Chin. Geogra. Sci. 2016 Vol. 26 No. 3 pp. 377–388
doi: 10.1007/s11769-015-0750-5

Springer Science Press
www.springerlink.com/content/1002-0063

From Sequential to Parallel Growth of Cities: Theory and Evidence from Canada

SHENG Kerong[1], FAN Jie[2, 3], SUN Wei[2, 3], MA Hailong[4]

(1. *Shandong University of Technology*, *Zibo* 255012, *China*; 2. *Institute of Geographic Sciences and Natural Resources Research*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*; 3. *Key Laboratory of Regional Sustainable Development Modeling*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*; 4. *Beifang University of Nationality*, *Yinchuan* 750021, *China*)

Abstract: This paper examines city growth patterns and the corresponding city size distribution evolution over long periods of time using a simple New Economic Geography (NEG) model and urban population data from Canada. The main findings are twofold. First, there is a transition from sequential to parallel growth of cities over long periods of time: city growth shows a sequential mode in the stage of rapid urbanization, i.e., the cities with the best development conditions will take the lead in growth, after which the cities with higher ranks will become the fastest-growing cities; in the late stage of urbanization, city growth converges according to Gibrat′s law, and exhibits a parallel growth pattern. Second, city size distribution is found to have persistent structural characteristics: the city system is self-organized into multiple discrete size groups; city growth shows club convergence characteristics, and the cities with similar development conditions eventually converge to a similar size. The results will not only enhance our understanding of urbanization process, but will also provide a timely and clear policy reference for promoting the healthy urbanization of developing countries.

Keywords: sequential city growth; Gibrta′s law; finite mixture model; convergence club; Canada

Citation: Sheng Kerong, Fan Jie, Sun Wei, Ma Hailong, 2016. From sequential to parallel growth of cities: theory and evidence from Canada. *Chinese Geographical Science*, 26(3): 377–388. doi: 10.1007/s11769-015-0750-5

1 Introduction

The paper investigates the city growth process that occurs as countries develop economically. There are two related aspects to this issue: the first concerns how cities of different size grow relative to each other, i.e., is it sequential where the initially large cities are the first to grow fastest, followed by medium and then small cities further down in the urban hierarchy, or is it parallel where cities of different size grow at similar rates, and the second concerns how city size distribution evolves and which theoretical distribution provides the best approximation. Understanding city growth dynamics is crucial to advancing our understanding of urbanization process and formulating effective policies for developing countries that face rapid urbanization (Henderson and Venables, 2009).

A large body of literature has been developed to address the two issues. Regarding the first aspect, most studies suggest that the relative size and rank of cities remain stable over time, consistent with the proportionate effect of Gibrat′s law. When studying city growth in France (1876–1990) and Japan (1925–1985), Eaton and Eckstein (1997) found that the large cities maintained their ranking over the entire reference time, which means that city growth is parallel, rather than divergent or convergent. Other studies that found supports for parallel growth are Dobkins and Ioannides (2001), Black and Henderson (2003), Sharma (2003), Resende (2004), Schaffar and Dimou (2012). But another strand of lit-

Received date: 2014-08-26; accepted date: 2014-11-21
Foundation item: Under the auspices of Key Program of Chinese Academy of Sciences (No. KZZD-EW-06-01)
Corresponding author: SUN Wei. E-mail: sunw@igsnrr.ac.cn

erature has shown that city growth presents sequential characteristics (Anderson and Ge, 2005; Henderson and Venables, 2009; Sheng and Sun, 2013). Empirically, Cuberes (2011) found that the average rank of the fastest-growing cities in a large set of countries tends to increase over time, indicating cities historically have follows a sequential growth pattern. El-Shakhs (1972), Wheaton and Shishido (1981), and Junius (1999) reported that urban concentration first increases then decreases across countries as income rises. The literature delivers contradictory evidence. Under what conditions is city growth sequential, and under what conditions parallel? Is there a relationship between sequential and parallel growth patterns? There is an urgent need for a consistency framework to reconcile these two streams of literature.

Turing to the second aspect, the dominating view is that the Pareto distribution fits best, which states that there is an inverse linear relationship between the logarithmic size of a city and its logarithmic rank. A special case of the rank-size distribution with the Pareto exponent q equal to one is known as Zipf's law (Zipf, 1949). Explanation for Zipf's Law has revolved around two lines: one economic (Brakman *et al.*, 2001) and one defined by random process (Gabaix, 1999a; 1999b). However, frequent departures from the Zipf's law have also been found (Rosen and Resnick, 1980). Among others, Soo (2005) found that only one-third of the sampled countries had a quadratic term significantly close to zero, indicating Zipf's law is rejected far more often than would be expected based on random chance. Some researchers have also proposed other distribution functions to fit city systems, i.e., double Pareto lognormal distribution function (Reed, 2002), lognormal distribution (Anderson and Ge, 2005), and piecewise functions (Garmenstani *et al.*, 2005; Garmenstani *et al.*, 2008). While city growth patterns are closely related to city size distribution, together constituting different aspects of urban system evolution in the urbanization process, there is little research explicitly building organic connection between them. It is needed to strengthen the study in this field to improve our understanding of urban system.

This paper attempts to reconcile the sequential with parallel city growth theories to enhance our understanding of the multi-dimensional aspects of city growth dynamics. The paper contributes to the debate on city growth patterns on three dimensions. First, this paper present a stylized fact that cities grow in sequential order until urbanization enters the late stage of development, when city growth converges to parallel behavior consistent with Gibrat′s law, i.e., there is a transition from sequential towards parallel patterns of city growth over long periods of time. Second, this paper investigates the linkages between the dynamic behaviors of city growth with city size distribution evolution, and examines the effect that a country′s urbanization process has on its urban hierarchy. Third, this paper relates spatial heterogeneity with the structural properties and processes of city size distributions, and introduces finite mixture models to investigate the size clusters in urban systems.

2 Model

2.1 Setup

The model builds on the Footloose Capital (FC) model, which is the Martin and Rogers (1995) adaption of Krugman (1991). The great merit of the FC model is its ability to deal with exogenous asymmetries such as market size and asymmetric production costs. Three extensions are made here: 1) the model is combined with the assumption that the upper-tier utility is quasi-linear rather than Cobb-Douglas drawing on a widely used specification (Pflüger, 2004); 2) negative feedbacks, or congestion effects, is incorporated to allow for the viability of small cities; 3) spatial inhomogeneities, which are emphasized by traditional economic geography, are taken into account to analyze the effects of city specific comparative advantages on city growth. The model can be used to study the evolution of urban system, which is composed of large and small cities.

