Zipf's Law for Cities and Beyond

The Case of Denmark

By THORBJØRN KNUDSEN*

ABSTRACT. Zipf's law for cities is one of the most conspicuous and robust empirical facts in the social sciences. It says that for most countries, the size distribution of cities must fit the power law: the number of cities with populations greater than *S* is inversely proportional to *S*. The present paper answers three questions related to Zipf's law: (1) does the Danish case refute Zipf's law for cities?, (2) what are the implications of Zipf's law for models of local growth?, and (3) do we have a Zipf's law for firms? Based on empirical data on the 61 largest Danish cities for year 2000, the answer to (1) is NO—the Danish case is not the exception which refutes Zipf's law. The consideration of (2) then leads to an empirical test of (3). The question of the existence of Zipf's law for firms is tested on a sample of 14,541 Danish production companies (the total population for 1997 with 10 employees or more). Based on the empirical evidence, the answer to (3) is YES in the sense that the growth pattern of Danish production companies follows a clean rank-size distribution consistent with Zipf's law.

PREDICTION IN ECONOMICS, and in the social sciences generally, is a rather scarce commodity (Reder 1999) and perhaps an unattainable ideal (Aumann 2000). According to Aumann, the value of a good theory lies in its usefulness in structuring reasoning and, therefore, one empirical fact to be cited in favour of a theory is its diffusion in some population of scientists. In other words, the more use of a particular theory, the better. As Reder notes, economists tend to place higher value on technique than content; clever theoretical ideas are valued over the assiduous gathering and careful presentation of data. And since mainstream

* The author is at the University of Southern Denmark, Main Campus: Odense University, School of Business and Economics, Department of Marketing, Campusvej 55, DK5230 Odense, Denmark. E-mail: tok@sam.sdu.dk. I appreciate inspiration and comments from Rabah Amir, Laurence S. Moss, and two anonymous reviewers. All remaining errors were produced without any help.

American Journal of Economics and Sociology, Vol. 60, No. 1 (January, 2001).

economics, in any case, has a sufficiently flexible theoretical basis to rationalize contrary empirical facts also, data do not play the prominent role they do in the natural sciences. As important reasons for the gap between theory and applied work, Reder points to the rather low status of empirical facts and the tendency to use the term "prediction" when "retrodiction" would be more suitable (pp. 27–29). In contrast to this generally disappointing state of affairs (see, e.g., Reder 1999), there exists one exceptional case, a notable empirical success story in which theory must bow to facts. This case is Zipf's law for cities, which has important implications for the admissibility of theoretical growth models. In economics one very rarely finds empirical relationships which deserve to be called laws. Zipf's law for cities, however, is one of the most conspicuous empirical facts in economics and in the social sciences in general (Brakman et al. 1999). It is surely an outstanding empirical regularity deserving the status of an experimental law (Gabaix 1999).

According to Zipf's law, the growth pattern of cities almost everywhere follows the power law—the number of cities with populations greater than *S* is proportional to *1/S*. Put differently, if we rank a sample of cities according to population size, and then place the log of the size on the *X*-axis and the corresponding log of the rank on the *Y*-axis, there should appear a straight line with slope –1. Should the numerical value of the slope exceed 1, cities are more dispersed than predicted whereas a slope less than one indicates that cities are more even sized than the prediction. Suprisingly, we actually see a slope of about 1 when data on American metropolitan areas are used. Both Gabaix (1999) and Krugman (1996) obtained a slope of –1.005 (std.dev. 0.010) and an R^2 of .986 for the 135 American metropolitan areas listed in the Statistical Abstract of the United States for 1991. Similar results have been reported for most countries in contemporary times (Rosen and Resnick 1980). The support of Zipf's law for previous periods has included samples of cities in India (Zipf 1949), China (Rozman 1990), the Netherlands (Brakman et al. 1999) and the United States (Krugman 1996; Zipf 1949).

Although most evidence corroborates Zipf's law, some evidence has been reported which seems to refute the prediction of a slope of –1. Thus, Brakman et al. (1999) compare data from the Netherlands in 1600, 1900, and 1990 and, despite a very good fit for all three regression models (R^2= 0.96 or better), obtain estimates that deviate from the slope of –1 predicted by Zipf's law. In their study, only data for 1900 fit this prediction. Both the estimates for 1600 and 1990 obtain a lower value, indicating that cities are more even-sized than predicted. Inspired by this deviation, we have sampled data to test the Danish case for the year 2000. Since Denmark and the Netherlands are both small countries and to some extent comparable in development, it is interesting to ask if the Danish case will show yet another deviation from Zipf's law for cities.

At this point it should be noted that Zipf's law is a special case of what is known as the rank-size distribution, which states an inverse linear relationship between the logarithmic size of a city and its logarithmic rank without any constraints on the slope. In the study of size distributions of firms, the term "Pareto distribution" is often used synonymously with the term rank-size distribution (see, e.g., Ijiri and Simon 1977). Furthermore, Zipf's law, or rather the rank-size distribution, applies to a much wider number of phenomena than size distributions of cities (Zipf 1949). For example, rank-size distributions have been shown to fit empirical observations of relative income (Zipf 1949), the relative size of business firms (Ijiri and Simon 1977), the number of biological species per genus (Zipf 1949) and the relative frequency of a word (Estoup 1916; Zipf 1935, 1949; Irmay 1997). Moreover, Irmay (1997) showed that Zipf's law is approximately equal to Benford's (1938) logarithmic distribution of first significant digits in a table of numbers. That Zipf's law fits size distributions of cities and business firms, relative frequencies of large texts of words (in several languages) and can account for first digits in various tables of numbers begs a second question: What is the explanation for Zipf's law and why do we see it in the case of cities?

The present paper suggests that, in view of the great success of Zipf's law in accounting for the growth of cities, it is not only possible but also plausible that the growth of firms may follow Zipf's law or at least a modified version of it. Note that the related proposition that size-distributions of firms can be explained by the Gibrat assumption has previously met strong criticism (Schmalensee 1989; Sutton 1999) and has, perhaps wrongly, been rejected. Thus, the third question to be addressed in the present paper is: Does the empirical evidence on asset distributions of business firms follow Zipf's law?

