

Portnov, Boris A.

Conference Paper

The Change of Support Problem (COSP) and its Implications for Urban Analysis: Some Evidence from a Study of the European Urban System

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Portnov, Boris A. (2010) : The Change of Support Problem (COSP) and its Implications for Urban Analysis: Some Evidence from a Study of the European Urban System, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118794>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Change of Support Problem (COSP) and its Implications for Urban Analysis: Some Evidence from a Study of the European Urban System

Boris A. Portnov

Department of Natural Resources & Environmental Management, University of Haifa, Israel

Abstract

The Change of Support Problem (COSP) reflects a possibility that the outcome of an urban analysis may critically depend on the researcher's choice of territorial units. To verify this assumption, the present study examines the association between population growth and population size of localities, using 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, when individual localities are considered, the growth of localities appears to be strictly proportional to population size, but 'dissipates' when the settlement system is disaggregated into two urban sub-systems, formed by well-positioned localities and poorly positioned ones. Concurrently, for urban areas, a strong positive association between population size and growth emerges both before and after controlling for location attributes. However, this association between population size and growth is not especially strong, if favorably and unfavorably located urban areas are looked at in separation.

Keywords: Urban growth; European urban system; urban areas; location attributes; Gibrat's Law

Introduction

The lack of significant association between pollution size of localities and the rates of their growth is known as Gibrat's Law for cities (Eeckhout 2004). According to this law, originally formulated for economic entities (such as firms and factories) and 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 the number of workers in the case of factories (Gibrat 1931, cited in Kalecki 1945). When applied to cities, growth rates are thus 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.¹

¹ The classification of a locality as either 'city' or 'town' depends on its population size and on the functions it performs, and may thus vary by country, depending on its land area, population size, level of economic development, etc. For simplicity's sake, in the present analysis, we will use interchangeably two generic terms - 'city' and 'locality,' - to describe all populated places covered by the analysis, irrespectively of their functions and population sizes. According to Eurostat (2009), for data collection purposes, these individual places are defined as LAU2 or Level 2 Local Administrative Units (formerly known as NUTS5) and consist of local municipalities or equivalent units in most European countries. Another term used throughout this study is 'urban area.' It refers to conglomerates of individual localities, forming territorial contiguities of urban development.

However, it seems patently obvious that for any urban system to develop, some places must attract more people and grow faster than others; otherwise a homogenous population distribution across space would have occurred. Indeed, with the considerable recent work on agglomeration economies and cumulative causation, it is generally acknowledged that the largest cities and metropolitan areas tend to grow faster, than their smaller counterparts (Pumain and Moriconi-Ebrard, 1997; Black and Henderson, 2003).

A skeptic may argue that the relationship between city size and growth is a statistical artifact with no particular importance, apart from a scholarly discussion. In our view, however, whether or not Gibrat's Law for cities is correct is *not* a purely theoretical question. If, for instance, the long-term growth of cities is indeed *strictly independent of their sizes*, urban development policies [aimed either at enhancing urban growth in priority areas or at restricting it in overpopulated regions] are doomed to failure. If, however, it is known that in some locations, settlements of particular sizes are more likely to achieve sustained population growth than their 'less fortunate' counterparts, that knowledge may help formulate effective development policies, aimed at restricting or encouraging urban growth, whenever one of these objectives is desirable.

According to a popular line of argument (see *inter alia* Eeckhout, 2004), the development path through which 'proportionate growth' can lead to 'city diversity,' may be explained in the framework of a general equilibrium theory, which postulates that every city periodically 'experiences exogenous technology shocks,' which change its 'equilibrium utility' and may accelerate population growth in the short run. However, this 'short-term exogenous shock' theory fails to explain why, over past decades, major cities and their metropolitan areas (Greater London, Madrid, Paris, etc.,) have grown much faster than the rest of the cities in their respective countries, thus essentially 'invalidating Gibrat's urban growth model' (Pumain and Moriconi-Ebrard, 1997).

Another possible explanation for the emergence of 'proportionate growth' in urban systems is the existence of different population size – population growth relationships in different groups of cities. For the sake of simplicity, let's assume that there are two kinds of *places* – one favorable to urban development and another one - less favorable (or unfavorable). For instance, these may be stagnating inner city rings vs. rapidly growing suburban belts, or booming coastal areas vs. less attractive inlands. Now assume that growth rates across each settlement group *do* depend on size (contrary to Gibrat's Law). However, the *direction of relationship differs*, e.g., monotonic increase in unfavorable *loci* vs. monotonic decrease (or no direct relationship) elsewhere. Over the entire settlement system, these opposite trends may cancel each other out, generating the 'no-trend' relationship, expected under Gibrat's Law.

Another possibility is that the relationship between the population growth of localities and their population sizes may depend on the territorial units used for the analysis, thus reflecting the change of support problem (COSP) in spatial data analysis (Gotway and Young, 2002). For instance, different trends may potentially emerge if individual localities (e.g., LAU's or municipalities) are used in the analysis, as opposed to integrated territorial units (i.e., contiguous urban areas), combining several adjacent LAUs.

In this paper, we attempt to examine these possibilities, using population growth data for 4,667 local administrative units (i.e., municipalities) and 2189 urban areas, formed by territorial contiguities of built-up areas in 40 European countries. Our findings *do not* support Gibrat's Law at either level of spatial resolution. In particular, there appears to be a *significant relationship between the towns' sizes and rates of growth, which emerges upon controlling for settlement proximity to the nearest major city, population density, latitude, etc.* Thus, at the aggregate (system-wide) level, no association of city size and growth of individual localities emerged and Gibrat's law appeared vindicated. However, Gibrat's law turned into a statistical artifact when the settlement system was *disaggregated into two urban sub-systems, formed by well-positioned localities vs. poorly positioned ones.* Concurrently, for *urban areas* a strong positive association between population sizes and their growth emerged both *before and after* controlling for location attributes. Yet, the association between population size and growth was not found to be especially strong once favorably and unfavorably located urban areas were looked at separately.

Previous studies of Gibrat's Law for cities

While several empirical studies (see *inter alia* Ioannides and Overman, 2003; Eeckhout 2004; Rose 2005) support Gibrat's Law for cities, other empirical studies (cf. e.g., Pumain and Moriconi-Ebrard, 1997; Black and Henderson, 2003) lead to the opposite conclusion, viz., that growth rates significantly depend on population sizes of individual localities and urban areas under analysis.

The simplest formulation of Gibrat's Law applied to city sizes is as follows:

$$\ln P_t = \mu + \ln P_{t-1} + u_t, \quad [1]$$

or alternatively:

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

where P_t is the population size of a locality at time t , μ represents an average rate of growth, and u_t is a random error term (see Anderson and Ge, 2005). (*Note:* The left hand side of [2] denotes the logarithm of the growth rate).²

Clark and Stabler (1991) use the following more complex specification for testing Gibrat's Law:

$$\ln(P_{it}) - \ln(P_{it-1}) = \Delta \ln P_{it} = c_i + \Theta \ln(P_{it-1}) + \sum \beta_{ij} \Delta \ln(P_{it-j}) + u_{it} \quad [3]$$

where P_{it} is the population size of city i at time t , and c , β and Θ and coefficients estimated by regression and u_{it} is a random error term. The summation in [3] represents past growth rates, implying that the current one depends on them, which is not presumed in [2]. Clark and Stabler (ibid.) test the null hypothesis that $\Theta=0$;

² Gibrat's Law is closely related to another empirical regularity, known as Zipf's (1932) Law. According to this Law, city sizes follow a Pareto distribution, with the rank (R) of a city i being proportional to its size (S): $R(i)=A*S^{-\alpha}$ or $\ln(R) = \ln(A)-\alpha*\ln(S)$, where α is a slope gradient or Pareto parameter, varying around 1 (Black and Henderson 2003; Nitsch 2005). In recent years, several urban and regional scientists (see *inter alia* Gabaix 1999; Eeckhout 2004; Batty 2006; Rossi-Hansberg and Wright, 2007; Córdoba, 2008) investigated the relationship between Gibrat's and Zipf's Laws, with Gabaix (1999) going as far as to suggest that Zipf's law is a 'natural' outcome of Gibrat's Law. The examination of the relationship between the two Laws is, however, beyond the scope of the present study, which focuses *solely* on the question whether the growth of cities is independent of (or proportionate to) their sizes and whether a particular urban setting may 'disrupt' this relationship.

otherwise the growth of the i^{th} city *does* depend upon its size (*ibid.*). As they concluded, 'Gibrat's law *cannot* be rejected for the seven largest Canadian cities.'

