Deviations from Zipf's Law for American cities: an empirical examination

Rafael, González-Val

Universidad de Zaragoza

23 January 2010

Deviations from Zipf's Law for American cities: an empirical examination

Rafael González-Val

Departamento de Análisis Económico

Universidad de Zaragoza

Abstract: This paper presents a simple method for calculating deviations between actual city size and the size which would correspond to it with a Pareto exponent equal to one (Zipf's Law). The results show two differentiated behaviours: most cities (80.25%) present a greater size than that which would fulfil Zipf's Law, while small cities (19.75%) tend to be too small. Our aim is to analyse the distribution element by element, using data about city characteristics from all American cities in 2000, and to explain the deviation between the size predicted by Zipf's Law and the actual size of each city. To do this a Multinomial Logit Model is used. The most important variables affecting the probability of a city presenting a negative or positive deviation are per capita income, human capital levels and the percentage of the population employed in some sectors.

Keywords: Cities, Zipf's Law, deviations, Pareto distribution, Multinomial logit.

JEL: C16, R00, R12.

Address: Facultad de CC. Económicas y Empresariales, Universidad de Zaragoza, Gran Vía 2, 50005 Zaragoza (Spain). Fax: 34976761996. E-mail: rafaelg@unizar.es.

1. Introduction

One of the stylized facts in Urban Economics is that the city size distribution in many countries can be approximated by a Pareto distribution, whose exponent is equal to one[1]. If this is the case, it can be concluded that there is evidence for Zipf's Law[2] (Zipf, 1949). This is an extensively studied empirical regularity in many countries: France (Guérin-Pace, 1995), Greece (Petrakos et al., 2000), China (Song and Zhang, 2002), Malaysia (Soo, 2007), or the United States (US) (Ioannides and Overman, 2003; Black and Henderson, 2003; Eeckhout, 2004; González-Val, 2010).

This empirical regularity has given rise to many theoretical developments explaining the fulfilment of Zipf's Law, justifying it analytically, associating it directly with an equilibrium situation and relating it to proportionate city growth (Gibrat's Law[3]), another well-known empirical regularity. Both are considered to be two sides of the same coin. While Gibrat's Law has to do with the population growth process, Zipf's Law refers to its resulting population distribution.

These models have been developed by researchers from various fields (Urban Economics, Statistical Physics, and Urban Geography) with different foundations: productivity or technology shocks (Duranton, 2007; Rossi-Hansberg and Wright, 2007), local random amenity shocks (Gabaix, 1999), asymmetric exchanges among cities based on the variation of offered commodities (Semboloni, 2008), or a random multiplicative growth (Benguigui and Blumenfeld-Lieberthal, 2007). The model due to Gabaix (1999) is especially important because it proves that Zipf's Law is the necessary steady-state city size distribution. Gabaix presents a model based on local random amenity shocks, independent and identically distributed, which through migrations between cities generate Zipf's Law. The main contribution of the work is to justify the fulfilment of Zipf's Law in that cities in the upper tail of the distribution follow similar growth processes, so that the fulfilment of Gibrat's Law involves Zipf's Law. Recently, Córdoba (2008) concludes that Zipf's Law is equivalent to Gibrat's Law under plausible conditions.

To summarize, the theoretical economic models rest on local externalities, whether amenities or shocks in production or tastes, which must be randomly distributed independently of size, and identify deviations from Zipf's Law with a distribution of these shocks which is not independent of size. Other works also show the empirical relevance of other variables distributed in a clearly heterogeneous manner, such as climate or geographical advantages (access to the sea, bridges, etc), in the growth rate of cities (Black and Henderson, 1998; Beeson et al., 2001).

These theoretical developments arise in response to numerous empirical works which explore the relationship between growth rate and Zipf's Law. In fact, a large part of this literature takes as a reference the case of the United States assuming a Pareto exponent equal to one, since Krugman (1996) used data from metropolitan areas from the Statistical Abstract of the United States and concluded that for 1991 Pareto's exponent is exactly equal to 1.005. For a dynamic analysis, Ioannides and Overman (2003) use data from metropolitan areas from 1900 to 1990 and arrive at the conclusion that Gibrat's Law holds in the urban growth processes and that Zipf's Law is also fulfilled approximately for a wide range of city sizes. However, their results suggest that local values of Zipf's exponent can vary considerably with the size of cities. Nevertheless, Black and Henderson (2003) arrive at different conclusions for the same period (perhaps because they use different metropolitan areas). Zipf's Law would be fulfilled only for cities in the upper third of the distribution, while Gibrat's Law would be rejected for any sample size. These results highlight the extreme sensitivity of conclusions to the geographical unit chosen and to sample size. To close the debate, Eeckhout (2004) demonstrates that, if we consider all cities for the period 1990 to 2000, the city size distribution follows a lognormal rather than a Pareto distribution, so that the value of Zipf's parameter is not one, as earlier works concluded, but is about 0.5, and Gibrat's Law is also fulfilled for the entire sample.

