Zipf's Law and the Spatial Interaction Models

Hsin-Ping Chen
Department of Economics
National Chengchi University

Received : Jan. 17 2001 ; Accepted : Nov. 12 2001

Abstract

It is well known that the size distribution of cities is surprisingly well described by Zipf's law. It is considered the criterion for the local growth model. The purpose of this paper is to explain Zipf's law through the use of a dynamic process based on a spatial interaction model derived from entropy. Empirical findings show that: (1) the purposed dynamic process can generate both stable and unstable patterns in accordance with the value of the parameters, (2) in the stable evolution, the model can generate both deterministic and stochastic growth processes, (3) both deterministic and stochastic growth processes converge in Zipf's pattern, and (4) evidence from cities in Taiwan shows a diminishing estimated intercept and slope as the proposed model predicted. Size distribution in Taiwan converges to Zipf's pattern.

空間交互模型與普瑞夫定理

陳 心 蘋

國立政治大學經濟系

（收稿日期：2001 年 1 月 17 日；修訂日期：2001 年 8 月 13 日；接受刊登日期：2001 年 11 月 12 日）

摘 要

本文的主要目的是以模擬方式分析最大熵概念為基礎所推導的都市交互與成長過程，是否會衍化出趨近普瑞夫定理（Zipf's Law）的城市分配。模擬結果顯示：(一)動態交互模型在不同的參數值下可產生穩定或不穩定的成長型式。(二)在穩定的衍化過程中，本文的模型可產生決定性的與隨機的成長過程。(三)決定性的與隨機的成長型式都會趨近於普瑞夫定理的分配型態。

關鍵字：普瑞夫定理，最大熵，城市分配

1.Introduction

It is widely recognized that the size distribution of cities is surprisingly well described by Zipf's law across countries with various economic structures and histories. Its use of robust empirical evidence and its significant regularity in economics makes Zipf's law the minimal criterion for any urban growth model. However, there is a lack of plausible theoretical models for explaining this empirically robust distribution.

The expression of Zipf's law can be visualized by taking cross sectional data on city size and rank and drawing a graph with the log of rank along the y-axis, and the log of the population along the x-axis. The resulting graph, based on the regression, will most likely show a straight line with the slope very close to −1. The linear relation between the log of size and the log of rank is explained as the famous Zipf's law. This amazing result is shown in various data sets: it is demonstrated in most modern countries by Rosen and Resnick (1980); in India in 1911 by Zipf (1949); in U.S. history by Dobkins and Ioannides (1998), Krugman (1996) and Zipf (1949); and in mid-nineteenth century China by Rozman (1990). Empirical evidence from different countries and periods shows the general explanatory power of Zipf's law. The following are some of the important attempts at trying to explain or resolve the puzzle of the rank-size rule: Losch(1954), Hoover (1954) and Beckman (1958) who make use of it in the economic model; spatial model in Fujita, Krugman and Venables (1999); and Simon's random-growth model (1955). Although these efforts do provide different ways to analyze the possible theoretical foundation, the essential puzzle remains.

Gabaix (1999) proposed Gibrat's law as an explanation of Zipf's law. He found that homogeneous growth processes in cities could lead the distribution to converge into a Zipf pattern. Homogeneity of growth processes refers to the common mean and common variance of city growth rate. Regardless of the driving forces behind the growth of cities and the economic structures of countries, as long as they satisfy Gibrat's law, Zipf distribution will appear. According to Gibrat's law, both mean and variance of growth rate are independent of the size of the city. Thus, randomly growing cities with the same expected growth rate and the same variance will comply to a Zipf pattern.

Gabaix's work proposes a general and neat interpretation for explaining that puzzling regularity in city distribution known as Zipf's law. Gabaix uses Eeaton and Eckstein's data to show that the variance of the growth rate does not seem to differ across city sizes. Eaton and Eckstein (1997) find that there is no correlation between the initial size and the growth rate in both Japanese and French cities. These empirical results show some evidence in support of Gabaix's finding. However, the interaction behavior among cities, that is the essential driving force of agglomeration, in the region is not expressed in Gabaix's work.

The purpose of this paper is to try to explain Zipf's law by a growth process, oriented from a spatial interaction model, which is theoretically derived from the concept of entropy in physics. Furthermore, this paper will elucidate the following questions regarding Zipf's law: What are the implications of Zipf's law in the distribution of cities? What is the significance of the slope of the curve consistently being close to -1? Is it possible to effect a change in this "law" in terms of the timing and the slope?

Central to the concept of entropy is the derivation of the maximum uncertainty estimator when faced with limited information. This feature has been widely applied in urban and regional modeling for commuting patterns and the location choice probability in transportation and location models. These urban and regional models focus primarily on the static solution derived from entropy, such as, probability distribution and the implied spatial interaction model. Nijkamp and Reggiani (1991) have derived a dynamic process for determining the location choice probability distribution derived from entropy, nevertheless, the evolution property and the size distribution have yet to be investigated. The legitimacy of the static probability estimator in explaining regional modeling, makes it essential to the investigation of the properties of the followed dynamic process and the converged distribution. Due to the overwhelming regularity of Zipf's law, in regard to the empirical size distribution of cities in a large number of countries over long periods of time, there is a strong impetus for examining the evolution process and the limiting size distribution from the perspective of entropy, and the possible relation between it and Zipf's law.

In section 2 the theoretical background of the proposed dynamic Logit model is explained. While section 3 investigates the properties of the proposed model through simulations, and examines the size distribution of cities in Taiwan. Section 4 presents the conclusion.

2 Residential location and the spatial interaction model

2.1 Entropy in a spatial interaction model

Wilson (1967) discusses the application of entropy-maximizing methods in trip distribution. The major concept is to maximize the "uncertainty" in terms of possible assignment, subject to all prior information, with respect to the trip distribution. The entropy theory, applied to this topic, aims to derive the most probable trip or migrant distribution given additivity conditions and transport cost budget constraints.