We address all three questions in turn. Due to the importance of the third question, we test it on data for what is a complete sample of all Danish production firms with 10 or more employees for 1997.

Zipf's Law for Cities: The Danish Case

WILL THE DANISH CASE SHOW yet another deviation from Zipf's law for cities? As noted by Brakman et al. (1999), this question may, due to its attention to a specific member of the the family of rank-size distributions (with slope –1), obscure the more interesting question whether some rank-size distribution will fit the city-size distribution of Denmark. Nevertheless, it is a curious fact that Zipf's law at present holds true in the United States. Clearly, it would be interesting if this were the case also in Denmark, since that would indicate some generality in the underlying dynamics of city growth across countries.

In accordance with Brakman et al.'s reservations, we expect some rank-size distribution to hold for the case of Denmark but would be surprised if Zipf's law was supported. That is, when the log of the size is placed on the *X*-axis and the corresponding log of the rank is placed on the *Y*-axis, there should appear a straight line. However, it would be surprising if its slope turned out to be –1, at least if we consider that the geographical size of Denmark is comparable to the Netherlands where Brakman et al. estimated a slope well below –1 as predicted by Zipf's law.[1]

Contradicting Brakman et al.'s scepticism, Gabaix (1999) has showed that Zipf's law can be viewed as the unique steady state distribution arising from Gibrat's law (originally stated in Gibrat 1931). Gabaix further showed that city size processes may converge to Zipf's law within a relatively short period of time (100 years) in dynamic urban systems. Thus, even very young urban systems may satisfy Zipf's law as, for example, empirical evidence on U.S. cities in 1790 has shown (Zipf 1949, cf. Gabaix 1999). So, if Gabaix is right, the Dutch case is an exception, and we should therefore expect Zipf's law with an exponent close to 1 to hold in the Danish case.

The more fundamental issue that the empirical evidence on Zipf's law and other rank-size distributions point to is the underlying cause for variation in the size of cities. Thus, for regional economists, the question of why cities vary in size is a fundamental one. There has been much progress in understanding the role, existence and growth of cities (Brakman et al. 1999) and Zipf's law plays an important role as a very tight constraint on the admissible models of local growth (Gabaix 1999). In short, any theoretical explanation for the growth of cities should, according to Gabaix, be consistent with Zipf's law or, according to Brakman et al., at the very least satisfy some rank-size distribution. In the following section, we shall address this issue. In the present section, we now turn to the Danish case as illustration of Zipf's law.

To avoid terminological confusion, we use Brakman et al.'s (1999) useful distinction between Zipf's law and rank-size distributions. Thus, equation (1), below, shows Zipf's law for cities and equation (2) shows the more general rank-size distribution.[2] The log-linear version shown in equation (3) was used for empirical estimation reported in the present paper. R_j refers to the rank of city j and S_j is its size. C and K are country-specific parameters to be estimated and E is the error-term.

(1) $$R_j S_j = C, \qquad j = 1, 2, \ldots, N$$

(2) $$(R_j)^K S_j = C, \qquad j = 1, 2, \ldots, N$$

(3) $$ln(R_j) = C - K\, ln(S_j) + E_j \qquad j = 1, 2, \ldots, N$$

We obtained a sample for of the 61 largest cities in Denmark for 2000 from the Danish Statistical Bureau. Equation four shows the estimates obtained for the Danish sample.

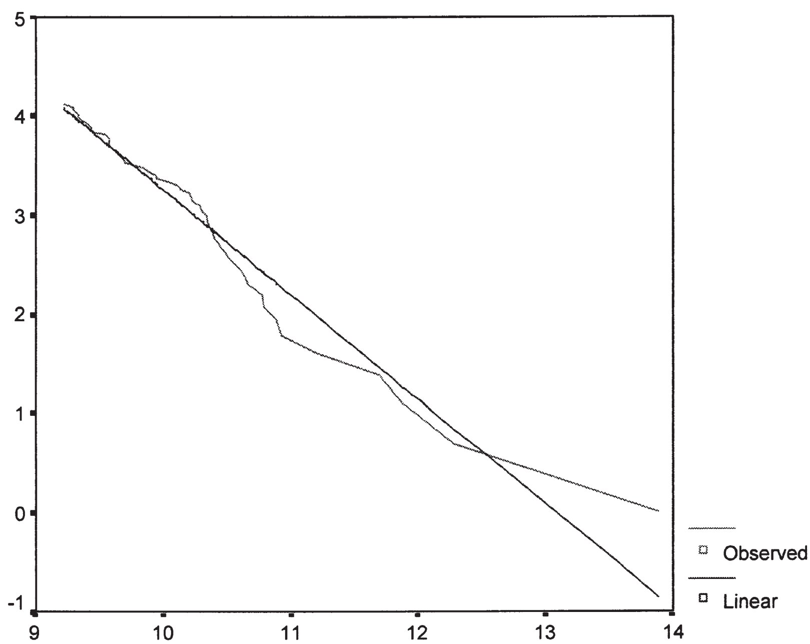
(4) $$ln(R_j) = 13.82 - \underset{(0.027)}{1.056}\, ln(S_j) + E_j \qquad j = 1, 2, \ldots, 61$$

The standard error of the estimate is in parentheses and the R^2 is 0.962, a value that indicates a good fit. Additionally, the estimated slope is reasonably close to –1.[3] According to equation (2), a useful interpretation of this finding is that the observed distribution for Danish cities is a specific instance of the rank-size distribution where the probability that the size of a given city is larger than some S is proportional to $1/S^K$ with K approximately 1 (Gabaix 1999). However, as can be seen in Figure 1, the largest Danish town, Copenhagen, as well as some mid-range towns deviate from the linear prediction. A plausible explanation is simply that capitals are peculiar objects driven by unique political forces (Ades and Glaeser 1995; Gabaix 1999). Therefore, the deviation of the Danish capital Copenhagen from the linear prediction is the exception to be expected.