With a similar specification (which however does not assume dependence on past growth rates), Black and Henderson (2003) investigated the relation between the population sizes of 194 major metropolitan areas in the continental U.S., using population census data for 10-year periods between 1900 and 1990. In all regression runs, the coefficients for the population size (\ln) variable were found to be significant and negative ($P < 0.01$), which led the authors to reject Gibrat's Law. The Law was also rejected for sub-samples of 'population-gaining' and 'population-losing' cities. In Black and Henderson's analysis, the 'population' variable also emerged as highly statistically significant ($P < 0.01$) upon controlling for confounders, such as coastal location, the average annual number of days during which indoor heating is required (\ln), annual precipitation (\ln), and the area's market potential. 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. However, the results of this study should be treated with caution, as the analysis only covered a small number of metropolitan areas (termed by the authors 'the sample of winners'), leaving out their less successful counterparts.

In a summary of studies, carried out using the *Geopolis* database, which contains information on ca. 26,000 towns and cities worldwide with population of 10,000+ residents for ten-year periods between 1950 and 1990, Pumain and Moriconi-Ebrard (1997) suggested that, over past decades, major metropolitan areas across the globe have grown 'systematically more rapidly' than the rest of their urban systems, thus 'invalidating Gibrat's urban growth model.'

However, in another empirical study, Eeckhout (2004) reached a different conclusion. In his analysis, based on 1990 and 2000 Census data and covering a similar number of populated places in the U.S.A. (some 25,000 localities, including 135 largest cities of the county), the growth of cities was found 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, may be due to his use of the entire population of American cities, whereas most previous studies used 'truncated' distributions, only covering the upper tail of major cities (*ibid.*).

In a separate study, Ioannides and Overman (2003) applied the technique of "non-parametric stochastic kernel smoothing" to U.S. data for 1900-1990 and concluded that growth rates are essentially independent of city size, even though their variances vary across the city size distribution. Rose (2005) reached a similar conclusion that the growth rate of a city's population is uncorrelated with its size, demonstrating that the relationship holds for both the population distribution of cities and that of countries. (For a more formal examination of Gibrat's Law, see Appendix 1).

Research Hypotheses

One possibility is thus that the locality's growth rate is indeed *independent of its population size*, so that any differences in population growth rates can be attributed either to the accumulation of random growth 'shocks' or to differences in attributes, that may be favorable (i.e., conducive to sustained growth) or unfavorable (i.e., discouraging such growth) (see e.g., Black and Henderson 2003).

Now assume that the above location attributes cluster across geographic space so that some areas are more 'favorable' than others. If there is indeed a link between the location of cities and their growth rates, as we hypothesize, then, the growth rates should huddle across geographic space in similar manner, with localities in some places exhibiting higher rates of growth than in others. If this hypothesis is correct (i.e., localities of the same size grow at different rates in different locations), then the population size variable should emerge as statistically significant upon controlling for the location confounders. In addition, different size-growth trends should also emerge, if the settlement list is disaggregated into subsets of 'favorable' and 'unfavorable' locations. We shall examine these possibilities in the following sections, using data for the European settlement system. Our approach is similar in spirit to that of Clark and Stabler (1991) (see [3]), but instead of examining the relationship of growth and size, while adjusting for the effect of past growth rates, we adjust for *location attributes which may affect growth*.

Research methods and data sources

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). The database covers localities spread over 40 European countries, ranging from 2,000 to 7,000,000 residents.³

The selection of observation units (either municipalities or built contiguities or metropolitan areas) is a critical consideration. While in some studies of Gibrat's Law for cities, individual localities were used (see *inter alia* Eeckhout, 2004; Ioannides and Overman, 2003), other studies investigated the applicability of this law to metropolitan areas, formed by several neighboring municipalities (see, for instance, Pumain and Moriconi-Ebrard, 1997; Black and Henderson, 2003). Each of these territorial units (that is, contiguous urban areas and individual localities) has its own advantages and disadvantages.

Thus, selecting metropolitan areas as the units of analysis may give the researcher considerable conceptual advantage because a metropolitan area is generally assumed to function as a whole and may thus be considered as an economically integrated unit, formed by individual municipalities, connected by commuting flows. However, such large aggregate units, spreading in some cases up to 100-150 km from the central city (see *inter alia* Bode, 2008), are likely to be extremely heterogeneous, due to the diversity of their population as well as to the uneven proximity of their subdivisions to location landmarks, such as major cities and other employment *loci*, thoroughfare roads, etc.

Thus, for instance, the *Ruhr* urban area in Northern Germany spreads over 4500 km² and hosts more than 80 individual municipalities with a total population of 7.3 million residents. Some of these municipalities have established populations, characterized by low fertility rates and relatively low migration turnover, while other localities are populated by large numbers of foreign immigrants and characterized by large families, elevated natural growth rates and high in-migration.

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.

The *Randstad* ‘urban crescent’ in the Netherlands, which consists of more than 50 individual municipalities and hosts about 7.5 million inhabitants (nearly half of the country's total population), is another heterogeneous urban area. Clearly, upon aggregating dozens of diverse geographic units into larger urban contiguities and subsequent averaging their population growth rates into a single estimate for the urban area as a whole, a substantial loss of information due to data aggregation may be expected (Salway and Wakefield, 2005).

On the other hand, it may be argued that there is no discernable logic in investigating growth patterns of more or less arbitrarily defined administrative units, such as municipalities, with any potential association between population size and population growth becoming a statistical artifact with no underlying conceptual basis. However, this assumption may be only partially correct, as *development disparities between local administrative units* may have a profound effect on their attractiveness and population growth rates in general. Thus, a local authority with a sound locally-generated income, attributed to its location advantages and attractiveness to newcomers and investors, may provide better services and facilities, which would further increase its attractiveness to migrants, residents or businesses. Concurrently, a poorly developed municipality, with no stable economic base, may face severe budget limitations and might have to downgrade its services, such as education, public order and street cleaning, which is likely to reduce even further its attractiveness to potential newcomers (Vaturi *et al.*, 2004).

In view of the above considerations, both LAUs and urban areas may have conceptual advantages as spatial units of the analysis. Therefore, in the present study, we will use separately population growth data for both local administrative units (i.e., municipalities) and urban areas, formed by territorial contiguities of built areas.

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) was calculated in the ArcGIS9.xTM 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.

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™ 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).

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 (G), as commonly done in Gibrat's Law studies:

$$G_{t,t-1} = \ln(P_t/P_{t-1}) = \ln(P_t) - \ln(P_{t-1}), \quad [4]$$

where P_t is population size of locality at time t ; ,the end of the study period P_{t-1} , is population size of locality at *the beginning of the study period* and G is growth rate.

For most countries covered by this 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.

In the analysis, the population growth rates were measured in two ways: first, as absolute rate of population growth, and second, as country-standardized population growth rate, i.e., population growth rate in a locality minus the growth rate of the country as a whole. [The latter transformation was required to take into account country differences in population growth rates, most prominent between the countries of northern and southern Europe. However, that the results of the analyses performed using these two types of growth rates turned out to be essentially similar, and the results for the country-adjusted growth rates are not reported in the following discussion, for brevity's sake, and can be obtained from the authors upon request].

Gibrat's Law may be interpreted as the convergence, in the long run, to a lognormal city size distribution. However, previous studies of this law investigated its applicability to urban localities, using relatively short-term data, mainly for past decades (see *inter alia* Clark and Stabler, 1991; Pumain and Moriconi-Ebrard, 1997; Eeckhout 2004). 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.

Explanatory variables

The following factors served as explanatory variables: population size of localities (\ln); population density (per km^2); distance to the sea shore (km); distance to the closest major city (km), and the interaction term between a place's latitude (decimal degrees) and its elevation above sea level (meters). [In the absence of more specific climatic data, the latter variable served as proxy for climatic harshness].

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, as briefly demonstrated below.