Let T.. be the number of trips (or migrants) and c the travel cost between zones i and j; let O_i be the total outflows from zone i, and D_j be the total inflows to zone j. The entropy $W(T_{ij})$ measures the uncertainty of assignments of individual units to an origin-destination matrix. Maximizing logarithm of $W(T_{ij})$, subject to the additivity conditions (1) and (2), and transport cost budget constraints (4) derive the most probable arrangement of spatial distribution of trips in the system.

$$\sum_j T_{ij} = O_i \qquad (1)$$

$$\sum_i T_{ij} = D_j \tag{2}$$

$$w(T_{ij}) = \frac{T!}{\prod_i \prod_j T_{ij}!} \tag{3}$$

The travel budget C is expressed as follows:

$$\sum_i \sum_j c_{ij} T_{ij} = C$$

The solved optimal trip estimator is: (4)

$$T_{ij} = A_i B_j O_i D_j \exp(-\beta c_{ij}) \tag{5}$$

Where A_i and B_i are balancing factors, and $\exp(-\beta c_{ij})$ is the distance friction function. Parameter β, in the distance friction function, is the marginal possible state (the objective function in an entropy problem) per unit of transport cost. And the parameter, c_{ij}, represents the general transport cost between location i and j. The function of this optimal flow corresponds to the gravity theory. (Please refer to the appendix for details on the deriving process and explanation of the variables). The probability of transport or migration from location i to j, derived from the gravity type migrant flow (equation (5)), is:

$$P_{ij} = \frac{T_{ij}}{O_i} = A_i B_j \overline{D}_j \exp(-\beta c_{ij}) = \frac{B_j \overline{D}_j \exp(-\beta c_{ij})}{\sum_j B_j \overline{D}_j \exp(-\beta c_{ij})} = \frac{W_j \exp(-\beta c_{ij})}{\sum_j W_j \exp(-\beta c_{ij})} \tag{6}$$

Where $W_j = B_j \overline{D_j}$ is the weight. This derived probability is similar to the qualitative choice model introduced in the following.

2.2 Qualitative choice model

Traditional location theory assumes that households maximize their utility, subject to budget constraints, in residential location decisions. Assuming V_{ij} as the systematic household utility and ε_{ji} as the error term, the household utility function is as follows:

$$U_{ij} = V_{ij} + \varepsilon_{ji}. \tag{7}$$

Under the consideration of household utility maximization, and given the distribution for the unsystematic part of utility (ε_{il})[1], the probability that the household will migrate from city i to city j is

$$P_{ij} = \Pr ob(U_{ij} > U_{il}, for \quad all \quad l, \quad l \neq j) = \frac{e^{V_{ij}}}{\sum_{l} e^{V_{il}}}, \tag{8}$$

This is the multinomial Logit model. The utility function serves as the location advantage. The larger the observed utility a household is capable of achieving in city j, the more attractive it is to the household. And, consequently, the higher the probability that the household would choose to migrate to city j. The probability that the household would migrate from city i to city j (equation (6)) is the relative location advantage.

2.3 Dynamic process of the discrete choice model

A simplified probability model with the time variable from equation (6) is:

$$P_{j,t} = \frac{\exp(V_{j,t})}{\sum_{l} \exp(V_{l,t})} \tag{9}$$

Where $\sum_{i=1}^{n} P_{i,t} = 1$. This is in a multinomial logit form based on the assumption that a household chooses alternative j to achieve the maximized observed utility V_j. The negative term, $-\beta c_{ij}$ (in equation (6)), represents the major criterion on which location choice is based on in this simplified model. It indicates the reduction of the possible number of states due to city i's location in the region. The shorter the distance between city i and other cities, the higher its location accessibility will be; and consequently, the higher its selection advantage. In an extended model, the location endowment, as well as the location differences, will be included in

[1] Assume that each ε_{il} is distributed independently, identically in accordance with the extreme value distribution.

this location advantage term. The discrete dynamic Logit model derived by Nijkamp and Reggiani (1991) is as follows:

$$P_{j,t+1} = (\dot{V}_j + 1)P_{j,t} - \dot{V}_j P_{j,t}^2 - P_{j,t}\sum_{j \neq l} \dot{V}_l P_{l,t} \tag{10}$$

The first two terms, on the right-hand side, is the logistic growth of choice probability P_j. The third term is the interaction effects within the region. This dynamic spatial interaction process, derived from entropy, expresses that the change of choice probability for city j is not only influenced by its current choice probability, in a decreasing rate, but that other cities' choice probabilities also play competitive roles in city j's growth.

The variable V_j is the observed utility, or location benefit, in city j. This systematic location benefit is assumed to consist of two parts according to the time variable: (1) Geographical advantage, Ψ_i, a non-temporal location advantage caused by known geographical endowments and benefits. It is the source of the deterministic force in the dynamic process. (2) Agglomeration advantage is defined by function $h(y_{i,t})$, where $y_{i,t}$ denote the size of city i at time t. It is a temporal location advantage caused by external effects from population and employment congregation. It is historically dependent and the source of possible stochastic forces in the growth of cities. Geographical advantage does not vary through time; it is the given endowment. On the contrary, agglomeration advantage depends on the city size or the scale of the industry in the city; it is not constant through time.

$$V_{i,t} = \Psi_i + h(y_{i,t}) \tag{11}$$

In general, entropy solves gravity type optimal flow estimator which generates a logit probability model. According to the logit probability model, a corresponding dynamic discrete logit probability model is derived. The growth of city size based on the proposed dynamic discrete logit probability model could generate certain growth process of each cities in the region. Consequently, the size distribution of the cities in the region in the steady state could be analyzed. The model proposed in the following section is the dynamic discrete logit probability model from the entropy.

3. The long-term location pattern of the spatial interaction model

3.1 The model

The discreste dynamic Logit model is simulated by assuming a region with "n" number of cities, where each city grows as a result of the inward immigration of industry and migrants from outside regions: assuming that there is no inter-city immigration. The growth process is based on the discrete dynamic location choice probability as in equation (10). This discrete logit choice model is derived from the gravity type flow estimator solved from entropy. We simulate the spatial interaction model to exam the property of the evolution process by varying the variables and the initial conditions in the following way: number of the city's rank (n), length of the time frame (t), change of utility (a), and the initial value of location choice probability (P_j).

Households choose residential locations where utility is maximized, and industries choose location where their profits are maximized. Both utility and profit, in corresponding locations, reflect the location advantage to the decision-makers. The location advantages are the major concern behind the decisions of the respective decision-makers, in this model. It is assumed that the location advantage breaks down into fixed and time varying parts, which correspond to the determined and stochastic forces of the growth process. The determined location advantage, called the geographical advantage, affects the growth of the city through the initial location choice probability, which in turn reflects upon the relative geographical advantage. Furthermore, the time varying location advantage influences the growth of the city through the change of utility in the model. The change of utility through time is the change of the time-dependent advantage (agglomeration benefit).

