Article

Urban Analytics and
City Science

EPB: Urban Analytics and City Science
2021, Vol. 48(2) 231–247

DOI: 10.1177/2399808319861974
journals.sagepub.com/home/epb
SAGE

As different as night and day: Scaling analysis of Swedish urban areas and regional labor markets

Deborah Strumsky and Jose Lobo
Arizona State University, USA

Charlotta Mellander
Jonkoping University, Sweden

Abstract

The urban scaling framework views cities as integrated socioeconomic networks of interactions embedded in physical space. A crucial property of cities highlighted by this approach is that cities act to mix populations, a mixing both facilitated and constrained by physical infrastructure. Operationalizing a view of cities as settings for social interactions and population mixing—assembling a set of spatial units of analysis which contain the relevant social aspects of urban settlements—implies choices about the use of existing data, the assignation of data to locations, and the delineation of the boundaries of urban areas, all of which are far from trivial research decisions. Metropolitan areas have become the spatial unit of choice in urban economics and economic geography for investigating urban life as they are seen as encompassing the distinct phenomena of "urbanity" (proximity, density) and social interactions indirectly captured through a unified labor market. However, the population size and areal extent of metropolitan areas, as most often defined, render opaque the distinction between two salient types of urban population: those who work and those who reside within a metropolitan area. These two sets of individuals, among whom of course there is great overlap, putatively engage in different economic and social interactions which are in turn differently embedded in physical space. Availing ourselves of Swedish micro-level data for two distinct spatial units, tätorts ("dense localites") and *local labor markets*, we can distinguish which types of populations and which types of spatial agglomerations are responsible for the observed scaling effects on productivity and physical infrastructure. We find that spatially contiguous labor markets are not enough to generate some of the most salient urban scaling phenomena.

Corresponding author:
Jose Lobo, Arizona State University, PO Box 875502, Tempe, AZ 85287-5502, USA.
Email: jose.lobo@asu.edu

Keywords

Urban scaling, metropolitan areas, tätorts, labor markets

Introduction

The perspective that all cities—across eras, geographies, and cultures—share some fundamental socioeconomic processes as well as certain predictable quantitative properties has recently coalesced into *urban scaling theory* (Bettencourt, 2013; Bettencourt and Lobo, 2016; Bettencourt et al., 2007, 2008). Urban scaling theory draws on insights from urban economics, economic geography, and regional science and shares with these disciplines a common explanation for the existence and development of cities as resulting from the interplay between centripetal and centrifugal "forces" (Colby, 1933; Fujita et al., 1999; Isard, 1956). The nucleation forces in turn result from the socioeconomic advantages of concentrating human populations in space after accounting for the costs (Glaeser, 2011). These are known as *agglomeration* or *scaling* effects and constitute the foundational concepts for explaining the formation and persistence of cities anywhere (Duranton and Puga, 2004; Rosenthal and Strange, 2001). Urban agglomeration effects are based on the observation of systematic changes in average socioeconomic performance, land-use patterns, and infrastructure characteristics of all cities as functions of their population size (O'Flaherty, 2005). Population size is arguably among the most important determinants and consequents of socioeconomic development and change (Boserup, 1981; Carneiro, 2000; Johnson and Earle, 2000). Such relations are known across the sciences as scaling relations, which relate macroscopic properties of a system—here a city—to its scale, or size (Barenblatt, 2003; Brock, 1999). For this reason, the systematic study of such relationships in cities is known as urban scaling.

The urban scaling framework views cities as integrated socioeconomic networks of interactions embedded in physical space (Bettencourt, 2013). A crucial property of cities highlighted by this approach is that cities act to mix populations: that is, even if people in the city explore different locations at different times, anyone can in principle be reached and interacted with by anyone else. The mixing of population is both facilitated and constrained by physical infrastructure, the presence of which is often used to demarcate urban from non-urban (Ratcliffe et al., 2016). But operationalizing a view of cities as settings for social interactions and population mixing—assembling a set of spatial units of analysis which contain the relevant social aspects of urban settlements—implies choices about the use of existing data, the assignation of data to locations and the delineation of the boundaries of urban areas, all of which are far from trivial research decisions (Uchida and Nelson, 2010).

The concern that observed or inferred relationships between socioeconomic variables and physical characteristics of urban areas are unduly restricted by the choice of spatial unit of analysis, or even more starkly, are artifacts of such a choice, has a long vintage in urban studies. This issue is often referred to as the *Modifiable Areal Unit Problem* (MAUP): "…the areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating" (Openshaw, 1983: 3). For studies of contemporary urban systems, the definition of *functional cities*, as integrated socioeconomic units, has become the common standard for many scientific analyses of the properties of cities and urban systems (Glaeser et al., 1992). The U.S. Census Bureau has a long-standing, and arguably the most consistent, definition of functional cities, known as *Metropolitan Statistical Areas* (MSAs), whose grounding in theoretical considerations was developed in the 1960s (Berry, 1967; Berry et al., 1969; Fox, 1968; Fox and Kumar 1965). MSAs are in effect unified labor markets reflecting the frequent flow of goods, labor, and information, which in turn is a proxy for intense socioeconomic interactions (Glaeser et al., 1995). An effort to define functional cities in a conceptually meaningful and empirically consistent manner has been recently undertaken by the Organization for Economic Cooperation and Development (OECD), in collaboration with the European Union (OECD, 2012). This has resulted in a new set of harmonized metropolitan area definitions across the EU and other OECD nations. At present, these definitions represent the most consistent attempt to define functional urban areas in Europe, making contact with those of other nations such as, for example, the USA, Mexico, and Japan.

Metropolitan areas (MAs) encompass the distinct phenomena of “urbanity” (proximity, closeness, and density) and social interactions indirectly captured through a labor market. However, the population size and areal extent of metropolitan areas, as most often defined, render opaque the distinction between two salient types of urban population: those who work and those who reside within a metropolitan area. These two sets of individuals, among whom of course there is great overlap, putatively engage in different economic and social interactions which are in turn differently embedded in physical space. Metropolitan areas are defined in accordance with commuter flow patterns, so very nearly all the employed people work and live in the same metropolitan area. It is only at geographical scales below the MAs level that differences in work and residential locations emerge. A business district within an urban core may be the setting for intense socioeconomic interactions during work hours but be devoid of people come nighttime. And the interactions among residents of a neighborhood might also be intense and of consequence but different from those engaged in by co-workers in the same firm or industry.