Population density, calculated as the total population of places located within a given commuting range from a specific locality (75km or ~1dd),⁴ is, in fact, a proxy for other development parameters, including quality of life and local productivity. Thus, according to Rappaport (2006), individuals are willing to endure severe crowding and high housing costs, in order to enjoy better commercial services and higher wages. In this sense, varying local population density may be perceived as the primary mechanism whereby local wages and house prices adjust to equate utility and profits across localities.

The *population size* of localities and urban areas should, expectedly, affect their attractiveness and growth rates, since frequently they have to reach a given threshold, to ensure sufficient employment diversity and adequate services (Alonso, 1971; Portnov and Erell, 2001).

Seashore proximity may also facilitate regional and international trade, allowing urban localities to grow in a more sustained way, improving their overall economic performance (Fujita and Mori, 1997). Seashore proximity may be especially important in countries lacking a developed inland transportation network (Gallup et al, 1999).

Large *distances to major population centers*, which tend to be the major markets and sources of employment, often imply economic weakness and limited job opportunities (Ades and Glaeser, 1995; Fujita and Mori, 1997). Thus, remote localities tend to grow slowly, being relatively unattractive to migrants and investors (Duranton, 1999).⁵

The *harsh climate* of some geographic areas places limitations on interurban exchanges, as well as on human comfort and access to urban amenities. Moreover, towns located on high elevations in northern latitudes, are often hindered in their access to national *loci* of employment and cultural life (Cheshire and Magrini, 2006).

In addition, to take into account differences between countries of 'Old Europe' vs. 'New Europe,' that is, Western Europe vs. the former Soviet bloc, reflecting differences in 'initial' economic systems (e.g., long-term effects of migration restrictions imposed on major population centers in the latter countries during the Soviet era), and subsequent specialization, etc., we included the "west-east" variable in the analysis. The variable takes on value 1 if a town or city is located in Western Europe and 0 – otherwise.

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, individual countries were represented in the 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 From the outset of the analysis, we set the 75 km range (ca. one decimal degree (dd)) as commuting threshold. It corresponds to the findings of previous studies of commuting patterns on the continent (see *inter alia* Schwanen, 2002; Karlsson and Olsson, 2006).

⁵ Whether a city may be labeled 'major' depends on the function it performs, and may thus vary by country, depending on land area, population size etc. In the analysis, we decided that 500,000 residents would be our population threshold for the 'major city' group.

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.

Finally, 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 a recent study of Portnov and Schwartz (2009b), there are several possible approaches to calculating the value of the 'location package' (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.

In the present study, we use the first and simplest approach to calculating the LP of a locality, as the overall 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.⁶

⁶ The relative importance of location attributes may also have an important *temporal* dimension: Whereas in the initial stages of economic development, connectivity, and proximity to basic resources are likely to dominate location decision-making, as an economy develops, new location-

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). In line with our initial assumption, we expected that among localities with a favorable 'location package' (i.e., the first location group), *the importance of population size would be larger than for localities or urban areas for which this factor may be the only advantage (i.e., in 'least favorable' locations).*

Statistical analysis

The analysis took place in several steps. First, we looked at the population growth rates of the whole set of localities under study to confirm whether *growth rates* show a) independency of size, using scatter-plots, and OLS regressions. In the next phase, we investigated how population size affects growth rates, by controlling for several location factors, using multiple regression analysis (MRA). As mentioned, the analysis was performed separately for two levels of geographic resolution: first, individual localities (or LAUs), and, then, for urban contiguities, formed by adjacent LAUs (see Fig. 1 and the sub-section on data sources for more detail).⁷ Lastly, we performed separate analyses for location subgroups of localities (favorable locations, unfavorable ones and the rest of the sample), to determine whether the size-growth relationship differ across them, in line with our initial research hypothesis.

<<< Figure 1 about here >>>

Results

Preliminary analysis

Table 1 reports the results of the analysis of variance (ANOVA) test of differences in the growth rates of localities of different population sizes. To run this test, the whole set of localities was split into *nine* groups based on their population sizes (ln) rounded to integer values. Although the whole sample of individual localities shows some growth rate variability, the Levene statistic⁸ *fails* to reject the null hypothesis that group variances are equal at the 0.01 probability level (see Table 1), thus pointing out that the variance of population growth rates in the entire sample of individual localities is, generally, consistent with Gibrat's Law. The Law is however rejected if urban areas are used as observation units (see the last two columns in Table 1), with average population growth rates in larger urban areas being significantly higher than in smaller ones (F=4.036; P<0.01), along with a significant difference in variances (Levene=3.963; P<0.01).

The scatter plots of growth vs. population size (Figs. 2 & 4A) and simple two-variable regression models reported in Table 2 generally lead to the same conclusion:

related elements may gain importance. They may include climatic differentials, environmental attractiveness, and proximity to unique urban functions (such as culture, education and services), which few urban centers may provide (Duranton, 1999; Cheshire and Magrini, 2006; Portnov and Schwartz, 2008). However, in this study, which covers a relatively short period (1990-2000) the above effects are assumed to be fixed.

⁷ For visualization purposes the scatter plots report growth in percentages while the statistical analysis follows the theoretical models (see formulae [2],[3]&[4]) and uses growth in logarithmic units.

⁸ The Levene statistic assesses whether variances across samples are equal. The statistic is reported as a probability value with values below a certain critical value (e.g., p<0.05) indicating that there are substantial variance differences.

for the whole sample of localities, *there is no significant relationship between locality sizes and growth rates* (Population (ln): $t=-0.37$; $P>0.9$; see Table 2), with the trend line being nearly parallel to the population axis (see Fig. 1). In contrast, for urban areas, population size and growth rate are significantly, although weakly, positively associated ($t=5.587$; $P<0.01$; $R^2=0.014$).

<<< Tables 1-2 and Figure 2 about here >>>

Multivariate analysis

The list of variables and the resulting models for the whole sample of individual localities are reported in Table 3, while models for urban areas are reported separately in Table 7. In Model 1 fixed effects (i.e., countries' dummies) are not included, while Model 2 includes them. Model 3 preserves the same setting but omits the population size variable. Model 4 is calculated for a sub-sample of localities with moderate growth rates (i.e., $\pm 2\%$), which make up 91% of the entire sample of settlements and comply fully with the normality assumption.⁹ Model 5 includes only localities with 10,000+ residents, which represent the middle part and the upper tail of the population size distribution and growth rates can be deemed most accurate in terms of volatility and measurement errors. Lastly, Models 6 & 7 (Table 3) are calculated using spatial lag (SL) regressions (based on three nearest neighbors' neighborhood matrix), better fitted for taking into account potential auto-collinearity of population growth rates in nearby localities.

<<Table 3 about here >>

As Table 3 shows, the population size variable emerges as highly statistically significant in all of the models ($|t|>2.7$; $P<0.01$), steadily exhibiting a negative sign, in both OLS and SL models. This indicates that, other things being equal, *bigger cities grow slower than smaller ones*, thus contradicting Gibrat's Law. Possible reasons for this trend are more or less clear: emergence of dis-agglomeration economies in big cities, such as high housings costs, traffic congestions, air pollution, etc.

Omission of the population size variable (see Model 3) reduces regression fit and generality (R^2 -adjusted=0.347; $F=57.428$ in Model 2 vs. R^2 -adjusted=0.355; $F=58,152$ in Model 2; Table 3), thus implying that the population size variable *does* belong in model. Characteristically, most location-related variables (proximity to the seashore, major cities, latitude, density, etc.) appear highly significant ($P<0.01$), thus confirming our initial hypothesis that for an urban place, location attributes are important growth determinants.