In the simple case, assuming the change of utility as a constant α_j :

$$P_{j,t+1} = (\alpha_j + 1)P_{j,t} - \alpha_j P_{j,t}^{\ 2} - P_{j,t}\sum_{j \neq l}\alpha_l P_{l,t} \tag{12}$$

If change of the utility (α_j) equals zero, the choice probability will be fixed through time and will converge with the real size proportion in the long run. The growth process reaches a steady state when the choice probability converges with the real size proportion. Equation (12) is the proposed dynamic discrete logit probability model from entropy concept. This model will be simulated to examine its limiting static size distribution feature by changing the parameter values in the following.

3.2 Simulation

The proposed dynamic logit probability model is simulated in this section to examine the property of the proposed growth process. The purpose of the simulation is not to calibrate the parameter by the true data; it is to examine the feature of the proposed dynamic discrete probability model through the change of parameter values. We assume the simulated cities in the region have the same initial population, and all cities grow based on the dynamic discrete model proposed in the previous section. The static size distribution of the region is observed after a given time periods. To sum up, the static limiting size distribution of cities in the region based on the dynamic growth process oriented from entropy concept is observed and examined in the simulation.

The evolution process, based on the dynamic spatial interaction model, may lead to two different kinds of dynamic pattern processes. Depending on the parameter values a stable or unstable process will arise.

(1) Stable process

Assuming that there are 50 cities in the region, and all are of a uniform initial size; and that the change of the utility is a constant ‘a’, which is generated from a random number within range

(0, 0.04), then the dynamic probabilities and city sizes are converging in a stable trajectory. The simulated time path for cities in the region is presented in Fig. 1 and Fig. 2.

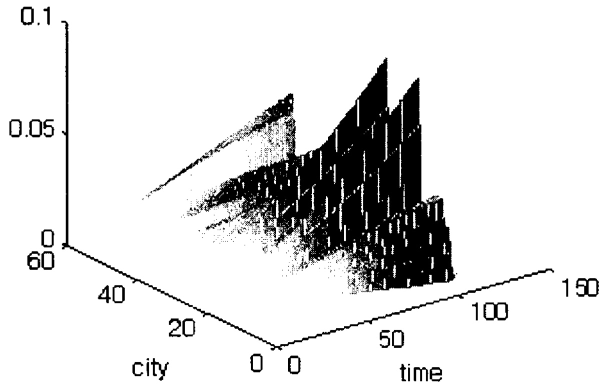

Fig. 1 The dynamic probability path of all cities in the region
(n=50, t=100, a=0.04)

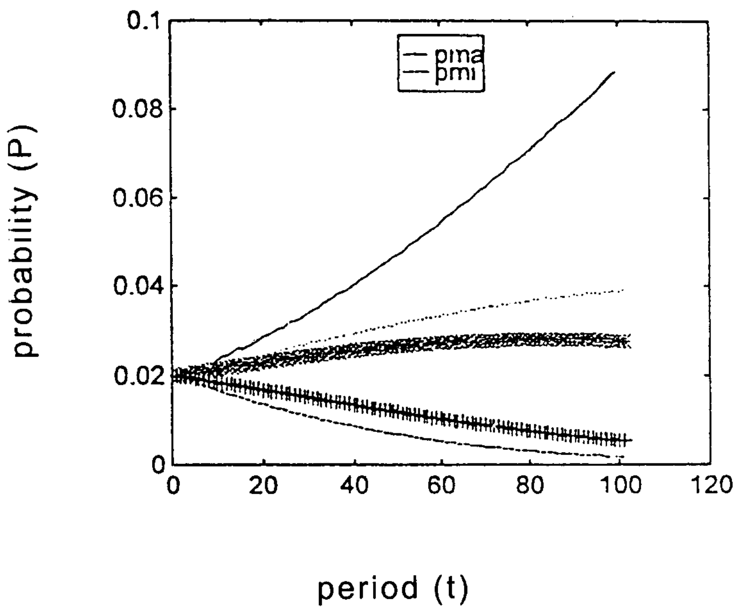

Fig. 2 The dynamic probability path of four cities in the region
(Including the cities with highest (pma) and lowest (pmi) choice probability)

(2) Oscillating process

Let us look at another experiment based on the same assumptions and initial conditions. The range of the random number, generated as the change of utility, becomes (0, 2.7). The simulated time path for all 50 cities is presented in Fig. 3 and Fig.4. The dynamic probabilities and city sizes are oscillating with the unstable trajectories. These two experiments show that the same dynamic interaction rule would lead to two essentially different evolution processes, due to the value of the parameter (scale of time varying location advantage).

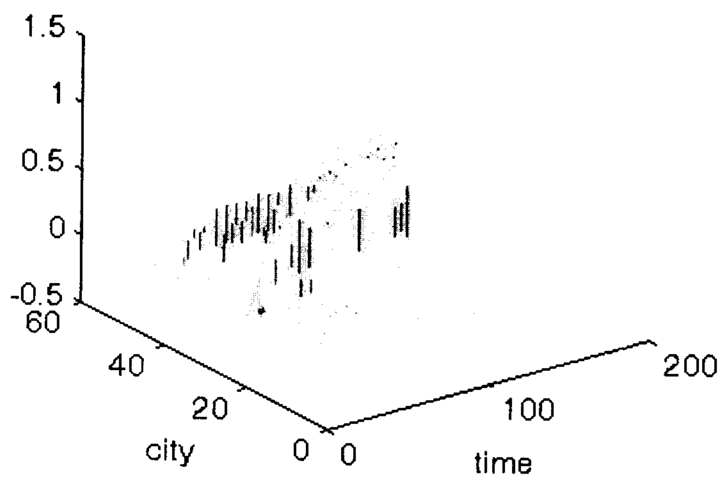

Fig. 3 The dynamic probability path of all cities in the region (n=50, t=100, a=2.7)

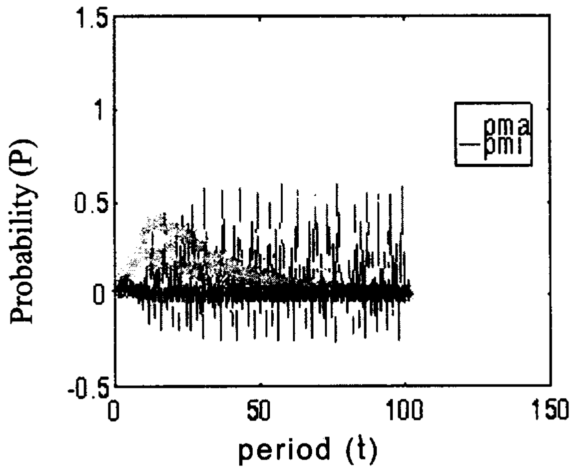

Fig. 4 The dynamic probability path of four cities in the region
(Including the cities with highest (pma) and lowest (pmi) choice probability)

(3) Features in stable evolution

Assume the same numbers of cities, (n=50), evolution time, (t =100), and the scale of time varying location advantage, (a =0.04), are all the same as in experiment (1). The simulated distribution of city size and rank is presented in Fig. 5. The Zipf plot that shows the distribution of log size versus log rank, is presented in Fig. 6. We run the regression of Zipf's law.