The distinction between residential and work locations is analytically important because metropolitan areas encapsulate both the residential and workplace uses of physical space in a single geographic unit obscuring the ability to test which interactions are generative to scaling results. In the case of urban scaling relationships, the fact that many of the reported empirical regularities have been obtained using data for metropolitan areas has animated the criticism that such regularities are an artifact of the choice of metropolitan areas as unit of analysis (Arcaute et al., 2015; Depersin and Barthelemy, 2018; Leitao et al., 2016). By increasing geographic granularity, an investigation may be able to distinguish which types of interactions are associated with different scaling outcomes. Furthermore, by separating metropolitan populations into its residential and working components, we contribute to ongoing efforts at elucidating how the distribution of specific urban attributes generates observed aggregate scaling patterns (see, e.g., Cottineau et al., 2017, 2018; Keuschnigg et al., 2019; Sarkar, 2018; Sarkar et al., 2016).

In the present discussion, we avail ourselves of micro-level data (individuals and their wage income, place of work and place of residence) provided by *Statistics Sweden* for two spatial units defined by *Statistics Sweden* which represent different socioeconomic uses of physical space, *tätorts* (“dense locality”) and *local labor markets* (LAs). Tätorts delineate spaces of an “urban” nature based on population density and the built environment while LAs trace regions based on labor flows and do not have a strong “urban” nature uniformly within their boundaries. Both types of spatial units are somewhat independent of administrative boundaries and, with the exception of the largest few, are smaller (population-wise) than metropolitan areas. For both types of spatial units we are able to track those who use the space for residential purposes (referred to here as “nighttime population”) and those who labor inside the space (referred to here as “daytime population”). The yearly data covers the period 1990 to 2015. Differentiating between types of population utilizing different physical spaces and infrastructures makes it possible in turn to isolate the role of different types of social interactions in generating modern urban scaling phenomenon.

The discussion is organized as follows. The next section provides a summary description of the urban scaling theoretical framework. The formalism that underlies urban scaling theory, and the proposed mechanisms for generating predicted scaling relations, are very general and are not tailored to the specific characteristics of modern cities or even restricted to settlements of a certain size.[1] The spatial units (in which social interactions are embedded), as well as the population and wage income data used in the investigation are described in the Estimation Framework and Results section. We conclude by discussing implications of our findings for urban scaling analysis.

Urban scaling framework

We proceed by explicitly stating the quantitative expectations and realm of applicability of urban scaling theory as a model for analyzing cities and urban areas (for a more detailed treatment see Bettencourt (2013) and for more discussion of the topological aspect's see Bettencourt and Lobo (2016)). Scaling relations—which imply proportionality between changes in the values of the independent and dependent variable—are naturally written in terms of power-law functions (Barenblatt, 1996). The urban scaling framework proposes that any city-wide property (e.g. total economic output or built-up area), denoted by Y, should be written as

$$Y_i = Y_0 N_i^{\beta} e^{\xi} \tag{1}$$

where N_i refers to the population of the ith city in an urban system, Y_0 is a baseline prefactor common to all cities, β is a dimensionless scaling exponent (or elasticity, in the language of economics) and ξ are statistical fluctuations accounting for deviations in each city from the expected (power-law) scaling relationship. (The variables can all be easily made time-dependent by modifying the notation). The term Y_0 captures system-wide socioeconomic development. The choice of a power-law function form assumes that the effect on the dependent variable of increasing population size is not additive but multiplicative which is to say that the increase in Y is driven by the interaction of many factors observationally summarized in an increase in population size (Bettencourt et al., 2013; Coffey, 1979). We note that equation (1) is formally equivalent to the production function used in Glaeser et al. (1995) in which capital and labor are assumed to move freely among urban areas and differences in urban output are attributed to location-specific differences in productivity and not to differences in capital-to-labor ratios.

There are different quantitative expectations for the β exponents corresponding to different urban metrics. To calculate the expected value of these exponents, urban scaling theory proposes a self-consistent model of socioeconomic networks embedded in urban built space, as decentralized infrastructure networks (Bettencourt, 2013). Urban scaling theory builds on a long analytical tradition which describes a city functionally as a (short-term) spatial equilibrium whose spatial extent is set by the balance of density-dependent socioeconomic interactions and transportation costs (Richardson, 1973) The general features of this equilibrium can be obtained, as in an Alonso–Muth–Mills type model (Alonso, 1964; Glaeser, 2007; Mills, 1967; Muth, 1969), via a simple argument equating the expected costs of mobility, c, to the per capita net benefits accruing from social and economic interactions in the city, y.

The costs of spatial movement are set by the typical length scale of the city, $L = \sqrt{A}$, where A is the area of the city, via a fractal dimension of movement (H) and a cost per unit length, ε, leading to

$$c = \varepsilon A^{1/2} \tag{2}$$

It can be argued that the intensity of socioeconomic interactions is set, on average, by the population density over the built area, and thus can be written as

$$y = G\frac{N}{A} \tag{3}$$

where G is a term translating interactions into benefits (which include but are not restricted to the value of economic transactions). The notion, that increasing productivity derives from the concentration, intensification, and differentiation of social interactions, goes back at least to Adam Smith and is the basic idea behind economics models of agglomeration effects (Jones and Romer, 2010; Storper and Venables, 2004). Human effort is bounded, of course, which requires that G be, on average, independent of N. Bounded effort is in general a function of human constraints and urban services and infrastructure. Equating these costs and benefits defines the area extent of the city in terms of its population size

$$A_N = \left(\frac{G}{\varepsilon}\right)^{2/3} N^{2/3} = \alpha N^{2/3} \tag{4}$$

The area A_N increases with more productive interactions, e.g., due to economic growth, and decreasing transportation costs, as is observed in world-wide patterns of urban sprawl over time (Angel et al., 2016).