Models 11&15, estimated for all urban areas and reported in Table 7, essentially lead to the same conclusion, with the regression coefficients for the population size variable being significant ($t>2.4$; $P<0.01$), albeit positive, and not negative, as in the models estimated for individual localities (see Table 3). Moreover, the population density variable (estimated, as mentioned, as the average density of population

⁹ Our analysis of the Q-Q (quantile-quantile) plots of the regression residuals from the models obtained for the entire sample of localities indicated that the distribution of the residuals appeared to differ from normality, which was confirmed by the Kolmogorov-Smirnov (KS) normality test ($KS=9.0$; $P<0.001$). This deviation from normality may be attributed, inter alia, to higher volatility of growth rates in smaller localities. Characteristically, the distribution was found to be fairly normal for the $\pm 2\%$ growth range interval ($KS=2.3$; $P>0.05$). The latter is an important consideration for regression analysis, as the above deviation from normality exhibited by the whole sample could bias regression estimates. To verify this possibility, a separate model for the localities with $\pm 2\%$ growth range (see Model 4) was run, and found to be qualitatively similar to 'original' model (Model 2).

residing within commuting distance from the localities forming urban areas) emerged as positive and highly statistically significant in the models estimated for localities (see Table 3) as well as in the models estimated for the whole set of urban areas ($t > 3.2$; $P < 0.01$; see Models 11 & 15; Table 7). The importance of these results will be discussed in the concluding section of the paper.

Location sub-groups

To verify the initial research hypothesis that the growth-size relationship should expectedly depend on a populated place's location setting (see the sub-section on explanatory variables), we split 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 (see Table 4 and Appendix 2).

<<Table 4 about here >>

As Table 4 shows, the Levene statistic rejects the null hypothesis that group variances are equal for either individual localities or urban areas ($P < 0.01$). Furthermore, the Kruskal–Wallis test of the significance of differences in mean rates of growth confirms that they are statistically significant ($\chi^2 > 12.0$, $P < 0.01$ for localities and $\chi^2 > 7.8$, $P < 0.05$, for urban areas; see Table 4).¹⁰ This further supports our hypothesis that growth rates *are not* distributed randomly, as expected under Gibrat's Law, but vary with the location characteristics of urban areas and localities.

Figure 3 features population size-growth relationship in each location subgroup covered by the study. Although regression fit (measured by R^2) in any of the groups is *not* especially high, the F -test confirms that the slopes of the fit lines for 'unfavorable' localities and 'favorable' localities appear to be *significantly different from zero* ($P < 0.001$), being *slightly* positive in the former group (see Fig. 3A) and *negative* – in the latter (Fig. 3B). Concurrently, the rest of the sample (Fig. 3C) shows the 'no-trend' relationship, like that detected for the entire sample of settlements (see Fig. 2). 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 may cancel each other out if the whole set of localities is considered.

The Chow test of regression residuals, reported in Table 5, also confirms that the slopes of 'size-growth' regression lines for 'favorably' and 'unfavorably' located towns are significantly different (Chow test=9.644; $P < 0.001$; Table 5), i.e., a negative slope in the 'unfavorable locations' group and a positive slope in the 'favorable locations' group.

<<<Figure 3 and Table 5 about here >>>

The positive association observed in unfavorable locations between rates of population growth and city sizes, 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

¹⁰ Given the heterogeneity of the variances (Levene Statistic > 8 ; $P < 0.01$) and a lack of symmetry in the growth rates distribution within each location group, the non-parametric Kruskal-Wallis test was used.

for unfavorably located settlements, 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.? To answer this question the regression analysis was rerun for each location group (1-3) separately. The results are reported in Table 6 (Models 8-10).

<<Table 6 about here >>

As Table 3 shows, 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 (Model 8) and negative elsewhere ($P < 0.01$; Models 9-10). 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 other location advantages, the importance of size for the development of individual towns may be less profound.

Characteristically, when *urban areas* are used as observation units, instead of individual localities (see Table 7), the association between population size and growth emerges as *positive*, albeit statistically weak, indicating that across all location groups, larger urban areas may tend to grow faster, all other things being equal, than their smaller counterparts. A difference in the strength of association between population growth rates and population density observed in the 'urban areas' models is also noteworthy: This association appears to be much stronger in the 'LP=2,3' group of urban areas than elsewhere ($t = 6.167$; $P < 0.01$; Model 18 vs. e.g., $t = 0.01$; $P > 0.5$, Model 17 for LP=4,5 (that is, 'favorably located') urban areas - see Table 7).

This difference may be explained as follows: In places with relatively few location advantages, dense urban network may increase the access to job opportunities available within daily commuting range, thus boosting the potential for further population growth (Portnov and Erell, 2001). In contrast, in places with numerous location advantages, and particularly around major cities, further increases in urban density may be less beneficial, or even detrimental, for individual localities and urban areas as a whole, due to overcrowding and increasing diseconomies of agglomeration, which any further increase in density are likely to entail.

<<Table 7 about here >>

Conclusion

It is not uncommon in urban and regional science that, upon data desegregation, several, often opposite, trends emerge. Thus, a zero migration balance in a locality may 'hide' an outflow of wealthy residents to suburban areas and an inflow of poor families to central cities, with the two opposite currents being of similar strength (see *inter alia* Ishikawa and Montanari 2004).

Similar explanations may apply to the emergence of the proportionate effect of growth-size relationship expected under Gibrat's Law at the aggregate settlement

system level and the 'disappearance' of such growth-size independence upon controlling for location features.

Even though the inter-factor relationship may *not* be altered completely when accounting for locational and other exogenous attributes, the strength of this relationship may vary considerably. For instance, in his study of U.S. cities, Glaeser (2005) noted that higher skills have a much greater impact on growth rates in cold-weather metropolitan areas than in warmer ones.

In this study, we considered a possibility that growth rates *may* depend on size, while the direction of relationship differ, e.g., a monotonic increase in unfavorable *loci* vs. a monotonic decrease elsewhere. Over the entire settlement system, these opposite trends will cancel out, with the 'no-trend' relationship, expected under Gibrat's Law, emerging.

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.

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 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.

According to our findings, when *individual localities* are considered, 'proportionate' growth (expected under Gibrat's Law) does emerge at the aggregate (system-wide) level, but 'dissipates' when the settlement system is disaggregated into two urban sub-systems, formed by well-positioned localities and poorly positioned ones. Concurrently, for *urban areas* a strong positive association between population sizes and growth emerges both before and after controlling for location attributes, albeit this relationship was not strong once we looked at the three location sub-groups separately. A possible explanation may be that at the level of integrated urban areas, population density may become a better measure of population concentration and proximity benefits, than population size of the integrated urban unit *per se*. Indeed, in our analysis, we saw that population density emerged as highly statistically significant and positively associated with population growth rates in most location sub-groups.

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.

The difference in the results obtained at different levels of geographic resolution (that is, individual localities vs. integrated urban areas) is, in fact, unsurprising. In our view, it reflects 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.

In our view, the positive association between population size and growth detected by the present study for urban areas and the negative association detected at the level of individual localities may both be attributed to the *effect of urban clustering on the population growth of localities*. While clustering of several places within urban areas may be beneficial for the *urban areas as a whole*, due to

agglomeration economies (see inter alia Ades and Glaeser, 1995), such concentration of individual settlements near each other may be *less beneficial for each of them separately*, due to inter-town competition for investors and migrants (Portnov and Erell, 2001; Portnov and Schwartz, 2009a).

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 (non location-related) 'seam-lines', such as established vs. transitional economies, local towns vs. metropolitan areas, etc. Such possibilities may deserve investigation in future studies.