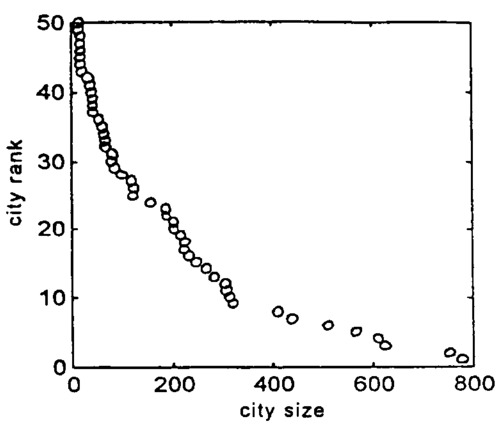

Fig. 5 City size versus rank. (n=50, t=100, a=0.04)

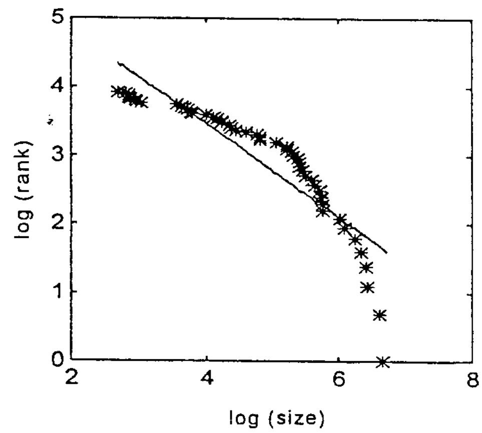

Fig. 6 Log size versus log rank

$\ln (Rank) = A - B \ln (Size)$,

The result is

$\ln (Rank) = 6.17 - 0.67 \ln (Size)$,

(0.77 0.57)

Where the 95% confidence interval of the estimated slope is in parentheses, and R^2 is 0.787. The estimated slope in the Zipf plot is different from 1, which is consistent with Zipf's law. The experiment results are in Table 1 and Table 2. The low value of the standard deviation, of both the estimated intercept and the slope, imply that a negative slope Zipf plot will invariably be generated from a dynamic Logit model. Also, given the same conditions and randomly generated change of utility parameters, different growth processes have similar estimated values for both intercepts and slopes in the Zipf plot.

Table 1
Simulation result of 'rank-size' regression (n=50) *
ln (*Rank*) = A – B ln (*Size*)

	A	B
	10.097	1.393
	10.637	1.483
	10.387	1.441
	9.952	1.362
	9.681	1.311
	10.947	1.548
	9.879	1.351
	9.174	1.211
	10.377	1.439
	11.815	1.712
	9.970	1.364
	9.896	1.343
Mean	10.234	1.413
Standard deviation	0.678	0.122

* City size (n=50), evolution time (t=50), scale of location advantage (a=0.04)

Table 2
Simulation result of 'rank-size' regression (n=100) *
ln (*Rank*) = A – B ln (*Size*)

	A	B
	6.221	0.679
	5.943	0.634
	6.192	0.679
	6.143	0.679
	6.636	0.761
	6.867	0.800
	5.946	0.633
	5.945	0.646
	6.535	0.757
	6.467	0.735
	5.899	0.639
Mean	6.295	0.702
Standard deviation	0.330	0.059

* City size (n=50), evolution time (t=100), scale of location advantage (a=0.04)

(3.1) Evolution time and region size

The simulations in this section are based on the same number of cities (n=50), and the scale of time-varying location advantage of (a=0.04). The only difference is the evolution time path (t). The regression result is in Table 3. The corresponding evolution graphs and the Zipf plots are in Fig. 7 to Fig. 14. The simulation results of the region of 100 cities (n=100) are in Table 4.

The absolute value of the estimated slope in the Zipf relation gets smaller over a longer evolution time. The longer the period of time the more divergence in the size of cities in the region is seen. This is due to the cumulated effect of the location advantage. A longer evolution time reduces the scale of the slope in the Zipf relation. At a certain time during evolution, the absolute value of the slope will be close to 1. In Table 3, the number of cities is 50, and the estimated slope is close to one at t=67; in Table 4, on the other hand, the number of cities is 100, and the estimated slope is close to one at t=75. A larger region, with a greater number of cities, does not change the fact that city sizes get less homogeneous over a longer evolution time. On the contrary, a larger number of cities in the region reduce the speed of the interaction process.

Table 3

Simulation result of 'rank-size' regression (n=50)

ln (*Rank*) = A – B ln (*Size*)

Time	A (Estimated constant)	B (Estimated slope)
50	3.62	1.28
60	8.69	1.13
65	8.36	1.07
67	7.96	0.998
70	7.73	0.96
75	7.54	0.93
80	7.13	0.85
100	3.09	0.70
200	2.54	0.34
300	2.28	0.26

Notes: Scale of location advantage (a=0.04)

Table 4
Simulation result of 'rank-size' regression (n=100)
ln (*Rank*) = A – B ln (*Size*)

Time	A (Estimated constant)	B (Estimated slope)
50	9.91	1.41
60	8.56	1.12
65	8.72	1.16
70	8.10	1.04
75	7.97	1.01
80	7.44	0.90
100	6.53	0.73
200	4.61	0.36
300	3.88	0.24

Notes: Scale of location advantage (a=0.04)

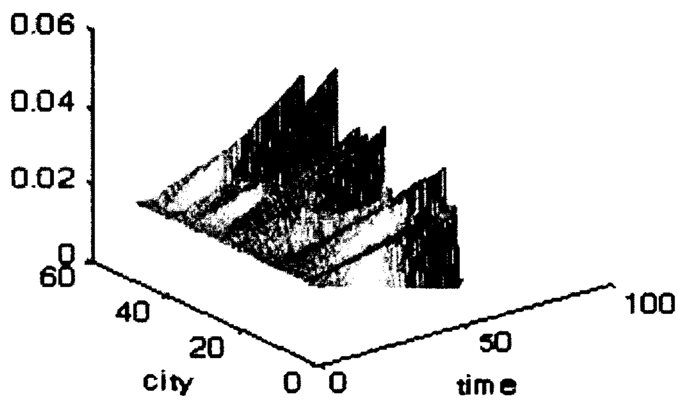

Fig. 7 The dynamic probability path of all cities in the region
(n=50, t=50, a=0.04)

