città et territorio.* Bologne: Pàtron (Cited in: Pumain, D (2000) Settlement systems in the evolution, *Geographical Analysis.,* 82 B (2): 73-87).

LE CHOIX DE L'UNITÉ URBAINE EST-IL FONDAMENTAL POUR LA VALIDATION DE LA LOI DE GIBRAT POUR LES VILLES ?

Résumé - *Dans cet article, nous admettons l'hypothèse que le choix des unités géographiques déterminant les aires urbaines est fondamental dans la validation de la loi de Gibrat pour les villes, d'après laquelle la croissance urbaine est indépendante des effets taille. Nous examinons, plus particulièrement, les différences d'interprétation selon que l'on utilise des localités individuelles ou des unités urbaines intégrées (comportant plusieurs localités). Cette analyse s'appuie sur une analyse empirique menée sur les changements démographiques entre 1990 et 2000 de 4667 unités administratives (communes) et de 2189 aires urbaines contigües de 40 pays européens. Nous montrons que les résultats varient selon le choix de l'unité urbaine utilisée, mais aussi selon leurs avantages de localisation : pour les unités urbaines où ces avantages sont importants la relation taille-croissance est moins importante que pour les unités où ces avantages sont faibles.*

Mots clés : CROISSANCE DÉMOGRAPHIQUE, AIRES URBAINES, AIRES ADMINISTRATIVES, SYSTÈME URBAIN EUROPÉEN, AVANTAGES DE LOCALISATION, LOI DE GIBRAT