Thus, if we accept that Zipf's Law is not fulfilled when considering the distribution of American cities, we can ask what factors explain this deviation from an empirical point of view. That is, we can analyse the distribution element by element and explain the deviation between the size predicted by Zipf's Law (associated with a Pareto exponent equal to one) and the actual size of each city, using data on per capita income, employment distribution among sectors, individuals' levels of education, etc; variables which attempt to capture the influence of local externalities. This is the objective of this work, and for this purpose, data from the year 2000 are used, the first census in which the US Census Bureau offers data on all cities (places) without size restrictions.

This question has already been dealt with in the literature, albeit indirectly. On one hand, Ioannides and Overman (2003) contrast the relationship between Zipf's and Gibrat's Laws for the United States using graphical and non-parametric methods, confirming the theoretical results of Gabaix (1999): the explanation for the smaller cities' having a smaller Pareto exponent is that the variance of their growth rate is larger (deviations from Zipf's Law appear due to deviations in Gibrat's Law). On the other hand, there are also works which explore the factors influencing growth rates. For the US, Glaeser and Shapiro (2003) study what factors influence the growth rate of American cities (cities of over 25,000 inhabitants and MSAs) in the decade 1990–2000 using a very wide range of explicative variables (per capita income, average age of the residents, variables about the education level of individuals, temperature, distribution of employment among sectors, public spending per capita, etc.). According to this work, the three most relevant variables would be human capital, climate and transport systems for individuals (public or private). For previous periods, Glaeser et al. (1995) examine the growth patterns of the 200 most populous cities in the US from 1960 to 1990 in relation to various initial characteristics of the cities in 1960, and conclude that income growth and city population growth are: (1) positively related to the initial education level, (2) negatively related to initial unemployment, and (3) negatively related to the initial percentage of employment in manufacturing industries.

The approach proposed in this work is simpler and empirically more direct. The only precedent would be the work of Soo (2005), insofar as it explains the differences in the Pareto exponent between different countries using such explicative variables as per capita income, area, population, transport costs, public spending, political variables, etc., with the important difference that as it uses Pareto's exponent per country as a dependent variable, it is comparing entire distributions, while we propose to study the deviation of each of the elements within a single distribution.

Much of the recent empirical work on the size distribution of cities has focused on deviations from Zipf's law (Soo, 2005; Garmestani et al., 2007; Garmestani et al., 2008; Rozenfeld et al., 2008; Michaels et al., 2009; Ioannides and Skouras, 2009; González-Val, 2010). Moreover, other statistical distributions have been proposed instead of the Pareto distribution: lognormal distribution (Parr and Suzuki, 1973; Eeckhout, 2004), q-exponential distribution (Malacarne et al., 2001; Soo, 2007), or even the double Pareto lognormal distribution (Giesen et al., 2009). This paper is in line with this literature.

The next Section sets out the method used to calculate deviations from Zipf's Law. Section 3 presents the variables used to try to explain the deviations. Section 4 shows the empirical model used, a Multinomial Logit Model (MNLM), and analyzes the results obtained. The work ends with our conclusions.

2. Calculating the deviations

Let S be the city size (its population), distributed according to a Pareto distribution. Then, following Eeckhout (2004), the density function $p(S)$ and the accumulated probability function $P(S)$ are:

$$p(S) = \frac{a\underline{S}^a}{S^{a+1}}, \qquad \forall S \geq \underline{S}$$

$$P(S) = 1 - \left(\frac{\underline{S}}{S}\right)^a, \qquad \forall S \geq \underline{S}$$

where $a > 0$ is the Pareto exponent, and $\underline{S}$ is the population of the city at the truncation point. The relationship with the empirically observed rank R is:

$$R = \underline{N} \cdot (1 - P(S)) = \underline{N} \cdot \left(\frac{\underline{S}}{S}\right)^a,$$

where $\underline{N}$ is the number of cities above the truncation point. Taking natural logarithms, we obtain the linear specification that is usually estimated:

$$\ln R = \ln \underline{N} + a \ln \underline{S} - a \ln S + u = K - a \ln S + u,$$

where u represents a random error which we suppose to meet the standard conditions, $E(u) = 0$ and $Var(u) = \sigma^2$, and K is a constant, $K = \ln \underline{N} + a \ln \underline{S}$.