As cities grow, space becomes occupied and transportation of people, goods, and information is channeled into decentralized networks (Batty, 2009; Kropf, 2018; Morris, 1979). The space created by these networks is different from the circumscribing area and can be thought of as an "infrastructural area," A_I, in which socioeconomic interactions occur. Assume that there is a quantity of infrastructure per capita, d, which is proportional to circumscribing area (A_N, the physical space in which social and economic interactions are embedded)

$$d \sim \left(\frac{A_N}{N}\right)^{1/2} \tag{5}$$

with total network (or infrastructure) proportional to population size

$$A_I \approx Nd = N\left(\tfrac{A_N}{N}\right)^{1/2} = A_N^{1/2} N^{1/2} = \left(\alpha N^{2/3}\right)^{1/2} N^{1/2} = \alpha^{1/2} N^{5/6} \tag{6}$$

Let us assume now that total urban socioeconomic outputs (Y) are proportional to local interactions embedded in infrastructure area so that

$$Y = yN = G\frac{N^2}{A_I} = G\frac{N^2}{\alpha^{1/2} N^{5/6}} = Y_0 N^{7/6} \tag{7}$$

Equation (7) embodies the perspective that cities are concentrations not just of people, but rather of social interactions (a point emphasized by Jane Jacobs, 1969). The urbanized area of a city is thus naturally sublinear on its population size, while socioeconomic outputs resulting from social interactions are superlinear. This superlinear phenomenon is a spatially embedded version of "network effects" in which the benefit of participating in a network is proportional to its number of links, not nodes (Rohlfs, 1974; Shapiro and Varian, 1998).

We have two explicit statements (predictions) as to what the urban scaling coefficients should be for the areal extent of an urbanized area and for a measure of socioeconomic outputs: $\beta_{area} = 2/3 \cong 0.66$ and $\beta_{output} = 7/6 \cong 1.16$. Note that these coefficients result not only from geometric considerations but also from assumptions regarding social interactions among individuals in the city.[2] These individuals do not form an undifferentiated mass but constitute sociologically and economically distinct populations (or sub-populations of a larger population): residents, economic agents, retirees, students, etc. The predicted urban scaling coefficients represent the presence and activity of different types of urban population interacting differently in different urban spaces. We use micro-level data for Swedish urban areas to capture the scaling effects of two different populations using urban space: the collection of individuals residing in an urban space ("nighttime population") and the collection of individuals working in an urban space ("daytime population").

Spatial units and data

Urban areas and labor markets

A tätort ("dense locality" in Swedish) is a geographical area defined by *Statistics Sweden* based on a certain number of population and building characteristics. It has a minimum of 200 inhabitants and the distance between buildings does not exceed 200 meters. A tätort may be a city, town, or larger village. Tätorts have been consistently defined since 1960 and were the focus of the first investigation of urban scaling (Nordbeck, 1971). The 200-meter-rule can sometimes be abrogated if the tätort influences a larger land area. In the case of a smaller tätort, without a proper "urban core," the distance between the buildings can be less than 200 meters. The built-up environment includes all residential units (houses, apartment blocks, uninhabited houses, and vacation homes), and structures only used as workplace activities (workplaces, offices, warehouses, etc.). Buildings related to farming activities are not included. The 200-meter-rule is not valid if the area between the buildings can be used for public purposes, e.g. infrastructure, parking lots, parks, sports facilities, and cemeteries (this is also the case for warehouses, railways and harbors). Tätort spatial boundaries are independent of administratively set boundaries, e.g. associated with political municipalities. The number of tätorts has changed over the years, today approximately 87% of the Swedish population lives in a tätort. The area of a tätort is not fixed but varies according to the criteria of density and built-up environment. We highlight here that the "urban" nature of a tätort resides (no pun intended) in certain types of buildings being located in close proximity to each other.

Another definition of a Swedish functional region is a *local labor market* ("lokal arbetsmarknad" in Swedish). LAs are an attempt at delineating geographical areas of integrated economic activity, such as labor markets connected by commuting flows. However, not all parts of an LA are built-up land but can also to a certain extent consist of rural areas, farm land, and forests. This implies that an LA includes a number of tätorts and also the land area surrounding them where there may be a lower population density or even farm land. In that sense, Swedish LAs are not equivalent to U.S. Metropolitan Statistical Areas or similar definitions of functional cities in OECD nations (OECD, 2012). Furthermore, Swedish LAs do not need to satisfy a minimum number of inhabitants threshold, but they do require a "core." The *core criteria* set by *Statistics Sweden* are: (1) the share of individuals that commute to work in another municipality must not exceed 20%, and (2) the "out-commuters" must not exceed 7.5% to another single municipality within the labor market. The municipality which fulfills these criteria is then described as the core of an LA and the municipality is considered "self-sufficient" when it comes to labor. (Note that an LA can have more than one core municipality.) The surrounding municipalities are connected to the LA core municipalities based on their commuting patterns. There is also a geographical maximum for an LA in the sense that only two outward "links" are allowed. So you would have the core municipality, a neighbor municipality, and then possibly a municipality adjacent to the neighbor.

In the majority of cases, an LA consists of a number of municipalities where one (or several) municipalities are defined as a core, while the other municipalities are the "periphery." There are some cases in which the municipality and the LA are the same spatial unit, mostly in the northern parts of the country where municipalities cover large tracts of land (as well as in the case of the island of Gotland). The definition of an LA builds on classic concepts of *central place theory* where the core (the central place) consists of the majority of the economic activity while the periphery is home to most of the labor force, which commutes to in and out of the center (Isard, 1956). The strength of the core and its need for labor determine the commuting distance. The LA definition has changed over time and the number of LAs has decreased. Based on LA definitions from *Statistics Sweden*, the number of LAs has decreased from 187 in 1970 to 81 in 2006. The results presented here are based on using the current (2006) definition and count of LAs and imposing these boundaries to all periods.

The set of LAs constitute a complete tiling of the territory of Sweden, and thus include land that is sparely populated and decidedly not urban by other measures, which is not the case for tätorts. Since LAs are not obligatorily defined around urban areas, these labor markets include farming communities and mining regions as well as major urban areas such as Stockholm. Figure 1 shows the partitioning of Sweden into LAs and tätorts.

