Departures from Gibrat’s Law, Discontinuities and City Size Distributions

Ahjond S. Garmestani, Craig R. Allen, Colin M. Gallagher and John D. Mittelstaedt

[Paper first received, May 2005; in final form, December 2006]

Summary. Cities are complex, self-organising, evolving systems and the emergent patterns they manifest provide insight into the dynamic processes in urban systems. This article analyses city size distributions, by decade, from the south-eastern region of the US for the years 1860–1990. It determines if the distributions are clustered into size classes and documents changes in the pattern of size classes over time. A statistical hypothesis test was also performed to detect dependence between city size and growth using discrete probability calculations under the assumption of Gibrat’s law. The city size distributions for the south-eastern region of the US were discontinuous, with cities clustering into distinct size classes. The analysis also identified departures from Gibrat’s law, indicating variable growth rates at different scales.

1. Introduction

Cities are complex, self-organising, evolving systems (White and Engelen, 1993). As the hierarchical centres of consumption in regional economies, cities are characterised by pulses of growth and decline through history (Huang, 1998). Huang (1998) views urbanisation as a process of spatial convergence of energy into the centres of cities. Cities store, for instance, capital, labour and land, while energy, goods and services flow through cities (Huang, 1998). In time, cities evolve and transform, entailing a change in dynamics leading to a different level of organisation (Huang, 1998).

Urban economic analysis has suggested that the size of a city is dependent upon numerous factors (such as human capital and specialised inputs) (Wheeler, 2003). These factors create a positive feedback loop, which acts as an attractor to further growth. Urban theory predicts that at a threshold size, agglomeration benefits decrease and diseconomies (such as crime, congestion and living expenses) increase; cities should grow to a certain point and then growth rates should decrease as size increases (Wheeler, 2003). There is evidence of this phenomenon, as larger cities tend to grow at a slower pace (Dobkins and Ioannides, 2001), which may be indicative of cities growing faster when their market potential is large relative to their size, but that growth slows as a city’s size nears a critical threshold (Duranton and Puga, 2000). Interestingly, this phenomenon is largely absent from studies on the growth of metropolitan area populations, but is extant in county-level data (Wheeler, 2003).
If growth rates vary as a function of size, then the growth of the largest cities in a dynamic system will exceed their ‘proportional’ growth, leading to domination of the system (Batty, 2001). When elements in a dynamic system grow in proportion to natural advantage, the natural advantage acts as an attractor for growth and creates a positive feedback loop (Arthur, 1988). Historical advantage also serves to reinforce natural advantage and, in areas that were developed first, cities have the capacity to attract growth at a faster rate than later additions to the urban landscape (Arthur, 1988; Dobkins and Ioannides, 2001). The ratio of white to blue-collar workers and the capital–labour ratio exert a positive influence on productivity in most industries (Rigby and Essletzbichler, 2002). Cities of varying sizes are likely to have different specialised inputs; larger cities have a wider variety of inputs than might be available in smaller cities (Holmes, 1999). Thus, population growth should lead to increasing per capita income growth, which is strongly at odds with the empirical evidence. Batty (2001) speculates that growth is proportional to size, but the growth rates are random, which would ensure that there cannot be an unchanging steady state (for example, Zipf’s law) in city size distributions. Importantly, Zipf’s law is not consistent with agglomeration economies, and associated diseconomies of scale (Cordoba, 2001).

The focus upon rank–size characteristics of city size distributions has led to the acceptance of simplistic models that downplay important economic and social forces, and has reduced research on departures from rank–size characterisations (Overman and Ioannides, 2001). Thus, models that assume urban systems are static or develop in one-dimensional space may not be the most effective tools for characterising urban system dynamics at every scale (Mun, 1997). Dynamic models incorporating spatial features (such as geographical location) are based on the assumption that random perturbations drive the self-organisation of urban systems (Chen, 1996; Berling-Wolff and Wu, 2004). It is likely that complex interactions in urban systems cannot be captured by linear models, as cities evolve according to non-linear growth processes (Bertuglia et al., 1987; Huang, 1998). Echoing this finding, Hadar and Pines (2004) have demonstrated that the process of bifurcation in urban systems can be discontinuous or continuous, with the land ownership regime of the region playing a key role in the nature of the bifurcation of city growth. Gao and Chen (2005) created a single-species discrete population model, which revealed a sequence of bifurcations with non-unique dynamics with several attractors, small amplitude oscillations, large amplitude multiannual cycles and chaos. Periodic pulsing in population growth destroyed stable equilibria and initiated chaos (Gao and Chen, 2005). Positive feedback reinforces some of the pulsing in an urban system and can intensify that disturbance (Anas et al., 1998).