References

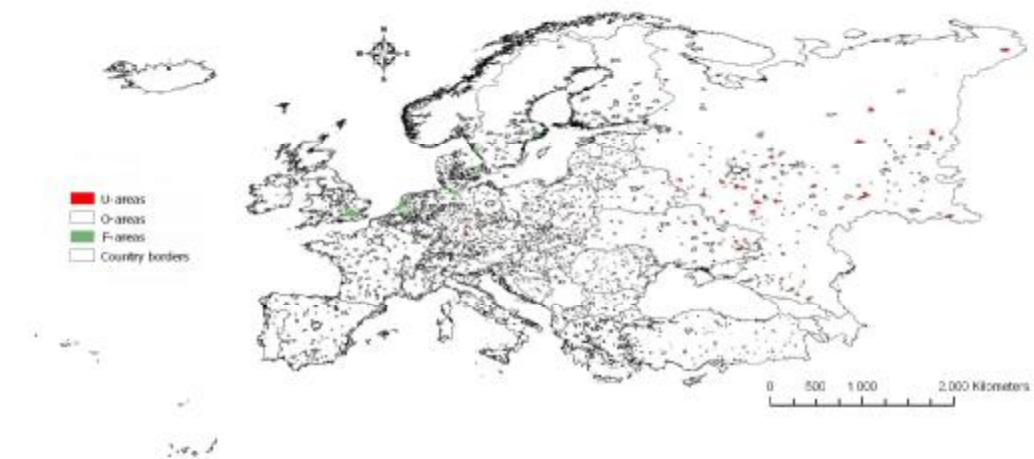
- Ades A. F., and Glaeser E.L. 1995. 'Trade and Circuses: Explaining Urban Giants,' *The Quarterly Journal of Economics*, 110(1), 195-227.
- Alonso, W. 1971 (1977 reprint), 'The Economics of Urban Size,' in J. Friedman and W. Alonso (eds.), *Regional Policy: Theory and Applications*, Cambridge, Massachusetts: The MIT Press, pp. 334-450.
- Anderson G. and Ge Y. (2005) The size distribution of Chinese cities, *Regional Science and Urban Economics*, 35(6): 756-776.
- Auerbach, F. (1913) Das Gesetz der Bevolkerungskonzentration, *Petermanns Mitteilungen*, 59: 74-76 (cited in Nitsch V. (2005) Zipf zipped, *Journal of Urban Economics*, 57: 86-100).
- Batty M. (2006) Hierarchy in cities and city systems, In: D. Pumain (Ed.), *Hierarchy in Natural and Social Sciences*, Heidelberg et al: Springer, pp. 143-168.
- Black D. and Henderson V. (2003) Urban evolution in the USA, *Journal of Economic Geography*, 3: 343-372.
- Bode E. (2008). "Delineating Metropolitan Areas Using Land Prices," *Journal of Regional Science*, 48(1): 131-163.
- 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.
- Córdoba J.C. (2008) On the distribution of city sizes, *Journal of Urban Economics*, 63: 177-197.
- 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.
- Eurostat (2008) Urban Audit Database - Inforegio (<http://www.urbanaudit.org/DataAccessed.aspx>; accessed in June 2008).
- Eurostat (2009). *Local Administrative Units* (http://ec.europa.eu/eurostat/ramon/nuts/lau_en.html; accessed in Dec. 2009).
- Fujita, M. and Tomoya M. 1997. 'Structural Stability and Evolution of Urban Systems,' *Regional Science and Urban Economics*, 27, 399-442.
- Gabaix, X. (1999) Zipf's law for cities: an explanation, *Quarterly Journal of Economics*, 114: 739-767.
- Gallup JL, 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*, Paris: Sirey (Cited in: Kalecki M. (1945) On the Gibrat Distribution, *Econometrica*, 13(2): 161-170).
- Glaeser E.L (2005) Smart Growth: Education, *Skilled Work & the Future of Cold Weather Cities*. Policy Brief. John F. Kennedy School of Government, Harvard University.
- Gotway C. and Young L.J. (2002). Combining incompatible spatial data, *Journal of the American Statistical Association*, 97 (48): 632 – 647.
- Henry, M.S., Barkley, D.L. and Bao S. (1997). The hinterland's stake in metropolitan growth: evidence from selected southern regions, *Journal of Regional Science*, 37(3): 479-501.
- Ioannides Y.M. and Overman H.G. (2003) Zipf's Law for cities: an empirical examination, *Regional Science and Urban Economics*, 33: 127-137.
- Ishikawa Y. and Montanari A. eds. (2004) *The New Geography of Human Mobility: Inequality Trends*, Rome: Societa Geographica Italiana.
- Kalecki M. (1945) On the Gibrat Distribution, *Econometrica*, 13(2): 161-170.
- Karlsson, C. and Olsson M. 2006. 'The Identification of Functional Regions: Theory, Methods, and Applications,' *Annals of Regional Science*, 40(1), 1-18.
- 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.
- Nitsch V. (2005) Zipf zipped, *Journal of Urban Economics*, 57: 86-100.
- Openshaw S. (1984) The modifiable areal unit problem, In: *Concepts and Techniques in Modern Geography*. Monograph Series. London: Geo Books; 38, 41 pp.

- Partridge M., Bollman R.D., Olfert M.R., and Alasia A. (2007) Riding the wave of urban growth in the countryside: spread, backwash, or stagnation?' *Land Economics*, 83(2): 128-152.
- Polese M. and R. Shearmur (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 (in press).
- 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.
- Reed W. J. (2002) On the rank-size distribution for human settlements, *Journal of Regional Science*, 42(1):1-17.
- Rose A.K. (2005) Cities and countries, *NBER Working Paper Series*, Working Paper 11762, National Bureau of Economic Research, Cambridge, MA.
- Rossi-Hansberg E. and Wright M.L.J. (2007) Urban structure and growth, *Review of Economic Studies*, 74: 597-624.
- Salway R. and Wakefield J. (2005) Sources of bias in ecological studies of non-rare events. *Environmental and Ecological Statistics*, 12: 321– 347.
- Schwanen T. 2002. 'Urban Form and Commuting Behaviour: A Cross-European Perspective,' *Tijdschrift voor Economische en Sociale Geografie*, 93(3), 336–343.
- Strutzer A and Frey BS (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.
- Vaturi A., Portnov B.A. & Schwartz M. (2004) Internal migration and fiscal capacity of local authorities: a case study of the Greater Tel Aviv Metropolitan Area, 1985-1997, *Space & Polity*, 8(1): 61-85.
- Zipf G.K. (1932) *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, Mass.