The country is composed of four regions, denoted by region 1, region 2, region 3 and region 4, two factors of production, labor (L) and capital (K), and two sectors, agriculture (A) and manufacturing (X). The four regions are all one-dimensional bounded location spaces, along which lies land of homogeneous quality. The aggregate labor force is assumed to be one unit and is partitioned between the four regions, of which the size in region 1 is 1/2, and those in regions 2, 3 and 4 are respectively $1/6+2\varepsilon$, $1/6-\varepsilon$ and $1/6-\varepsilon$, where the exogenous variable ε ($\varepsilon \to 0$) is used to characterize the advantage in market potential of region 2 relative to regions 3 and 4. In other words, region 1 has the largest market potential, while regions 3 and 4 are most inaccessible. Capital accumulates over time: the aggregate capital at time t is $K(t)$, and capital growth rate is $\mathrm{d}K(t)/\mathrm{d}t = g(t)$. For simplicity, it is assumed that the capital is evenly distributed among labor, i.e., the capital in regions 1, 2, 3 and 4 are 1/2, 1/6 + 2ε, 1/6–ε and 1/6–ε, respectively. In the long run, capital is mobile between regions, while labor is only intersectorally mobile.

Individuals have identical preferences given by:

$$U = \gamma \ln C_{\mathrm{X}} + C_{\mathrm{A}}, \quad C_{\mathrm{X}} = \left(\sum_{j=1}^{2} \sum_{i=1}^{n_j} c_{ji}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)},$$

$$\gamma > 0, \quad \sigma > 1 \tag{1}$$

where C_{X} is the manufacturing aggregate; c_{ji} is the quantity consumed of variety i produced in region j; n_j ($n_j = s_jK$) is the number of varieties in region j; C_{A} is the consumption of agricultural good; σ is the elasticity of substitution between manufacturing varieties.

Regions are assumed to have identical technology and trade cost. Agriculture production is not mobile and is characterized by perfect competition and zero transport costs. The agricultural good is chosen as numéraire. It is further supposed that production of one unit of agricultural good require one unit of labor such that the price of agricultural good and the wage of agricultural labor are normalized to 1. The existence of a local, immobile labor force is an important spreading factor because it ensures that there is always a positive demand in each region.

The monopolistic competitive sector X produces differentiated goods with capital as the only fixed input and labors the only variable input. For simplicity, the cost function $C(q)$ involves one unit of capital and a constant marginal cost a_m, given by:

$$C(q) = \Pi + a_m q \tag{2}$$

where Π is the nominal return on capital, q is the producer's output. The equilibrium price for variety i is $p_i \equiv p = a_m\sigma/(\sigma - 1)$. Choosing units and letting $a_m = (\sigma - 1)/\sigma$, the price of manufactured goods is also normalized to unity. The trade of manufactured goods incurs no cost within the same region while inhibited by iceberg costs between regions. Without loss of generality, it is assumed that cities are set in a hypothetical space where any two cities are equidistant. This implies between any pair of regions only $1/\tau$ of a unit of a variety arrives for consumption and that the price of the imported variety rises by τ times, where $\tau > 1$ is a parameter to characterize transport costs.

2.2 City formation

We assume that at each point of time, manufacturing industries are concentrated in the geographical center of each region, called 'cities'. Assume also the proportion of capital in cities 1, 2, 3 and 4 is s_1, s_2, s_3 and s_4, respectively. Then the numbers of firms and thus varieties are s_1K, s_2K, s_3K and s_4K in these four cities respectively, and from Equation (2) the population sizes N_1, N_2, N_3 and N_4 are s_1Ka_m, s_2Ka_m, s_3Ka_m and s_4Ka_m. Let the aggregate demand in region m of region j be d_{jm}. From Equation (1), utility maximization yields the demand functions:

$$d_{jj} = \gamma \frac{s_j}{P_j^{1-\sigma}}, \; j = 1,\, 2,\, 3,\, 4 \tag{3a}$$

$$d_{jm} = \gamma \frac{s_m \theta}{\tau P_m^{1-\sigma}}, \; j = 1,\, 2,\, 3,\, 4;\; m = 1,\, 2,\, 3,\, 4;\; j \neq m \tag{3b}$$

where, P_j and P_m are the consumer price indexes in regions j and m respectively.

$$P_j = \left(s_j K + \sum_{m=1, m\neq j}^{4} \left(s_m K \theta \right) \right)^{1/(1-\sigma)} \tag{4}$$

where, $\theta = \tau^{1-\sigma}$ $(0 \le \theta \le 1)$ is interregional trade freeness, which increases with the decline in τ (lower transportation costs) or σ (more preferred to diversified products). When $\tau = 1$ and $\theta = 1$, interregional trade involves no cost; and when $\tau \to \infty$ and $\theta = 0$, interregional trade cost is infinite.

Two factors affecting city growth are emphasized. First, there are differences in the comparative advantages of natural conditions across cities. Many various causes have led to the localization of industry, but the chief causes have been physical conditions (Marshall, 1920; Ellison and Glaeser, 1999; Kim, 1999). Specifically, Ellison and Glaeser (1999) found that the percentage of agglomeration that is predicted by the natural advantage proxies is roughly 20%.

Second, the expansion of city size leads to an increase in congestion costs (Brakman *et al*, 1996). They assert that, compared with the spreading forces originated from immobile consumption market and increase in agriculture product price, congestion effect constitutes a more important spreading factor. This constitutes a spreading force for manufacturing activities. A note of caution is that the transformation of urban structure from monocentric to polycentric pattern to some degree eases the pressure of rising rents (Fujita and Ogawa, 1982; Anas and Small, 1998). And the introduction of communication technologies also increases the incentive to agglomerate (Desmet and Rossi-Hansberg, 2009). Thus, net congestion may have resumed decreasing during the past several decades.

The focus is on analyzing the consequence of city-specific comparative advantages and congestions rather than their origins. Following Brakman *et al*. (1996), this paper capture the essence by simply assuming $\Pi_j = \pi_j/[\beta_i g(s_j, K)]$, where π_j is the real return on capital, $\beta_i > 0$ and $g(s_j, K)$ represent the comparative advantages and congestion-related negative feedbacks in city j, respectively. Specifically, suppose that $g(s_j, K) = \varphi/[(1+s_j)K]$, where $\varphi(\varphi<(1+s)K)$ is a constant. Combining the conditions of market clearing and zero profit, the real return on capital is given:

$$\pi_j = \beta_j \phi \left(d_{jj} + \sum_{m=1, m\neq j}^{4} \left(\tau d_{jm}\right) \right) / \left[\sigma\left(1+s_j\right)K\right], j = 1,\ 2,\ 3,\ 4 \tag{5}$$

Under FC model context, the profits of capital are all repatriated to the regions where their owners reside, thus capital flows in the four cities will depend on the real returns they can offer. From Equation (5), home market effect and comparative advantages are the agglomeration forces, while the immobile agricultural labor and congestion effect constitute the spreading forces. In our model, it is the interaction of these two opposing forces that leads to capital flow and the evolution of city size. A long-run equilibrium is only reached when capital rewards are equalized.