Populations and wage income

The Swedish micro-data (*Microdata Online Access*, MONA is the Swedish acronym) is a dataset that consists of all individuals in Sweden with a "right of residence."[3] The individual-level characteristics registered in the dataset include age, gender, immigration background, educational attainment and background, occupation, and wage incomes (which includes wages from employment and income from self-employment). Wage income, by excluding any governmental transfer, rental income, stock dividends, or investment interest, can be considered a measure of individual-level productivity. Individuals' identity is censored to protect privacy, but such characteristics can be tallied at the level of a tätort and an LA. The micro-level data also cover all establishments (physical places of work) in Sweden enabling the matching of individuals to the establishments they work in and to data on self-employment.

Using the Swedish MONA data, we are able to track the exact residential location of each individual and also the exact location of their workplace. Based on this information, we can measure the wage income that individuals bring home ("nighttime population"), and also the geographical source of these incomes ("daytime population"). By examining the scaling effects, the distinction between residential and working populations will not reveal a

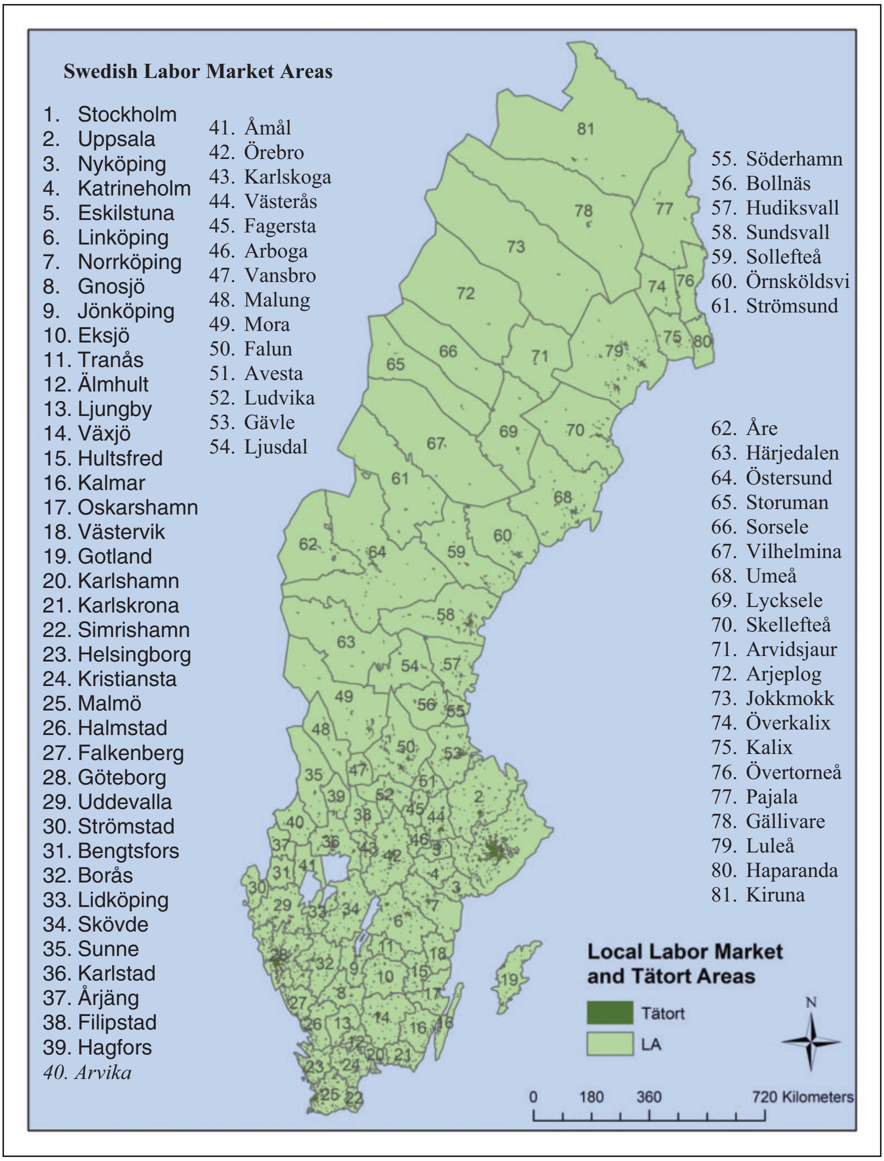

Figure 1. Map showing local labor markets (LAs) and tätorts in Sweden (2017).

distinction for LAs since the daytime and nighttime activities are contained within the same geographical boundaries. At the LA scale, with the exception of a small proportion of commuters, a person's workplace and residence are geographically overlapping, but there are major differences when we examine the much smaller tätort areas, since fewer individuals simultaneously live and work within the same tätort.

Table 1 shows the summary statistics of the variables used in the regressions for the year 2015—we show the statistics for only one year since the salient feature of the summary statistics is the great variability exhibited by the data across the spatial units. As is easily gleaned by looking at Figure 1, the areal extent of LAs and tätorts differs significantly across

Table 1. Summary statistics (year = 2015).

	Daytime			Nighttime		
	Employment	Total wages	Land area	Population	Total wages	Land area
Tätorts						
Mean	2070	6,798,473	275	3533	6,798,473	275
Std dev	22,645	71,650,615	1086	31,189	71,650,615	1086
Min	1	141,613	10	139	141,613	10
Max	892,027	2,920,596,277	38,163	1,219,382	2,920,596,277	38,163
CoV	10.94	10.54	3.95	8.83	10.54	3.95
# obs	1956	1956	1956	1956	1956	1956
las						
Mean	55,176	179,259,491	5011	100,120	190,717,780	5011
Std Dev	144,700	525,864,101	4317	235,200	528,387,448	4317
Min	907	2,133,416	788	2160	3,137,151	788
Max	1,184,900	4,389,355,278	19,992	1,876,490	4,363,298,077	19,992
CoV	2.62	2.93	0.86	2.35	2.77	0.86
# obs	81	81	81	81	81	81

Land area is measured in hectares; wages measured in current Swedish kronor.

Sweden. Consequently, the daytime and nighttime populations, and total wages, differ hugely across the spatial units. Do any of the predicted relationships between area and population and output and employment emerge from such a sea of variability?