Zipf's Law is an empirical regularity, which appears when the Pareto exponent of the distribution is equal to unity ($a = 1$) and means that, ordered from largest to smallest, the size of the second city is half that of the first, the size of the third is a third of the first, and so on (see the excellent surveys of Cheshire, 1999, and Gabaix and Ioannides, 2004, for further explanation; a more interdisciplinary review is given by

Newman, 2006). When $a=1$ the above expression can be formulated in deterministic terms:

$$\ln R = \ln \underline{N} + \ln \underline{S} - \ln S^{Z} \rightarrow \ln S^{Z} = \ln \underline{N} + \ln \underline{S} - \ln R, \qquad (1)$$

where S^{Z} is the deterministic value of the population of the city when Zipf's Law is fulfilled. This expression can be directly brought back to the rank-size rule $\left(S = \frac{\overline{S}}{R} \right)$, where $\overline{S}$ is the population of the largest city:

$$\ln S^{Z} = \ln \underline{N} + \ln \underline{S} - \ln R \rightarrow \ln S^{Z} = \ln(\underline{S} \cdot \underline{N}) - \ln R = \ln \overline{S} - \ln R = \ln\left(\frac{\overline{S}}{R} \right),$$

although it is preferable to leave it in terms of the size of the smallest city, as the most populous is always bigger than predicted by Zipf's Law (Soo, 2005), for various reasons (especially political[4]).

However, if the estimated parameter is other than one and the errors are not white noises, we would obtain an estimated size for each city:

$$\ln R = \ln \underline{N} + \hat{a} \ln \underline{S} - \hat{a} \ln S + \hat{u} \rightarrow \ln S = \frac{1}{\hat{a}} \cdot \ln \underline{N} + \ln \underline{S} - \frac{1}{\hat{a}} \cdot \ln R + \frac{1}{\hat{a}} \cdot \hat{u}. \quad (2)$$

One of the key points of this methodology is that residues ($\hat{u}$) are not white noises because, by definition, they represent the difference between the actual distribution and the power law fit. Although there are not many works on this issue, a classical reference is the method of residues by Coleman (1964). The method of residues has been applied to urban modelling by Batty and March (1976) when looking at deviations from a null gravity hypothesis for real data.

Subtracting (2) from (1), we obtain a relationship between the size which fulfils Zipf's Law ($\ln S^{Z}$ and $a=1$), and the actual size of the city and the estimated value of the Pareto exponent ($\ln S$ and $\hat{a}$):

$$\ln S^Z = \ln S + \left(\frac{1}{\hat{a}} - 1\right)(\ln R - \ln \underline{N}) - \frac{1}{\hat{a}} \cdot \hat{u} \, ,$$

$$\ln\left(\frac{S^Z}{S}\right) = \left(\frac{1}{\hat{a}} - 1\right)(\ln R - \ln \underline{N}) - \frac{1}{\hat{a}} \cdot \hat{u} \, . \qquad (3)$$

Graphically, $\ln(S^Z/S)$ represents the distance between the two distributions, the actual and the Pareto distribution corresponding to Zipf's Law. Figure 1 displays histograms for both distributions, for the year 2000. Figure 2 represents the estimation of both density functions through an adaptive kernel. Finally, Figure 3 shows the sample values of $\ln(S^Z/S)$. The calculation is done by applying (3), and using the entire distribution, 25,000 cities, from New York City with a population of 8,008,278 inhabitants to Paoli town with 42 inhabitants. The OLS[5] estimate of the Pareto exponent is $\hat{a} = 0.534$ [6].

As Eeckhout (2004) shows, the actual city size distribution comes close to being a lognormal when the whole sample is considered, and is found to be above the Pareto density function for almost all sample sizes. In fact, Wilcoxon's Rank-sum test offers a p-value of 0.4168 when considering the entire distribution, offering evidence in favour of the null hypothesis of lognormality[7].