As can be seen from Figure 1, there is a smaller deviation which lies in the range of about 10.2 to 10.9 for $\ln(S_j)$. Thus there are more medium-sized cities in Denmark than predicted by Zipf's law and it is this effect which is responsible for raising the exponent K slightly above 1. A further but even smaller deviation is that (the many) Danish cities

Figure 1

The 61 largest Danish cities, year 2000: X: ln(size), Y: ln(rank).



with a population less than 27,000 ($\ln(S_j)<10.2$) has an exponent less than 1. So, according to Zipf's law, there are too few small Danish cities (with 27,000 inhabitants or less), a finding that is consistent with data for the United States (Gabaix 1999). In the following sections, we shall return to an explanation provided by Gabaix (1999) for this phenomenon.

The important point to convey here is that the empirical data for the size distribution of Danish cities for 2000 provides yet another reasonable fit with Zipf's law. This statement must be supplemented with the observation that two of the three most important deviations from the linear prediction observed in the Danish case have also been reported to be present in other studies (Gabaix 1999). In sum, according to the Danish case of Zipf's law for cities for year 2000, there is no grounds for concluding that we have encountered a refuting case.[4]

Explaining Zipf's Law for Cities

ZIPF'S LAW IS ONE OF THE OUTSTANDING EMPIRICAL success stories in economics and in the social sciences in general. We have seen that the Danish case is consistent with its prediction. Not only is Zipf's law an empirical success story, it also has a rather surprising empirical regularity. What could be the explanation for the large number of empirical studies that arrive at these results? Or is the attention to Zipf's law as a specific realization of the rank-size distribution, as Brakman et al. suggest, unwarranted since the exponent is likely to reflect both time- and sample-specific variation?

From the viewpoint of social theory, the problem with Zipf's law is its experimental character. As Brakman et al. note, we still do not have a proper understanding of the underlying explanation for its observed empirical regularities. Many explanations can produce Zipf's law and many have been proposed (Carroll 1982). This is not the place to review the wide range of explanations given for Zipf's law. Rather, we shall concentrate on the most important ones. According to Gabaix (1999), arguably the two most successful models have been Steindl's (1965) and Simon's (1955) path-breaking and now classic works. We shall therefore briefly consider the rationale for Zipf's law offered by these models as well as some difficulties which have been pointed out as limits to their plausibility.

We first consider Steindl's model (1965) in which existing cities grow at a rate G and new cities are born at a rate V. Accordingly, the size distribution of cities will follow a power law with exponent $K=V/G$ such that the number of cities with size greater than S will be proportional to S^{-K} (Gabaix 1999). This explanation is clearly problematic since it demands the implausible condition $V=G$ to be satisfied. For example, this condition is clearly violated in the Danish case as well as for most other mature urbane systems where we observe $V<G$ (Gabaix 1999).

Simon's (1955) model is, according to Gabaix (1999), the most successful stochastic growth model for Zipf's law; however, it is not without its difficulties. Briefly, in Simon's model, migrants will form a new city with probability P and will go to an existing one with probability $(1-P)$. Since P is proportional to the size of the city, this model generates the familiar power law but now with the steady state exponent $K= 1/(1-P)$. There are two major difficulties with this model. As pointed out by Krugman (1996), the speed of convergence to Zipf's law is infinitely slow (since this requires P very close to 0, which again requires that existing cities are infinite at the limit). The second problem, pointed out by Gabaix (1999), is the model's implication that the rate of growth of the number of cities has to be larger than the growth rate of the population of the existing cities, a consequence which is refuted by empirical data.

Were it not that empirical data repeatedly show an exponent close to 1, both Steindl's and Simon's models would be possible contenders as an explanation for size distributions of cities. Note here that Simon's model, as Gabaix points out, may be viewed as a special case of Steindl's model, with the steady state exponent $K= V/G= 1/(1-P)$. The two models, however, imply competing empirical predictions. According to Steindl's model, $K<1$ (since $V<G$ for most urban systems) whereas Simon's model suggests $K>1$. One solution is the approach taken by Brakman et al. who view the exponent of 1 as a special case. Arguably, a more attractive solution is the one provided by Gabaix's model, which delivers Zipf's law as the limit of a stochastic process whose assumptions are consistent with empirical observations.

Now, consider Gabaix's (1999) model, which has a number of young migrants deciding in which city (populated by a number of seasoned agents) to locate. The agents maximize wage w_t and amenities a_t (which are independent and identically distributed), so the equilibrium of "utility adjusted" wages will be the same across cities ($w_t\ a_t = u_t$). Given that the increase in population of city i is ΔN_t and that agents die with probability P_d, the normalized growth G_t of city i will be:

(5a) $$G_t = \Delta N_t / N_t = f'^{-1}(u_t / a_t) - P_{d,}$$

As can be seen, Gabaix's model of city growth is an expression of what is commonly known as Gibrat's law, which simply states that growth is scale-invariant and independent of initial size, i.e., $G_{t+1} = S_{t+1} / S_t$. Since Gabaix's model implies a common and size-independent variance, it is necessary to justify this implication. This is done by assuming that the total variance of the growth rate $q^2(S)$ can be broken down into three components of which only one is size-dependent:

(5b) $$q^2(S) = q^2_{policy} + q^2_{region} + q^2_{industries}/S_,$$

It is thus assumed that the variance (but not the *level*) of the city-specific provision of public goods is size-independent and that regional shocks affect all cities of the region equally. Whereas policy and regional shocks are seen as size-independent, the shocks experienced by a city's industries should, according to Gabaix, depend on size. Again, this assumption seems reasonable because larger cities may hedge the industry-specific growth risk by diversifying their industrial portfolio. All this boils down to the needed justification that the variance of the growth process is size-independent at least for large cities (since for a large S, $q^2_{industries}/S$ tends to 0). Thus, cities in the upper tail of the size distribution should follow Zipf's law whereas small cities should deviate from it, a point which, according to the previous section, was supported by the data on Denmark.