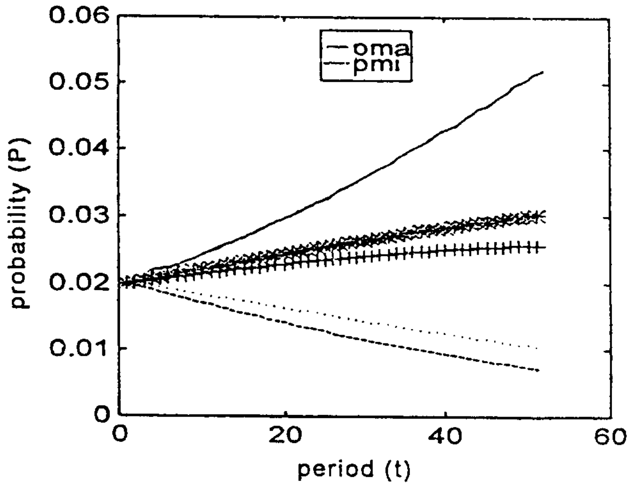

Fig. 8 The dynamic probability path of four cities in the region
(Including the ' ιty with highest (pma) and lowest (pmi) choice probability)

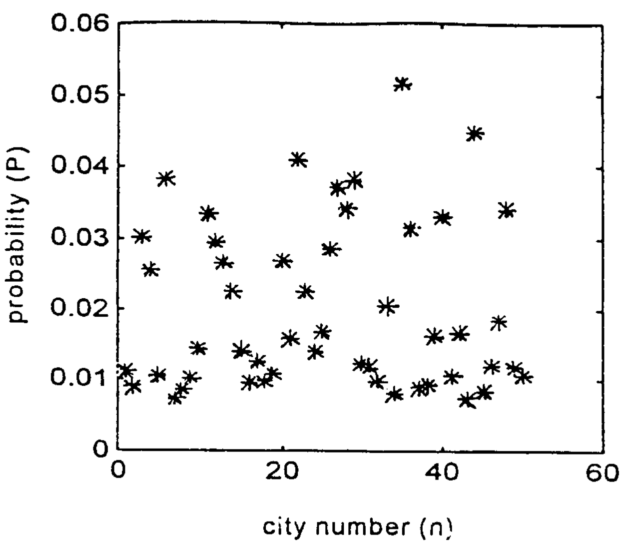

Fig. 9 Cities versus choice probability at t=50

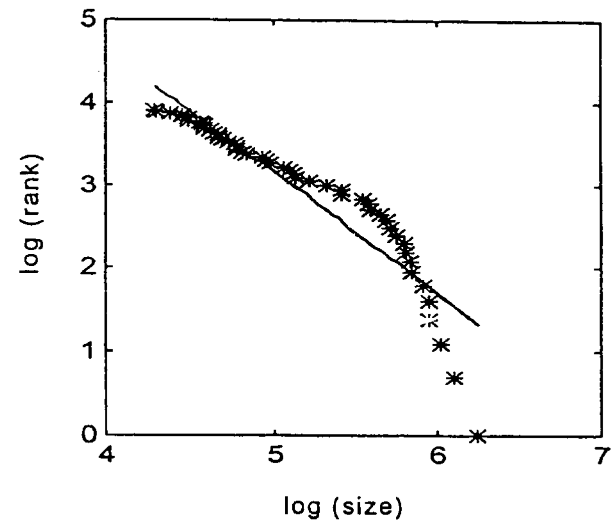

Fig. 10 log size versus log rank

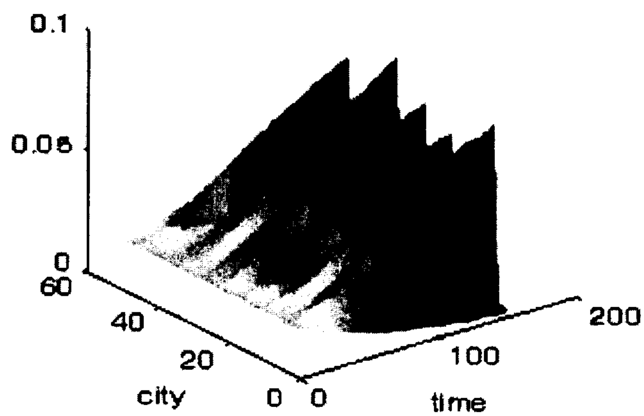

Fig. 11 The dynamic probability path of all cities in the region

(n=50, t=150, a=0.04)

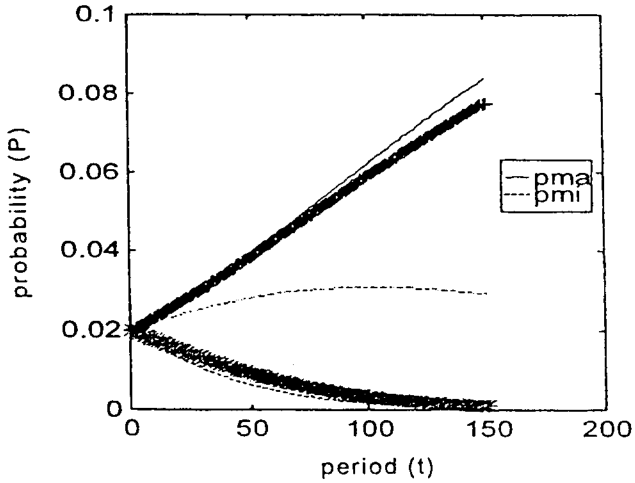

Fig. 12 The dynamic probability path of some cities in the region
(Including the cities with highest (pma) and lowest (pmi) choice probability)