Estimation framework and results

Equation (1) is an average statement that cannot be obeyed exactly in every instance. This is not only because all cities have specific local characteristics and urban indicators fluctuate over time but, more fundamentally, because a continuous scaling relation must break down in the limit of small discrete numbers (Bettencourt et al., 2010; Gomez-Leviano et al., 2012). Assuming that ξ behaves approximately as Gaussian noise (with zero mean) justifies using the simplest fitting procedure for Y versus N, a linear relation in logarithmic variables and minimizing ordinary least squares (OLSs)

$$\ln Y_i = \ln Y_0 + \beta \ln N_i + \xi_i \tag{8}$$

where i indexes different cities in the urban system, the dependent variable denotes either total wages or land area, and N is either the nighttime or daytime population. This implies that the exponent β is the slope of the linear regression of $ln Y_i$ on $lnNi$, and the prefactor, $ln Y_0$, is its ordinate at the origin ($N = 1$). This implies that the scaling relation $Y = Y_0 N^{\beta}$ is the expectation value of the approximately lognormally distributed stochastic variable Y for a city, given its $<N> |_N = Y_0 N^{\beta}$.

The regression results—obtained using yearly data—are shown in Tables 1 and 2. The regressions were estimated using a control for heteroskedasticity and the software package Stata SE version 14. For each estimated fit, the R^2 square is around 0.96. Note that since the data used represent a statistical population, and not a sample, confidence intervals and standard errors are not meaningful. Of course, the distance of any one observation from the fitted line is a source of information, inviting an investigation as to why anyone spatial

Table 2. Scaling results for tätorts (urban areas).

Tätort	Daytime		Nighttime	
Year	Beta (TW)	Beta (Area)	Beta (TW)	Beta (Area)
1990	1.11	0.37	1.02	0.67
1991	1.11	0.37	1.02	0.58
1992	1.11	0.38	1.02	0.59
1993	1.13	0.38	1.02	0.59
1994	1.14	0.39	1.02	0.59
1995	1.14	0.39	1.02	0.60
1996	1.13	0.40	1.02	0.60
1997	1.13	0.41	1.03	0.60
1998	1.13	0.41	1.02	0.60
1999	1.13	0.41	1.02	0.60
2000	1.12	0.41	1.02	0.60
2001	1.10	0.41	1.02	0.61
2002	1.11	0.42	1.02	0.61
2003	1.11	0.42	1.02	0.61
2004	1.12	0.43	1.02	0.61
2005	1.13	0.43	1.02	0.62
2006	1.13	0.43	1.02	0.62
2007	1.13	0.43	1.02	0.62
2008	1.12	0.43	1.02	0.62
2009	1.12	0.45	1.02	0.62
2010	1.13	0.45	1.02	0.63
2011	1.13	0.46	1.02	0.71
2012	1.10	0.53	1.03	0.71
2013	1.10	0.52	1.03	0.71
2014	1.10	0.52	1.03	0.70
2015	1.10	0.52	1.03	0.70
Mean	1.12	0.43	1.02	0.63
Std Dev	0.01	0.05	0.00	0.04
CoV	0.01	0.11	0.00	0.07

unit exhibits behavior different from the expected average (Bettencourt et al., 2010); however, analysis of these deviations from the pattern are beyond the scope of the present study.

Table 2 shows the results for tätorts and Table 3 for LAs. The headings Daytime and Nighttime distinguish the two population variables: daytime working population and resident population, respectively. Beta(TW) and Beta(Area) denote whether the dependent variable is *total wages*, accrued by those who work or reside in a specific spatial unit, or the *land area* of the specified spatial unit. Since in every case the data use all of the available observational spatial units and include all adult legal residents of Sweden, once more the regression parameter estimates are population-based values, not samples, and thus there is no standard error to be reported. The estimated values for the scaling coefficients remained remarkably stable over a 25-year period as evidenced by the very small coefficient of variation for the time series of coefficients.

Given that the preponderance of the Swedish population that reside within tätorts, the LA and the LA population contained within tätort boundaries are very nearly identical population counts. The distinction becomes relevant when considering regressions with land

Table 3. Scaling results for LAs (local labor markets).

LAs	Daytime		Nighttime	
Year	Beta (TW)	Beta (Area)	Beta (TW)	Beta (Area)
1990	1.05	0.75	1.03	0.79
1991	1.05	0.75	1.03	0.75
1992	1.05	0.76	1.03	0.75
1993	1.05	0.76	1.03	0.75
1994	1.05	0.76	1.03	0.75
1995	1.05	0.76	1.03	0.75
1996	1.04	0.75	1.03	0.74
1997	1.04	0.76	1.03	0.74
1998	1.04	0.75	1.04	0.74
1999	1.05	0.74	1.04	0.73
2000	1.06	0.74	1.04	0.73
2001	1.05	0.74	1.04	0.72
2002	1.04	0.74	1.04	0.72
2003	1.04	0.75	1.04	0.72
2004	1.04	0.75	1.04	0.72
2005	1.05	0.75	1.04	0.72
2006	1.05	0.75	1.04	0.73
2007	1.04	0.75	1.04	0.73
2008	1.04	0.74	1.04	0.72
2009	1.05	0.74	1.04	0.72
2010	1.05	0.74	1.03	0.72
2011	1.05	0.74	1.04	0.75
2012	1.05	0.76	1.04	0.75
2013	1.05	0.76	1.04	0.75
2014	1.05	0.76	1.05	0.81
2015	1.05	0.76	1.05	0.81
Mean	1.05	0.75	1.04	0.74
Std Dev	0.01	0.01	0.00	0.03
CoV	0.01	0.01	0.00	0.03

area, as the amount of unpopulated land area with in LA can be significant (see Figure 1). Since the results for the LAs and urbanized LAs are basically the same, for ease of exposition, we discuss only the results for urbanized LAs. It is not surprising that scaling results for tätorts and LAs are different, after all they represent different ways of partitioning physical space using socioeconomic criteria and thus represent different units of spatial analysis (Spiezia, 2003).

The scaling coefficients obtained for tätorts—spatial units intended to capture "urbanity" through the built environment—conform to the expectations of the urban scaling framework. Total wage income accrued by the individuals who work in tätorts scales superlinearly with employment population by a superlinear scaling coefficient of 1.13 (averaged across the 25 years of data), indicating the presence of interactions among individuals generative of increasing returns to scale. The total wages earned by the population residing in tätorts scales nearly linearly with resident population, suggesting that the interactions among this specific population does not generate increasing returns to scale. The total area of tätorts scales sublinearly with residential population (with an average coefficient of approximately 0.63) and working population (with an average coefficient of approximately 0.43).