A



B

Figure 1: Location subsets covered by the analysis

A – Individual localities; B – Urban areas

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

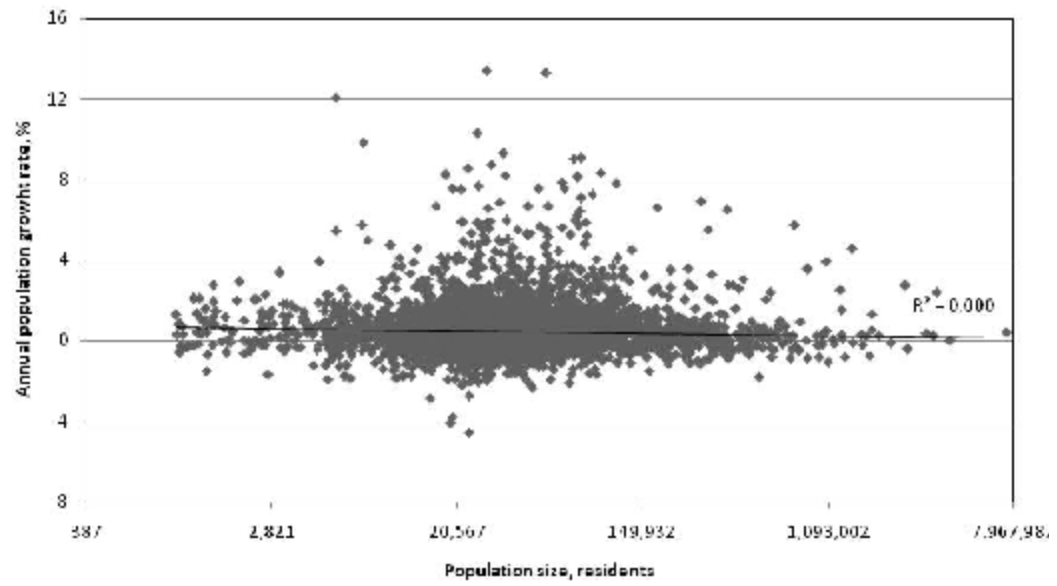
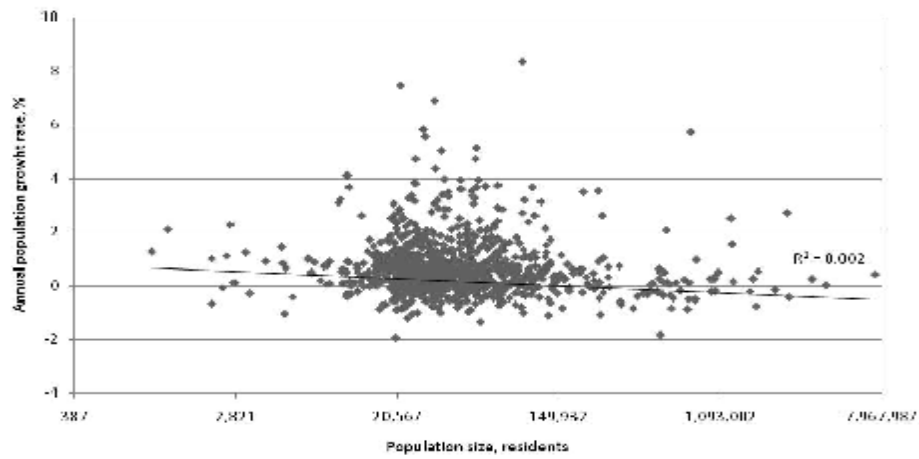
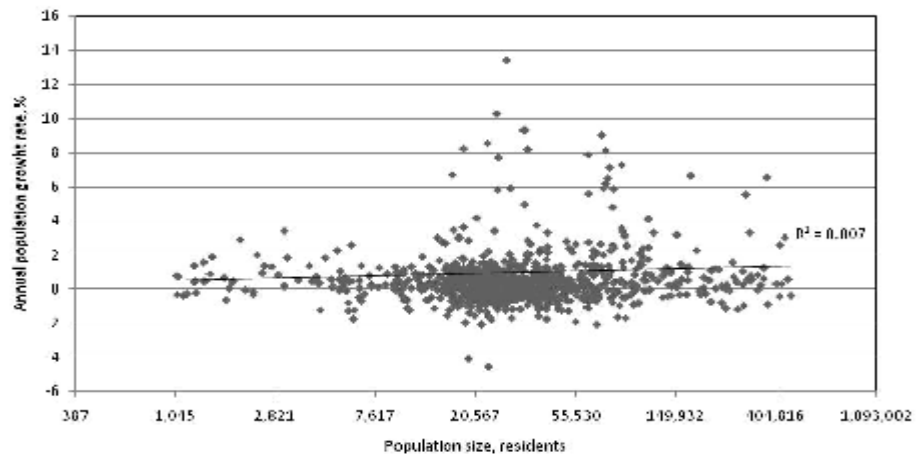
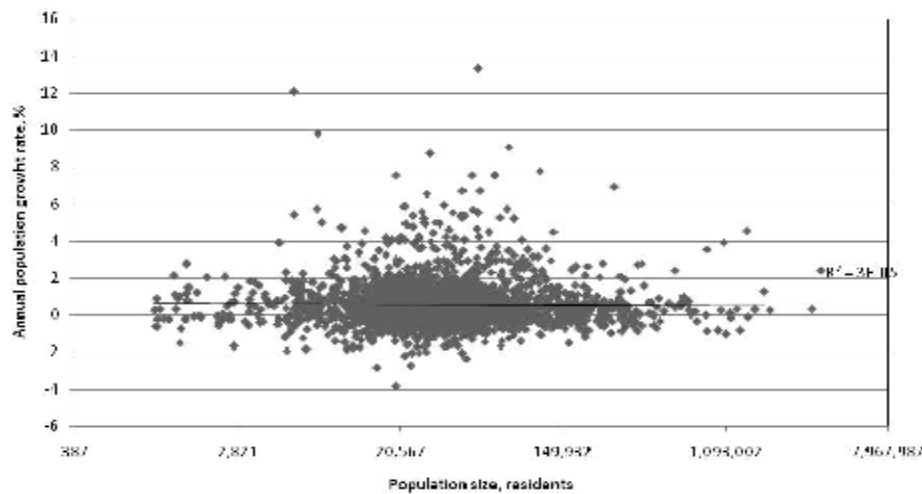


Figure 2: Population growth in Europe in 1990-2000 against population size (ln) of localities



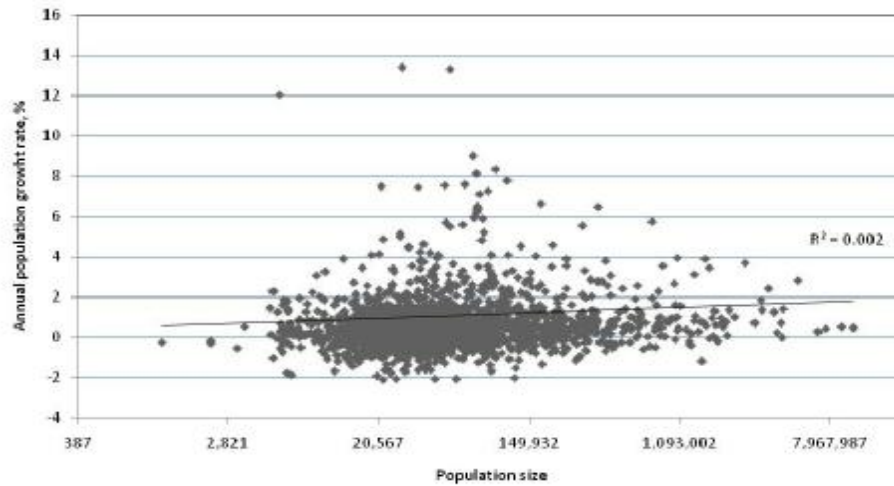
B



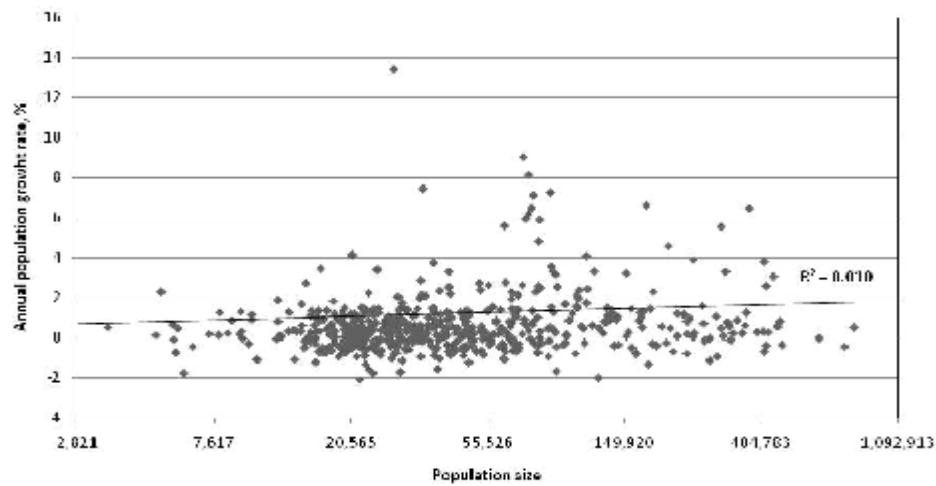
C

Figure 3: Annualized population growth rates vs. population size (ln) of localities stratified by location subsets

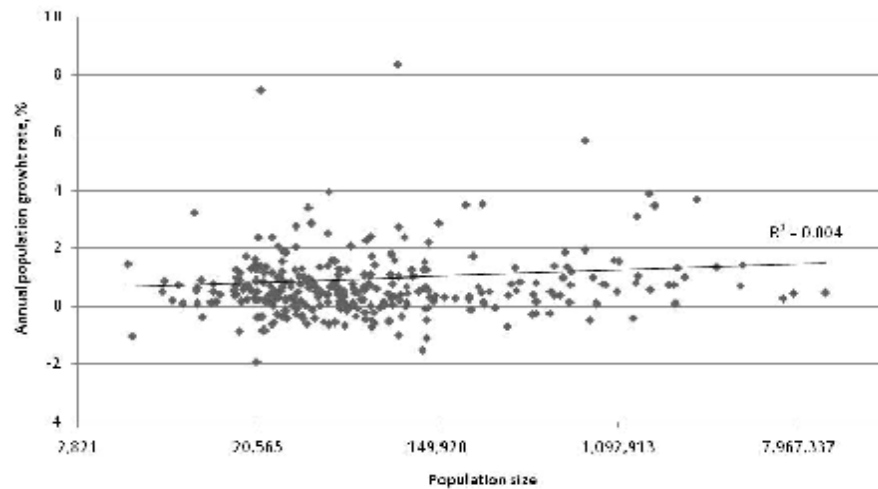
A – U[nfavorable]-locations (LP=0.1); B – F[avorable]-locations (LP=4,5); C – O[ther]- locations (LP=2,3)