Much of urban theory has developed from central place theory (Garmestani et al., 2005). A central place is characterised as an attractor which can have a number of small towns at equal distances from it, where the smaller towns make use of the central place’s shops and services (Christaller, 1933). Christaller (1933) theorised that the differences in central places and their satellites produced two rules: the larger the central place, the fewer central places there are; and, the larger the central place, the greater the ‘sphere of influence’ of that place. Zipf (1949) identified a rank–size relationship for cities. This distribution manifests when all central places in an urban hierarchy have size-invariant growth rates (Gibrat, 1957). This rank–size relationship for urban systems, as described by Zipf’s law, is believed to be a reflection of a steady-state condition (Gabaix, 1999). For Gibrat (1957), growth is conceived as occurring within a large number of individual time-periods (Marshall, 1997). Within each time-period, growth rates are independent of city size and, more importantly, the rates of
growth within a size class vary randomly around an average growth rate (Marshall, 1997). Several researchers assert that Gibrat’s law is compromised, in that no urban system exhibits growth rates as conceived by Gibrat, which means that Gibrat’s law rests upon assumptions that are known empirically to be false (Marshall, 1997; Pumain and Moriconi-Ebrard, 1997). Thus, the contention is that urban growth rates are neither independent of city size nor homoscedastic with respect to city size (Marshall, 1997). This is important, because city sizes of a geographical region must exhibit random growth rates (Gibrat’s law) in order for the distribution of these cities to satisfy Zipf’s law (Gibrat, 1957; Gabaix, 1999).

Garmestani et al. (2005), Dendrinos (1992) and Dendrinos and Sonis (1990) have proposed that discontinuously distributed variables of inherently different speeds entrain growth processes to create clustering of cities in aggregations based on the similarity of their sizes. The evidence comes in the form of constraints on the size-density function, revealed as aggregations of cities, separated by gaps, that cannot be satisfactorily explained by developmental, historical or other factors (Allen and Holling, 2002; Bessey, 2002). These size classes may represent scales of opportunity for cities within the context of growth on a regional scale. Additionally, this research has a broad-scale connection to the characterisation of discontinuously distributed variables in complex systems in general.

In this study, we test city size distributions for discontinuities. If the distributions are discontinuous, we will test the growth rates of cities for departures from Gibrat’s law, which would serve as an explanation for the generation of discrete size classes in city size distributions.

2. Methods

We define a city as a human settlement above a threshold population size that satisfies the functional requirements of that population (Bessey, 2002). The cutoff for determining what is urban is arbitrary and arises from practical rather than theoretical considerations (Marshall, 1989). This analysis used a US census dataset incorporating the urbanised area (UA) definition. A UA comprises a central place and the urban fringe, which includes other ‘places’ (Bessey, 2000). The Bureau of the Census officially defines a ‘place’ as a concentration of population, which must have a name and be locally recognised, although it may or may not be legally incorporated under the laws of its state (Bessey, 2002).

Many US Bureau of the Census classifications have evolved over the past 120 years. Regional systems theory conceives of cities as the central places in regional, social and economic systems, nested within a larger hierarchy of cities and regions (Skinner and Henderson, 1999). US urban development in the 20th century was characterised by distinct regional patterns (Overman and Ioannides, 2001). Bureau of Economic Analysis (BEA) regions comprise defined entities whose boundaries hold historically. Additionally, aggregating cities on a national scale masks discontinuous patterns that manifest at a regional scale (Skinner and Henderson, 1999). Functional economic regions are likely to capture scale effects better than convenient political divisions (Rigby and Essletzbichler, 2002). Analysing the data based on BEA regions allowed for research along smaller and more uniform biophysical, economic and socio-cultural characteristics (Bessey, 2002).

We ranked cities in order of population size to determine whether size classes existed within the city size distribution. This study used a BEA dataset of cities in the southeastern region (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia and West Virginia) of the US. City size distributions were analysed with simulations that compared actual data with a null distribution established by calculating a kernel density estimate of the log-transformed data (Hall and York, 2001). Significance of size classes in the data was determined by
calculating the probability that the observed discontinuities were chance events by comparing observed values with the output of 1000 simulations from the null set (Restrepo et al., 1997). Because \( n \) in our 14 datasets varied from 50 cities in 1860 to 310 cities in 1990, we maintained a constant statistical power of \( \sim 0.50 \) for detecting discontinuities (Lipsey, 1990). Maintaining constant power rather than constant alpha levels (i.e. keeping type II error rates constant, rather than type I error rates) is a more robust approach when the focus is the detection and comparison of pattern among datasets with greatly varying \( n \) (Holling and Allen, 2002). We confirmed our results with cluster analysis based on variance reduction (SAS/STAT software, SAS Institute, 1999). A discontinuity was defined as an area between successive city sizes that significantly exceeded the differences between adjacent city sizes generated by the continuous null distribution (Allen et al., 1999). A size class was a grouping of three or more cities with populations not exceeding the expectation of the null distribution (Allen et al., 1999). City size classes were defined by the two end-point cities that defined either the upper or the lower extremes of the size class (Allen et al., 1999).