However, for very small cities the behaviour is opposite. This indicates that, in general, cities will have a larger size than would guarantee the fulfilment of Zipf's Law, except for the smallest cities, whose size is much smaller than would correspond to a Pareto distribution. This can be seen in Table 1, showing the values of S^Z/S for the largest and smallest cities. It is also notable that for larger cities, in the upper-tail of the distribution, deviations are reduced until they almost[8] disappear, agreeing with the general consensus that Zipf's Law is a phenomenon which mainly appears when considering larger cities. Recently, Eeckhout (2009) has shown that in the upper tail both distributions, Pareto and lognormal, can be valid[9]. Figure 2 also shows that there is a point where both density functions cross, after which the actual size is always larger than the size which would fulfil Zipf's Law, although in the upper tail of the distribution both density functions again become closer[10].

Some recent papers give an economic explanation for deviations from Zipf's Law. Rossi-Hansberg and Wright (2007) identify the standard deviation of industrial productivity shocks as the key parameter which determines dispersion in the city size distribution, Eeckhout (2004) presents a model which also relates the migration of individuals between cities with productive shocks, obtaining as a result a lognormal and non-Pareto distribution of cities, and Duranton (2007) offers a model of urban economics with detailed microeconomic foundations for technology shocks, which are the fundamental drivers of the distribution of city sizes in the steady state.

3. Data description

There are two main issues in every study on city size distribution: city definition (Rosen and Resnick, 1980; Soo, 2005) and sample size (Parr and Suzuki, 1973; Eeckhout, 2004). Any study that deals with issues relating to city size distribution faces the problem of what is meant by the term "city", as there are various ways of defining a city. Moreover, the geographical unit chosen is closely related to sample size. For example, if data come from Metropolitan Areas, you are imposing an implicit truncation point, because in the US, to qualify as a Metropolitan Statistical Area a city needs to have 50,000 or more inhabitants, or the presence of an urbanised area of at least 50,000 inhabitants, and a total metropolitan population of at least 100,000 (75,000 in New England).

Cheshire (1999) summarizes three possible criteria for sample size selection: a fixed number of cities or a size threshold (Rosen and Resnick, 1980) or a size above which the sample accounts for some given proportion of a country's population (Wheaton and Shishado, 1981). Against this background, Eeckhout (2004) demonstrates the statistical importance of considering the whole sample. If the underlying distribution is the lognormal distribution, then the estimate of the parameter of the Pareto distribution is increasing in the truncation city size and decreasing in the truncated sample population.

Following Eeckhout (2004), we use data for all cities in the United States, without imposing any minimum population truncation point, as our proposal is to cover the entire distribution. Although MSAs make more economic sense[11], the reason to consider all cities (with legally determined boundaries) is purely statistical: any (explicit or implicit) truncation point leads to biased estimates of the Pareto exponent. We identify cities as what the US Census Bureau calls places[12]. Since the 2000 census, this generic denomination includes all incorporated and unincorporated places. The source of data is the 2000 census[13].

The US Census Bureau uses the generic term 'incorporated place' to refer to a type of governmental unit incorporated under state law as a city, town (except for the New England states, New York, and Wisconsin), borough (except in Alaska and New York), or village and having legally prescribed limits, powers, and functions. On the other hand, there are 'unincorporated places' (which were renamed Census Designated Places, CDPs, in 1980), which designate statistical entities, defined for each decennial census according to Census Bureau guidelines, comprising a densely settled concentration of population that is not within an incorporated place, but is locally identified by a name. Evidently, the geographical boundaries of unincorporated places may change if settlements move, so that the same unincorporated place may have different boundaries in a different census. They are the statistical counterpart of incorporated places. The difference between them in most cases is merely political and/or administrative. Thus for example, due to a state law of Hawaii there are no incorporated places in that state; they are all unincorporated.

The US Census Bureau has established size restrictions for the inclusion of unincorporated places, with the main criterion being that they have more than 1,000 inhabitants. The 2000 census is the first to include them all, without size restrictions, and this is why we take only this year. However, there are no data for some of the explicative variables for all cities, slightly reducing the sample size to 23,519 cities. However, the range of city sizes is as wide as possible, from cities of 76 (like Weatherby town, MO) or 91 inhabitants (Alton city, KS) to the largest, New York City, NY, with a population of 8,008,278 inhabitants. Medium size cities are, for example, Colville city, WA (5,004 inhabitants), Lexington city, NE (10,014), or Lake Forest city, IL (20,018).

