JOURNAL OF REGIONAL SCIENCE, VOL. 39, NO. 1, 1999, pp. 183–213

THE RETURN OF ZIPF: TOWARDS A FURTHER UNDERSTANDING OF THE RANK-SIZE DISTRIBUTION*

Steven Brakman

University of Groningen, Department of Economics, P.O. Box 800, 9700 AV Groningen, The Netherlands. E-mail: brakman@eco.rug.nl

Harry Garretsen

University of Nijmegen, Department of Economics, P.O. Box 9108, 6500 HK, Nijmegen, The Netherlands. E-mail: H.garretsen@bw.kun.nl

Charles Van Marrewijk

Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands. E-mail: marrewijk@few.eur.nl

Marianne van den Berg

University of Groningen, Department of Economics, P.O. Box 800, 9700 AV, Groningen, The Netherlands. E-mail: m.van.den.berg@eco.rug.nl

ABSTRACT. We offer a general-equilibrium economic approach to Zipf's Law or, more generally, the rank-size distribution—the striking empirical regularity concerning the size distribution of cities. We provide some further understanding of Zipf's Law by incorporating negative feedbacks (congestion) in a popular model of economic geography and international trade. This model allows the powers of agglomeration and spreading to be in long-run equilibrium, which enhances our understanding of the existence of a rank-size distribution of cities.

1. INTRODUCTION

Why do cities exist, and why do they vary in size? These fundamental questions have received a considerable amount of attention from regional and urban economists in recent years. Consequently, a large and growing body of literature exists today in which many factors contributing to our understanding of the role, existence, and growth of cities are examined. Cities are now thought to arise, for example, to give consumers easy access to a large variety of goods

*We would like to thank three anonymous referees, conference participants at the 1996 WEA meeting in San Francisco, The Royal Economic Society 1997 in Stoke-on-Trent, the 1997 EEA conference in Toulouse, Monash University, Australia, the University of Osnabrueck, Germany, Andrew Bernard, Catrinus Jepma, Hans Kuiper, Jan Oosterhaven, Teun Schmidt, Dirk Stelder and Tom Wansbeek for useful suggestions.

Received May 1997; revised October 1997; accepted November 1997.

Blackwell Publishers, 350 Main Street, Malden, MA 02148, USA and 108 Cowley Road, Oxford, OX4 1JF, UK.

(either public or private), or because of the "external" effects (such as closeness to friends and relatives) of consumer location, or because of the advantage for consumers of proximity to their workplace (Fujita, 1989). However, consumers who live and work in cities are also confronted with competition for space, environmental problems, etc. The trade-off between the pros and cons of cities determines the location choice of consumers. The existence of spillovers may also play a role in their location decision of the producers. Cities provide firms with a large clientele, an efficient location for knowledge spillovers or technological externalities, maintenance services, financial services, and legal aid (Fujita and Thisse, 1996; Glaeser et al., 1992; Henderson, 1977). Political systems can also be important in the formation of cities; Ades and Glaeser argue that an urban giant may "ultimately stem from the concentration of power in the hands of a small cadre of agents living in the capital," (1997, p. 224).

The factors above are no doubt important but they only give a rationale for the existence of single cities. In general, they do not explain why cities are spread unevenly across space nor do they explain why systems of cities exist. However, the latter is studied in Central Place Theory (Christaller, 1933; Lösch, 1940; Stewart, 1948; Henderson, 1974; Eaton and Lipsey, 1982; Fujita, Ogawa, and Thisse, 1988; Abdel-Rahman and Fujita, 1993; Abdel-Rahman, 1994; Quinzii and Thisse, 1990; see Mulligan, 1984, for a survey). For a long time it proved difficult to provide a sound economic rationale for systems of cities. It was up to Eaton and Lipsey (1982) to develop a model with multipurpose shopping that creates a demand externality causing clustering. It remained analytically difficult to determine the optimal geographic pattern of agglomeration, so Eaton and Lipsey *verified* whether the equilibrium is consistent with a central place structure. Such hierarchical systems can be shown to be socially optimal given some specific conditions (Quinzii and Thisse, 1990; Suh, 1991). Subsequent research showed that systems of cities may also evolve due to differences in fixed costs or variations in industrial structure, that is, specialization versus diversification (Abdel-Rahman and Fujita, 1993; Abdel-Rahman, 1994) or differences in the type of goods produced, that is, substitutability (Fujita, Ogawa, and Thisse, 1988).

Although the aforementioned studies have clearly contributed to our understanding of systems of cities "it is fair to say that the microeconomic underpinnings of central place theory are still to be developed," (Fujita and Thisse, 1996, p. 343). Sometimes the hierarchical structure is assumed beforehand, and the purpose of the study is to show that such a system is "optimal" or an "equilibrium" outcome in some sense. Sometimes there is no interaction between location decisions, market structure, and price-setting behavior to determine the evolution of the system of cities. In particular this state of our theoretical knowledge about the size-distribution of cities is not very satisfying because we still do not have a proper understanding of the outstanding *empirical* regularity concerning city-size distributions: the so-called rank-size distribution. The rank-size distribution 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 is known as Zipf's Law. Even though the rank-size distribution does not hold exactly in reality, it does perform surprisingly well for the (historical) size distribution of cities in most industrialized countries. Nonetheless, a convincing theoretic microeconomic foundation for its existence is still lacking, so it seems that the rank-size distribution is an empirical regularity in search of a theory in which it can be grounded on the behavior of the individual consumers and producers.[1] As Mulligan put it "At this scale of analysis, research emphasizes the behavior of the central place *system*, rather than the motivations and actions of economic agents," (1984, p. 25).

Zipf's Law can be derived from a stochastic or self-organizational approach, sometimes borrowed from physics (Simon, 1955; Read, 1988; Bura et al., 1996; Krugman, 1996a, 1996b; Roehner, 1995). Haag and Max (1995) explicitly introduced "push" and "pull" factors, which are, however, simply assumed to exist and not derived from microeconomic principles. Eaton and Eckstein (1997) further extended this approach by identifying the push and pull factors, for example, distance and human capital. The purpose of this paper is to show how the rank-size distribution can be derived from a general-equilibrium economic model based on microeconomic principles. The model we use is developed in Brakman et al. (1996). It incorporates negative feedbacks (congestion) in Krugman's (1991a, 1991b) geography model. The presence of negative feedbacks explains the simultaneous existence of large and small cities. This is crucial because the rank-size distribution only makes sense in a world in which cities of various sizes coexist. We show that the model can generate a size distribution of cities that meets the requirements of the rank-size distribution.

The paper is organized as follows. In Section 2 we describe the rank-size distribution and discuss the influence of some economic variables on the size distribution of cities using The Netherlands as an example. In Section 3 we introduce the model and show how changes in parameters, such as transportation costs, the degree of industrial activity, returns to scale or congestion costs, influence the rank-size distribution. In Sections 4 and 5 we investigate some structural aspects of the model and summarize our conclusions.

2. THE RANK-SIZE DISTRIBUTION

Although there seems to be some confusion in the use of terminology (see Read, 1988, for a short survey of the literature) Zipf's Law (Zipf, 1949) was initially stated as in Equation (1), in which R_j is the rank of city j, M_j is its size,

[1]In other words "taking an even longer view the growth of cities in increasingly integrated markets raises another intriguing possibility. In the US, cities follow the so-called rank-size rule . . . No one knows quite why this is," (*The Economist*, Survey of Cities, July 29th 1995, p. 18). Furthermore, there is a fractal aspect to the rank-size distribution as it also applies at different levels of aggregation.

and Co is a constant. This elementary version can be generalized (as suggested by Zipf) to Equation (2), in which q is a (positive) constant.[2]

$$R_j M_j = Co \quad ; j = 1,..,r \tag{1}$$

$$R_j^q M_j = Co \tag{2}$$

Henceforth we refer to Equation (1) as Zipf's Law and Equation (2) as the rank-size distribution. Thus, using this terminology Zipf's Law is a special case of the rank-size distribution with the parameter q, which plays an important role in the sequel, equal to one.[3] The size of each city is measured by its population, and the city with the largest population is given rank 1, the second-largest city rank 2, and so on. Under Zipf's Law (q=1) the largest city is precisely k times as large as the k^{th} largest city; the graph of Equation (1) is a rectangular hyperbola. In empirical tests the log-linear version of Equation (2) is estimated

$$\log(M_j) = \log(Co) - q \log(R_j) \tag{3}$$

Under Zipf's Law $q = 1$ and Figure 1 results. If $0 < q < 1$ the slope of the curve would be flatter and a more even distribution of city sizes results than predicted by Zipf's Law. If $q > 1$ large cities are larger than Zipf's Law predicts, resulting in a wider dispersion of city sizes.

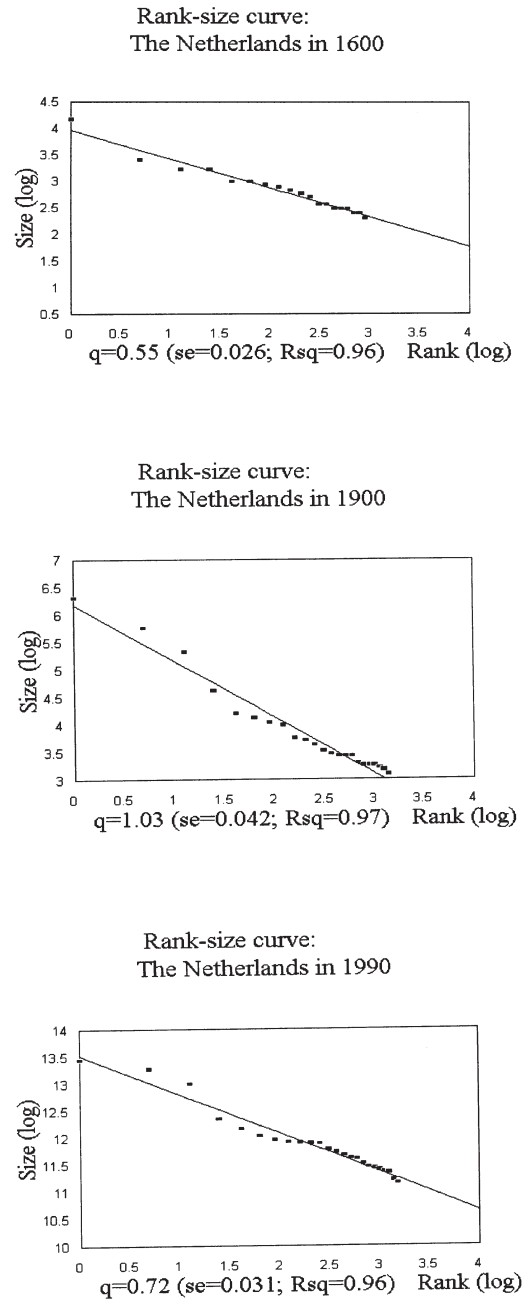
In general the various estimations of the rank-size distribution for the city-size distribution of individual countries fit the data remarkably well (see Read, 1988). The following critical remarks are worth mentioning.