Armed with the justification for constant variance of the growth process, Gabaix then develops a stochastic version of the equation for city growth simply by taking the continuous limit of equation (5a), which gives

(5c) $$dS_t/S_t = md_{\mathrm{t}} + qdB_{\mathrm{t}},$$

Here B_t is reflected Brownian motion (random walk with a barrier) which depends on the model parameters shown in equation 5(a) (u_t, a_t, $P_{d,}$). The expected growth in normalized size m is the difference between the growth rate $G(S)$ of city with size S_t and the mean growth rate ($m = G(S)\text{-}G_{mean}$). We are now in a position to convey Gabaix's central result, which is that Zipf's law necessarily emerges as the steady state size distribution. In terms of the above model (5c), for $S_t > S_{min}$ and $dS_t = s_t\ max[md_t + qdB_t, 0]$, Gabaix shows that the distribution converges to a Zipf distribution with exponent $K = 1/(1\text{-}S_{min}/S_{mean})$. Thus, when the minimal allowable city size tends to 0, K tends to 1 from above. According to this result, the existence of some lower barrier will induce city size to be distributed according to the power law. All that is needed for the emergence of Zipf's law, then, is the existence of some repelling force that keeps cities from becoming too small. Having presented the core of Gabaix's model, we shall briefly note that the problem of slow convergence and the possibility of deviations from an exponent of 1 is explicitly dealt with. Thus, Gabaix shows how the convergence to a steady state will be reasonably fast if the variance $q^2(S)$ is not too low and further provides an elaborate and useful analysis of the two possible causes of deviation from an exponent of 1, the mean and the standard deviation of the growth process.

In sum, Gabaix's paper presents a very useful and convincing stochastic explanation for Zipf's law and the necessary emergence of an exponent of 1 within a reasonably short time frame. The model is driven by the condition that the ratio of wages and amenities equilibrate across cities and the (infinitesimal) lower barrier to city size, which helps a size-invariant random walk converge to a power distribution with an exponent of 1 at the limit. It should be noted that Gabaix shows this result to hold for other stochastic processes as well. The crucial point is that a wide range of stochastic growth models may produce Zipf's law from the Gibrat assumption. This result is also surely important for growth models of business firms where Gibrat's law plays an important role (Sutton 1999). We shall return to this issue in the ensuing section, where a complete sample of Danish production companies for 1997 is used to test Zipf's law for business firms. However, first we shall briefly consider Brakman et al.'s model and how it relates to Gabaix's result.

Brakman et al.'s claim that the time-distribution of the exponent will follow an *n*-shaped pattern is clearly at odds with Gabaix's result. So, how does Brakman et al. arrive at this conclusion? Their main point is that the reported structural economic changes rather than random growth underlies the size distribution of cities. This size distribution might well follow the rank-size distribution; however, we should expect the exponent to change over time according to an *n*-shaped pattern. The reason for this shape is that the force of agglomeration has its strongest effect in the industrialization period. For different reasons, spreading forces have relatively higher effects in both the pre- and post-industrialization periods, with the result that small cities will have a tendency to grow more quickly. Therefore, the exponent of the rank-size distribution will peak in the industrialization period. Brakman et al. support this claim with data from the Netherlands which show an exponent of 0.55 in 1600, 1.03 in 1900 and 0.72 for 1990 and further refer to Parr's (1985) results as support for the *n*-shaped time-pattern of the exponent. Then, based on Krugman's (1991a, 1991b) general equilibrium location model, Brakman et al. proceed to develop an equilibrium model of industrial location which can mimic the data observed for the Netherlands. The model is built on reasonable assumptions and, based on simulations, quite successfully reconstructs the historical trends observed for the Dutch case. The value of this simulation exercise lies in its ability to use the parameters found important by economic historians to reproduce the expected effects in terms of the rank-size distribution as a structural outcome of the modelling exercise. A major drawback, however, is that the parameter settings, although plausible, are completely arbitrary. Nevertheless, the question remains whether an exponent of 0.72 presents an unsurmountable difficulty for stochastic models as Brakman et al. claim.

As indicated above, Gabaix's stochastic model has no difficulty in handling this problem. Gabaix's general explanation for an exponent of 0.72 would be that either the mean or the variance of the growth process deviates from Gibrat's law. Thus, Gabaix shows that if a range of cities has a high growth rate, its distribution will decay more slowly than in the pure Zipf case (the exponent will decrease) because small cities constantly feed the stock of big cities. The second cause for a small exponent Gabaix gives is that the variance of the growth process is size-dependent. The data reported by Brakman et al. for the Netherlands does not allow evaluation of these possible alternative explanations for the observed low exponent for 1990. The point is that, at least in principle, Gabaix's stochastic model can account for these deviations as deviations from Gibrat's law.

In sum, we have seen that the puzzling empirical regularity of Zipf's law with exponent 1 can be explained as the expression of the steady state size distribution arising from Gibrat's law. Given that the process has had time enough to reach a steady state (which should be the case for all mature urban systems), deviations from the exponent of 1 may be explained as deviations from Gibrat's law. Alternatively, one may side with Brakman et al. and attempt to account for the expression of power law rank-size distributions as the outcome of structural economic changes. Although the latter approach is very attractive due to its attempt to align with actual historical data, the equilibrium models provided so far essentially rely on rather arbitrary parameter settings. This is not the place to pass a verdict over the comparative value of the two approaches, only to note the differences and some of the associated pros and cons. Having presented the most important explanations for Zipf's law, this paper's next section examines the size-distribution of Danish production companies. Since Zipf's law can be viewed as the outcome of Gibrat's law, it should be interesting to see whether this also holds for the size-distribution of business firms and, if not, whether a deviation from Zipf's law may be explained within the framework of Gabaix's stochastic model.

I

Zipf's Law for Firms?