2.3 City growth

It is well-known that this type of model can not be solved analytically due to its strong nonlinear nature. Nevertheless, basic understanding of city growth and the consequent city size distribution evolution can be obtained by means of numerical simulations. In the simulations below, this paper focuses on the parameters θ, ε, and β.

Trade freeness θ increases over time. In agricultural society, the transportation costs are high, and the small manufacturing sector produces close substitutes, which implies that trade freeness remains high and changes little. During industrialization, the emergence of modern transportation leads to an unprecedented decline in transportation costs (Bairoch, 1988). Meanwhile, large-scale machine-based production supplants small-scale craft production, and eventually dominates manufacturing sector. The spectacular decreases in transportation costs and increasing importance of economies of scale imply a significant increase in θ. In post-industrialization, transport costs remain low with little change, and as before the industrial sector is characterized by differentiated products and increasing returns to scale, which implies trade freedom tends to be stable. This paper stimulates the historical change of trade freeness by choosing a logistic growth function of θ over time t. Specifically, it is assumed that $\theta = 0.80/[1+e^{-0.20(t-20)}]$, where t is the time variable.

Parameters β and ε remain constant over time. Without loss of generality, it is assumed that $\varepsilon = 0.008$. Parameter β varies across cities. To investigate the growth dynamics of cities with different development conditions, it is assumed that $\beta_1 = \beta_2 = \beta_3 = 1$, $\beta_4 = 0.85$. Thus, in our model, city 1 represents the city with the largest market access and the best development conditions, cities 2 and 3 represent the cities with similar development conditions, and city 4 represents the city with the poor development conditions in the fringe areas.

Figure 1 depicts the evolution of long-run equilibrium population shares in the four cities as trade freedom increases. From the dynamic equation of city growth $[dN_j(t)/dt]/N_j = [ds_j(t)/dt]/s_j + g(t)$, the growth rate of city population depends on the rate of change in population shares s_1, s_2, s_3 and s_4. It is apparent from Fig. 1 that, at the early stages of the urbanization process, city population shares are closely related to local market demands; with the increase of trade freeness, the landscape of urban systems undergoes tremendous changes. Our theory has four clear-cut and testable predictions for a given country.

First, cities grow in a sequential order. In the early stage of industrialization, the agglomeration force of economic activity outweighs the spreading force, and manufacturing firms are strongly encouraged to concentrate

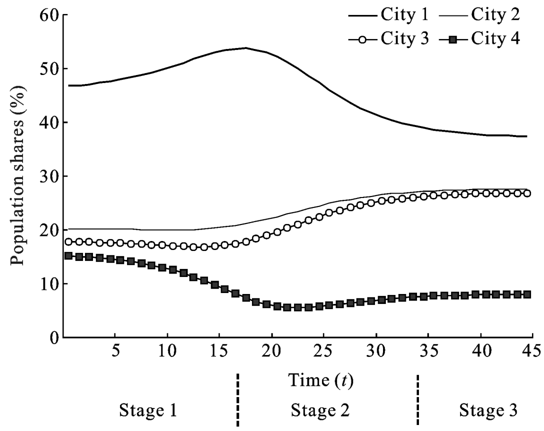


Fig. 1 Evolution of population shares in four cities

production in the location with the best development condition. As a result, city 1 becomes the first to grow fastest and eventually become the primate city in the country. With the deepening of the industrialization process, congestion effect starts to dominate the agglomeration force in the primate city. The growth rate of the primate city declines, and the cities with relatively better development conditions (city 2 and city 3) become the fastest-growing cities. Therefore, the model predicts that cities grow sequentially during the evolution process of urban systems. The result also implies that the size distribution of cities against their ranks (at least in the upper tail) should become more flatter with increase in the level of urbanization.

Second, city size distribution exhibits structure characteristics. Cities with similar development conditions (city 2 and city 3) tend to have similar population sizes, and form specific clusters in city size distribution, resulting in clusters in city size distribution. The empirical city size distributions have previously been approximated by single distribution function such as Pareto, lognormal, or double Pareto lognormal function, which implies a continuous pattern exists across multiple scales, and that common processes, or mechanisms propagate across a wide range of scales. Nevertheless, clustering structures in city size distribution implies that there are processes operating at distinct scales, and these processes create scale-specific patterns deviating from the general rank size rule or Zipf's law. Individual distribution function can not effectively capture the discontinuities or lumpy patterns in urban systems, and thus it is required to develop mixture distribution functions to fit city size distribution.

Third, city growth eventually converges to patterns consistent with Gibrat's law. In the late stage of urbanization development (stage 3), the forces of agglomeration and spreading are in balance. Cities with favorable development conditions have large population size and consequent great congestion costs, while cities with small market potential index have small population size and are less trouble by congestion effects. Eventually, the growth rates of these cities tends to be stable (s_1, s_2, s_3 and s_4 values are no longer changed), and cities of different size show growth patterns consistent with the law of proportionate effect. Thus, it is plausible to conclude that sequential pattern and parallel pattern are manifestations of city growth in different stages of urbanization process. In other words, when examining the dynamics of cities in the accelerated stage of urbanization, city growth should show a sequential pattern; when examining city growth in the final stage of urbanization, Gibrat-like behavior should be detected. The model implies that Gribat's law is valid for describing city growth as a long-run regularity.

Fourth, city growth exhibits behavior that is characterized via convergence clubs. As trade freeness rises, comparative advantages in regional development conditions outweigh home market effect to become the dominating factor influencing city growth, leading to great divergence in growth trajectories between city 4 and cities 2 and 3. On the other hand, sequential city growth implies that, with the deepening of regional development process, the size difference of city 1 from cities 2 and 3 will eventually be shrinking, and cities in the upper tail of the size distribution show a convergence trend. Thus, when urbanization reaches its final stage, these four cities are self-organized into two clubs, in which club 1 includes cities 1, 2 and 3, and club 2 includes only city 4. City growth can manifest multiple stable steady states via differential growth rates. Club convergence of city growth indicates that, structures in city size distribution, or discontinuity between size classes, is a stylized fact over the whole urbanization process.