First, the rank variable is a transformation of the size variable, which inevitably creates a (negative) correlation between the two variables.

Second, a distribution pattern as predicted by the rank-size distribution is often found only when very small cities are excluded from the sample. If the size of the city drops below a certain level (which is neither constant through time nor the same for every country) there is hardly any negative correlation between size and rank left for this group of small cities. For instance, Krugman argues that the rank-size distribution works best for US cities "over a range of two orders of magnitude from cities of around 200,000 up to metropolitan New York, with almost 20,000,000" (1996a, p. 95). A possible rationale for this procedure is that very small cities are indistinguishable from rural areas and can be omitted from the data. In other word, there is a threshold value for urbanization, see

[2]Equation (2) is the Lotka form, see Parr (1985) or Roehner (1995). These authors also show that the problem has been considered prior to Zipf's contribution, for example by Auerbach (1913), Goodrich (1925), and Lotka (1925).

[3]As pointed out by an anonymous referee 'rank-size rule' or 'Zipf's Law' is now standard usage if $q = 1$, whereas 'rank-size function' or 'rank-size distribution' is used if q may differ from unity.



Source: Kooij (1988) and CBS (1995) Stand van de Bevolking

FIGURE 1: Rank-Size Rules for The Netherlands.

Roehner (1995) for a discussion. Various methods are used to determine this threshold (Parr, 1985).

Third, for the city-size distribution in some countries (notably the USA at present) Zipf's Law holds because q is not statistically different from 1. However, for most other countries and at different times, q is often found to be *different* from 1. Therefore unnecessary attention to the value $q = 1$, is unwarranted (see Krugman (1996b).

Fourth and most importantly, Equation (2) does not deal with the comparative static effects of Zipf's Law. The value of q in Equation (2) is not constant over time. That is, urban growth is not proportional. Sometimes, the urban population becomes more concentrated in the larger cities (which increases q) whereas during other periods the urban population becomes more evenly distributed across the various cities (which decreases q).[4] We now illustrate this point.

A Brief Look at Urbanization and Economic Change

What causes cities to become larger or smaller relative to one another, or what accounts for changes in q over time? If one believes, as we do, that the reason for these changes is not necessarily to be found in random growth, but instead in structural economic changes, a model that links these changes to Zipf's Law is called for. This is the topic of the next section. Parr (1985) investigates the size-distribution of cities for various countries since about 1900 and argues that over time a nation tends to display an n-shaped pattern in the degree of interurban concentration.[5] This also holds for The Netherlands, one of the earliest intensively urbanized countries in Europe, which we use as an illustration. The discussion is based on Kooij (1988), who distinguishes three stylized periods:[6]

1. *Pre-industrialization (ca. 1600–1850)* characterized by high transportation costs and production dominated by immobile farmers.

2. *Industrialization (ca. 1850–1900)* characterized by declining transportation costs and the increasing importance of "footloose" industrial production with increasing returns to scale.

3. *Post-industrialization (ca. 1900–present)* characterized by a declining importance of industrial production and an increased importance of negative feedbacks such as congestion.

As early as 1600 The Netherlands contained 20 cities with more than 10,000 inhabitants. In terms of the rank-size distribution the size distribution of these

[4]It must also be emphasized that the ranking of *individual* cities is not constant over time. The history of urban development is very much a story of the rise and fall of particular cities.

[5]Parr (1985) finds a U-shaped pattern for the variable $1/q$, which therefore translates into an n-shaped pattern for q.

[6]De Vries (1981, pp. 96–104; 1984) gives additional supporting historical data on Europe as a whole.

cities was relatively even. There was no integrated urban system at the national level until halfway through the nineteenth century, which marks the start of the era of industrialization in The Netherlands. Apart from the fact that the industrialization process had by and large yet to begin, the relatively high transportation costs between cities is thought to have provided an additional important economic reason for the lack of a truly national urban system.

The estimation of Equation (3) for The Netherlands reveals that q increases from 0.55 in 1600 to 1.03 in 1900 (per our own calculations, see also Figure 1).[7] In the second half of the nineteenth century an integrated urban system was formed. Two interdependent economic changes were mainly responsible for this formation. First, the development of canals, a railroad network, and, to a lesser extent, roads significantly lowered transportation costs between cities which enhanced trade between cities. Second, due to lower transportation costs, the industrialization process really took off and cities often became more specialized which stimulated trade between cities. Although the industrialization process did not lead to dramatic changes in the overall rank-size order, it can be concluded nevertheless that as time went by the initially large cities gained a relatively larger share of the urban population.

Structural changes in the rank-size distribution take decades to materialize, so it is only well into the twentieth century that most Western industrialized countries, including The Netherlands, gradually entered the post-industrialization era. The share of the services sector in total employment becomes ever more important at the expense of the industrial sector. Comparing the Dutch rank-size distribution for 1900 with 1990 it is evident that the size distribution of cities has become more "flat." In 1900 q was 1 compared to 0.7 in 1990. The declining importance of industry (and hence of production characterized by increasing returns to scale) may be one factor contributing to this change in the size distribution. Increased congestion, especially in the large cities, is thought to have stimulated the decline of such cities as Amsterdam, Rotterdam, and The Hague.

From the above discussion we conclude that Period 2, the industrialization period, was special in its power of agglomeration, as also noted by Kooij ". . this was the era of the large cities," (1988, p. 363). However, for all three periods the rank-size distribution holds. This is illustrated in Figure 1 which shows the results of estimating Equation (3) for our sample of Dutch cities in 1600, 1900, and 1990: at least 96 percent of the variance in city size is explained by the rank-size distribution. Industrialization apparently leads to an increase of q and it is during this period (around 1900) that the Dutch rank-size distribution mimics Zipf's Law (q = 1). Finally, these three periods in which changes in economic variables demonstrably have an impact on the rank-size distribution enable us to simulate the impact of such changes in the sequel of the paper.

[7]Standard errors are 0.026 and 0.042 and R^2s are 0.96 and 0.96, respectively. The sample consisted of 19 cities for 1600 and 23 cities for 1900.

3. A GENERAL EQUILIBRIUM MODEL OF INDUSTRIAL LOCATION AND ZIPF'S LAW

The Model

In Brakman et al. (1996) we extend Krugman's (1991a, 1991b) general equilibrium location model. Krugman's basic location model usually produces only a few industrial cities of equal size, or even monocentric (one city) industrial production, and it is therefore not suited to derive the rank-size distribution. The extension we offer is the possibility of negative feedbacks or negative externalities.[8] By doing so we can explain the viability of small cities or small industrial clusters and thus in principle are able to derive rank-size distributions.

There are N cities producing manufactured goods that consist of many individual varieties, n_j in city j, and a homogeneous agricultural good, which serves as numeraire. The production of manufactured goods is characterized by firm-specific increasing returns to scale, and is assumed to be of the so-called "footloose" type: firms are able to change their production location without costs. The increasing (internal) returns to scale are responsible for the fact that each variety is only produced by a single firm. Concentration of production reinforces further concentration because total city income grows if workers decide to move to that specific city. This pecuniary externality or growth of city income then attracts more industrial firms and so on. Agricultural production is not mobile, is characterized by perfect competition and, for simplicity, zero transport costs.[9] 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. If all industrial production is concentrated in just one location a firm may still consider relocation to another city in the hope of gaining a large part of that city's local market by not having to charge transportation costs.

On the demand side it is assumed that each city spends a fixed amount of its income on both types of goods. This is the outcome of the well-known nested Cobb-Douglas or CES utility function given in Equations (4) and (5), where α is the share of income spent on manufactured goods and σ is the elasticity of substitution between different varieties of manufactured goods. If we let γ denote the share of industrial workers in the total work force and λ_j the share of these in city j, then the total number of industrial workers in city j, L_j, is given in Equation (6). Finally, let Q_{Aj} denote production of agricultural goods in city j and ϕ_j the share of agricultural workers in city j, then normalization of the production of agricultural goods, which takes place under constant returns to scale, leads to Equation (7)

[8] The properties of this model and a complete description can be found in Brakman et al. (1996).

[9] Positive transport costs for the agricultural sector are investigated in Fujita and Krugman (1995).

(4) $$U = C_m^{\alpha} C_a^{1-\alpha}$$

(5) $$C_m = \left[\sum_{i=1}^{n_j} \sum_{j}^{N} c_{ij}^{1-1/\sigma} \right]^{\frac{\sigma}{\sigma-1}}$$

(6) $$L_j = \lambda_j \gamma L$$

(7) $$Q_{aj} = \phi_j (1-\gamma) L$$

To analyze location aspects, positive transportation costs of the "iceberg" type are introduced in the manufacturing sector. That is, if one unit of manufactured goods is shipped from city j to i only $0 < t_{ij} < 1$ units arrive in city i. An increase in the parameter t_{ij} implies a reduction in transport costs from city j to city i.