The chosen explicative variables coincide with those of other studies on urban growth in the United States and city size. These are variables whose influence on city size has been tested empirically by other works (Glaeser et al., 1995; Glaeser and Shapiro, 2003), although our endogenous variable is completely different. We can group them into three types of variables: city characteristics and local external effects variables, human capital variables and productive structure variables. Table 2 presents the variables and gives some descriptive statistics. It is notable that, in general, the standard deviations are fairly high, showing great heterogeneity among the variables chosen when considering all places.

The first group of variables basically aims to gather some of the city characteristics and costs of urban congestion. In the first place we monitor the economic size of the city using per capita income in 1999; it would make no sense to include the population again, as it has already been used to calculate deviation. We also include two variables which reflect the age of the city: the variable "percent housing units built 1939 or earlier" which we use as a proxy for the physical age of cities, and the variable "total population: median age", which reflects the age of the city's inhabitants.

One of the most typical congestion costs is the increased cost of housing as the city size increases (taking into account that the supply of housing tends to be fairly rigid and responds slowly to increases in demand). Glaeser et al. (2006) analyse the role of the housing supply in urban and productivity growth in the US. We attempt to capture this effect through the variable "percent owner-occupied housing units with a mortgage; contract to purchase; or similar debt", as it is to be expected that as housing prices rise, more individuals will be obliged to incur mortgages or other debts. Commuting costs are another characteristic congestion cost of urban growth and are included explicitly in some theoretical models. That is, the idea that as the population of a city grows, so do the costs in terms of time for individuals to get from their homes to their places of work. To capture this effect we use the variable "workers 16 years and over who did not work at home: Median travel time to work (in minutes)".

The last two variables in this group refer to the division produced in United States cities depending on whether they are built around public transport or private cars. As Glaeser and Kahn (2004) show, in the last few decades, the model of United States cities has been characterised by being built around private cars, while public transport loses importance.

Regarding human capital variables, there are many works demonstrating the influence of human capital on city size, as cities with individuals with higher levels of human capital tend to grow more. Simon and Nardinelli (2002) analyse the period 1900–1990 for the US and conclude that cities with individuals with greater levels of human capital tend to grow more, and Glaeser and Saiz (2003) analyse the period 1970–2000 and show that this is due to skilled cities being more productive economically. We take two human capital variables: "percent population 25 years and over: High school graduate (includes equivalency) or higher degree" and "percent population 25 years and over: Some college or higher degree". The former represents a wide concept of human capital, while the second centres on high educational levels (some college, Associate degree, Bachelor's degree, Master's degree, Professional school degree, and Doctorate degree).

The third group of variables, referring to productive structure, contains the sectorial distribution of employment. The distribution of labour among different productive activities provides valuable information on other aspects of a city. Thus, the level of employment in the primary sector (agriculture; forestry; fishing and hunting; and mining) also represents by proxy the natural physical resources of the city (farming land, sea, etc.). This is also a sector which, like construction, is characterised by constant or even decreasing returns to scale.

Employment in manufacturing informs us of the level of local economies of scale in production, as this is a sector which normally presents increasing returns to scale. The level of pecuniary externalities also depends on the size of the industrial sector. Marshall put forward that: (i) the concentration of firms of a single sector in a single place creates a joint market of qualified workers, benefiting both workers and firms; (ii) an industrial centre enables a larger variety at a lower cost of concrete factors needed for the sector which are not traded, and (iii) an industrial centre generates knowledge spillovers. This approach forms part of the basis of economic geography models, along with circular causation: workers go to cities with strong industrial sectors, and firms prefer to locate nearer larger cities with bigger markets. Thus, industrial employment also represents a measurement of the size of the local market. Another proxy for the market size of a city is employment in commerce, whether retail or wholesale.

4. Empirical model and results

4.1 Empirical model

Unfortunately we have data for a single period only, the year 2000, as the census for the year 2000 is the first to include all places without size restrictions, and we wanted to consider the entire sample[14]. Also, our endogenous variable, deviation from the size which satisfies Zipf's Law, presents two clearly differentiated behaviours, so that the interpretation of the influence of either of the explicative variables cannot be unequivocal, as happens with standard regressions. We define the deviation from the size of Zipf's Law size as $\ln\left(S^{Z}/S\right)$, which is calculated with equation (3). This specification implies that the deviation for cities with a larger size than would fulfil Zipf's Law will be negative. This is the majority case, as shown in Figure 3. To be concrete, of 23,519 cities in the sample, 18,874 present a negative deviation (80.25%). Meanwhile, for the remaining 4,645 cities (19.75% of the sample), the deviation is positive as their sizes are less than would fulfil Zipf's Law.