SINCE ZIPF'S LAW TURNS OUT to be the steady state distribution of the familiar Gibrat's law commonly used as a foundation for growth theory, it would not be unreasonable to expect Zipf's law also to hold for the growth of firms. As Sutton (1999) notes however, the development of the empirical literature since Gibrat (1931) has indicated that attempts to make simple generalizations about the shape of firm size distributions have been rather dubious. Sutton's reason for this is that a model of firm growth needs a rational basis. Therefore, a stochastic model has no chance of mimicking an empirical size distribution of business firms. Also, according to Gabaix (1999), the empirical question of size distributions of firms is still unresolved.

Motivated by the insights presented about Zipf's law and the need to access a comprehensive data set of size distributions of firms, we use a sample of Danish production companies for 1997. The sample is made up of the total population of 14,541 firms with more than ten employees. One suspicion regarding the failure of previous empirical tests to support the Gibrat assumption, at least for some industries, is that these failures were due to bias introduced by missing data and/ or industry-specific samples. The obvious way to rule out this problem is simply to use a sample that includes the total population rather than some more or less arbitrarily chosen subset of firms. This is the approach we followed in the present study. Furthermore, in contrast to most previous studies, we use data on assets rather than sales as a proxy for firm size. The reason for this is that we think asset accumulation comes much closer to agglomeration than sales growth does. Again, we use equation (3) to obtain the empirical estimate on basis of the full sample:

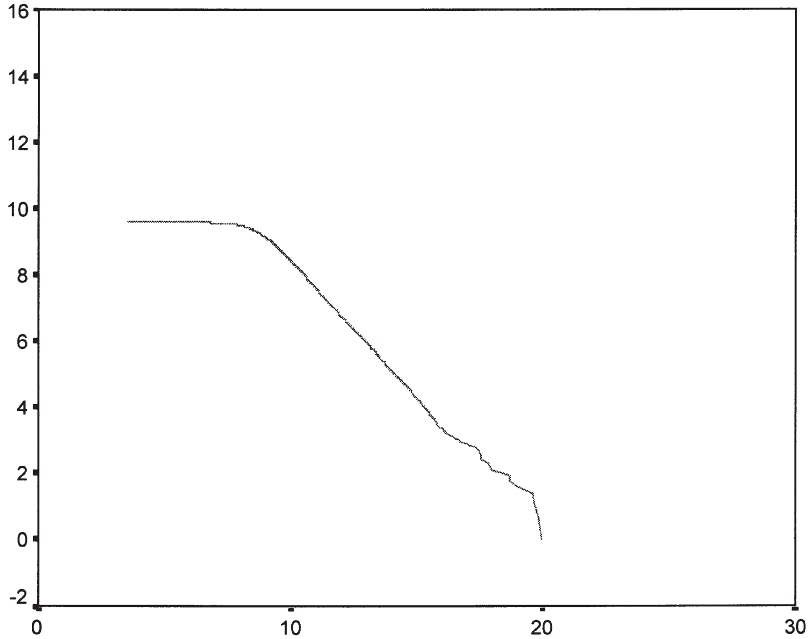
$$\text{(6a)} \qquad ln(R_j) = 14.98 - \underset{(0.0013)}{0.669}\, ln(S_j) + E_j = 1, 2, \ldots, 14{,}541$$

The standard error of the estimate is in parentheses and the R^2 is 0.942, a value that indicates a reasonable fit. The estimated slope clearly seems too far from –1 to warrant any speculation that we have stumbled over yet another instance of Zipf's law. The good fit, however, indicates that the size distribution might well be a rank-size distribution that follows the power law. If we take a look on the plot of the size distribution in Figure 2, below, we see that it has two very distinct parts.

As can be seen from Figure 2, the distribution consists of a first part with almost even-sized small firms (across all industries) and a second part (from about ln size 9.5), which portrays a straight line with negative slope. The linear fit of the second part is striking, as can be seen from the following regression which includes the 13,543 firms (93%) with assets (measured in Danish kroner) equal to or higher than ln(9.5).

Figure 2

The 1997 population of 14,541 Danish production companies with ten or more employees: X: ln(size), Y: ln(rank).



(6b) $ln(R_j) = 15.73 - 0.741\ ln(S_j) + E_j = 1, 2, ..., 13{,}543$

(0.0017)

Considering the sample size, the R^2 of 0.985 indicates an extremely good fit. There is no doubt whatsoever that the size distribution of Danish production companies excluding the 7% of smallest firms follows a rank-size distribution with exponent 0.741.

This result is quite remarkable in view of the widespread scepticism towards models of firm behaviour which are merely statistical and not based on maximizing behaviour (Sutton 1999). Thus Schmalensee (1989) notes the failure of statistical models such as Ijiri and Simon's (1977) to provide a satisfactory description of size distributions for some industries. And as noted by Sutton, there has been a long-standing concern about the theoretical basis of statistical growth models. According to Sutton (1999: 245), most authors now claim only that the size distribution will be skew, but do not specify the extent of the skewness, or its particular shape.

Sutton himself develops a model which characterizes the minimum degree of industry level inequality of firm size. He develops Ijiri and Simon's (1977) model by substituting its Gibrat law assumption with the much weaker assumption that the probability of seizing a new market opportunity is size-independent. By further assuming constant probability that the next market opportunity will be filled by a new entrant, as Ijiri and Simon assumed, Sutton arrives at a benchmark case which defines a limiting Lorenz curve. It is defined in two-dimensional space in terms of the normalized rank k and the k-firm concentration ratio C_k:

(7) $$C_k \geqq k/N\,(1\text{-}ln(k/N)).$$

The point is that this limiting Lorenz curve places a lower bound on firm inequality (in terms of the k-firm concentration ratio). Based on a game-theoretic "island model," assuming strictly independent submarkets, Sutton shows that this result also holds in the presence of strategic effects. That is, the effect of independence *between* submarkets is strong enough to override any strategic effects *within* submarkets so a minimal degree of inequality emerges in the limit. The size distribution obtained from the game-theoretic analysis is at least as unequal as the limiting Lorenz curve. So how does the present paper's very clean result obtained for the size distribution of Danish firms relate to the criticism of the empirical and theoretical failure of statistical models raised by Sutton and others? And how does the obtained evidence for Danish firms compare to Sutton's benchmark case, the limiting Lorenz curve?