3 Empirical Evidence

3.1 Data description

Canada is selected to conduct the empirical analysis due to its long record of city populations, its large land area, as well as its well-developed market economy system. The dataset covers a 115-year period from 1871 to 1986. The data are collected from two sources: information from 1871 to 1966 is obtained from the historical Canada Yearbook compiled by Statistics Canada (http://www66.statcan.gc.ca/acyb_000-eng.htm); and information from 1971 to 1986 is reported in the website created by Jan Lahmeyer (http://www.populstat.info/). The term city here is defined as urban area. According to Statistics Canada, an urban area is a human settlement with a population of at least 1000 and a population density of no less than 400/km^2. According to this definition, Canada′s urbanization rate rise from 19% in 1781 to 76% in 1986.

This paper focuses on the growth of cities in the upper tail of the distribution. A total of 112 cities with a population over 76 000 inhabitants in 1986 are selected as the complete sample. At the same time, a balanced panel including the 42 most populous cities is also constructed to explore the growth pattern of existing cities. A parallel study of both the balanced and complete samples is employed. Figure 2 describes the evolution of rank size curves for the complete and balanced samples in 1871–1986. It is apparent from Fig. 2a in the accelerated stage of urbanization, the rank size curve shifts upward, and becomes more flatter; Fig. 2b in the final stage of urbanization, especially in 1971–1986, the curves remains almost stable.

3.2 Sequential city growth

Relative growth rate is firstly used to analyze the sequence of city growth. Relative growth rate is defined as the ratio of the population growth of individual cities to the average growth of cities in the sample. The distributions of relative growth rates for the balanced and complete samples are skewed to the right in all five-year intervals except for 1976–1981 and 1981–1986 periods, indicating some of the cities grow much faster than the rest during the transition to the steady state. The evolution trend of relative growth rates of individual cities is further estimated using the following specification:

$$[\ln(N_{j,\,t+1}) - \ln(N_{j,\,t})] \;/\; [N^{-1}\Sigma(\ln(N_{j,\,t+1}) - \ln(N_{j,\,t}))] = a_j + b_j t + \varepsilon_{j,\,t} \quad (6)$$

where $N_{j,t}$ is the population size of city j at time t; $N_{j,t+1}$ is the size of the same city in the subsequent period; $\varepsilon_{j,t}$ is the residual. Table 1 reports the estimated coefficients for the balanced sample. It is apparent from Table 1 that, the relative growth rates of cities with low rankings (Montréal, Toronto, and Vancouver) decreases over time, while those with high rankings (London, Brampton, Kitchener, Oshawa, and Guelph) increases.

To facilitate the interpretation of the finding, Fig. 3 depicts the evolution of the average relative growth rates for different subsamples. As is evident in Fig. 3, in the early stage of urbanization, the most populous cities grow fastest (Fig. 3a), while in the late stage, the medium-sized cities become the fastest-growing ones (Fig. 3c). And it can also be found that, the growth rates of cities further down in the urban hierarchy remains relatively constant (Fig. 3d). This implies that, in the process of urbanization, the low ranked cities grow first, followed by a decline in growth rates, and then high-ranked cities become the fastest growing ones, which constitutes direct evidence of sequential city growth hypothesis.

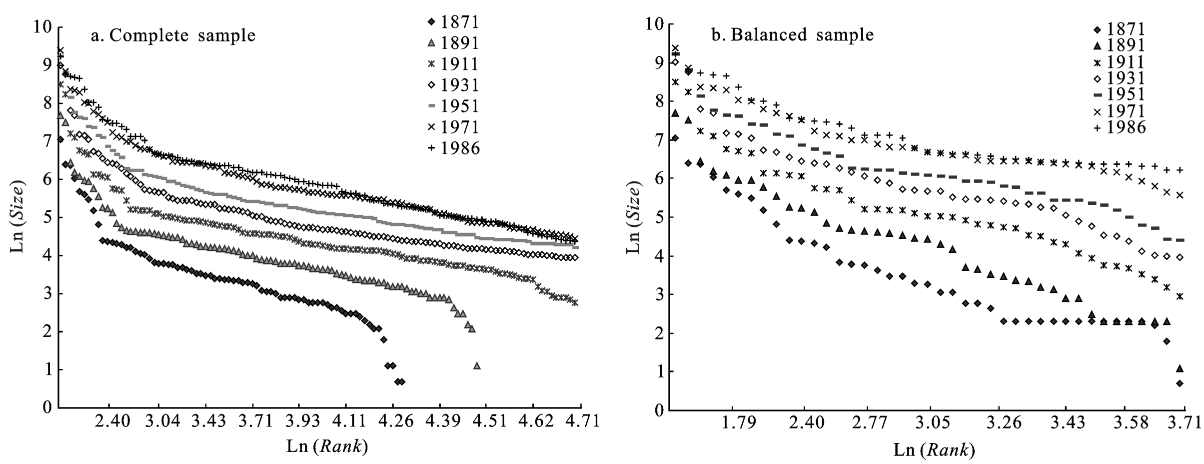


Fig. 2 Rank size plots for Canadian cities, 1871–1986

Table 1 Relative growth rates of cities in balanced sample

Rank	City	*t*	Rank	City	*t*	Rank	City	*t*
1	Montréal	–0.149*	15	Kitchener	0.097***	29	Dartmouth	–0.008
2	Calgary	0.160	16	Longueuil	0.145	30	Kamloops	0.106
3	Toronto	–0.135***	17	Oshawa	0.133***	31	Chicoutimi	0.077
4	Winnipeg	–0.145	18	Saint Catharines	0.026	32	Peterborough	–0.014
5	Edmonton	0.074	19	Halifax	–0.033	33	Verdun	–0.335***
6	Vancouver	–0.208*	20	Sudbury	–0.205*	34	Lethbridge	0.113
7	Hamilton	–0.030	21	Sault Sainte Marie	–0.034	35	Hull	–0.112
8	Ottawa	–0.019	22	Guelph	0.137***	36	Waterloo	0.216***
9	London	0.096***	23	Saint John	–0.518	37	Jonquière	–0.071
10	Windsor	–0.086	24	Brantford	0.105	38	Moncton	–0.136*
11	Brampton	0.681***	25	Sherbrooke	–0.089	39	Kingston	–0.051
12	Saskatoon	0.113	26	Niagara Falls	0.002	40	North Bay	–0.105
13	Regina	0.013	27	Saint-Laurent	–0.052	41	Trois-Rivières	–0.096**
14	Québec	–0.091*	28	Victoria	–0.032	42	Sarnia	0.0845

Note: *, **, and ***denote significance at 0.1, 0.05, and 0.01 level, respectively