It is important in this model that industrial workers (and hence the associated demand for manufactured products) are located where the production of manufactured products is located, and vice versa. The agglomeration process is fueled by the interaction of increasing returns, transportation costs, and the mobility of the industrial workers. Increasing returns are responsible for industrial concentration. However, the fact that some factors of production are immobile provides for the spreading forces in the model: there is always demand for manufactures in the periphery, that is in the rural areas.

So far the model is simply a summary of Krugman (1991a, 1991b). Our model differs from his because of the role of negative feedbacks, external diseconomies, or congestion costs. These costs provide an additional spreading force and thereby enable an equilibrium outcome with not only a few cities, but a reasonable number of cities of different sizes—a necessary condition for a theoretical foundation of Zipf's Law. The importance of negative feedbacks was noted by Balassa (1961, pp. 202–204); although economies of scale tend to concentrate production in centrally located cities, there can still be a "spreading" effect towards rural cities. In modern times external diseconomies may arise because of limited physical space, limited local resources (such as water for cooling processes), environmental pollution (which may require extra investment), and other congestion effects such as heavy usage of roads, communication channels, and storage facilities.

Our aim is to analyze the consequences of congestion rather than its origin. We capture the essence of negative feedbacks in Equation (8), where l_{ij} represents the amount of labor necessary to produce x_{ij} units of a particular variety i in city j

(8) $$l_{ij} = f_j(n_j) + \beta_j(n_j) x_{ij} \text{ ; with } f_j', \beta_j \geq 0$$

Note that, in contrast to the more commonly used formulation in imperfect competition models, the fixed costs (f_j) and the variable costs (β_j) depend positively on the number of firms in the city of location (n_j).[10] Although a novel feature in this type of model, Mulligan (1984, pp. 22-23) discusses models with similar characteristics, and a justification is in Arnott and Small (1994). Firms located in city j take both the fixed and variable costs as given and thus do not take congestion externalities into account when maximizing their objective function (profit). Note that these costs can be different for different cities. This may be used to model differences between cities, such as differences in production technology (Brakman and Garretsen, 1993).

Consumers maximize utility, taking income and prices of manufactured goods as given. Producers maximize profits by setting the price of their good (i.e., their variety) in a monopolistically competitive environment, taking factor prices, congestion externalities and production functions as given. The price elasticity of demand for a manufactured variety is constant so, provided the number of varieties is large, this gives rise to the familiar pricing rule in Equation (9) that price is a constant mark-up over marginal cost, where P_j is the price of a variety produced in city j. If P_{ij} denotes the price charged in city i for a variety produced in city j this price is given in Equation (10). Given the distribution of the industrial workers over the various cities and applying a zero-profit condition, the production of a representative manufacturing firm in city j, x_j, is derived in Equation (11)

$$P_j = \left(\frac{\sigma}{\sigma - 1}\right) \beta_j (n_j)\, w_j \tag{9}$$

$$P_{ij} = \frac{P_j}{t_{ij}} \tag{10}$$

$$x_j = \frac{f_j\left(n_j\right)}{\beta_j\left(n_j\right)} (\sigma - 1) \tag{11}$$

First, the short-run model is solved for the real wages in the various cities *given* the distribution of the labor force. Second, the cities with high real wages will attract mobile workers from the cities with lower real wages, up to the point where either (a) real wages are equal for those cities with mobile workers, or (b) all mobile workers are concentrated in only one city. The central equations [see Appendix 1 for a derivation of Equation (14)] are

$$Y_j = \phi_j (1 - \gamma)L + \lambda_j \gamma L w_j \tag{12}$$

$$I_j = \left[\sum_k n_k P_{jk}^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{13}$$

[10]In the simulations the specific functional form is $f = an^{\tau}$.

$$w_j = \left[\alpha\sigma^{-\sigma}(\sigma-1)^{\sigma-1}\right]^{\frac{1}{\sigma}}\left[\beta_j(n_j)\right]^{\frac{1-\sigma}{\sigma}}\left[f_j(n_j)\right]^{\frac{1}{\sigma}}\left[\sum_i Y_i\left(I_i t_{ij}\right)^{\sigma-1}\right]^{\frac{1}{\sigma}} \tag{14}$$

Equation (12) defines income in city j as the sum of income from agriculture (the numeraire) and income from the mobile work force. Equation (13) defines the exact price index of manufactures needed to determine the real wage. Equation (14) gives the nominal wage in terms of the numeraire in city j. It follows from the condition that demand equals supply in all markets. The left-hand side of Equation (14) represents the cost of producing in city j. The right-hand side determines demand for varieties from city j, which is a function of the other cities' income, their exact price indices, and the cost of transporting goods from city j to the city in question. Given the distribution of the labor force (λ_j) the number of varieties produced in each city (n_j) can be determined. Prices (P_{jk}) depend on wages (w_k) and transport costs (t_{jk}), so that Equations (12) to (14) may be solved simultaneously for income, wages, and price indices of manufactures in all cities.[11] The price indices and wages can be used to determine the real wages ω_j and the average real wage ϖ. The distribution of the mobile labor force is adjusted until the real wage of each city is close to the average real wage of all cities (see also Appendix 1). In our simulations below all cities are located on a circle at equal distances. Thus space is one dimensional and neutral: it does not inherently favor any specific location. Therefore, if the rank-size distribution results in the long-run equilibrium it is a feature of the model and not of the preassumed spatial structure. In principle workers can move anywhere, not only to the next city. This means in the case of 24 cities, for example, that the distance between cities 1 and 23 is 2, etc. For ease of reference we call this the "equidistant-circle."

Due to its nonlinear nature the model cannot be solved analytically and we have to use computer simulations to derive the equilibrium distribution of cities for a given set of parameters. Different initial conditions may lead to a different long-run equilibrium. In this sense our rank-size distribution (see Section 4) will be path-dependent. The simulations below show that negative feedbacks are of crucial importance in explaining industrial location and the economic viability of small industrial centers. In general, monocentric equilibria are not an outcome of our model. These results may be explained as follows. Other things being equal, agglomerating forces become stronger without congestion, such that the initially largest city generally attracts all industrial workers. With congestion the forces of agglomeration and spreading are in balance in the long-run equilibrium. Zipf also believed that the balance between two opposing powers, which he called the *Force of Unification* and the *Force of Diversification*, is crucial

[11]Thus, in the simulations below where we analyze 24 or 25 cities, one short-run equilibrium is the solution to 72 or 75 simultaneous nonlinear equations. Such a solution has to be found for a number of iterations until a long-run equilibrium is reached.

for understanding the rank-size distribution, for example to economize on transport costs of materials

> "One course is to move the population to the immediate sources of raw materials in order to save the work of transporting the materials to the persons; the effect of this economy, which we shall call the *Force of Diversification*, will be to split the population into a larger n number of small, widely scattered and largely autarchic communities that have virtually no communications or trade with one another.
>
> The other course of economical action, which we shall call the *Force of Unification*, operates in the opposite direction of moving the materials to the population, with the result that all production and consumption will take place in one big city where the entire population of C persons will live. In practice, therefore, the actual location of the population will depend upon the extent to which persons are moved to materials and materials to persons in a given system" (Zipf, 1949, p. 352).

Simulating Zipf

To mimic the relationship between economic parameters and the level of agglomeration illustrated in Section 2 we start with 24 cities located on an equidistant-circle.[12] Initially each city receives a random share of the industrial labor force. In the subsequent analysis (Figure 2) we only include cities with a long-run industrial sector; pure agricultural areas are left out as they do not represent a city. In this respect the number of cities is endogenous. Based on the stylized periods distinguished by Kooij (1988) for The Netherlands (discussed in Section 2), we now discuss changes in economic parameters for each of these three periods.

1. *Pre-industrialization*. The small industrial sector in this period produces close substitutes and production is dominated by immobile farmers. We simulate the small industrial sector by choosing a relatively high value for the share of agricultural workers in the total labor force (γ = 0.5) and the almost homogeneous industrial sector by choosing a relatively high value for the elasticity of substitution between varieties (σ = 6, which simultaneously implies that increasing returns to scale are relatively unimportant). The low level of regional integration (high transportation costs) is described by choosing t = 0.5. Negative feedbacks are not very important in this period, but they are not

[12]Unless otherwise indicated parameter values are identical to the base-run values given in Table 1. Although arbitrary in principle, the variables are reasonable; for example, the share of mobile workers in the labor force equals 0.60, the elasticity of substitution equals 4, the share of income spend on manufacturing equals 0.60, and so on. The chosen parameter values not only give rise to the rank-size distribution, but also allow us to vary them within a reasonable range and thus analyze the effects of these changes (see Section 4 and Appendix B). We emphasize throughout the paper that parameters change over time and thus affect the size distribution of cities. Moreover, interactions between parameters (mutual dependence) are likely to determine the final outcome, see Brakman et al. (1996).

absent (think of the disease-ridden large cities in the Middle Ages) which is simulated by choosing a moderate value for τ , $\tau = 0.25$.

2. *Industrialization*. The basic characteristic in this period is the spectacular decrease in transportation costs and the increasing importance of footloose industrial production with increasing returns to scale. At the same time negative feedbacks are not absent, but also not very important in the sense that they prevent large cities becoming even larger. In the model we simulate these factors by lowering transport costs to $t = 0.8$ (remember, a high value of t implies a low level of transportation costs) and increasing the share of the industrial labor force in total employment, $\gamma = 0.6$. The increased importance of economies of scale and differentiated industrial products are represented by choosing $\sigma = 4$. In this period the strong industrialization leads to the disappearance of small cities. This corresponds with the idea that agglomerating forces dominate during the era of big-city growth.

3. *Post-industrialization*. In this period transportation costs remain low and as before the industrial sector is characterized by differentiated products and increasing returns to scale. The notable difference with earlier periods is congestion, such as the growing traffic jams, air pollution, and rising land rents in cities. Smaller cities are less troubled by such effects and therefore have a tendency to grow faster. In the model we simulate this by increasing the congestion parameter τ to 0.5.