Growth rates from 31 cities in the southeastern US that existed from 1860 through 1990 were analysed to detect for statistical departures from Gibrat’s law. Growth is defined in this paper as the change in the population over time. We performed a statistical hypothesis test to detect dependence between city size and growth using discrete probability calculations under the assumption of Gibrat’s law. As in Gabaix (1999), we denote the normalised size of city \( i \) at time \( t \) (population of city divided by total population of all 31 cities) by \( S_i^t \) and the growth rate over the next 10 years by \( \gamma_{t+1}^i = S_{t+1}^i / S_t^i \). Gibrat’s law states that \( S_i^t \) and \( \gamma_{t+1}^i \) are statistically independent. We can think of statistical independence in the following way: given a set of normalised city sizes, the growth rates are randomly assigned to cities according to some growth distribution—or, vice versa, given a set of growth rates, city sizes are randomly assigned to the growth rates.

3. Results

The decadal city size distributions for the south-eastern region of the US were discontinuous. Distinct size classes of cities were identified in each decade. We observed 3–6 size classes in each decadal dataset (Figure 1). This structure is significant, as random draws of the same \( n \) from the null model revealed that 91 per cent of the outputs randomly generated were either unimodal or bimodal in their distribution and fewer than 1 per cent had over three size classes (Allen et al., 1999). Further, despite great change in the number of cities across decades, the dominance of a few cities and the number of size classes were relatively static over time. For each time-period analysed, there is a range of city sizes, a different number of cities represented and a different hierarchical relationship of the cities.

For the south-eastern region of the US, there were 50 cities in 1860 and 310 cities in 1990. Within each decade, the range of city sizes ranged from 2546 to 181 000 in 1860 to 10 005 to 2 157 806 in 1990. Beginning in 1860 and continuing until 1950, the largest city in the south-eastern region of the US was New Orleans. From 1960 to 1970, Miami–Hialeah was the largest city in the region and then Atlanta from 1980 to 1990. From 1860 to 1900, New Orleans and Louisville were the two dominant cities of the south-eastern region of the US. Atlanta was the fourth-largest city in the region and had begun its gradual ascent to the dominant city of the region, while Wheeling, West Virginia, was also a member of the size class of the largest cities at the upper tail of the distribution and was the ninth-largest city in the region in 1900. Over the course of the 20th century, Atlanta grew into the urban hub of the south-eastern region, while Wheeling continued a slow descent until it settled as the 72nd-largest city in 1990 and a member of the large size class of small cities in the lower tail of the city size distribution.
From 1950 to 1990, the state of Florida became a focal point of growth not only in the south-eastern region, but in the US in general. By 1950, Miami was the 4th-largest city, Jacksonville was the 10th-largest and Tampa was the 12th-largest city in the region. Thereafter, Florida cities increased in population and became some of the largest cities in the region. By 1990, Miami was the 2nd-largest city in the region and Tampa, Fort Lauderdale, Orlando, West Palm Beach and Jacksonville were all members of the size class with the largest cities in the region at the upper tail of the city size distribution.

The data indicate departures from Gibrat’s law in the ‘tails’ in the sense that the largest growth rates are associated with small normalised city size, while the large normalised size greatly increases the probability of below-average growth. This indicates a non-linear but interesting dependence between growth rates and city size (Figure 2). The 10 largest growth rates all occurred in cities with normalised size less than the average (horizontal line in Figure 2). Furthermore, of the 10 largest normalised sizes, 9 have growth rates below the average (vertical line in Figure 2). Combining these 20 points represents the 5 per cent most extreme (2.5 per cent largest sizes and 2.5 per cent smallest growth rates; the fact that these two groups are mutually exclusive is the better part of the point). One could attempt to use a chi-squared test to test if above-average growth rates were associated with smaller than average cities and vice versa; however, this test cannot pick up the most remarkable attribute of Figure 2: all of the very largest cities have small growth rates and all of the smallest cities have large growth rates. We test our research hypothesis (statistical alternative hypothesis) that the 2.5 per cent largest cities tend to grow at below-average rate and the 2.5 per cent smallest cities tend to grow at an above-average rate as follows. Our statistical null hypothesis states that city size and growth rates are independent. Following the general theory of statistical hypothesis testing one can develop a test which, unlike the chi-squared test, requires no additional assumptions as follows. In the data represented in Figure 2, 49.1 per cent of normalised sizes are below 0.0178, while 61.6 per cent of normalised sizes are below average.
Using a standard binomial probability calculation, the probability that the 10 largest growth rates would be randomly assigned (Gibrat’s law) to city sizes less than 0.0178 is about 0.0008. The probability that 9 of the 10 largest city sizes would be assigned below-average growth rates under statistical independence is < 0.049. Under the null hypothesis (Gibrat’s law), calculating the chance that both of these events would occur simultaneously results in a p-value smaller than 0.00004. Since our calculated p-value is well below any reasonable significance level, we reject the hypothesis that city size and growth rate are independent with a high level of significance. Both the graph and the probability calculations indicate strong evidence for departures from Gibrat’s law.