The scaling coefficient obtained when using the nighttime population corresponds to that predicted by the scaling framework for how areal extension should expand as population increases; this is an expansion driven by rates of social interaction and decreasing costs of moving within physical space. The increase in the area extent of tätorts induced by an increase in the number of individuals working in them is more compact than that predicted, either as the result of expanding the area of movement or the expansion in infrastructure. The land area of a tätort is set by the built environment, both related to daytime and nighttime activities, but it may be the case that most of the built environment (especially in the outskirt of the tätort) is primarily residential housing, which in turn means that a larger share of the land area is used for residential purposes than for employment purposes.

In the case of LAs, which recall are regional labor markets, the scaling coefficients are practically the same whether daytime or nighttime population data is used. The increase in total wage income (a measure of socioeconomic output) with either working or residential population is nearly linear indicating proportionality between population size and accrued wage income (which, prima facie, signifies the absence of increasing returns). The scaling of area with population is approximately 0.75, a value close to the predicted 5/6 for area as shaped by decentralized networks for the flow of people, goods, and information.[4]

Discussion

The productivity advantage conferred by large population size and density (through a variety of mechanisms) is a stylized fact that has long been studied in urban economics (see, e.g., Abel et al., 2012; Ciccone and Hall, 1996; Echeverri-Caroll and Ayala, 2010; Glaeser, 1999; Sveikauskas, 1975). The relationship between population size and the areal extent of cities has also been a long-standing topic of interest in urban studies (see, e.g., Batty and Kim, 1992; Craig and Haskey, 1978; Marshall, 2007). The contribution of the urban scaling framework is that such relationships are generated in a predictive manner, and the underlying processes producing scaling effects are conceived as social interactions embedded in networks.

Urban scaling theory builds on over a century of urban models in geography, economics, and complex systems. The central idea is that urban areas result from self-consistently balancing incomes and costs, especially transportation, thus defining functional cities implicitly through this budget constraint. The “urban” in “urban scaling” refers to the same attributes that Edward Glaeser does when he describes cities as “the absence of physical space between people and companies. They are proximity, density, closeness.” (Glaeser, 2011: 6) What makes a settlement inherently “metropolitan”? A textbook definition (O’Sullivan, 2011) defines cities as geographical areas with a concentration of individuals and activities higher relative to the surrounding area. These characterizations illustrate the general principle that the essence of urbanism is not physical space per se, but the frequent and intense social interactions (“mixing”) among a diversity of individuals, occupations, and organizations within a given space (Smith, 2019). However for “proximity, density, closeness” to facilitate social interactions in a sustained and cumulative fashion does require physical infrastructure and social networks.

One of our main points is not new—just like we would not approach data in physics or biology without theory and expect it to make sense, so too in urban science. The choice of spatial unit of analysis must be informed by a theoretical framework capable of generating hypotheses and expectations (Muthukrishna and Henrich 2019). For spatial units to exhibit scaling patterns, it is not enough for population to be aggregated or even connected via physical infrastructure (Arcaute et al., 2015). What is more potentially revealing is to choose spatial units which have some of the features associated with "functional cities" but which also allow for disaggregation of the interacting social agents.

Tätorts might at first seem like very different entities than metropolitan areas, but some of their scaling behavior resembles scaling results obtained for metropolitan areas reported elsewhere (Bettencourt, 2013; Bettencourt et al., 2007). The residential populations of tätorts densify in accordance with the urban scaling model, but the population size of the tätort has no effect on per capita wage earnings of the people who sleep there at night. The superlinear scaling effect with wages is visible for the population who actually work in a tätort during the day which must be the result of commuter flows into tätorts, and it suggests that this flow is higher in larger tätorts than smaller ones. The result suggests further that the degree of gain is set by the access network that interactions happen over, whereas residential arrangements are more set by physical space.

That tätorts exhibit scaling behavior reminiscent of metropolitan areas is not that surprising once one considers that the analytical unit constructed by focusing on tätorts' wage earning population is profoundly urban in nature. Since the very start of the process which gave rise to metropolitan areas as data collection units, they have been seen as a way to empirically represent cities in a functional, and not simply structural, way.[5] The concept of functional cities is multidimensional, just like cities are. It includes consideration of population size and density, but also of economic activity, political organization, and geographic characteristics requiring one or more large and dense set of diverse employment. The definition of metropolitan areas (as a stand-in for a functional urban area) followed from urban theory, which in turn guided data collection and organization within a standardized national effort by the census. Certainly, the built environment and the physical infrastructure are crucial for facilitating the myriad of interactions which together express "urban life" (Parr, 2007). The scaling results for tätorts reveal that *who* is using a space conformed by physical infrastructure matters for understanding which settings can be expected to exhibit urban scaling and which one not. A glaring question is why the scaling behavior of LAs does not match that observed for metropolitan areas in different parts of the world (recall that metropolitan areas are in effect labor markets revealed through commuting flows). The Swedish local labor markets provide a sort of "Black Swan" to the perspective that spaces delineated by commuting flows are inherently "urban" or "metropolitan." For there to be agglomeration economies, of the sort captured by the scaling framework, requires "proximity, density, closeness" as facilitators of interactions. Presumably, if the LA spatial units encompass different types of settlements, then the scaling of wage income will approach the national average total output per capita, the scaling would be linear (as is the empirical case), and the differences between daytime and nighttime vanish. Why is the area scaling for LAs approximately 3/4, reminiscent of the scaling of metabolic rate with size for organisms? Is it that the scaling of the area of LAs with population size reflects the relationship between transportation networks and population size? We hope to elucidate these questions in future research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Recent efforts have extended the empirical reach of urban scaling theory to pre-modern and even non-urban settlement systems (Cesaretti et al., 2016 (in press); Hanson and Ortman, 2017; Hanson et al., 2017; Ortman and Coffey, 2017; Ortman et al., 2014, 2015, 2016; Ossa et al., 2017).
2. Nordbeck (1971) also derived a scaling coefficient of 2/3 for the relationship between area and population, using data for Swedish tätorts. Nordbeck's argument is an elegant instance of dimensional analysis: assuming that urban areal extent is a function of population size, the dimensionality of area and of population, linked by an equal sign, must be the same. Since area is two-dimensional, and population utilize space three-dimensionally, area must scale with population to the 2/3 power ($A = aN^{2/3}$).
3. For an English-language description of MONA, go to https://www.scb.se/en/services/guidance-for-researchers-and-universities/mona–a-system-for-delivering-microdata/
4. Interestingly, the value of 0.75 matches that of the coefficient for the scaling of electricity use with population size for U.S. metropolitan areas (Fragkias et al., 2016).
5. An effort that has been subject to criticism since the very beginning (see, e.g., Rosenwaike, 1970) and which contineus to be the subject of suggested improvements (e.g., Isserman (2005)).