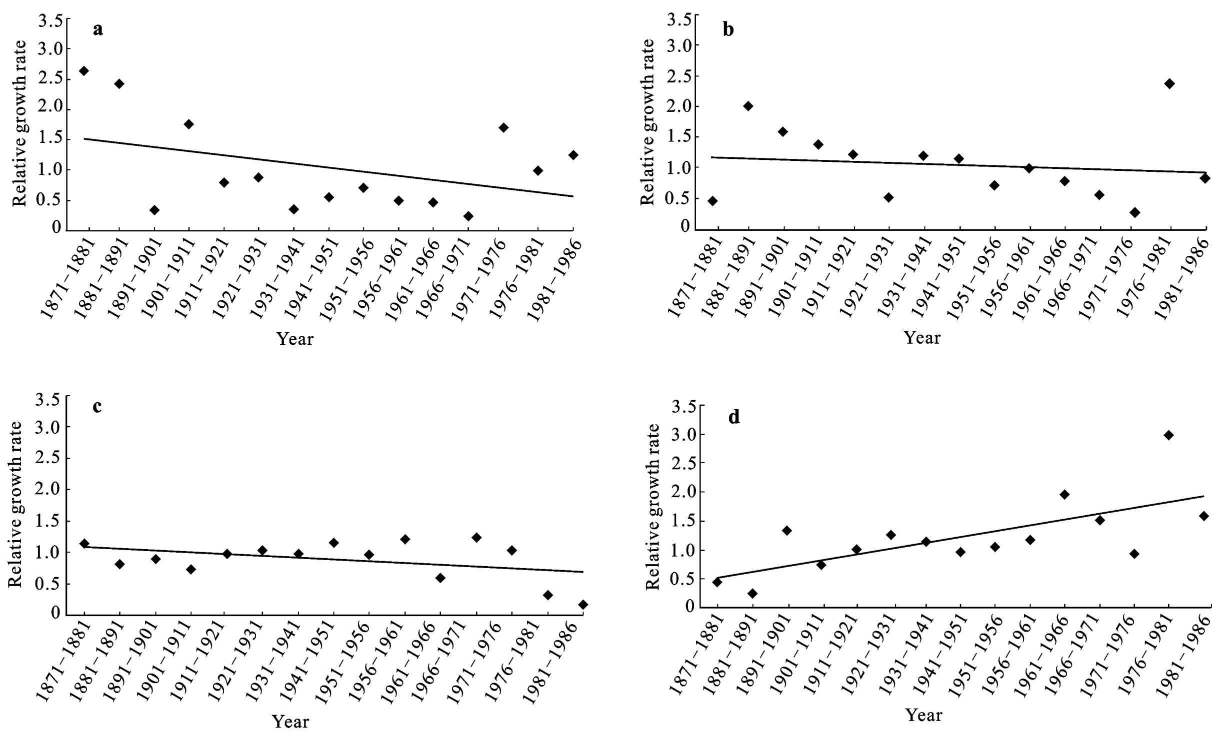


Fig. 3 Evolution of average of relative growth rates, panel a includes Montréal, Calgary, Toronto, Winnipeg; panel b includes Edmonton, Vancouver, Hamilton, Ottawa, and London; panel c includes Windsor, Brampton, Saskatoon, Regina, Québec, Kitchener, Longueuil, Oshawa, Saint Catharines, Halifax; panel d includes Sudbury, Sault Sainte Marie, Guelph, Saint John, Brantford, Sherbrooke, Niagara Falls, Saint-Laurent, Victoria, Dartmouth, Kamloops, Chicoutimi, Peterborough, and Verdun

The evolution of Pareto exponents confirms the pattern revealed in the above analysis. The log-linear equation is estimated: $\ln N_j = \ln A - q\ln(R_j)$, where $\ln A$ is the constant, N_j and R_j are the population and rank of city j. The large value of the exponent implies the high degree of convergence of population in large cities. For every five-year data, the exponents for the complete and balanced sample are estimated respectively to study how the population concentration in large cities changes with the growth of urban population (Fig. 4). In Fig. 4, Pareto

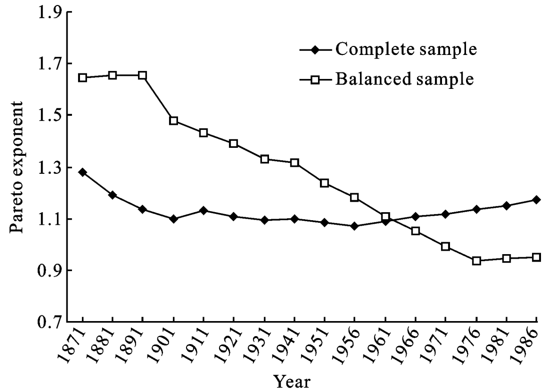


Fig. 4 Evolution of Pareto Exponent, 1871–1986

exponents for the complete sample increases slowly from 1.110 in 1911 to 1.172 in 1986, while those for the 42 most populous cities declines dramatically from 1.647 in 1871 to 0.952 in 1986. This pattern indicates that, while concentration of population towards the upper tail of the distribution remains, populations are no longer as convergent on large cities after 1871, and more populations concentrate on medium size cities (their ranks are still within the top 42).

3.3 Structure in city size distribution

Finite mixture model (FMM) is used to capture the structural characteristics of city size distribution (McLachlan and Basford, 1988). Mixtures of normal distribution are flexible to accommodate various shapes of continuous distributions and able to capture skewed and multimodal characteristics of economic data. Thus each component in the FMM is assumed to take the same form of normal distribution density. The FMM for fitting city size distribution with c components is thus given by:

$$f(\ln(N)\,|\,\ln(Rank);\Theta)=\sum_{j=1}^{C}\pi_h\frac{1}{\sqrt{2\pi}\sigma_h}\exp\left[-\frac{\ln(N)-\left(a_h+q_h\ln(Rank)\right)}{2\sigma_h^2}\right] \quad (7)$$

where q_h is the coefficient associated with ln(*Rank*) in component h (Pareto exponent in component h), π_h is the weight. Table 2 reports the analysis results for the complete sample by decade.

The size distribution of cities in every decade is composed of 2–3 components, with the estimated coefficients q_h all significant at the 0.01 level. It is worthy to note that, there are significant differences in the estimated q_h between different components. For instance, in 1881 the estimated coefficient for the second component q_2 is less than 16 standard errors away from the estimate for the first component q_1; in 1921 the estimated q_2 exceeds q_1 by 16 times the standard errors, but is less than q_3 by 2 times the standard errors; in 1981 the estimated q_2 is more than 5 standard errors away from q_1, but less than 10 standard errors from q_3. The statistical difference between the estimated coefficients for individual components, not only indicates that the overall Pareto exponent does not capture evidence of processes affecting city size at a finer scale of analysis, but also that Canadian city system exhibits persistent structure characteristics.