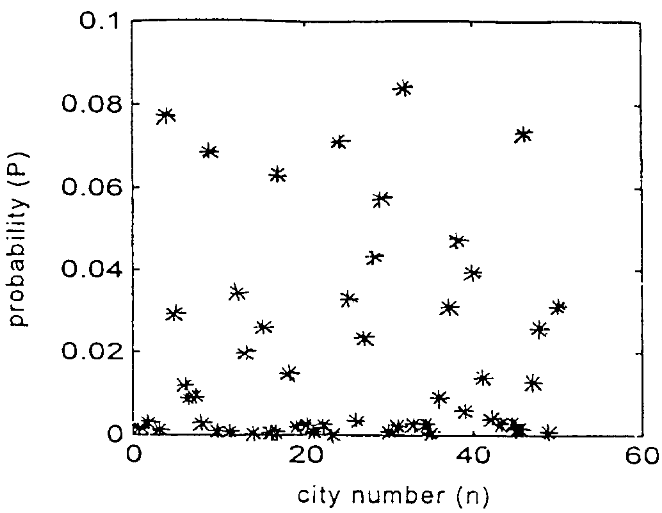

Fig. 13 Cities versus choice probability at t=150

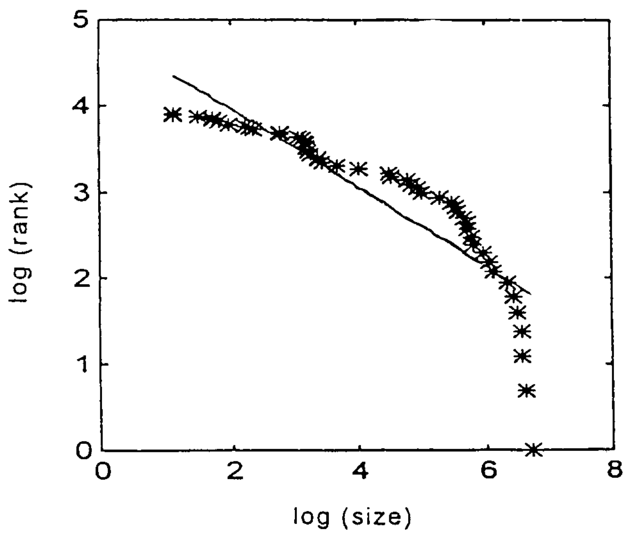

Fig. 14 Log size versus log rank

(3.2) The scale of the time-varying location advantage

In equation (12), the change of the time-varying location advantage is assumed to be a constant α_i for each city through time. Table 5 lists simulation results given different values of parameter α_i: The larger the value of parameter α_i, the smaller the absolute value of the slope.

Table 5

Change of scale of location advantage (a)*

a	A	B
0.004	68.723	12.415
0.04	9.762	1.324
0.4	2.765	0.115
0.9	2.318	0.030

* Number of cities (n=50), evolution time (t=50)

This implies that the more significant the difference of each city's change of utility, which affects migrants' choices, the more divergent the city sizes are within the region.

(3.3) Test of lock-in effect

In this experiment, we change the value of parameter α_i of city 3 into three times the original scale at time equals fifty (t=50), and examine whether the final choice probability distribution (t=100) will change. The correlation coefficient of the final distribution, both with and without the change of the parameter, is 0.86. The dominant city will maintain its dominance even if city 3 has relatively higher time varying location advantage than it did in the middle of the evolution. This result implies the possible "lock-in" property of the dynamic process. This property is one of the essential features of the self-organization system.

(3.4) The average and variance of the growth rate across city sizes

The criteria required by Gibrat's law are examined using the initial arbitrary probability distribution. Fig. 15 and Fig. 16 show the plot of growth rate versus normalized population size.

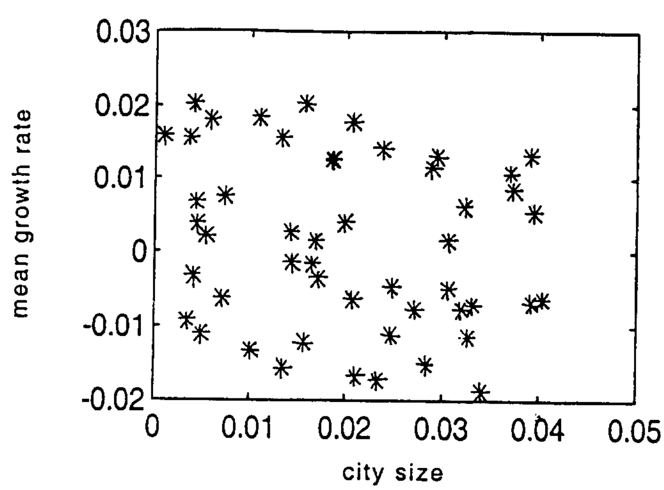

Fig. 15 Mean growth rates versus city sizes at t=1

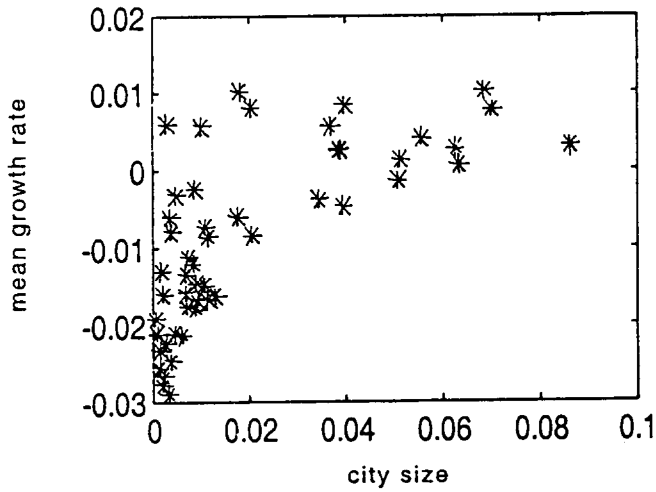

Fig. 16 Mean growth rates versus cities size at t =100

The mean growth rates in the first period are clearly independent of city sizes; and the mean growth rates at period 100 show no significant relation to the city sizes. Eeaton and Eckstein (1997) provide empirical evidence in support of this feature. The mean and variance of the average growth rates are in Table 6. The variance of growth rate across city sizes is the same. The

Table 6

Means and variances of the average and variance of growth rates

	Average growth rate	Variance of growth rate
Mean	-0.0042	8.8277e-006
Variance	0.000132	1.1357e-039
Minimum	-0.0242	8.8277e-006
Maximum	0.0149	8.8277e-006
Observations	50	50

average growth rate across city sizes, on the other hand, does not appear to be the same.[2] However, the differences between average growth rates across cities are within 0.0391.

(3.5) Determinism versus chance

Distinguishing the time-dependent property in the location advantage V_j, as in equation (11), allows for the introduction of both deterministic and stochastic features into the growth process. The geographical advantage is determined by the given location benefit, which is fixed across time. A city with a higher geographical advantage has a selection advantage. Agglomeration advantage depends on the current size of the population and the employment rate, and it changes through time and is historically independent.