References

Abel JR, Dey I and Gabe TM (2012) Productivity and the density of human capital. *Journal of Regional Science* 52: 562–586.

Alonso W (1964) *Location and Land Use*. Cambridge: Harvard University Press.

Angel S, Blei AM, Parent J, et al. (2016) *Atlas of Urban Expansion: 2016 Edition Volume 1: Areas and Densities*. New York: New York University, Urban Expansion Program.

Arcaute E, Hatna E, Ferguson P, et al. (2015) Constructing cities, deconstructing scaling laws. *Journal of the Royal Society Interface* 12: 1–8.

Barenblatt GI (1996) *Scaling, Self-Similarity, and Intermediate Asymptotics*. Cambridge: Cambridge University Press.

Barenblatt GI (2003) *Scaling*. Cambridge: Cambridge University Press.

Batty M (2009) Cities as complex systems: Scaling, interaction, networks, dynamics and urban morphologies. In: Meyers R (eds) *Encyclopedia of Complexity and Systems Science*. New York: Springer.

Batty M and Kim KS (1992) Form follows function: Reformulating urban population density functions. *Urban Studies* 29: 1043–1069.

Berry BJL (1967) Generalization of the metropolitan area concept. *Proceedings of the Social Statistics Section of the American Statistical Association* 10: 86–91.

Berry BJL, Goheen PG and Goldstein H (1969) *Metropolitan Area Definition: A Re-Evaluation of Concept and Statistical Practice*. Washington, DC: United States Bureau of the Census.

Bettencourt LMA (2013) The origins of scaling in cities. *Science* 340: 1438–1441.

Bettencourt LMA and Lobo J (2016) Urban scaling in Europe. *Journal of the Royal Society Interface* 13: 20160005.

Bettencourt LMA, Lobo J, Helbing D, et al. (2007) Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the USA* 104: 7301–7306.

Bettencourt LMA, Lobo J, Strumsky D, et al. (2010) Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS One* 5(11): e13541.

Bettencourt LMA, Lobo J and West GB (2008) Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics. *The European Physical Journal B* 63: 285–293.

Bettencourt LMA, Lobo J and Youn H (2013) *The hypothesis of urban scaling: Formalization, implications and challenges*. SFI Working Paper 2013-01-004. Santa Fe: Santa Fe Institute.

Boserup E (1981) *Population and Technological Change: A Study of Long-Term Trends*. Chicago: University of Chicago Press.

Brock WA (1999) Scaling in economics: A readers' guide. *Industrial and Corporate Change* 8: 409–446.

Carneiro RL (2000) The transition from quantity to quality: A neglected causal mechanism in accounting for social evolution. *Proceedings of the National Academy of Sciences of the USA* 97: 12926–12931.

Cesaretti R, Lobo J, Bettencourt LMA, et al. (2016) Population-area relationship for medieval European cities. *PLoS One* 11(10): e0162678.

Cesaretti R, Lobo J, Smith ME, et al. (in press) Increasing returns to scale in the provincial towns of Tudor England.

Ciccone A and Hall RE (1996) Productivity and the density of economic activity. *American Economic Review* 86: 54–70.

Coffey WJ (1979) Allometric growth in urban and regional social-economic systems. *Canadian Journal of Regional Science* 11: 49–65.

Colby CC (1933) Centrifugal and centripetal forces in urban geography. *Annals of the Association of American Geographers* 23: 1–20.

Cottineau C, Finance O, Hatna E, et al. (2018) Defining urban clusters to detect agglomeration economies. *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/2399808318755146.

Cottineau C, Hatna E, Arcaute E, et al. (2017) Diverse cities or the systematic paradox of Urban Scaling Laws. *Computers, Environment and Urban Systems* 63: 80–94.

Craig J and Haskey J (1978) The relationship between the population, area, and density of urban areas. *Urban Studies* 15: 101–107.

Depersin J and Barthelemy M (2018) From global scaling to the dynamics of individual cities. *Proceedings of the National Academy of Sciences of the USA* 115: 2317–2322.

Duranton G and Puga D (2004) Micro-foundations of urban agglomeration economies. In: Henderson JV and Thisse J-F (eds) *Handbook of Regional and Urban Economics*. Vol. 4. Amsterdam: Elsevier, pp. 2063–2117.

Echeverri-Carroll E and Ayala SG (2010) Urban wages: Does city size matter? *Urban Studies* 48: 253–271.

Fox KA (1968) Functional economic areas and consolidated urban regions of the United States. *Statistical Reporter* 68: 145–148.

Fox KA and Kumar K (1965) The functional economic area: Delineation and implications for economic analysis and policy. *Papers of the Regional Science Association* 15: 57–85.

Fragkias M, Lobo J and Seto K (2016) A comparison of nighttime lights data for urban energy research: Insights from scaling analysis in the US system of cities. *Environment and Planning B: Urban Analytics and City Science* 44: 1077–1096.

Fujita M, Krugman P and Venables AJ (1999) *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge: The MIT Press.

Glaeser EL (1999) Learning in cities. *Journal of Urban Economics* 46: 254–227.

Glaeser EL (2007) *The economics approach to cities*. NBR Working Paper 13696. Cambridge: National Bureau of Economic Research.

Glaeser EL (2011) *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. New York: Penguin Press.

Glaeser EL, Kallal HD, Scheinkman JA, et al. (1992) Growth in cities. *Journal of Political Economy* 100: 1126–1152.

Glaeser EL, Scheinkman JA and Shleifer A (1995) Economic growth in a cross-section of cities. *Journal of Monetary Economics* 36: 117–143.