Table 2 Regression results for finite mixture models

Year	Component 1		Component 2		Component 3	
	q_1	π_1	q_2	π_2	q_3	π_3
1871	1.379	0.206	1.223	0.794		
1881	1.439	0.171	1.090	0.829		
1891	1.415	0.099	1.001	0.261	1.084	0.640
1901	1.128	0.568	1.078	0.432		
1911	1.048	0.736	1.291	0.175	1.228	0.089
1921	1.040	0.649	1.168	0.305	1.209	0.045
1931	1.078	0.741	1.198	0.259		
1941	1.062	0.639	1.188	0.361		
1951	1.012	0.537	1.149	0.463		
1961	1.081	0.497	1.157	0.425	1.139	0.078
1971	1.105	0.590	1.041	0.083	1.198	0.327
1981	1.089	0.624	1.025	0.126	1.269	0.250

Note: all coefficients are significant at 0.01 level

3.4 Convergence towards parallel growth

Parallel growth means cities keep their initial hierarchical ranks during the growth process. Firstly, two methods, Rank Mobility Index and Markov chain, are used to explore the trends of intra-distributional movements of cities. The Rank Mobility Index is formulated as $RMI = (Rank_{t+1} - Rank_t)/(Rank_{t+1} + Rank_t)$, where $Rank_t$ represents a city′s rank at time t. The value of RMI varies on (–1, 1). When a city moves up (or down) in the urban hierarchy, its RMI is negative (or positive). When the rank of a city does not change, its RMI is zero. The RMI for every city based on its decadal rank change is estimated. Figure 5 illustrates the distribution of RMI of all cities using box plots. The range of RMIs between 25% and 75% percentiles become very small in recent decades, indicating the city system has become increasingly stable.

Markov chain is then used to estimate the intra-distributional dynamics of the individual cities over the periods of 1911–1961 and 1971–1986, respectively. The use of Markov chain techniques requires the discretization of the distribution by assigning each city to one of a predetermined number of groups based on its relative size (Quah, 1993; Eaton and Eckstein, 1997; Anderson and Ge, 2005). Following Eaton and Eckstein (1997), the chosen intervals were based on 0.30, 0.50, 0.75, 1 and 2 times the sample mean at each time. Each transition probability, m_{ij}, in the transition matrix is estimated by maximum likelihood, i.e.

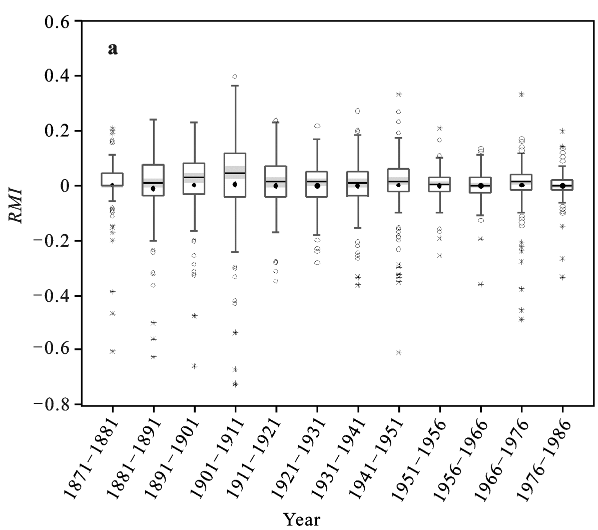
$$m_{ij} = \left(\sum_{t=1}^{T-1} n_{it,jt+1}\right) / \left(\sum_{t=1}^{T-1} n_{it}\right) \quad (8)$$

where $n_{it,jt+1}$ denotes the number of cities moving from the group i in the five-year interval to group j in the subsequent five-year interval, n_{it} is the number of cities in group i in the period t. Panels A and B in Table 3 report the average transition matrix, each element of which is the probability that a city initially in the cell corresponding to its column will join the cell corresponding to the row in the subsequent period. The most striking result is that where in early stage of urbanization the all off-diagonal elements are significantly different from zero, the pattern changes in the late stage of urbanization when almost all off-diagonal elements are not significantly different from zero. This movement can be seen as evidence of a convergence tendency of the distribution toward Gibrat-like pattern of growth through time, suggesting all cities eventually grow at the same rates stating at different levels.

Secondly, the following logarithmic specification of Gibrat′s law is estimated to examine the relation between city growth rates and their initial sizes:

$$\ln(N_{j,\,t+1}) - \ln(N_{j,\,t}) = a_t + b_t \ln(N_{j,\,t}) + \varepsilon_{j,\,t} \quad (9)$$

where the estimated coefficient b_t is the key indicator: if $b_t = 0$, then the growth rate and initial size are independently distributed and Gibrat′s law is in operation; by contrast, if $b_t < 0$, smaller cities grow at a systematically higher rate than do their larger counterparts, while the opposite is the case if $b_t > 0$. The estimate results are shown in Table 4. The results for the balanced sample reveal that, prior to 1971, the estimated b_t is significantly negative except for the period of 1931–1941, but after 1971, it is not statistically different from zero. The



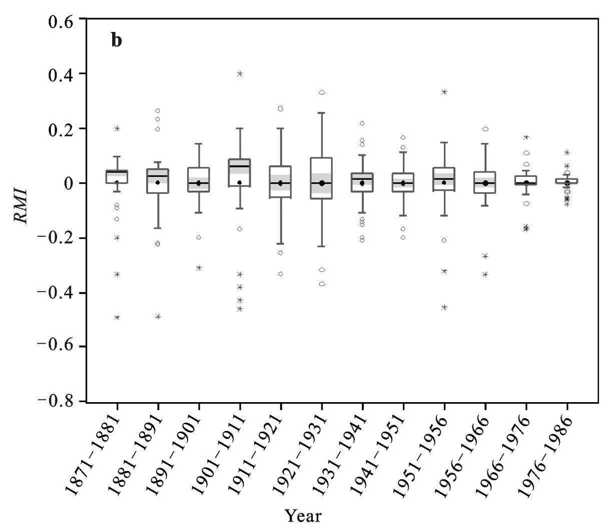


Fig. 5 Distributions of Rank Mobility Index (RMI) of cities: a) the complete sample; b) the balanced sample