Figure 2 presents some simulation results for the above three periods. At least 93 percent of the variance in city size is explained by the rank-size distribution. More importantly, these simulations suggest that the n-shaped pattern of q over time, identified by Parr (1985), depends on the economic parameter changes. Thus, in principle, we can reproduce rank-size distributions by varying those parameters that have been identified in the literature to be relevant for understanding the changes in the size distribution of cities.

4. STRUCTURAL ANALYSIS

In Section 3 we demonstrate that it is possible to derive a size distribution of cities generating the rank-size distribution based on an explicit general-equilibrium location model. The question arises whether or not rank-size distributions are a structural outcome of our model. First, we analyze the migration dynamics inherent in the adjustment process of the model for a particular example using the base-run parameter values and a random initial distribution of the mobile labor force. Second, we analyze whether the economic model is important in increasing the explanatory power of the rank-size distribution. Third, we investigate the impact of changes in economic parameters on q and the power of agglomeration. Finally, the complex nature of the impact of transportation costs on agglomeration induces us to analyze a somewhat more general spatial structure.

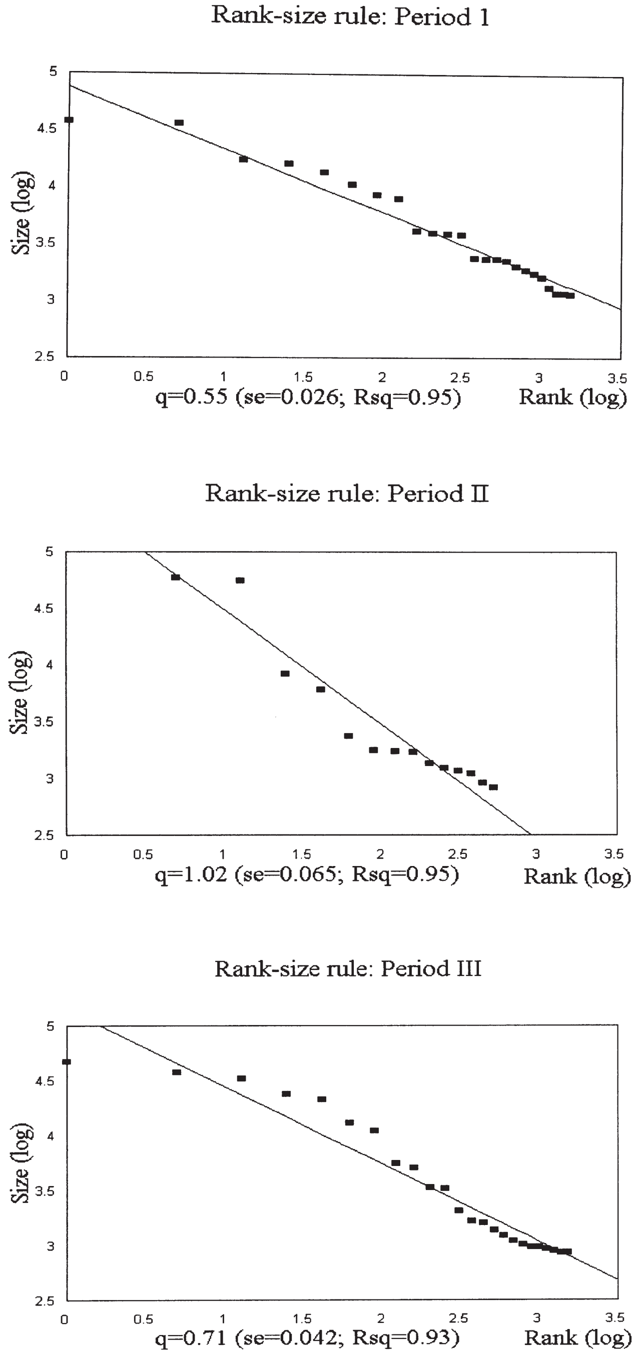


FIGURE 2: Simulating Zipf.

Adjustment

Analyzing more closely the adjustment over time of a typical simulation example is the best way to get an intuitive feel for the adjustment process.[13] Figure 3a depicts a random initial distribution of city size over the 24 cities on the equidistant circle.[14] Given this initial distribution we solve for the short-run equilibrium. The real wages for industrial workers differ between the 24 cities which starts a migration process of workers from cities with low real wages to cities with high real wages according to Equation (23) in Appendix 1. The redistribution of workers determines a new short-run equilibrium and starts a second migration process because the real wages are still not equal for all cities. The process continues until the long-run equilibrium is reached, in this simulation example, after 16 migration process. Figure 3b depicts the final distribution of city size over the 24 cities at the long-run equilibrium. Ultimately there are two agglomeration centers of economic activity: one center around City 1 and one center around City 12.

Inspecting both panels of Figure 3 shows that the two final centers of economic agglomeration are close to the initial centers of high economic activity. At the same time, the two final centers are rather evenly spread over the equidistant circle, that is they are not too close to each other. This is also clear from Figure 4a, showing the evolution of size over time as a result of the migration process for Cities 1, 2, 11, and 12 (ultimately ranked number 1–3 and 5 in size). After two migration processes, Cities 2 and 11 are the largest cities. However, as agglomeration centers these two cities are too close to each other; therefore City 2 ultimately becomes smaller than City 1 and City 12 becomes substantially larger than City 11. As demonstrated by City 11, the adjustment process is not monotone over time.[15] The prosperity of individual cities does not depend only on its own size, but also on that of its neighbors. City 14, for example, is initially the largest city. However, it is surrounded by smaller cities so it ultimately drops in the rankings to number 7. Nonetheless, initial size does matter: eight out of the ten ultimately largest cities were in the initial top ten list.

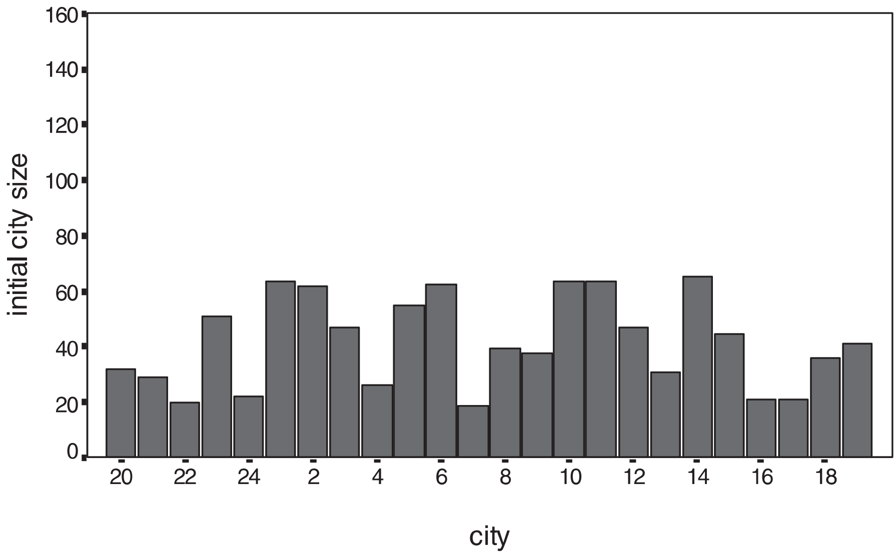
Finally, Figure 4b suggests that the model increases the predictive power of the rank-size distribution: as the city size distribution is adjusting to the long-run equilibrium the share of the variance explained by the rank-size distribution is increasing (until the level $R^2 = 0.95$ is reached). Simultaneously, the level of agglomeration as measured by the q-value is increasing over time

[13]This approach was suggested to us by an anonymous referee. The parameter values we use are given in Table 1.

[14]The uniform random distribution is used to determine the industrial labor force for the 24 cities. However, note that Figure 3 depicts total city size, that is the sum of industrial and agricultural workers.

[15]The size of Cities 1, 12, 13, and 24 is monotone increasing; of Cities 5, 6, 8, 14, 15, 18, 19, 20, 21, and 23 is monotone decreasing; and of the remaining ten cities is first increasing and then decreasing.

(a) Initial City Size



(b) Final City Size

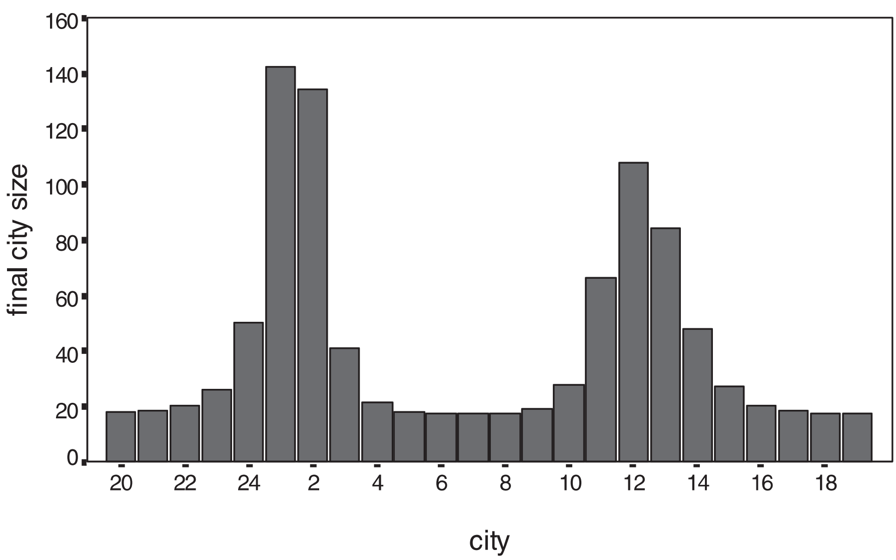
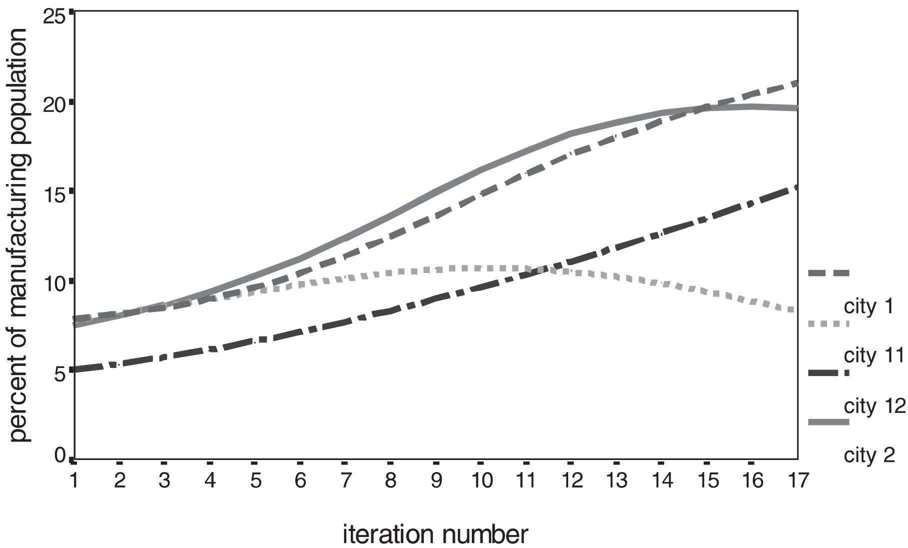