Gomez-Leviano A, Youn H and Bettencourt LMA (2012) The statistics of urban scaling and their connection to Zipf's Law. *PLoS One* 7(7): e40393.

Hanson JW and Ortman SG (2017) A systematic method for estimating the populations of Greek and Roman settlements. *Journal of Roman Archaeology* 30: 301–324.

Hanson JW, Ortman SG and Lobo J (2017) Urbanism and the division of labour in the Roman Empire. *Journal of the Royal Society Interface* 14(136): 20170367.

Isard W (1956) *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, and Urban Structure*. Cambridge: MIT Press.

Isserman A (2005) In the national interest: Defining rural and urban correctly in research and public policy. *International Regional Science Review* 28: 465–499.

Jacobs J (1969) *The Death and Life of Great American Cities*. New York: Random House.

Johnson AW and Earle T (2000) *Evolution of Human Societies: From Foraging Group to Agrarian State*. Palo Alto: Stanford University Press.

Jones CI and Romer PM (2010) The New Kaldor facts: Ideas, institutions, population, and human capital. *American Economic Journal: Macroeconomics* 2: 224–245.

Keuschnigg M, Mutgan S and Hedstrom P (2019) *Urban scaling and the regional divide, Science Advances* 5: 1–7.

Kropf K (2018) *The Handbook of Urban Morphology*. West Sussex: John Wiley & Sons.

Leitao JC, Miotto JM, Gerlach M, et al. (2016) Is this scaling nonlinear? *Royal Society Open Science* 3: 150649.

Marshall JD (2007) Urban land area and population growth: A new scaling relationship for metropolitan expansion. *Urban Studies* 44: 1889–1904.

Mills ES (1967) An aggregative model of resource allocation in a metropolitan area. *American Economic Review* 57: 197–210.

Morris AEJ (1979) *History of Urban Form: Before the Industrial Revolutions*. New York: John Wiley & Sons.

Muth R (1969) *Cities and Housing*. Chicago: University of Chicago Press.

Muthukrishna M, Henrich J (2019) A problem in theory. *Nature Human Behaviour* 3: 221–229.

Nordbeck S (1971) Urban allometric growth. *Geografiska Annaler* 53: 54–67.

O'Flaherty B (2005) *City Economics*. Cambridge: Harvard University Press.

Openshaw S (1983) *The Modifiable Areal Unit Problem*. Norwick: Geo Book.

Organisation for Economic Cooperation and Development (2012) *Redefining Urban: A New Way to Measure Metropolitan Areas*. Paris: OECD Publishing.

Ortman SG, Cabaniss AHF, Sturm JO, et al. (2014) The pre-history of urban scaling. *PLoS One* 9(2): e87902.

Ortman SG, Cabaniss C, Sturm JO, et al. (2015) Settlement scaling and increasing returns in an ancient society. *Science Advances* 1(1): e1400066.

Ortman SG and Coffey GD (2017) Settlement scaling in middle-range societies. *American Antiquity* 82: 662–682.

Ortman SG, Davis KE, Lobo J, et al. (2016) Settlement scaling and economic change in the Central Andes. *Journal of Archaeological Science* 73: 94–106.

Ossa A, Smith ME and Lobo J (2017) The size of plazas in Mesomerican cities and towns: A quantitative analysis. *Latin American Antiquity* 28: 457–475.

O'Sullivan A (2011) *Urban Economics*. New York: McGraw-Hill Education.

Parr JB (2007) Spatial definitions of the city. *Urban Studies* 44: 381–392.

Ratcliffe M, Burd C, Holder K, et al. (2016) *Defining Rural at the U.S. Census Bureau*. American Community Survey and Geography Brief, www2.census.gov/geo/pdfs/reference/ua/Defining_Rural.pdf (accessed 27 June 2019).

Richardson HW (1973) *Regional Growth Theory*. New York: John Wiley & Sons.

Rohlfs J (1974) A theory of interdependent demand for a telecommunications service. *Bell Journal of Economics and Management Science* 5: 16–37.

Rosenthal SS and Strange WC (2001) The determinants of agglomeration. *Journal of Urban Economics* 50: 377–393.

Rosenwaike I (1970) A critical examination of the designation of standard metropolitan statistical areas. *Social Forces* 48: 322–333.

Sarkar S (2018) Urban scaling and the geographic concentration of inequalities by city size. *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/2399808318766070.

Sarkar S, Phibbs P, Simpson R, et al. (2016) The scaling of income distribution in Australia: Possible relationships between urban allometry, city size, and economic inequality. *Environment and Planning B: Urban Analytics and City Science* 45: 603–622.

Shapiro C and Varian HR (1998) *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business Review Press.

Smith ME (2019) The generative role of settlement aggregation and urbanization. In Gyucha A (ed) *Coming Together: Comparative Approaches to Population Aggregation and Early Urbanization*. Albany: State University of New York Press, pp. 37–58.

Spiezia V (2003) *Measuring regional economies*. OECD Statistics Brief No. 6. Paris: Statistics Directorate of the OECD, Organization for Economic Cooperation and Development.

Storper M and Venables AJ (2004) Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography* 4: 351–370.

Sveikauskas L (1975) The productivity of cities. *Quarterly Journal of Economics* 89: 393–413.

Uchida H and Nelson A (2010) *Agglomeration index: Towards a new measure of urban concentration*. United Nations University World Institute for Development Economics Research Working Paper 2010/029. Helsinki: UNU-WIDER.

Deborah Strumsky is an assistant professor in the School of Sustainability and the School for the Future of Innovation at Arizona State University (Tempe, AZ, USA). Trained as a Regional Scientist, her research focuses on the interplay between invention and the characteristics of urban areas, and how invention and innovation drive urban economic development.

José Lobo is a faculty in the School of Sustainability at Arizona State University (Tempe, AZ, USA). Trained in physics and urban economics, his main research interest is the relationship between urbanization and socioeconomic development across eras and geographies. He has been involved in various efforts pertaining to studying urban and settlement scaling relationships in ancient and contemporary urban/settlement systems.

Charlotta Mellander is a professor of economics at Jönköping International Business School (University of Jönköping, Sweden). She has done pioneering work on the location patterns of creative individuals and firms and the effect of such locational choices on regional development.