Table 3 Average decade transition matrix for complete sample

Panel A: Early stage of urbanization (1911–1961)						
	N<0.3u	0.3u<N<0.5u	0.5u<N<0.75u	0.75u<N<u	u<N<2u	N>2u
N<0.3u	0.870	0.130	0	0	0	0
0.3u<N<0.5u	0.141	0.718	0.127	0.014	0	0
0.5u<N<0.75u	0	0.188	0.687	0.125	0	0
0.75u<N<u	0	0	0.222	0.500	0.278	0
u<N<2u	0	0	0	0.091	0.818	0.091
N>2u	0	0	0	0	0.133	0.867
Panel B: Late stage of urbanization (1971–1986)						
	N<0.3u	0.3u<N<0.5u	0.5u<N<0.75u	0.75u<N<u	u<N<2u	N>2u
N<0.3u	0.980	0.010	0.010	0	0	0
0.3u<N<0.5u	0.084	0.874	0.032	0.011	0	0
0.5u<N<0.75u	0	0.026	0.947	0.026	0	0
0.75u<N<u	0	0	0.235	0.735	0.029	0
u<N<2u	0	0	0	0.133	0.833	0.033
N>2u	0	0	0	0	0.027	0.973

Notes: N denotes population size of a particular city; u is sample mean

Table 4 Regression results of urban growth on their initial sizes

Time interval	Balanced sample	Complete sample	Time interval	Balanced sample	Complete sample
1871–1881	–0.267***	–0.210**	1951–1956	–0.053***	–0.020*
1881–1891	–0.031	–0.096**	1956–1961	–0.085**	0.003
1891–1901	–0.109**	–0.103**	1961–1966	–0.078**	–0.008
1901–1911	–0.157**	–0.123***	1966–1971	–0.097***	0.000
1911–1921	–0.031	–0.054**	1971–1976	–0.063	0.009
1921–1931	–0.055**	–0.041**	1976–1981	0.008	0.012
1931–1941	–0.008	–0.003	1981–1986	0.015	0.018
1941–1951	–0.082***	–0.023*			

Note: *, **, and ***denote significance at 0.1, 0.05, and 0.01 level, respectively

results for the complete sample show similar pattern with b_t significantly negative prior to 1956 (except for 1931–1941 period) and not significant from 1971 onwards. It indicates that, the growth of Canadian cities display a consistent convergence trend, until the final stage of urbanization when they converges to the parallel patterns consistent with Gibrat′s law. The puzzling debate about cities growth patterns is thus recomposed: although sequential patterns can be detected when considering the overall evolution of cities in the process of urbanization, Gibrtat′s law is an accurate description of the city growth in the long run.

3.5 Club convergence in city growth

In order to analyze the transitional behavior of cities and detect the convergence club, the log t test developed by Phillips and Sul (2007) is applied. The test focuses on the ultimate convergence of city sizes allowing for transitional divergence and heterogeneity in convergence speed across panel members. The following regression is estimated:

$$\log\left(H_0 / H_t\right) - 2\log\left(\log\left(t\right)\right) = \beta_0 + \beta_1 \log\left(t\right) + \varepsilon_t \qquad (10)$$

where H_t is the transition distance in period t, defined by

$$H_t = M^{-1}\sum_{i=1}^{M}\left(h_{jt} - 1\right)^2,$$

$$h_{jt} = \log\left(N_{jt}\right) / \left(M^{-1}\sum_{j=1}^{M}\log\left(N_{jt}\right)\right) \qquad (11)$$

where h_{jt} is the relative transition path tracing out the trajectory of city j to the average, N_{jt} is the population size of city j at time t, and M denotes the number of cities.

The complete sample is used to address this issue. Because the test is based on balanced panel data, a little data transformation is carried out by assigning a population of 1 to the cities that did not exist in the very early years. The transformation means that theses cities have a zero log-population when they did not exist. Phillips and Sul have recommended that, the initial observation in Equation (10) should be $[rT]$ with $r = 0.3$, so that the log t test regressions are based on time series data in which the first 3% of the data is discarded. Considering the structure of our sample, the analysis is conducted using the city sample after 1911, i.e., H_0 is the transition distance in the base year 1911.

The null hypothesis of convergence is rejected at 5% level if the t statistic of the coefficient on log (t) term in Equation (10) is less than –1.65. If convergence is rejected for the overall sample, the clustering mechanism test procedure is applied to determine if there exists club convergence. Following the above procedure, 4 convergence clubs are finally identified. Club 1 consists of 39 large cities, such as Montréal, Calgary, Toronto, Winnipeg, and Edmonton; club 2 includes 39 cities, such as Sherbrooke, Victoria, Peterborough, Verdun and Hull; club 3 is compose of 32 cities, such as Outremont, Westmount, Lindsay, Pembroke and Lauzon; and club 4 includes only 2 cities of Trail and Yarmouth. Figure 6 shows the relative transition paths of the 4 convergence clubs from 1871–1986. Therefore, the club convergence hypothesis of city growth passes the test.

4 Conclusions

The paper′s findings can be summarized as follows. The first is that cities grow in a sequential order in the early and accelerated stages of urbanization process, with cities with the most favorable conditions the first to grow

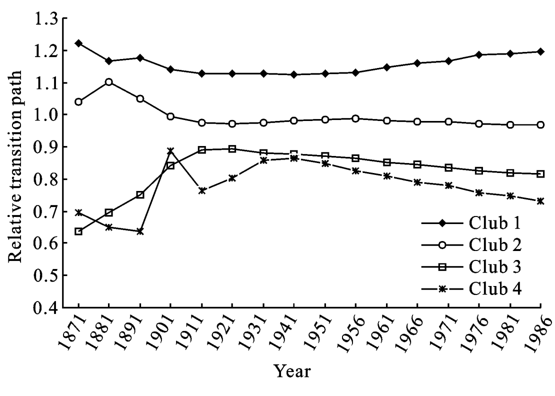


Fig. 6 Average relative growth paths of the four size clubs

rapidly. Second, the growth of cities (at least in the upper tail of the distribution) converges to parallel patterns consistent with Gibrat′s law of proportionate effect in the final stage of urbanization. Third, cities tend to self-organize into discrete size classes, leading to persistent structures, or discontinuities, in cities size distribution. Finally, the growth of cities exhibits behavior that is best characterized via convergence clubs, in which cities with similar developmental conditions have similar development paths and converge to similar stable sizes, resulting in multiple steady states.

The findings give evidences that reconcile the sequential city growth theories with parallel growth theories, showing that there exists a transition from sequential towards parallel growth patterns and Gibrat′s law is a reflection of a steady state condition of urban growth. Additionally, the two stylized facts of structures in size distribution and club convergence of city growth presented in this paper would presumable be valuable inputs to develop new theories of urban growth or extending existing ones to improve their goodness of fit. These findings are also valuable to policy makers, especially in countries that are urbanizing rapidly. For example, our study suggests that large cities should be supported in priority in regional development, and then policy focus should be shifted to the lower-ranked medium-sized cities. And the clustering structure and inverted U-shaped evolutionary path of city size distribution should also be fully considered. Policy makers can thus make informed decisions on where and when to invest in urban infrastructure.