The first thing to note is that our findings are consistent with, but much stronger than, the prediction offered by Sutton's model. If we compute the Lorenz curve of the size-distribution of Danish firms in the two-dimensional space used by Sutton, we find that the Lorenz curve bends much further away from the diagonal than predicted. In other words, the inequality of the firms in our distribution is much greater than indicated by Sutton's model. The next thing to note is that the part of the size distribution reported in the present paper, which excludes very small firms, clearly follows a rank-size distribution with exponent 0.741. As can be seen from Figure 2 and the R^2 of 0.985, this is a very clean result. Against widespread claims to the contrary, there is no doubt that, apart from the extreme lower tail, the size distribution of Danish production companies can adequately be summarized by a rank-size distribution. Moreover, the result is obtained on a complete sample of firms with 10 or more employees. A crucial reason for our result may well lie in the breadth of our sample. Previous studies have predominantly used more narrow samples focused on the industry level, which may explain why these studies have failed to obtained similar results. But how can we make sense of the obtained result? And what about the deviation in the lower tail and the slight deviation in the upper tail? Armed with insights provided by Gabaix, some answers to these question can be given.

The failure of statistical models to provide a satisfactory description of size distributions for some industries.

As Ijiri and Simon (1977) note, the observed regularities in business firm size distributions usually fit closely to the Pareto distribution, an outcome that follows from the Gibrat assumption that expected growth is proportional to size. Given the observed regularity, the Gibrat assumption then works as a criterion of admissibility for the class of models that aspire to explain firm growth. It is therefore of great importance if we dismiss the Gibrat assumption. As noted by Schmalensee (1989), statistical models (most based on the Gibrat assumption) have encountered persistent difficulties in providing an adequate description of the size distribution of some industries. Combined with the search for sound economic explanations of size distributions, this led to a widespread dismissal of the Gibrat assumption. The data presented in the present paper, however, suggests that the dismissal of Gibrat's law may have been premature—previous studies have analysed more narrow samples, typically defined in terms of industry. If we accept the reasonable assumption that entry fees vary greatly between industries, the left tail of the distribution, which consists of small even-sized firms, will be unevenly distributed across industries. Accordingly, an industry-level analysis will always fail to confirm Gibrat's law for some industries. The reason for the failure of previous studies may well be due to the fact that population-wide samples have not been used.

Explaining the observed size distribution.

The empirical size distribution of the population of Danish production companies with ten or more employees reported in Figure 2 has three parts. Part 1 consists of small even-sized firms with attained size ln(9.5) or less. This part contains 7% of the population. Part 2 consists of the 93% medium- and large-sized firms with a degree of inequality that shows a very clean fit with the Gibrat assumption— a rank-size distribution with exponent 0.741. Concerning this part of the distribution, Figure 2 shows some deviation from the almost perfect linear fit for the extreme upper tail of the distribution where size is ln(16) or more. We shall refer to this extreme upper tail as part 3 of the distribution and note that it contains the 30 largest Danish production companies. These 30 firms comprise 0.2% of the population in terms of number but in terms of their combined asset mass they account for 73.14% of the population's assets. Now to the outline of an explanation for the observed size distribution. We start with part 2 of the distribution. Since this part has an almost perfect fit with the rank-size distribution, it follows the Gibrat assumption. Hence, the expected growth and the variance of the growth process is size-invariant. In other words this 93% of firms experience the same shocks to the growth process.

(8) $$q^2(S) = q^2_{international} + q^2_{national} + q^2_{regional} + q^2_{industry}/S_, + q^2_{strategy}(S_k)$$

In line with the explanation given for size-independent variance of the growth process, we can break down the total variance into its components. For simplicity, we assume independent variance; however, this assumption is not essential since specifying the relevant interaction terms is straightforward. We may reasonably assume that international, national and regional shocks to the growth process will be experienced independent of firm size. By contrast, since we may adopt the quite reasonable assumption that large firms can hedge against industry-specific risk, we find that the effect of industry specific shocks decrease in size. Finally, we assume that for the *m* very large firms (rank 1, 2, ..., *k-m*), firm-specific strategic effects will introduce idiosyncratic variance into the growth process.

Thus, the proposed explanation for the observed size distribution is that the variance of the growth process for very small firms depends on industry-specific effects. With increasing size, these effects gradually trail off. This proposition is consistent with the curve observed in Figure 2 in the size range between ln(7.0) and ln(9.5). For part 2 of the size distribution, it is thus proposed that industry-specific shocks to the growth process have a minor effect compared to international, national and regional shocks. That is, excluding the largest 30 firms in the extreme right tail of the distribution, the mean and the variance of the growth process is size-independent. The size distribution observed in part 2 of the distribution is thus consistent with an underlying stochastic process which can be described in terms of equation (5c) as a random walk with a lower barrier, i.e., reflected Brownian motion.[5] Regarding the 30 largest firms found in the extreme right tail of the distribution, we suggest that firm-specific strategic effects of the sort typically studied in theories of industrial organization (see, for example, Tirole 1988) influence the variance of the growth process. Therefore for about 93% of the Danish production companies, their growth process may almost perfectly be described in terms of an underlying dynamic stochastic process which is consistent with the Gibrat assumption.

As observed by Ijiri and Simon (1977), the Gibrat assumption can be derived from the postulate that access to internal and external investment funds is proportional to size, without assuming rational choice as part of the causal mechanism. As we have seen, it is also consistent with a breakdown of variance approach.