FIGURE 3: Change in City Sizes.

(a) Adjustment of Some Cities Over Time



(b) Adjustment of q-Value and R^2 Over Time

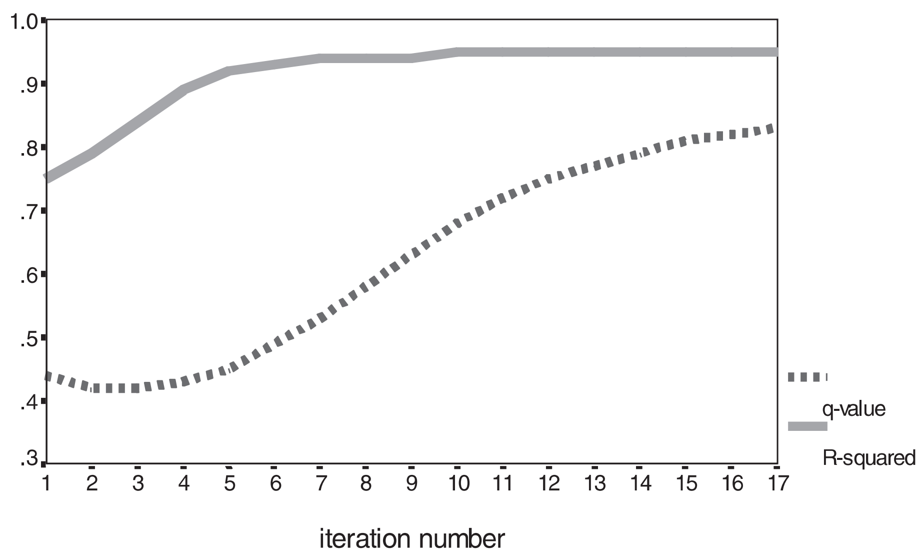


FIGURE 4: Stepwise Evolution.

after a small initial drop. However, this is only one example. We know that in this type of nonlinear model initial conditions may be very important in determining the final equilibrium. We now examine this in more detail.

Does the Model Matter?

We first test the rank-size distribution without using the model, that is by drawing (10,000 times) a random distribution of the mobile labor force for each of the 24 cities from the uniform distribution and calculating the q value for each draw.[16] This experiment results in a single-peaked distribution of q values with a wide range (from 0.3 to 1.7), an average fit (R^2) of only 0.54 and an average t-value of 5.3, see Table 1. Second, we repeatedly draw (240 times for the base scenario) an initial random distribution of the mobile labor force (also using the uniform distribution) to derive the concomitant long-run equilibrium for each draw applying the model. This procedure allows us to estimate the rank-size distribution, that is, we derive 240 q's. The resulting frequency distribution of q is depicted in Figure 5, which leads us to distinguish between two possible sets of spatial outcomes: agglomeration ($q > 0.725$) and spreading ($q < 0.725$). See below for a discussion of this criterion. Remarkably, the economic model generates the rank-size distribution fairly adequately both if agglomeration or spreading occurs, although the former slightly outperforms the latter. See Table 1 for these results and the selected base-run values for the parameters. More specifically, under agglomeration the average share of the explained variance in city size (i.e., R^2) equals 0.94 with an average t-value of 19.8, whereas if spreading occurs the explained variance in city size equals 0.88 with an average t-value of 14.5. Thus, inclusion of the model considerably increases the explanatory power of the rank-size distribution. As we observed in Figure 2, the model is able to explain changes in q, which is not the case with the stochastic approach.

Figure 5 illustrates three main points regarding the sensitivity of the rank-size distribution with respect to changes in initial conditions. First, a large range of long-run equilibrium outcomes is possible (q ranges from 0.24 to 0.88).[17] Second, the simulated density function is double-peaked. The lower peak at the left for low values of q, and thus relatively equally-sized cities, arises if the initial distribution of the mobile labor force is relatively even.[18] The much higher peak

[16]Alternative initial random distributions, with the obvious exception of the Pareto distribution (see Read, 1988), may also be used to investigate whether our model adds to the explanatory power of the rank-size distribution. No alternative initial distribution, not even the Pareto distribution, can be used for more fundamental questions, such as the changes in the distribution resulting from parameter changes.

[17]As noted in Appendix A, the range of long-run equilibrium outcomes is affected by the variable ε in Table 1. In the simulations a long-run equilibrium is reached if the relative deviation of the real wage does not exceed the value ε for each city.

[18]The reader should keep in mind that an exactly even distribution of the mobile labor force over the 24 cities of the equidistant-circle is a long-run equilibrium for all parameter settings in the sequel.

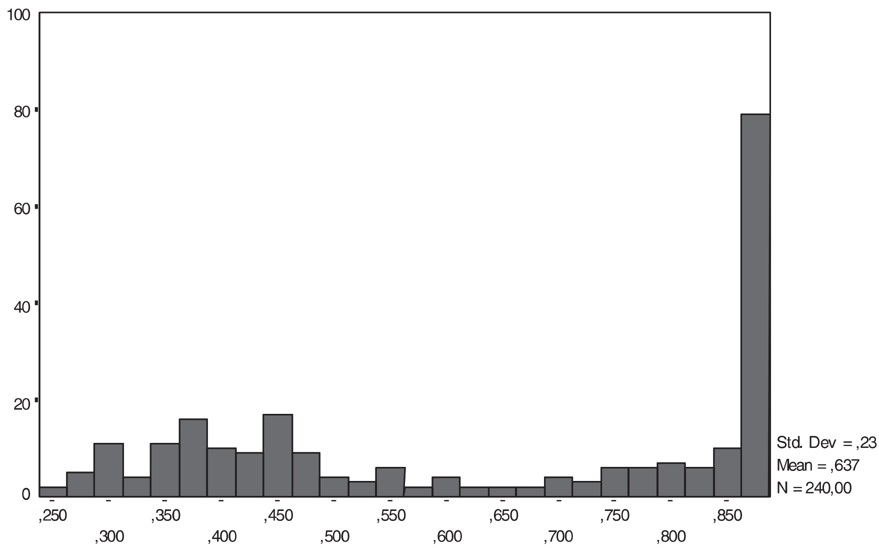


FIGURE 5: Histogram of Base-Run Scenario.

at the right arises if the initial distribution is more uneven, which triggers a process of agglomeration. Third, Figure 5 suggests there is a dominant underlying value of q characteristic of the base-run parameter setting (approximately $q = 0.87$). These three features do not depend on the specific base-run parameter values but also apply for other parameter values (see Appendix 2 for some of the corresponding simulated density functions). The double-peak and narrow range of Figure 5 is a reflection of the underlying economic structure of our model and a reminder of the interaction between economic parameter values and path-dependency.

After inspection of the double-peaked density functions we decided to subdivide all simulation experiments into two groups: agglomeration and spreading. Henceforth, agglomeration takes place if the predicted value of the largest city is at least *ten* times as large as the smallest city. This, admittedly arbitrary, rule translates into a value of $q = 0.725$ and is based on the bi-modal feature of Figure 5. However, the value $q = 0.725$ is an intermediate value for q found in empirical studies (Parr, 1985; De Vries, 1981). In short, if q exceeds 0.725 we will say agglomeration takes place, otherwise spreading occurs. For example, the mean q-value for the agglomeration subgroup of the base scenario ($q = 0.853$, see Table 2) leads to an average predicted value of the largest city equal to *fifteen* times the smallest city if agglomeration takes place. In contrast, the mean q-value for the spreading subgroup of the base scenario ($q = 0.437$, see Table 3) leads to an average predicted value of the largest city equal to only *four* times the smallest city if spreading occurs.