References

Anas A, Small K A, 1998. Urban spatial structure. *Journal of Economic Literature*, 36(3): 1426–1464.

Anderson G, Ge Y, 2005. The size distribution of Chinese cities. *Regional Science and Urban Economics*, 35(6): 756–776. doi: 10.1016/j.regsciurbeco.2005.01.003

Bairoch P, 1988. *Cities and Economic Development*: *From the Dawn of History to the Present*. Chicago: University of Chicago Press.

Black D, Henderson J V, 2003. Urban evolution in the USA. *Journal of Economic Geography*, 3(4): 343–372. doi: 10.1093/jeg/lbg017

Brakman S, Garretsen H, Marrewijk C V, 2001. *An Introduction to Geographical Economics*: *Trade*, *Location and Growth*. Cambridge: Cambridge University Press.

Brakman S, Garretsen H, Marrewijk C V *et al*., 1996. Negaive feedbacks in the economy and industrial location. *Journal of Regional Science*, 36(4): 631–651. doi: 10.1111/j.1467-9787.1996.tb01122.x

Cuberes D, 2011. Sequential city growth: Empirical evidence. *Journal of Urban Economics*, 69(2): 229–239. doi: 10.1016/j.jue.2010.10.002

Desmet K, R Rossi-Hansberg E, 2009. Spatial growth and industry age. *Journal of Economic Theory*, 144(6): 2477–2502. doi: 10.1016/j.jet.2008.11.001

Dobkins L H, Ioannides Y M, 2001. Spatial interactions among US cities. *Regional Science and Urban Economics*, 31(6): 701–731. doi: 10.1016/S0166-0462(01)00067-9

Eaton J, Eckstein Z, 1997. Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics*, 27(4): 443–474. doi: 10.1016/S0166-0462(97)80005-1

Ellison G, Glaeser E, 1999. The geographical concentration of an industry: Does natural advantage explain agglomeration? *American Economic Association Papers and Proceedings*, 89(2): 311–316. doi: 10.1257/aer.89.2.311

El-Shakhs S, 1972. Development, primacy, and systems of cities. *Journal of Developing Areas*, 7(1): 11–36.

Fujita M, Ogawa H, 1982. Multiple equilibria and structural transition of nonmonocentric urban configurations. *Regional Science and Urban Economics*, 12(2): 161–196. doi: 10.1016/0166-0462(82)90031-X

Gabaix X, 1999a, Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, 114(3): 739–767. doi: 10.1162/003355399556133

Gabaix X, 1999b, Zipf's law and the growth of cities. *American Economic Review*, 89(2): 129–132. doi: 10.1257/aer.89.2.129

Garmestani A S, Allen C R, Bessey K M, 2005. Time-series analysis of clusters in city size distributions. *Urban Studies*, 42(9): 1507–1515. doi: 10.1080/00420980500185314

Garmestani A S, Allen C R, Gallagher C M, 2008. Power laws, discontinuities and regional city size distributions. *Journal of Economic Behavior & Organization*, 68(1): 209–216. doi: 10.1016/j.jebo.2008.03.011

Henderson J V, Venables A J, 2009. The dynamics of city formation. *Review of Economic Dynamics*, 12(2): 233–254. doi: 10.1016/j.red.2008.06.003

Junius K, 1999. Primacy and economic development: Bell-shaped or parallel growth of cities? *Journal of Economic Development*, 24(1): 1–22.

Kim S, 1999. Regions, resources and economic geography: Sources of US regional comparative advantage, 1880–1987. *Regional Science and Urban Economics*, 29: 1–32.

Krugman P, 1991. Increasing returns and economic geography. *Journal of Political Economy*, 99 (3): 483–499. doi: 10.3386/w3275

Martin P, Rogers C A, 1994. Industrial location and public infrastructure. *Journal of International Economics*, 39(3–4): 335–351. doi: 10.1016/0022-1996(95)01376-6

Marshall A, 1920. *Principles of Economics*. London: MacMillan Publishers.

McLachlan G J, Basford K E, 1988. *Mixture Models*: *Inference and Application to Clustering*. New York: Marcel Dekker.

Pflüger M P, 2004. A simple, analytically solvable Chamberlinian agglomeration model. *Regional Science and Urban Economics*, 34(5): 565–573. doi: 10.1016/S0166-0462(03)00043-7

Phillips P C B, Sul D, 2007. Transition modeling and econometric convergence tests. *Econometrica*, 75(6): 1771–1855. doi: 10.1111/j.1468-0262.2007.00811.x

Quah D, 1993. Empirical cross-section dynamics in economic growth. *European Economic Review*, 37(2): 426–434. doi: 10.1016/0014-2921(93)90031-5

Reed W J, 2002. On the rank-size distribution for human settlements. *Journal of Regional Science*, 42(1): 1–17. doi: 10.1111/1467-9787.00247

Resende M, 2004. Gibrat's Law and the growth of cities in Brazil: A panel data investigation. *Urban Studies*, 41(8): 1537–1549. doi: 10.1080/0042098042000226993

Rosen K, Resnick M, 1980. The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 8(2): 165–186. doi: 10.1016/0094-1190(80)90043-1

Schaffar A, Dimou M, 2012. Rank-size city dynamics in China and India, 1981–2004. *Regional Studies*, 46(6): 707–721. doi: 10.1080/00343404.2010.521146.

Sharma S, 2003. Persistence and stability in city growth. *Journal of Urban Economics*, 53(2): 300–320. doi: 10.1016/S0094-1190(02)00515-6

Sheng Kerong, Sun Wei, 2013. A study of sequential city growth patterns: theoretical model and empirical evidence of U.S. urban system. *Acta Geographic Sinca*, 12(68): 1632–1642. (in Chinese)

Soo K T, 2005. Zipf's Law for cities: A cross-country investigation. *Regional Science and Urban Economics*, 35(3): 239–263. doi: 10.1016/j.regsciurbeco.2004.04.004

Wheaton W, Shishido H, 1981. Urban concentration, agglomeration economies, and the level of economic development. *Economic Development and Cultural Change*, 30(1): 17–30. doi: 10.1086/452537

Zipf G K, 1949. *Human Behaviour and the Principle of Least Effort*: *An Introduction to Human Ecology*. Addison-Wesley, Cambridge.