Allowing only known geographical advantage in the location advantage (utility or profit), without time-varying location advantage, will cause the regional growth patterns to become deterministic. The dominant city will always be the one with highest geographical advantage. The inclusion of the time-varying location advantage, while assuming constant value (constant change of agglomeration and other time-dependent advantage through time), will also lead to a deterministic long-term pattern. The long-term distribution is based on the initial known location advantage in conjunction with the known effect from agglomeration.

Relaxing the assumption of the constant change of utility into a time-varying variable will incorporate the stochastic features into the dynamic process. After the stochastic features have been incorporated, assume that the agglomeration advantage is bounded. The simulation results will now show possible multiple dominant cities in the steady state. These dominant cities are not necessarily endowed with the highest geographical advantage or the largest time-varying location advantage. In this case, the known geographical advantage dominates the historical dependent force, which implies the defining feature of this model; that a deterministic rule may

[2] An F-test evaluates the equality of the average growth rate of N cities show significant differences across N cities in both initial distributions. The F-statistic is F=1056 given initial uniform distribution, and F=1493 given initial arbitrary distribution.

lead to a stochastic long-term pattern. Experiment results indicate that the stochastic growth process may generate Zipf's pattern in the steady state.

3.3 Evidence on the size distribution of Cities in Taiwan

We collected data on the populations of 209 to 216 cities in Taiwan for the years 1971, 1974, 1977, 1980, 1983, 1986, 1989, 1992, 1995, and 1998, with the criterion for selection being a population of at least 20,000 inhabitants. The regression results for all 10 years are in Table 7. Data, from cities in Taiwan, shows that over time the absolute value of the estimated slope decreased and converged at a value of 1. The adjusted R^2 was 0.96 in 1971 and increased over

Table 7*

'Rank-size' regression of cities in Taiwan

ln (*Rank*) = A – B ln (*Size)*

Year	Observation	A	B	Adj- R^2
1971	216	21.138	1.478	0.96
1974	216	19.491	1.412	0.97
1977	216	18.932	1.355	0.97
1980	216	18.430	1.304	0.98
1983	216	18.055	1.265	0.98
1986	213	17.748	1.234	0.99
1989	210	17.351	1.196	0.99
1992	207	16.946	1.154	0.99
1995	207	17.060	1.163	0.99
1998	209	16.822	1.140	0.99

* Source: Statistics Annals by Ministry of Interior

time, demonstrating that the size distribution of cities in Taiwan tends to converge with Zipf's law. Similar to the results of previous simulations (see Table 3 and Table 4), both estimated intercept and slope have diminishing absolute values over time. This indicates that the urban system in Taiwan converges to a less homogeneous city size distribution. This may be due to the cumulated effect of location advantage including both fixed geographical and time-varying location advantages.

4.Conclusions

In this paper, we examined the properties and long-term distribution patterns of the growth process derived from the concept of entropy. The proposed model suggested the generation of both deterministic and stochastic growth processes. We determined that both deterministic and stochastic processes comply with Zipf's pattern over the long-run. Zipf's law indicates certain degrees of the combination of different sizes of cities. The decreasing absolute value of the slope over time, demonstrated by both empirical data and simulation results, indicates that cities grow from a more homogeneous state into a more heterogeneous distribution. Zipf's law shows that, in the evolution process, the region will not evolve beyond a certain degree of "heterogeneous distribution". That is to say that the absolute value of the slope will not decrease infinitely. The level of convergence is at a certain distribution, which corresponds to slope equals -1. The converged state is at the balance point of the two contradicting forces of positive and negative agglomeration effects in cities. The timing of the convergence depends mainly on the following conditions: the initial location differences and endowments, which affect decision makers' perceived location advantage; and the change of the location advantage through time. The change of the location advantage essentially indicates the change of the net agglomeration effects (positive and negative agglomeration effects) in cities. A change of the interaction effect and the structure of both positive and negative agglomeration effects may change the converging distribution (the slope).

Some findings about the properties of the growth process of this model are as follows: (1) The purposed dynamic process possible generates both stable and unstable patterns according to the value of the parameters. (2) In the stable evolution process, the purposed model possibly generates both deterministic and stochastic growth processes. (3) The longer the evolution time the less homogeneous are the cities in the region, and the smaller the absolute value of the slope in Zipf's plot. This is due to the accumulated effects of location advantages. (4) The number of cities in the region affects the speed at which Zipf's pattern is reached; the larger the size of the region (number of cities) the slower the evolution process. (5) The larger the change of utility through time the faster the speed of the evolution process. (6) Evidence from cities in Taiwan shows the diminishing estimated intercept and slope, as the proposed model predicted. Size distribution in Taiwan converges with Zipf's pattern.

Simulation findings correspond to the findings from previous studies that, for most modern countries regardless of their economic and social structures, the distribution of city size tends to follow Zipf's law. Although, there are countries or urban systems that do not currently comply with a Zipf pattern, they show tendencies of future convergence to Zipf distribution. Given the assumption of the dynamic process of location choice probability and the randomly generated geographical advantage, an urban system, with a certain parameter value, will evolve to a Zipf pattern in the long run. The time required to reach the Zipf pattern depends on the number of cities and the relative location advantage. The location advantage includes both fixed and time-dependent advantages, which helps explain why most countries with different properties converge to the same long-term Zipf pattern. The influence from the change of relative location advantage (or related parameters) on the amount of time it takes to reach a steady state for Zipf distribution is an important question for future research in policy implication.

APPENDIX:

Entropy Theory in Spatial Interaction and the Dynamic Logit Model

Let T_{ij} be the number of trips (or migrants) and c_{ij} the travel cost between zones i and j; let O_i be the total outflows from zone i, and D_j be the total inflows to zone j. The entropy $w(T_{ij})$ measures the uncertainty of assignments of individual units to an origin-destination matrix. Maximizing logarithm of $w(T_{ij})$, subject to the additivity conditions (2) and (3), and transport cost budget constraints (4) derive the most probable arrangement of spatial distribution of trips in the system.

$$w(T_{ij}) = \frac{T!}{\prod_i \prod_j T_{ij}!} \tag{1}$$

$$\sum_j T_{ij} = O_i \tag{2}$$

$$\sum_i T_{ij} = D_j \tag{3}$$

The travel budget C is expressed as follows:

$$\sum_i \sum_j c_{ij} T_{ij} = C \tag{4}$$

The following consistency condition should also hold:

$$\sum_i \sum_j T_{ij} = \sum_i O_i = \sum_j D_j = T \tag{5}$$

The optimal flow T_{ij} is derived:

$$T_{ij} = A_i B_j O_i D_j \exp(-\beta c_{ij}) \tag{6}$$

where $A_i = \{\sum_j B_j D_j \exp(-\beta c_{ij})\}^{-1}$

and $B_j = \{\sum_i A_i O_i \exp(-\beta c_{ij})\}^{-1}$ (7)

The term $-\beta c_{ij}$ represents the reduction in total numbers of possible states induced from transport cost between i and j, the term $\exp(-\beta c_{ij})$ is the distance friction function, and A_i and B_i are balancing factors. The function of this optimal flow appears corresponding to the idea of gravity theory.