TABLE 1: Base-run Simulation Parameters and Statistics

Base-Run Values				
$\alpha = 0.6$	$\gamma = 0.6$	$\sigma = 4$	$\tau = 0.4$	$\phi_j = 1/24$
$L = 1000$	$t = 0.8$	$a = 0.1$	$b = 0.2$	$\varepsilon = 0.00001$

Goodness-of-Fit for Rank-Size Rule

	Random Distribution	Base-run Simulations: Agglomeration	Base-run Simulations: Spreading
Average R^2	0.54	0.94	0.88
|Average t-value|	5.3	19.8	14.5

Tables 2 and 3 summarize the statistical information on estimated q-values for the base-run parameter setting and some interesting alternative parameter settings. Table 2 focuses on agglomeration ($q > 0.725$) and Table 3 on spreading ($q < 0.725$). If spreading occurs the variance (standard error for the mean of 0.010) is larger than if agglomeration occurs (standard error for the mean of 0.004) for the base-run scenario, see Figure 5. The base-run simulated probability of agglomeration, defined as the number of times agglomeration occurs divided by the total number of simulations, is approximately equal to the simulated probability of spreading (0.48 and 0.52, respectively)

Changes in Economic Parameters

Investigating the influence of changes in economic parameters, it is clear from Tables 2 and 3 that the probability of agglomeration increases if the elasticity of substitution between varieties (σ) decreases because the importance of the positive externality associated with varieties becomes more important. This makes it more attractive to locate where everyone else is located. Similarly, and not surprisingly, the simulated probability of agglomeration also increases if the mobile sector is more important, or if the costs of congestion are lower. Finally, a reduction in transport costs (an increase in t) also seems to increase the probability of agglomeration.

The advantage of having a general equilibrium economic model that can be used for simulations in explaining agglomeration is the ability to change a parameter without simultaneously changing the other parameters of the model. In contrast, economic historians often point at a number of factors that change simultaneously when evaluating changes in agglomeration over time, thereby focusing primarily on the reduction in transport costs and the simultaneous increase of the importance of returns to scale and the growth of the mobile sector. The simulations reported in Table 2 seem to indicate that the reduction in transport costs and the increase in returns to scale basically increase the probability of agglomeration, whereas the increased importance of the mobile sector affects the extent of agglomeration, that is, the average value of q increases (from 0.825 to 1.038). This is in accordance with the special nature of the industrialization period identified in Section 2 (see Figure 1.)

TABLE 2: Statistical Simulation Information on Rank-Size Distribution; Agglomeration*

	Base	High Elasticity of Substitution ($\sigma = 5 > 4$)	Low Elasticity of Substitution ($\sigma = 3.5 < 4$)	Mobile Sector More Important ($\alpha = \gamma = 0.7 > 0.6$)
Mean q	.853 (.004)	.825 (.022)	.838 (.006)	1.038 (.012)
Simulation Probability of Agglomeration	$\frac{115}{240} = 0.48$	$\frac{3}{50} = 0.06$	$\frac{32}{50} = 0.64$	$\frac{35}{51} = 0.69$
	High Transport Costs ($t = 0.65 < 0.8$)	Low Transport Costs ($t = 0.9 > 0.8$)	Low Cost of Congestion ($\tau = 0.25 < 0.4$)	Manhattan Circle; Base
Mean q	.840 (.008)	.842 (.001)	.884 (.004)	.872 (.002)
Simulation Probability of Agglomeration	$\frac{25}{59} = 0.42$	$\frac{39}{50} = 0.78$	$\frac{59}{102} = 0.58$	$\frac{44}{50} = 0.88$

*Standard errors in parenthesis. The mean value of q is calculated for the agglomeration subgroup of simulation outcomes ($q > 0.725$); the simulated probability of agglomeration is the number of times agglomeration occurs divided by the total number of simulations.

Although economic historians generally argue that a reduction in transport costs enables an increased level of agglomeration there is no a priori reason to believe this is always true. After all, a reduction in transport costs also makes it more attractive to produce at low cost in the periphery and (cheaply) transport the products to the cluttered center. In this respect the effect of a reduction in transport costs as reported in Table 2 seems to confirm the general perception that this increases the probability of agglomeration. This perception needs to be qualified. It is true that over a broad range of transport cost reductions there is a general tendency to increase the probability of agglomeration (roughly from zero at $t = 0.25$—not reported in Table 2—to 0.78 at $t = 0.9$) but it is also true that the rise in the probability of agglomeration is not monotone and abruptly reverses after reaching a peak.[19] For $t = 0.95$ the simulated probability of agglomeration equals zero (not reported in Table 2). The fact that both very high and very low transport costs result in a spreading of economic activity over the various cities confirms earlier findings, see Krugman and Venables (1995) and Brakman et al. (1996). It is easy to understand because extremely high transport costs make each city an almost autarkic entity that can only be served from

[19]For example, if t rises from 0.75 to 0.8 the simulated probability of agglomeration slightly drops from 0.52 (not reported in Table 2) to 0.48.

TABLE 3: Statistical Simulation Information on Rank-Size Distribution; Spreading*

	Base	High Elasticity of Substitution ($\sigma = 5 > 4$)	Low Elasticity of Substitution ($\sigma = 3.5 < 4$)	Mobile Sector More Important ($\alpha = \gamma = 0.7 > 0.6$)
Mean q	.437 (.010)	.421 (.011)	.389 (.018)	.559 (.021)
Simulation Probability of Spreading	$\frac{125}{240} = 0.52$	$\frac{47}{50} = 0.94$	$\frac{18}{50} = 0.36$	$\frac{16}{51} = 0.31$
	High Transport Costs ($t = .065 < 0.8$)	Low Transport Costs (t - 0.9 > 0.8)	Low Cost of Congestion ($\tau = 0.25 < 0.4$)	Manhattan Circle; Base
Mean q	.524 (.018)	.300 (.040)	.480 (.014)	.583 (.040)
Simulation Probability of Spreading	$\frac{34}{59} = 0.58$	$\frac{11}{50} = 0.22$	$\frac{43}{102} = 0.42$	$\frac{6}{50} = 0.12$

*Standard errors in parenthesis. The mean value of q is calculated for the spreading subgroup of simulation outcomes ($q < 0.725$); the simulated probability of spreading is the number of times spreading occurs divided by the total number of simulations.

within the city limits. On the other hand, for extremely low transport costs prices more readily reflect production costs including the cost of congestion in crowded areas. This makes the periphery more attractive and leads to spreading.

Manhattan Circles: Nonneutral Space

The occasionally ambivalent nature of a reduction in transport costs on agglomeration makes it appealing to look at this variable from a different perspective. In the equidistant-circle each city enters symmetrically and there is no structural bias toward either agglomeration or spreading at any particular location. In contrast, nature often provides specific locations which do favor agglomeration, such as valleys or natural harbors. Perhaps the so-called “Manhattan circles,” introduced into spatial economics by Kuiper, Paelinck, and Rosing (1990), are the most straightforward way to analyze such structures. Figure 6 depicts a Manhattan circle with radius 2. The distance between any two locations is measured stepwise; thus the distance between Locations 1 and 6 equals 2, and the distance between Locations 13 and 4 equals 3. By construction Location 1 is the most favorable location for agglomeration as the average distance to the other locations is minimal.

Tables 2 and 3 also report the agglomeration and spreading results for simulations of a Manhattan circle with radius 3 that consists of 25 locations

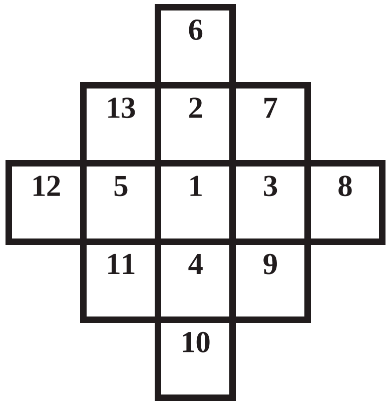


FIGURE 6: Manhattan Circle with Radius 2.

using the base-run parameter setting.[20] As expected, the structural bias toward agglomeration considerably increases the simulated probability of agglomeration (from 0.48 to 0.88, see Table 2). However, two aspects are more surprising. First, agglomeration is not automatic because occasionally spreading occurs. Second, despite its structural advantage in the center of the Manhattan circle Location 1 is not always the largest location in the long-run equilibrium. This is illustrated in Figure 7 where panel *a* depicts a typical outcome in which Location 1 is the largest and panel *b* depicts a more exceptional outcome in which an off-center location is the largest.[21] It is a reminder of the fact that the structural advantage of Location 1 is not always decisive in making it the largest location; sometimes path-dependency may be more important. None of the simulations with the Manhattan circle achieve the circular-like, transport-cost–minimizing agglomeration results in which Location 1 is smaller than equal-sized Locations 2, 3, 4, and 5 as reported in Kuiper, Kuiper, and Paelinck (1993), not even if we start with an even distribution of the mobile labor force.

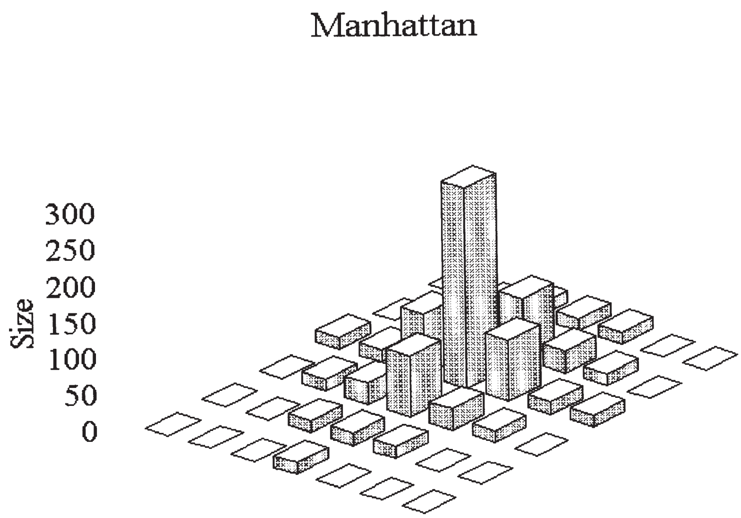
5. CONCLUSIONS

One rarely finds empirical relationships in economics which deserve to be called "laws." Zipf's Law is a noteworthy exception. Despite impressive progress in the theory of city formation and city systems in recent decades, the theoretic economic foundation is still poorly understood. In this paper we provide some further understanding of Zipf's Law by incorporating congestion in an economic location model. In our analysis economic parameters play an explicit role in determining the rank-size distribution and describing its evolution over time. The model is able to derive the rank-size distribution because the equality of

[20]The number of locations of a Manhattan circle with radius R equals $2\,R(R + 1) + 1$.