We also note apparent departures from Zipf’s law in the census data mentioned earlier, echoing Bessey (2002). Although there is much empirical evidence for power law behaviour at the national level, for this set of regional cities Zipf’s law is not satisfied. Consistent with Eeckhout (2004), we find empirically that a log-normal distribution provides an excellent fit for the normalised city size distribution data at a regional scale (Figure 3). Using any of the statistical tests for normality, we would conclude that the log city sizes for the entire dataset (i.e. all years and times) can be considered a sample from the normal distribution.

4. Discussion

The results of this analysis demonstrate that the hierarchical structure of regional urban systems is discontinuous, as theorised by Bessey (2002) and expanded upon by Garmestani et al. (2005). This is true despite significant variability in the growth dynamics of individual cities (Bessey 2002). Our results indicate that growth rates differ by city size and that cities in the south-eastern region of the US cluster into size classes, in contrast to the distribution expected if Gibrat’s law

Figure 2. Graphical display of $S_i$ versus $g_{i+1}$, which reveals departures from Gibrat’s law. The vertical line represents the sample average (over all years and cities) growth rate and the horizontal line represents the sample average normalised city size.
held for this dataset. At a regional level, our results indicate that city growth is not driven by small, random growth forces. Rather, our results suggest that growth is correlated to size, with smaller cities exhibiting higher growth rates and larger cities exhibiting lower growth rates. Importantly, our results are specific to the south-eastern region of the US and may not hold at a different scale.

Ioannides and Overman (2003) assert that deviations from Zipf’s law are easily explained by deviations from Gibrat’s law. Zipf’s law is better characterised as a general model which permits a simplified explanation of the hierarchical organisation of urban systems (Guerin-Pace, 1995). The rank–size rule operates as a guide, but it may mask the most interesting dynamics in city size distributions (Guerin-Pace, 1995). Guerin-Pace (1995) has argued that deviations from the rank–size rule are likely to be driven by intramodal growth processes (such as autocorrelation of growth within modes).

Soo (2002) reported that Zipf’s law was rejected for a majority of countries (73 per cent) using standard OLS and a small, but significant, number of countries (40 per cent) using a Hill estimator. This is significant, as deviations from Zipf’s law imply deviations from random growth (Soo, 2002). Larger scale economies and lower transport costs entail evenness in city sizes (Soo, 2002). Ellis and Andrews (2001) reported that Australia’s small towns followed Zipf’s law, which suggests that population growth is similar for small towns. This result is unique, as Zipf’s law, when it holds, has been found to hold for cities in the upper tail of a city size distribution rather than the lower tail (Ellis and Andrews, 2001). A relative lack of transport infrastructure may encourage the concentration of population in Australia into a few very large cities on a regional basis (Ades and Glaeser, 1995). Rossi-Hansberg and Wright (2004) state that, in order to generate Zipf’s law, growth.

Figure 3. Normal probability plot for the natural log of the normalised city sizes in the south-eastern region of the US.
processes at the city level must be scale-invariant. They claim that this is possible via two scenarios: one where the current productivity shocks are the only stochastic force in growth, and are permanent, producing permanent increases in the level of the marginal product of labour, so that the growth rate of the marginal product is independent of scale. The second case occurs when productivity shocks are temporary, but have a permanent effect on the product of labour via the accumulation of physical capital in a linear production setting. If neither of these conditions is satisfied, and a city is large, it will grow more slowly than average in the future and the opposite would be true for small cities (Gabaix and Ioannides, 2004).