According to Gabaix's argument, the exponent of 1 predicted by Zipf's law should emerge as the steady state distribution of Gibrat's law. So, why the deviation from the expected exponent of 1 predicted by Zipf's law? Perhaps the steady state has not been reached yet. Although it is possible that the process is not in a steady state due to ongoing pertubations, the difference beween part 1 and 2 of the observed distribution suggests an alternative explanation. As demonstrated by Gabaix, deviations from a Zipf exponent of 1 can be due to deviation in the expected growth rate and in the variance of the growth process for some range of firms. Both effects will result in an exponent less than 1. If a range of firms has a high growth rate, its distribution will be flatter. This is because the distribution will decay less quickly than in the pure Zipf case since small firms feed the stock of big firms (Gabaix 1999). It is reasonable to assume that the smallest Danish production companies (the 7% contained in part 1 of the observed distribution shown in Figure 2) have a very high expected growth rate, so it is not surprising to observe an exponent of 0.741. A second reason that could explain the observed deviation from the Zipf case, would be that a range of firms had high variance, which, due to the higher mixing of small and large firms, also results in a flatter distribution. Since this effect may well be present in the range of small firms at the extreme left tail of the distribution, there is an additional reason why we observed an exponent that was significantly less than the Zipf case.

In sum, we have presented data for the size distribution of the entire population of Danish production companies with ten or more employees and found a striking fit with the rank-size distribution with exponent 0.741 when the extreme left tail of the distribution was excluded. We presented the data for 1997 but also posess data for the four previous years. The data for the four years prior to 1997 do not show any deviation from the results reported here. Thus for about 93% of the Danish production companies, the growth process may almost perfectly be described in terms of an underlying dynamic stochastic process consistent with the Gibrat assumption. Inspired by Gabaix, we have further provided the outline of an explanation for this result. We propose that the growth process for very small firms depends strongly on industry-specific effects. With increasing size, these effects gradually trail off and what remains is international, national and regional shocks that hit all firms with equal force. Therefore, the variance of the growth process will be size-independent for all but the smallest firms. The few Big Players will, however, show firm-specific deviations due to strategic effects of the sort studied in the theory of industrial organization.

The obtained empirical result is very clean and quite surprising in view of the widespread scepticism towards statistical explanations, in part fed by the failure of previous studies to support the Gibrat assumption for some industries. As already mentioned, the present study suggests that this may well be due to sampling bias. Since the ratio of small firms to each industry is likely to vary due to the costs associated with entry, industry-level sampling and analysis is likely to face some inexplicable cases. To remedy this problem, it is recommended to use population-wide samples, as in the present study, or at the very least to use unbiased cross-industry samples.

Conclusion

THE PRESENT PAPER HAS EXAMINED one of the outstanding empirical regularities of the social sciences, Zipf's law for cities. The empirical data presented for Danish cities for year 2000 showed a Zipf exponent close to 1. This fact, combined with a sufficiently good fit, led to the conclusion that yet again we have encountered a case that is consistent with the prediction of Zipf's law.

We have further conveyed the most important explanations offered for the puzzling fact that city growth follows Zipf's law. According to Gabaix (1999), Zipf's law can be explained as the steady state distribution arising from Gibrat's law. Thus, when a stochastic process follows the Gibrat assumption, it will eventually converge to Zipf's law. And, as Gabaix showed, all that is needed to ensure that convergence will happen sufficiently quickly to make the emergence of Zipf's law plausible is that some impurity be introduced into the process, e.g., in terms of an infinitesimal lower barrier to city size. This result not only holds when city growth is modelled as reflected Brownian motion but also holds for a much more general class of models, namely any Markov process with repelling force. Alternatively, one may side with Brakman et al. (1999) and attempt to account for the expression of power law rank-size distributions as the outcome of specific structural economic changes. Even if this approach is attractive due to its attempt to align theory with actual historical data, such structural models may, as theirs does, suffer from arbitrary parameter settings due to the inclusion of unobservable variables. In sum, the two approaches may best be seen as complementary, each contributing an important facet of explanation.

Nevertheless, there is a crucial difference between Gabaix's stochastic model with its general prediction that city growth will converge to Zipf's law and Brakman et al.'s model devised to mimic the structural changes observed in a specific country. According to Brakman et al. the time-distribution of the Zipf exponent will be *n*-shaped with a peak value of 1 or more. By contrast, Gabaix's model implies that the exponent will converge to 1 unless there is deviation in the mean or variance of the growth process for some range of the distribution. This indicates that one of the most important implications for future research is to obtain better estimates of the time-distribution of the Zipf exponent in order to substantiate the veracity of the competing claims made by Gabaix and Brakman et al.

Inspired by Gabaix's result that Zipf's law can be viewed as the outcome of the Gibrat assumption, we went beyond city growth and examined the empirical size distribution of firms. Our sample included the entire population of Danish production companies with ten or more employees. Excluding the 7% smallest firms, the size distribution (assets) showed a striking fit with the rank-size distribution with exponent 0.741. This result indicates that the growth process of 93% of the Danish production companies can almost perfectly be described in terms of an underlying dynamic stochastic process consistent with the Gibrat assumption.

This result is very clean and, therefore, rather surprising in view of the widespread scepticism towards statistical explanations of size distributions of firms. A possible reason for the failure of previous studies to consistently obtain similar results may well be due to sampling bias. Since the ratio of small firms to each industry is likely to vary due to entry costs, industry-level samples will probably show apparent deviations from the Gibrat assumption, at least for some industries. Since the result obtained in the present study suggests that this problem disappears at the population-level size distribution, the obvious remedy would be to obtain population-level samples or at least unbiased cross-industry samples. Thus, the present study suggests that examination of population-level samples should be high on the agenda of empirical research on the size distributions of firms. It should also be noted that we find assets to be better proxies for size in growth models than the sales data typically used in previous studies (Sutton 1999).

Therefore, it is possible that the very clean empirical result reported in the present study partly reflects this fact.