[21]Location 1 was the largest location in the long-run equilibrium in 42 of the 50 simulations.

(a) Center Location is Largest



(b) Center Location is Not Largest

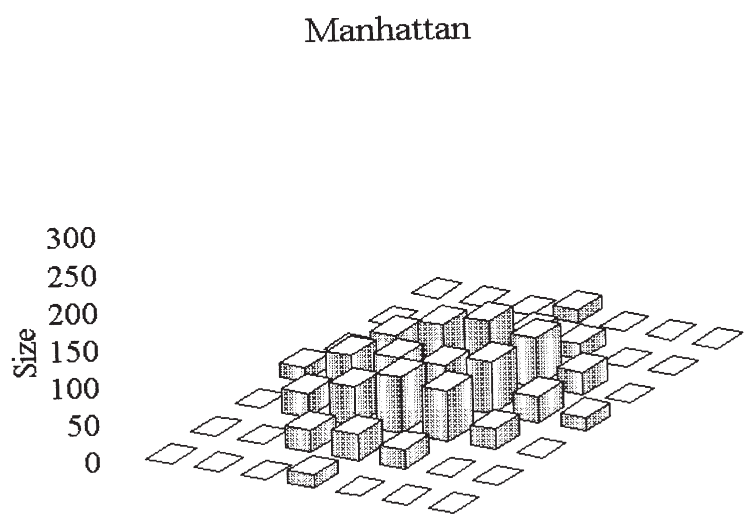


FIGURE 7: Simulations with Manhattan Circles.

agglomerating and spreading forces allows for the simultaneous existence of large and small cities.

Using historical data for The Netherlands as an illustration, we are roughly able to reconstruct historical trends with respect to the rank-size distribution by varying parameters that represent specific economic factors (such as the share of the industrial labor force in total employment, the level of integration between cities, congestion, and the type of goods). The parameters we use in our simulations are found to be important by economic historians for describing (changes in) the size distributions of cities. Simulations show that these parameters have the expected effects.

Furthermore, we show that the rank-size distribution is a structural outcome of our model. By repeatedly drawing an initial distribution of the mobile labor force from a uniform distribution we determine simulated long-run equilibrium density functions for q. This results in a double-peaked density function; a lower peak if the initial distribution of the mobile labor force is relatively even and a higher peak if the initial distribution is more uneven. It is demonstrated that our model considerably increases the explanatory power of the rank-size distribution. All in all, it seems possible to derive Zipf's Law from a general-equilibrium location model.

REFERENCES

Abdel-Rahman, Hesham M. 1994. "Economies of Scope in Intermediate Goods and a System of Cities," *Regional Science and Urban Economics*, 24, 497–524.

Abdel-Rahman, Hesham M. and Masahisa Fujita. 1993. "Specialization and Diversification in a System of Cities," *Journal of Urban Economics*, 33, 189–222.

Ades, Alberto and Edward Glaeser. 1997. "Trade and Circuses, Explaining Urban Giants," *Quarterly Journal of Economics,* 110, 195–227.

Arnott, Richard and Kenneth Small. 1994. "The Economics of Traffic Congestion," *American Scientist*, 446–455.

Auerback, Felix, 1913. "Das Gesetz der Bevoelkerungskoncentration," *Petermanns Geographische Mitteilungen,* 59, 74–76.

Balassa, Bela. 1961. *The Theory of Economic Integration*. New York: Allen and Unwin.

Brakman, Steven and Harry Garretsen. 1993. "The Relevance of Initial Conditions for the German Unification," *Kyklos*, 46, 163–181.

Brakman, Steven, Harry Garretsen, Richard Gigengack, Charles van Marrewijk, and Rien Wagenvoort. 1996. "Negative Feedbacks in the Economy and Industrial Location," *Journal of Regional Science*, 36, 631–652.

Bura, Stéphane, France Guérin-Pace, Hélène Mathian, Denise Pumian, and Lena Sanders. 1996. "Multiagent Systems and the Dynamics of a Settlement System," *Geographical Analysis*, 28, 161–178.

Christaller, Walter. 1933. *Central Places in Southern Germany* (Fischer, transl.). London: Prentice Hall.

de Vries, Jan. 1981. "Patterns of Urbanization in Pre-Industrial Europe, 1500–1800," in H. Schmal (ed.), *Patterns of European Urbanization since 1500,* London: Croom Helm, pp. 79–109.

———. 1984. *European Urbanization 1500–1800*, London: Methuen.

Eaton, Curtis B. and Richard G. Lipsey. 1982. "An Economic Theory of Central Places," *The Economic Journal*, 92, 56–72.

Eaton, Jonathan and Zvi Eckstein. 1997. "Cities and Growth: Theory and Evidence from France and Japan," *Regional Science and Urban Economics,* 27, 443–474.

Fujita, Masahisa. 1989. *Urban Economic Theory; Land Use and City Size*. Cambridge: Cambridge University Press.

Fujita, Masahisa, Hideaki Ogawa, and Jacques-François Thisse. 1988. "A Spatial Competition Approach to Central Place Theory: Some Basic Principles," *Journal of Regional Science*, 28, 477–494.

Fujita, Masahisa and Paul R. Krugman. 1995. "When Is the Economy Monocentric?: von Thünen and Chamberlin Unified," *Regional Science and Urban Economics,* 25, 505–528.

Fujita, Masahisa and Jacques-François Thisse. 1996. "Economics of Agglomeration," *Journal of the Japanese and International Economies*, 10, 339–378.

Glaeser, Edward L., Hedi D. Kallal, José Scheinkman, and Andrei Schleifer. 1992. "Growth in Cities," *Journal of Political Economy*, 100, 1126–1152.

Goodrich, Ernest. 1925. "The Statistical Relationship between Population and the City Plan." in E. R. Burgess (ed.), *The Urban Community*. Chicago: The University of Chicago Press, pp. 144–150.

Haag, Günther and Horst Max. 1995. "Rank-Size Distribution of Settlement Systems," *Papers in Regional Science*, 74, 243–258.

Henderson, Vernon J. 1974. "The Sizes and Types of Cities," American Economic Review, 64, 640–656.

———. 1977. *Economic Theory and the Cities*. New York: Academic Press.

Kooij, Pim. 1988. "Peripheral Cities and Their Regions in the Dutch Urban System until 1900," *Journal of Economic History*, 48, 357–371.

Krugman, Paul R. 1991. *Geography and Trade*, Cambridge, MA: Massacusetts Institute of Technology.

———. 1991b. "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

———. 1996a. *The Self-Organizing Economy.* Cambridge: Blackwell.

———. 1996b. "Confronting the Mystery of Urban Hierarchy," *Journal of the Japanese and International Economies*, 10, 399–418.

Krugman, Paul R. and Anthony Venables. 1995. "Globalization and the Inequality of Nations," *Quarterly Journal of Economics,* 110, 857–880.

Kuiper, Friso J., J. Hans Kuiper, and Jean H. P. Paelinck. 1993. "Tinbergen-Bos Metricised Systems: Some Further Results," *Urban Studies*, 30, 1745–1761.

Kuiper, J. Hans, Jean H. P. Paelinck, and Ken E. Rosing. 1990. "Transport Flows in Tinbergen-Bos Systems," in K. Peschel (ed.), *Infrastructure and the Space Economy*, Heidelberg: Springer Verlag, pp. 29–52.

Lösch, August. 1940. *The Economics of Location* (Fischer, transl.). New Haven: Yale University Press.

Lotka, Alfred J. 1925. *The Elements of Physical Biology*. Baltimore: Williams and Wilkins.

Mulligan, Gordon F. 1984. "Agglomeration and Central Place Theory: A Review of the Literature," *International Regional Science Review*, 9, 1–42.

Parr, John B. 1985. "A Note on the Size Distribution of Cities over Time," *Journal of Urban Economics*, 18, 199–212.

Quinzii, Martine and Jacques-François Thisse. 1990. "On the Optimality of Central Places," *Econometrica*, 58, 1101–1119.

Read, Cambell B. 1988. "Zipf's Law," in S. Kotz, N.L. Johnson, C.B. Read (eds.), *Encyclopedia of Statistical Sciences*, New York: Wiley.

Roehner, Bertrand Marie. 1995. "Evolution of Urban Systems in the Pareto Plane," *Journal of Regional Science*, 35, 277–300.

Simon, Herbert. 1955. "On a Class of Skew Distribution Functions", *Biometrika*, 425–440.

Stewart, John Q. 1948. "Demographic Gravitation: Evidence and Applications," *Sociometry*, 11, 31–58.

Suh, Seoung H. 1991. "The Optimal Size Distribution of Cities," *Journal of Urban Economics*, 30, 182–191.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*, New York: Addison Wesley.