All of this leads us to use a Multinomial Logit Model (MNLM), which solves the problems described above. It consists of transforming our dependent variable into categories, enabling us to differentiate specifically between the two behaviours observed (positive and negative deviations). As a consequence, that the results of the estimations will give us information about the probability (but not causality) of each variable affecting each category.

Based on the deviations $\ln\left(S^Z/S\right)$ calculated from equation (3), we construct four categories $\left(K=1,2,3,4\right)$ by applying the following criterion, which allows for the way that $\ln\left(S^Z/S\right)$ ranges from -2 to 2, as shown in Figure 3:

$$\begin{cases} \text{Strong negative deviation} \rightarrow & K=1 \text{ if } -2<\ln\left(S^Z/S\right)<-1.2 & \text{(7,039 places)} \\ \text{Medium negative deviation} \rightarrow & K=2 \text{ if } -1.2<\ln\left(S^Z/S\right)<-0.6 & \text{(7,007 places)} \\ \text{Weak negative deviation} \rightarrow & K=3 \text{ if } -0.6<\ln\left(S^Z/S\right)<0 & \text{(4,828 places)} \\ \text{Positive deviation} \rightarrow & K=4 \text{ if } 0<\ln\left(S^Z/S\right)<2 & \text{(4,645 places)} \end{cases}$$

This criterion enables us to differentiate between those cities presenting a negative deviation (80.25%), whose size is greater than that predicted by Zipf's Law – grouped in categories 1, 2 and 3 – and those cities (19.75%) for which the deviation is positive, as their size is less than the size fulfilling Zipf's Law. This particular grouping also ensures that the groups are as homogeneous as possible in size.

With the MNLM, we estimate a separate binary logit for each pair of categories of the dependent variable. Formally, the MNLM can be written as:

$$\ln \phi_{m|b} = \ln \frac{\Pr\left(K=m|\mathbf{x}\right)}{\Pr\left(K=b|\mathbf{x}\right)} = \mathbf{x}'\beta_{m|b} \qquad \text{for } m = 1 \text{ a } J\text{ ,} \qquad (4)$$

where b is the base category (in our case this is category 1, as it contains more cities), $J=4$ and $\mathbf{x}$ is the vector of explicative variables, reflecting local external effects, human capital or productive structure[15]. We propose to study how these explicative variables affect the probability of a city being in one or another category, that is, presenting a positive or negative deviation (greater or smaller). For example, if the percentage of the population with higher education (some college or higher degree) increases, does this increase the probability of the city size being larger than the size it would have if Zipf's Law were fulfilled? And, if so, will we be able to know if this is a strong, medium or weak deviation (which of the three categories with a negative deviation will be most likely)?

To deal with these questions we use odds ratios (also referred to as factor change coefficients). Holding other variables constant, the factor change in the odds of outcome m versus outcome n as x_i increases by δ equals:

$$\frac{\phi_{m|b}(\mathbf{x}, x_i + \delta)}{\phi_{n|b}(\mathbf{x}, x_i)} = e^{\beta_{i,m|n}\delta}. \qquad (5)$$

Thus, if the amount of change is $\delta = 1$, the odds ratio can be interpreted as follows: for a unit change in x_i, it is expected that the odds of m versus n change by a factor of $e^{\beta_{i,m|n}}$, holding all other variables constant.

4.2 Results

The estimated values of the β coefficients are shown in Table 3; 1 is the base outcome. This model includes many coefficients, making it difficult to interpret the effects for all pairs of categories. To understand the effect of a variable, one needs to examine the coefficients for comparisons among all pairs of outcomes.

To simplify the analysis, odds-ratio plots have been developed, shown in Figures 4, 5 and 6 for each of the three groups of variables. An odds-ratio plot makes it easy to quickly see patterns in results for even a complex MNLM. Moreover, to analyze the effect of each variable on the change in probability of a city being in one category or another, Table 4 shows the marginal effects for each category and the absolute average change in probability.