It is further proposed that the deviation from the Zipf exponent of 1, which should be expected according to Gabaix's model, was caused by the strong influence of industry effects on very small firms. With increasing size, it is argued, this effect gradually trails off and what remains are international, national and regional shocks that hit all firms with equal force. Therefore, the variance of the growth process will be size-independent for all but the smallest firms. A second possible source of deviation from Zipf's law was the strategic interactions of the few Big Players (only 30 firms or 0.2 % of the present sample).

In sum, the present study has raised the implication that a simple stochastic model based on the Gibrat assumption adequately describes the underlying causes of persistent inequality of firm size for most firms. It may thus be premature to dismiss such models despite the scepticism raised by previous authors (see for instance Sutton 1999). Notably, this conclusion is based on empirical data sampled at the population level. Had industry-level samples been used, some industries would have shown a relatively bad fit. The important exception to this conclusion is the growth of Big Players, which may better be understood in terms of the standard tools of the theory of industrial organisation (see e.g., Sutton 1999; Tirole 1988).

Regarding theory, the value of the empirical support for Zipf's law, which may be seen as the steady state outcome of Gibrat's law, lies in the strong bounds placed on the set of models that may be used to explain city and firm growth. One could formulate this condition in terms of an impossibility theorem which states that the set of possible models of population-level city and firm growth should be consistent with Zipf's law in the long run.

Notes

1. Brakman et al. (1999) reports a slope of –0.72 (R2 = 0.96).

2. There is some disagreement in terminology here. Gabaix (1999) refers to the rank-size rule, which states an inverse linear relationship between the actual rank and size (the size of the city of rank *i* varies with *1/i*). For Brakman et al. (1999), the rank-size distribution denotes the inverse linear relation between log rank and log size.

3. The standard deviation is –0.027 and the estimate is –1.056. If we

choose significance level 0.05, it must be rejected that the true estimate is –1.000. But if we choose significance level 0.01, we *cannot* reject that the true estimate is –1.000. Therefore, I say that the estimated coefficient is reasonably close to 1 and conclude that Denmark cannot be viewed as a refuting case.

4. The empirical regularity of Zipf's law raises the question of how cities are defined and why this does not seem to make a difference. In the face of the arbitrary manner in which cities are defined, the persistent regularity of Zipf's law is rather perplexing. The city as a part of its overall Standard Metropolitan Area (SMSA) differs between and within countries. As long as cities are defined so their relative sizes are internally consistent, between-country differences can be handled by the country-specific parameter C in equation (3). Within- country differences could in principle be handled by adding a dummy to equation (3) if necessary, but this is not widely done. But why has this inconsistency in the data not made a big difference in previous estimations? The reason is probably that the estimation of Zipf's law according to equation (3) is very robust regarding variation in relative size if two conditions hold: (1) the ranking of cities must be preserved; and (2) *systematic* bias must not be introduced.

5. This requires redefining models (5a) and (5c) in terms of firms maximizing, for example, market share and quality so the equilibrium of "profit adjusted" market shares will be the same across industries.

References

Ades, A. and Glaeser, E. (1995). Trade and Circuses: Explaining Urban Giants. *The Quarterly Journal of Economics*, CX: 195–228.

Aumann, R. J. (2000). What is Game Theory Trying to Accomplish? In: Aumann, R. J. (Ed.), *Collected Papers*, Volume I: 5–46. Cambridge, MA: The MIT Press.

Benford, F. (1938). *The Law of Anomalous Numbers*. Proceedings of the American Philosophical Society, 78: 551–572.

Brakman, S., Garretsen, H., van Marrewijk, C., van den Berg, M. (1999). The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution. *Journal of Regional Science*, 39(1): 182–213.

Carroll, G. (1982). National City-Size Distribution: What Do We Know After 67 Years of Research? *Progress in Human Geography*, VI: 1–43.

Estoup, J. B. (1916). *Gammes Sténographiques*. Paris: Gauthier Villars.

Gabaix, X. (1999). Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics*, August: 739–767.

Gibrat, R. (1931). *Les Inégalités Économiques*. Paris: Librairie du Recueil Sirey.

Ijiri, Y. and Simon, H. A. (1977). *Skew Distributions and the Sizes of Business Firms*. Amsterdam: North-Holland.

Irmay, S. (1997). The Relationship Between Zipf's Law and the Distribution of First Digits. *Journal of Applied Statistics*, 24 (4): 383–393.

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

Krugman, P. (1991a). *Geography and Trade*. Cambridge, MA: The MIT Press.

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

———. (1996). *The Self-Organizing Economy*. Cambridge, MA: Blackwell.

Parr, J. B. (1985). A Note on the Size Distribution of Cities over Time. *Journal of Urban Economics*, 18: 199–212.

Reder, M. W. (1999). *Economics. The Culture of a Controversial Science*. Chicago and London: The University of Chicago Press.

Rosen, K. and Reznick, M. (1980). The Size Distribution of Cities: An Examination of the Pareto Law and Primacy. *Journal of Urban Economics*, VIII: 165–186.

Rozman, G. (1990). East Asian Urbanization in the Nineteenth Century. In Van der Woude, A. and Shishido, H.(eds.) *Urbanization in History*, pp. 61–73. New York: Oxford University Press.

Schmalensee, R. (1989). Sunk Costs and Market Structure: A Review Article. *Journal of Industrial Economics*, 40: 125–133.

Simon, H. A. (1955). On a Class of Skew Distribution Functions. *Biometrika*, XLII: 425–440.

———. (1979). On Parsimonious Explanations of Production Relations. *Scandinavian Journal of Economics*, 81: 459–474.

———. (1997). Models of Bounded Rationality, Volume 3, empirically Grounded Economics Research, Cambridge, MA: The MIT Press.

Steindl, J. (1965). Random Processes and the Growth of Firms. New York: Hafner.

Sutton, J. (1999). Technology and Market Structure. Theory and History. Cambridge, MA: The MIT Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: The MIT Press.

Zipf, G.K. (1935). *The Psycho-biology of Language*. Cambridge, MA: The MIT Press.

———. (1949). *Human Behavior and the Principle of Least Effort*. Reading, MA: Addison-Wesley.