Fujita and Mori (1997) modelled city emergence as a result of catastrophic bifurcation of the existing urban system. They claimed that, as the number of cities increases, the urban system manifests into a central place system. The inherent limitation in Fujita and Mori’s model reflects the inherent limitation in central place theory: cities are assumed to exist in a homogeneous environment. However, the geographical location and transport network associated with a particular city may play a critical role in determining city size in a heterogeneous environment (Mun, 1997). Macro-regional systems developed within river basins, with social and economic ties along natural transport corridors. Skinner and Henderson (1999) found that the line of attraction between two competing cities is not at an equal surface or cost distance, but is reflected in orientation to central places at the next higher level of the urban hierarchy and reinforced by the physical landscape. Geographical space is modified by the transport network, which in turn leads to greater access to information, effectively reducing the distance between cities (Guerin-Pace, 1995). Improvements in transport networks cause concentration of firms and houses in a particular city, which in turn concentrates resources (Mun, 1997). Larger cities then manifest a greater ‘sphere of influence’ and stunt the growth of smaller cities, which leads to increasing inequality between cities (Guerin-Pace, 1995). For the period 1836–1990 in France, there is a strong correlation between city size and growth rates, characterised by a non-linear relationship between the population of cities and their growth rates. Guerin-Pace relates this non-linear relationship between city sizes and growth rates to the spatial pattern of urban systems. In particular, the non-Pareto behaviour of medium-sized cities was characterised by a positive relationship between growth rates and city size and a negative correlation of growth rates over time.

Urban hierarchy arises from critical productivity thresholds determined by non-linear external and internal economies (Rosser, 1994). The entrainment of non-linear business oscillations (i.e. perturbations) across levels of a hierarchy can trigger a phase transition and the emergence of a new level of the hierarchy (Rosser, 1994). Emergent levels of an urban hierarchy therefore arise from either an expansion of trade due to lower transport costs, or from an alteration of the internal structures of cities, allowing cities to take advantage of agglomerative economies (Rosser, 1994). Slower-moving, higher levels in the hierarchy ‘slave’ faster-moving, lower levels of the hierarchy (Haken, 1996; Rosser et al., 1994). Higher-level variables carry the bounded range of lower-level oscillations to their level (Rosser et al., 1994). Occasionally, a discontinuous change at a lower level triggers a discontinuous change at a higher level, which manifests in a dramatic systemic transformation throughout the hierarchy (Rosser et al., 1994). This phase transition is likely to occur at the point where a highly conserved system is near a critical threshold (for example, growth, release) (Rosser et al., 1994).

With respect to urban systems, discontinuous change could occur if one firm (i.e. microeconomic variable) adopts a new technology, then other firms adopt the technology triggering spin-offs and, ultimately, triggering a macroeconomic surge (Rosser et al., 1994). Malecki (1980) attributes departures from rank–size distributions to differential growth in different size classes. The interaction between endogenous comparative advantages and exogenous
trade and transport patterns triggers discontinuities in city growth rates, which manifest in cities clustering into distinct size classes (Dendrinos and Rosser, 1992). The evolution of urban systems is a highly non-linear process (Ioannides and Overman, 2004). Ioannides (2003) implies a law of motion for human capital that exhibits multiple equilibria differing in terms of stability. The distribution of human capital does not collapse into a single value partly due to non-linear neighbourhood interactions (Ioannides, 2003). Dendrinos and Rosser (1992) argue that urban evolution is driven by sudden change (discontinuities), rather than a gradual, continuous process. Changes in non-uniform comparative advantages trigger smooth or discontinuous changes in city size (Dendrinos and Rosser, 1992). We argue that clustered city size distributions offer a snapshot of a dynamic process in which size-dependent growth manifests in cities clustering into size classes, due to multi-level entrainment.

This analysis has identified departures from Gibrat’s law, indicating variable growth rates at different scales. Growth is not independent of size, as smaller cities grow faster and larger cities grow more slowly—implying, simply, that scale matters. The analysis has also determined that the city size distributions for the south-eastern region of the US are discontinuous, with cities clustering into distinct size classes, as exogenous, but continuous, changes in external transport costs and technological innovation can manifest in a discontinuous pattern of urban growth (Dendrinos and Rosser, 1992). In conclusion, size-dependent city growth explains the emergence of discrete size classes in city size distributions (Garmestani et al., 2005).

This article is an attempt to advance our understanding of the organisation of complex systems. We have attempted to achieve this goal via application of tenets derived from ecology and now applied in a cross-disciplinary manner to questions typically dealt with in economics and geography. Specifically, our goal was to advance our understanding of self-organisation in complex systems, with the focus of the research on determination of the existence of and characterisation of discrete size classes in the rank–size function of variables in a regional urban system.

References