A



B



C

Figure 4: Annualized population growth rates vs. population size (ln) of urban areas stratified by location subsets

A – All urban areas; B – U[nfavorable]-locations (LP=0.1); C – F[avorable]- locations (LP=4,5)

Table 1:

Analysis of variance (ANOVA) test of differences in the growth rates (ln) of localities and urban areas of different population sizes

Population size (ln), residents	Population range	Localities		Urban areas	
		Mean ^a	SD ^b	Mean ^a	SD ^b
≤7	<1,097	0.005	0.009	-	-
8	1,098-8,103	0.007	0.010	0.001	0.007
9	8,104-22,026	0.006	0.013	0.005	0.013
10	22,027-59,874	0.006	0.012	0.004	0.010
11	59,875-162,755	0.007	0.013	0.007	0.014
12	162,756-442,413	0.006	0.010	0.007	0.011
13	442,414-1,202,604	0.004	0.012	0.007	0.011
14	1,202,605-3,269,017	0.006	0.015	0.010	0.013
≥15	>3,269,017	0.007	0.013	0.011	0.010
F		0.770		4.036**	
Levene Statistic		2.160*		3.963**	

^a Absolute (unstandardized) rate of growth; ^b standard deviation

* indicates a 0.05 significance level; ** indicates a 0.01 significance level

Table 2:

Two-variable regressions (The whole sample of localities; Dependent variable – population growth rate (ln)^a)

Variable	Localities	Urban areas
(Constant)	0.007 (3.673**)	-0.008 (-3.117**)
Population (ln)	-6.5E-05 (-0.370)	1.25E-03 (5.587**)
No of observations	2702	2189
R Square	0.000	0.014
F	0.137	31.210**

^a Absolute (unstandardized) rate of growth; *t* values are in parentheses; ** indicates a two-tailed 0.01 significance level

Table 3: Factors affecting the annual rates of population growth of urban localities (Models – OLS and SL (moving averages); all localities;
Dependent variable - Absolute rates of population growth (LN))

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a
(Constant)	0.041 (15.792**)	0.050 (11.895**)	0.039 (9.717**)	0.033 (11.729**)	0.050 (9.599**)	0.045 (13.715**)	0.050 (11.931**)
Population size (ln)	-4.79E-04 (-2.716*)	-1.22E-03 (-7.651**)	-	-7.01E-04 (-6.818**)	-1.38E-03 (-7.533**)	-9.84E-04 (-6.793**)	-1.23E-03 (-7.708**)
Distance to sea shore	4.34E-04 (1.808)	3.03E-04 (1.496)	2.48E-04 (1.216)	1.42E-04 (1.065)	2.91E-04 (1.398)	-2.71E-05 (-0.083)	3.01E-04 (1.483)
Distance to major city	-1.18E-03 (-4.325**)	1.31E-04 (0.515)	1.73E-04 (0.674)	-3.83E-04 (-2.267*)	1.26E-04 (0.478)	-1.16E-03 (-2.737**)	1.33E-04 (0.520)
Climatic harshness	-6.30E-04 (-2.918**)	-6.55E-04 (-3.574**)	-5.06E-04 (-2.759**)	-4.29E-04 (-3.671**)	-7.25E-04 (-3.846**)	-5.05E-04 (-2.218*)	-6.56E-04 (-3.583**)
Latitude	-5.44E-04 (-18.595**)	-6.72E-04 (-10.242**)	-6.53E-04 (-9.901**)	-4.24E-04 (-9.970**)	-6.92E-04 (-10.340**)	-5.49E-04 (-10.776**)	-6.71E-04 (-10.224**)
Density	-5.25E-04 (-3.152**)	1.31E-03 (6.459**)	1.15E-03 (5.667**)	7.39E-04 (5.616**)	1.31E-03 (6.183**)	-4.18E-06 (-0.018)	1.31E-03 (6.473**)
West-East	2.26E-04 0.526	- -	- -	- -	- -	-3.33E-04 (-0.474)	- -
Country dummies (39)	no	yes	yes	yes	yes	no	yes
No of cases	4667	4667	4667	4261	4350	4667	4667
R ²	0.072	0.362	0.353	0.247	0.377	-	-
Adjusted R ²	0.071	0.355	0.347	0.239	0.370	-	-
SEE ^b	0.012	0.010	0.010	0.006	0.010	0.010	0.010
F	51.758**	58.152**	57.428**	30.802**	57.800**	-	-

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Rho ^c	-	-	-	-	-	0.178	0.008
Log-likelihood	-	-	-	-	-	1785.939	1941.872

* Indicates a 0.05 significance level; ** Indicates a 0.01 significance level;

^a Regression coefficients and *t*-statistics (in parentheses); ^b standard error of the estimate; ^c autoregression coefficient.

Model 1: All localities; fixed effects are not included; method – Ordinary Least Squares (OLS) regression;

Model 2: All localities; fixed effects are added; method – OLS;

Model 3: All localities; population size variable is omitted; method – OLS;

Model 4: Localities with -2% to 2% growth rates; method – OLS;

Model 5: Localities with population of 10,000 residents; method – OLS;

Model 6: All localities; fixed effects are not included; method – Spatial Lag (SL) regression; covariance family – moving averages (MA); neighborhood matrix – three nearest neighbors;

Model 7: All localities; fixed effects are added; method – SL/MA regression; neighborhood matrix – three nearest neighbors.

Table 4: Kruskal -Wallis Test for the significance of growth rates differences across location groups

Location group	N of localities	Localities ^a		Urban areas ^a	
		Mean	SD	Mean	SD
1	904	0.0056	0.0148	0.0061	0.0143
3	1061	0.0065	0.0102	0.0069	0.0109
2	2702	0.0061	0.0118	0.0056	0.0112
Levene Statistic		12.290**		8.441**	
Chi-Square ^b		17.969**		7.856*	

Location groups: 1-'Unfavourable localities' (LP=0,1); 2-'Favourable localities' (LP=4,5) and 3 - Rest of the sample (LP=2,3).

^a Absolute rate of growth; ^b Kruskal Wallis Test; * indicates a two-tailed 0.05 significance level; ** indicates a two-tailed 0.01 significance level.

Table 5: Chow's test of similarity of regression coefficients (model – two-variable regression; dependent variable – annualized population growth rates (ln); predictor – population size (ln))

Location group	No of cases	B ₀	t	B ₁	t	Chow test
'Unfavorable locations'	904	-7.47E-03	-1.482	1.27E-03	2.610**	
'Favorable locations'	1061	1.58E-02	4.685**	-8.70E-04	-2.767**	
Both groups	1965	4.08E-03	1.390	1.92E-04	0.688	9.644**

**Indicates a two-tailed 0.01 significance level.

Table 6: Factors affecting the annual rates of population growth of localities (Model – OLS; location subgroups; dependent variable - absolute rates of population growth (LN))

Variable	Model 8		Model 9		Model 10	
	b/(t) ^a	Tol. ^b	b/(t) ^a	Tol. ^b	b/(t) ^a	Tol. ^b
(Constant)	0.061 (8.589**)		0.048 (9.603**)		0.037 (11.084**)	
Population size (ln)	2.92E-05 (0.063)	0.866	-1.48E-03 (-4.729**)	0.931	-5.05E-04 (-2.209*)	0.938
Distance to sea shore	2.73E-03 (4.256**)	0.865	-3.65E-04 (-0.634)	0.630	-6.68E-05 (-0.227)	0.709
Distance to major city	1.61E-03 (2.462*)	0.779	-1.65E-03 (-2.697**)	0.841	-2.41E-03 (-6.360**)	0.746
Climatic harshness	-1.24E-03 (-2.381*)	0.840	-7.97E-04 (-1.293)	0.688	-2.63E-04 (-0.982)	0.736
Latitude	-1.16E-03 (-13.869**)	0.901	-3.77E-04 (-6.712**)	0.910	-4.13E-04 (-11.249**)	0.966
Density	-6.77E-04 (-1.487)	0.773	-4.99E-04 (-1.340)	0.729	-8.11E-04 (-3.912**)	0.678
West-East	-3.17E-03 (-2.821**)	0.628	-2.89E-03 (-2.629**)	0.863	1.42E-03 (2.528*)	0.817
Country dummies (39)	no		no		no	
No of cases	904		1061		2702	
R Square	0.221		0.084		0.061	
Adjusted R Square	0.215		0.078		0.059	
SEE ^c	0.013		0.010		0.011	
F	36.313**		13.749**		25.036**	

* Indicates a 0.05 significance level; ** Indicates a 0.01 significance level;

^a Regression coefficients and *t*-statistics (in parentheses); ^b tolerance (collinearity diagnostic); ^c standard error of the estimate.