APPENDIX 1

We derive Equation (14) of the main text. Full employment of industrial workers gives Equation (15). Given λ_j Equation (15) determines n_j, which in turn determines f_j. Income in city j and the exact price index is given by Equations (12) and (13) in the text. Let C_{ij} denote the consumption in city i of a commodity produced in city j; this gives Equation (16) the spending equilibrium. The first-order conditions for two representative products for a city k is given in Equation (17). Combining Equations (16) and (17) yields Equation (18)

$$L_j = n_j \sigma f_j \tag{15}$$

$$\sum_j n_j P_{ij} C_{ij} = \alpha Y_i \tag{16}$$

$$\frac{C_{ki}}{C_{kl}} = \left(\frac{P_{kl}}{P_{ki}}\right)^{\sigma} \tag{17}$$

$$C_{lk} P_{lk} = \alpha Y_i \frac{P_{lk}^{1-\sigma}}{\sum_j n_j P_{lj}^{1-\sigma}} \tag{18}$$

Furthermore, let s_{lk} represent expenditure in city l on commodities from city k; using Equation (18) we arrive at Equation (19). The income of manufacturing labor in city i, given in Equation (20), must equal total sales so Equation (21) follows, where city 1 is used as an example. After rearrangement Equation (22) follows (similar equations hold for the other cities)

$$S_{lk} = \frac{\alpha Y_l n_k P_{lk}^{1-\sigma}}{\sum_j n_j P_{lj}^{1-\sigma}} = \alpha Y_l n_k P_{lk}^{1-\sigma} I_l^{\sigma-1} \tag{19}$$

$$\lambda_i \gamma L w_i = n_i \sigma f_i w_i \tag{20}$$

$$\begin{aligned} n_1 \sigma f_1 w_1 &= \sum_i s_{i1} = \alpha \sum_i Y_i n_1 \left[\frac{\sigma \beta_1 w_1}{(\sigma-1) t_{i1}}\right]^{1-\sigma} I_i^{\sigma-1} \\ &= \alpha \left[\frac{\sigma \beta_1 w_1}{(\sigma-1)}\right]^{1-\sigma} n_1 \sum_i Y_i \left(I_i t_{i1}\right)^{\sigma-1} \end{aligned} \tag{21}$$

$$w_1 = \left[\alpha \sigma^{-\sigma} (\sigma-1)^{\sigma-1}\right]^{\frac{1}{\sigma}} \left[\beta_1\right]^{\frac{1-\sigma}{\sigma}} \left[f_1(n_1)\right]^{-\frac{1}{\sigma}} \left[\sum_i Y_i \left(I_i t_{i1}\right)^{\sigma-1}\right]^{\frac{1}{\sigma}} \tag{22}$$

Finally, we note that $\omega_j = w_j I_j^{-\alpha}$, where ω_j is the real wage in location j. The change in city j's share of mobile labor is given in Equation (23) and driven by the extent to which its wage deviates from the overall average $\overline{\omega} = \sum_j \lambda_j \omega_j$. In the simulations a long-run equilibrium is reached if the relative deviation of the real wage does not exceed the value ε for each city.

$$\lambda_j(s + 1) = \lambda_j(s) + \rho\lambda_j(s)\left[\omega_j(s+1) - \overline{\omega}(s)\right] \tag{23}$$

APPENDIX 2. SIMULATED HISTOGRAMS

This appendix displays the simulated histograms of estimated q coefficients for low costs of congestion, low transport costs, low elasticity of substitution, high mobility, high elasticity of substitution, and high transport costs, respectively.

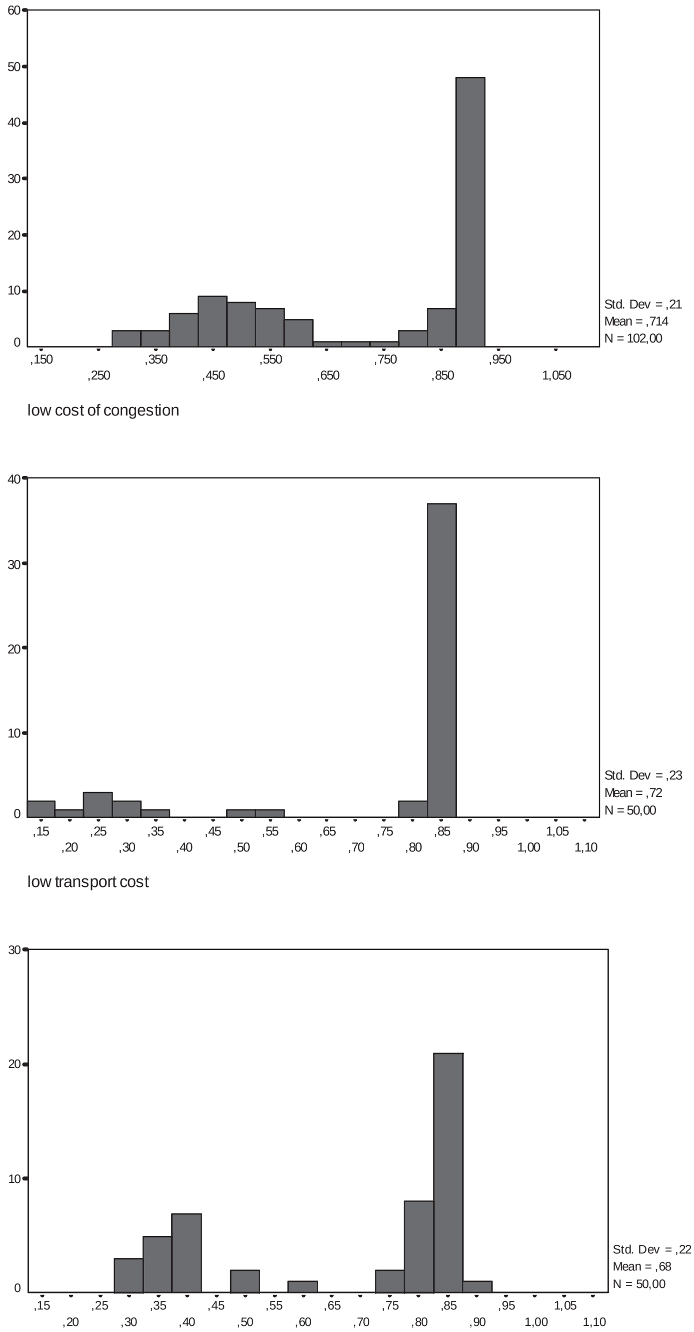
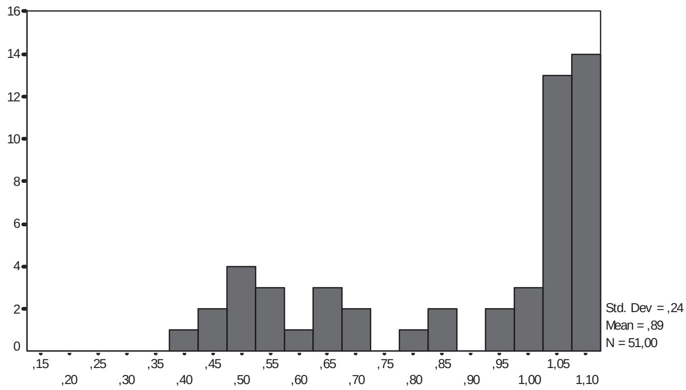
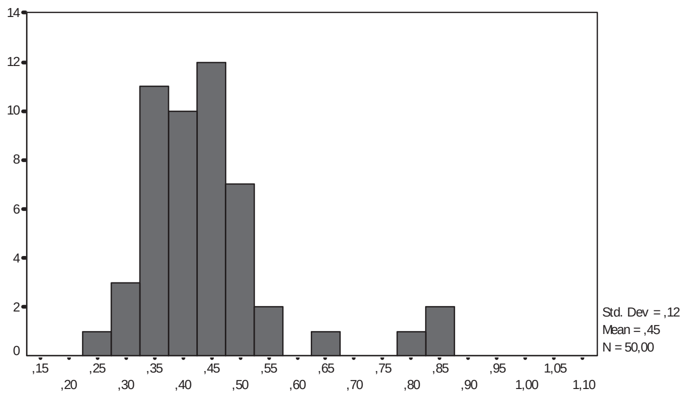


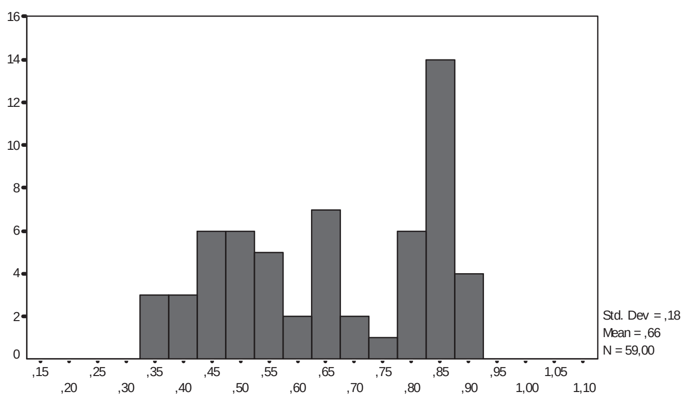
FIGURE 8: Histograms for Other Parameters.



high mobility



high elasticity of substitution



high transport costs

APPENDIX 3. LIST OF VARIABLES

U	=	utility
L	=	total labor force
C_a	=	consumption of agriculture, which is the numeraire
C_m	=	consumption of manufactures (or industrial goods)
σ	=	elasticity of substitution between manufactures
α	=	share of income spent on manufactured goods
γ	=	share of labor force working in the industrial sector
λ_j	=	share of industrial labor force working in city j
l_{ij}	=	labor required to produce variety i in city j
N	=	number of cities
n_j	=	number of varieties of manufactures in city j
x_{ij}	=	amount of variety i in city j
x_j	=	total production of manufactures of a representative producer in city j
f_j	=	fixed labor cost in city j
β_j	=	marginal labor cost in city j
ε	=	threshold value for real wage differences in simulations
P_j	=	price of a variety of manufactures in city j
w_j	=	nominal wage in city j
P_{ij}	=	P_j/t_{ij}
C_{ij}	=	consumption in city i of a variety produced in j
Y_j	=	income of city j
τ_j	=	congestion parameter with respect to n_j in fixed cost functions in city j
R_j	=	rank of city j
M_{j}	=	size of city j (size of total labor force in city j)
a_j	=	constant in the fixed (labor) cost function for city j
ϕ_j	=	fraction of agricultural labor in city j
ω_j	=	real wage in city j
ϖ	=	average real wage, $\Sigma\lambda_j\omega_j$
ρ	=	speed of adjustment of λ_j as a function of the real wage in j compared to the average real wage
t_{ij}	=	transport cost ("iceberg" type) of a shipment from j to i (t_{ii}= 1)
L_j	=	manufacturing labor in city j
Q_{Aj}	=	production of agricultural goods in city j
I_j	=	exact price index of manufactures in city j
c_{kj}	=	consumption in k of manufactured goods produced in city j
s_{ij}	=	expenditures in city i on goods from city j