The derived gravity type migrant flow from i to j (Equ. (6)) gives the probability of a destination choice from i to j as the following:

$$P_{ij} = \frac{T_{ij}}{\overline{O}_i} = A_i B_j \overline{D}_j \exp(-\beta c_{ij}) = \frac{B_j \overline{D}_j \exp(-\beta c_{ij})}{\sum_j B_j \overline{D}_j \exp(-\beta c_{ij})} = \frac{W_j \exp(-\beta c_{ij})}{\sum_j W_j \exp(-\beta c_{ij})} \quad (8)$$

where $W_j = B_j \overline{D_j}$ is the weight. In considering the possible time varying probability, add the time variable into equation (8):

$$P_{ij,t} = \frac{W_{j,t} \exp(-\beta c_{ij,t})}{\sum_j W_{j,t} \exp(-\beta c_{ij,t})} \quad (9)$$

where $c_{ij,t}$ represents the distance between i and j at time t.

This equation is transformed into a simpler form by omitting the symbol of the origin i and assuming weight $W_{j,t} = 1$, and $-\beta c_{ij,t} = u_{j,t}$.

$$P_{j,t} = \frac{\exp(u_{j,t})}{\sum_l \exp(u_{l,t})} \quad (10)$$

Where $u_{j,t}$ could be interpreted as a choice factor, which is the utility achieved by choosing alternative j. The above probability is the formula of multinomial Logit models in discrete choice models, which assume a household chooses alternative j to achieve the maximized utility u_j.

The evolution of the dynamic multinomial Logit model is expressed by the change of probability $P_{j,t}$ with respect to time t:

$$\frac{dP_{j,t}}{dt} = \dot{P}_{i,t} = \frac{d}{dt}\left[\frac{\exp(u_{j,t})}{\sum_n \exp(u_{n,t})}\right] \quad (11)$$

$$\dot{P}_j = \dot{u}_j\, P_j(1 - P_j) - P_j \sum_{n \neq j} \dot{u}_n P_n \tag{12}$$

Where the symbol t is omitted for the sake of simplicity. The term u_j represents the change of utility through time; it is assumed to be a constant α_j. Expression (12) is a system of the Lotka-Volterra type. The first term, on the right-hand side, is the logistic growth of probability P_j, and the second term is the interaction effects among probabilities. Equation (12) is approximate by discrete time and derives:

$$P_{j,t+1} = (\alpha_j + 1)P_{j,t} - \alpha_{jj} P_{j,t}{}^2 - P_{j,t} \sum_{j \neq l} \alpha_l P_{l,t} \tag{13}$$

where $\alpha_j = \dot{u}_j$

References

Alperovitch, G., "Scale Economies and Diseconomies in the Determination of City Size Distribution," *Journal of Urban Economics*, XII, 1982, 195-228

Beckman, M. "City Hierarchies and Distribution of City Size, *Economic Development and Cultural Change,* VI, 243-248, 1958.

Black, Duncan and Henderson, Vernon. "A Model of Urban Growth", *Journal of Political Economy*, April 1999, 107(2), 252-284.

Curry, L., "The Random Spatial Economy: An Exploration in Settlement Theory," *Annals of the Association of American Geographers*, LIV, 1946, 138-146.

Dobkins, L., and Y. Ioannides, "Dynamic Evolution of the U.S. City Size Distribution," in J.M. Huriot and J.F. Thisse eds., *The Economics of Cities,* New York, NY: Cambridge University Press, 1998.

Eaton, J., and Z. E., "City and Growth: Theory and Evidence from France and Japan", *Regional Science and Urban Economics*, XXVII (1997), 443-474

Fujita, M., P. Krugman, and A. Venables, *The Spatial Economy*, Cambridge, MA:MIT Press, 1999.

Gabaix's, X,. "Zipf's Law and the Growth of Cities," *American Economic Review Papers and Proceedings*, LXXXIX, 1999, 129-132.

Hill, B., and M. Woodroofe, "Stronger Forms of Zipf's Law", *Journal of the American Statistical Association*, LXX (1975), 212-219.

Hoover, E.M.. "The Concept of a System of Cities: A Comment on Rutledge Vining's Paper, *Economic and Cultural Change*, III, 196-198, 1954.

Krugman, P., *The Self-Organizing Economy*, Cambridge, MA: Blackwell, 1996.

Losch, A., *The Economics of Location*, Yale University Press, 1954.

Nijkamp, P. and A. Reggiani, "Chaos Theory and Spatial Dynamics", *Journal of Transport Economics and Policy*, Vol. XXV, No. 1, 1991, 81-96.

Rosen and Resnick, "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy," *Journal of Urban Economics*, VIII , 1980, 165-186.

Rozman, G., "East Asian Urbanization in the Nineteenth Century: comparisons with Europe," in A. D. van der Woude, J. de Vries & A. Hayami, eds, *Urbanization in History: a process of dynamic interactions,* Oxford University Press, 1990, 61-73.

Simon, H., "On a Class of Skew Distribution Function," *Biometrika*, XLII, 1955, 425-440.

Stanley, M., et al., "Zipf Plots and the Size Distribution of Firms," *Economics Letters*, XLIX (1995), 453-457.

Suarez-Villa, L., "Metropolitan Evolution Sectoral Economic Change, and the City Size Distribution", *Urban Studies*, XXV (1988), 1-20.

Sutton, J., "Gibrat's Llegacy", *Journal of Economic Literature*, XXXV (1997), 40-59.

Wilson, A. G., A Statistical Theory of Spatial Distribution Models, *Transportation Research*, vol. 1, 1967, 253-269.

Zipf, G., *Human Behavior and Principle of Last Effort*, Cambridge, MA: Addison-Wesley, 1949.