Model 8: 'Unfavorable' locations (LP=0 and 1); absolute rates of growth (ln); Country-normalized location values;

Model 9: 'Favorable' locations (LP=4 and 5); absolute rates of growth (ln); Country-normalized location values;

Model 10: The rest of locations (LP=2 and 3); absolute rates of growth (ln); Country-normalized location values;

Table 7: Factors affecting the annual rates of population growth of urban areas (Models– OLS; location subgroups; dependent variable - absolute rates of population growth (LN))

Variable	Model 11	Model 12	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18
	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a	b/(t) ^a
(Constant)	0.025 (6.915**)	0.048 (6.041**)	0.019 (2.260*)	0.017 (3.929**)	0.033 (6.104**)	0.047 (4.626**)	0.026 (1.790)	0.028 (4.183**)
Population size (ln)	6.28E-04 (2.831**)	7.56E-04 (1.390)	2.98E-04 (0.662)	4.89E-04 (1.781)	4.94E-04 (2.761**)	4.88E-04 (1.002)	2.32E-04 (0.630)	3.98E-04 (1.772)
Distance to sea shore	7.64E-04 (2.273*)	3.34E-03 (5.116**)	-3.80E-04 (-0.371)	-4.27E-04 (-1.009)	4.23E-04 (1.636)	4.34E-03 (5.924**)	7.62E-06 (0.008)	-1.47E-04 (-0.453)
Distance to major city	2.14E-04 (0.546)	2.74E-03 (3.557**)	-2.63E-03 (-2.280*)	-6.09E-04 (-1.211)	9.91E-04 (3.195**)	2.60E-03 (3.948**)	-2.27E-03 (-2.020*)	9.27E-04 (2.250*)
Climatic harshness	-6.82E-04 (-2.154*)	-5.81E-04 (-1.036)	-1.29E-03 (-0.978)	-2.52E-04 (-0.599)	-4.02E-04 (-1.657)	-2.36E-04 (-0.475)	-1.41E-03 (-1.268)	1.02E-04 (0.311)
Latitude	-5.19E-04 (-12.617**)	-1.16E-03 (-12.976**)	-2.08E-04 (-1.854)	-3.23E-04 (-6.597**)	-6.63E-04 (-8.207**)	-1.06E-03 (-6.748**)	-1.55E-04 (-0.551)	-6.10E-04 (-5.689**)
Density	7.49E-06 (3.275**)	1.76E-05 (1.605)	2.02E-06 (0.468)	7.23E-06 (2.692*)	1.49E-05 (6.899**)	4.03E-05 (3.503**)	4.81E-08 (0.010)	1.59E-05 (6.167**)
West-East	-1.89E-03 (-3.417**)	-7.50E-03 (-5.610**)	-2.94E-03 (-1.758)	-2.59E-04 (-0.371)	-	-	-	-
Country dummies	no	no	no	no	yes	yes	yes	yes
No of cases	2189	592	305	1292	2189	592	305	1292
R ²	0.090	0.313	0.083	0.052	0.497	0.585	0.530	0.481
Adjusted R ²	0.087	0.305	0.061	0.047	0.486	0.560	0.467	0.462
SEE ^b	0.012	0.012	0.011	0.011	0.009	0.009	0.008	0.008
F	30.863**	37.979**	3.827**	10.062**	47.008**	23.083**	8.404**	23.68**

See footnote to Table 6.

- Model 11: All urban areas; absolute rates of growth (ln); country-normalized location variables; country indicators excluded;
- Model 12: 'Unfavorable' urban areas (LP=0,1); absolute rates of growth (ln); Country-normalized location values; country indicators excluded;
- Model 13: 'Favorable' urban areas (LP=4,5); absolute rates of growth (ln); Country-normalized location values; country indicators excluded;
- Model 14: 'Other' urban areas (LP=2,3); absolute rates of growth (ln); Country-normalized location values; country indicators excluded;
- Model 15: All urban areas; absolute rates of growth (ln); country-normalized location variables; country indicators included;
- Model 16: 'Unfavorable' urban areas (LP=0,1); absolute rates of growth (ln); Country-normalized location values; country indicators included;
- Model 17: 'Favorable' urban areas (LP=4,5); absolute rates of growth (ln); Country-normalized location values; country indicators included.
- Model 18: 'Other' urban areas (LP=2,3); absolute rates of growth (ln); Country-normalized location values; country indicators included.

APPENDIX 1

Statistical examination of the Gibrat law

The proportionate growth premise, underlying Gibrat's Law for cities, implies that for a given locality:

$$P_t = (1 + \epsilon_t) P_{t-1}, \quad [4]$$

where the ϵ_t are identically independently distributed random variables representing the proportionate changes between time t and $t-1$ such that $|\epsilon_t|$ is small compared with 1. Setting $\mu = E\{\ln(1 + \epsilon_t)\} \approx 1 + E\{\epsilon_t\}$, [2] follows.¹¹

Thus, strictly speaking, the proportionate growth assumption implies that growth *does not* depend on city size in *either expected value or variance*.¹² Negative ϵ_t imply decreasing population size in the time interval, while positive values imply increasing size. In addition for large t , the distribution of $\ln P_t$ is expected to be normal, as widely discussed in the literature (e.g. Gibrat 1931; Kalecki 1945; Eeckhout 2004). Frequently this log-normal specification is tested empirically (e.g., Eeckhout 2004; Anderson and Ge 2005).¹³

Log normality can however hold even where proportionate growth does not. Hence, Kalecki (1945) introduced the possibility of a negative correlation between growth and size and found that the asymptotic log-normality assumption remain valid. Thus log-normality *does not* provide sufficient evidence for Gibrat's Law.

Alternatively, proportionate growth can be examined empirically by directly modeling the growth as a function of population size and testing the resulting effect on size. If the proportionate changes ϵ_t instead of being purely random shocks actually have an expected value depending on size, then the μ of [2] can be considered as a function of P_{t-1} , the simplest specification being:

$$\mu = \alpha + \beta \ln P_{t-1}. \quad [5]$$

Whether growth depends on city size can be examined by testing the hypothesis that $\beta=0$ (Eeckhout 2004). If $\beta=0$ is considered to hold, we regard μ as a constant denoting the average log growth. In addition, one can find out whether the variance of the growth process depends on city size (Gabaix 1999) As discussed above, more complicated specifications allowing for autocorrelation or other effects are sometimes considered (see *inter alia* Clark and Stabler 1991).

¹⁰ Using computer simulations, Batty (2006) showed that even for an initial uniformly distributed population and *randomly* chosen growth rates, the size distribution of individual localities tends to show 'every sign of being log-normal.' .

¹² Gabaix (1999) takes a strict view of Gibrat's Law, requiring that the entire probability distribution of the growth process (and not only the expected value and variance) should *not* depend on city size.

¹³ As Reed (2002) argued, the asymptotic argument for log-normality fails to take into account that cities differ in the time elapsed since foundation (T). This leads to a mixture of log-normal distributions depending on the assumed distribution of T. Rossi-Hansberg and Wright (2007) and Cordoba (2008) use microeconomic theory to construct stochastic urban models which can produce proportional growth under some assumptions. This paper focuses on the relation between city growth and size, not on distributional aspects.

APPENDIX 2

Selected characteristics of location groups of settlements

Variable	Unfavorable locations (LP=0,1)*	Favorable locations (LP=4,5)*	The rest of locations (LP=2,3)*
A. Individual localities			
Population growth per 1000 residents (ln)	5.63E-03	6.50E-03	6.14E-03
Average population size, residents (ln)	10.33	10.66	10.42
Number of localities	904	1061	2702
B. Urban areas			
Population growth per 1000 residents (ln)	6.11E-03	6.87E-03	5.59E-03
Average population size, residents (ln)	10.65	11.17	10.80
Number of localities	592	305	1292