In an odds-ratio plot, the independent variables are each represented on a separate row, and the horizontal axis indicates the relative magnitude of the β coefficients (see Table 3) associated with each outcome. The numbers which appear (1, 2, 3 or 4) are the four possible outcomes, the categories which we previously constructed. The additive scale on the bottom axis measures the value of the $\beta_{i,m|n}$ s. The multiplicative scale on the top axis measures the $e^{\beta_{i,m|n}}$ s. The 1s are stacked on top of one another because the plot uses outcome 1 as its base category (as this has most cities) for graphing the coefficients.

These plots reveal a great deal of information (for more details, see Long and Freese, 2006). To begin, if a category is to the right of another category, it indicates that increases in the independent variable make the outcome to the right more likely. Also, the distance between each pair of categories indicates the magnitude of the effect. And when a line connects a pair of categories this indicates a lack of statistical significance for this particular coefficient, suggesting that these two outcomes are “tied together”.

City characteristics and External local effects variables

Table 4 shows that the variable presenting the greatest absolute average change in probability (0.0428) is the per capita income in 1999. Also, Figure 4 shows how, given an increase of one unit in the logarithm of per capita income, the most likely category is, by a long way, 1 (strong negative deviation). This means that increases in the per capita income of the city increase the probability of a strong negative deviation, that is, that larger cities in economic terms will probably be cities with a much larger population than predicted by Zipf’s Law.

Regarding the variable which we use to try to reflect commuting costs, “median travel time to work (in minutes)”, at first glance the effect is the opposite of what we expected. In principle, the larger the size of the city, the longer the median travel time which workers must bear. However, Figure 4 shows category 4 as more likely, which would indicate that, given an increase of one unit of the median travel time, the most likely outcome is that the size of this city will be less than would correspond with a Pareto exponent equal to one. Therefore, this probability must be interpreted the other way around: the probability of the median travel time to work increasing is greater in smaller cities, as in very large cities it is very possible that commuting costs are close to their maximum value.

It should also be noted that the two variables used as proxies for age of the cities present very similar behaviour. In both cases, the greater the average age of the total population or “percent housing units: Built 1939 or earlier”, the greater the probability of the city presenting a positive deviation (category 4). That is, the cities with the oldest inhabitants or which were founded before “1939 or earlier” have a greater probability of having a population lower than would correspond to a Pareto exponent equal to one. One possible explanation[16] is administrative; if boundaries are very slowly adjusted via annexation, the older cities are more likely to be hemmed in and less likely to adjust, and more likely to be below their size as predicted by Zipf’s Law.

Neither do the two variables introduced to reflect the division produced in US cities depending on whether they were built around public transport or private cars show clearly differentiated behaviour, and for some categories they even lack of statistical significance. The signs of marginal effects by category coincide in both variables, although the variable “public transportation” presents a higher absolute average change in probability (0.0056 versus 0.0016). Figure 4 shows that for both variables the most likely outcome is a medium negative deviation (category 2).

Finally, the variable “percent owner-occupied housing units with a mortgage; contract to purchase; or similar debt”, which we use as a proxy for urban congestion costs through housing prices, presents the expected behaviour. As the price of housing increases, more individuals are obliged to resort to mortgages or similar debts. Figure 4 shows category 1 (strong negative deviation) as the most likely outcome; this indicates that the cities with very high housing prices, and thus a high congestion cost, are a long way from the size predicted by Zipf’s Law.

Human capital variables

The results show opposing behaviour for the two human capital variables we introduced. Thus, the signs of marginal effects by category (Table 4) are opposite and Figure 5 shows how the most likely outcomes are the opposite categories (1 and 4), a strong negative deviation and a positive deviation. Thus, increases in the more educated percentage of the population makes it more likely that the city size is much larger than the size fulfilling Zipf's Law, while if there is a higher percentage with further education in its wider human capital sense (high school graduate or higher degree) the most likely outcome is that the city will be smaller than would correspond with a Pareto exponent equal to one[17]. This result must be seen in relation to that obtained for per capita income, as education is usually closely related to per capita income. We have seen how, with increases in both variables, the most likely outcome is that city size will be much higher than predicted by Zipf's Law.

Productive structure variables

Table 4 shows that the sector of activity presenting the greatest absolute average change in probability (0.0131) is the primary sector (agriculture; forestry; fishing and hunting; and mining). If we interpret this variable as a proxy for the natural physical resources available to the city (farmlands, sea, etc.), Figure 6 shows category 4 (positive deviation) as the most likely outcome by a large margin. That is, more natural resources and higher employment in the primary sector mean a higher probability of city size being lower than would fulfil Zipf's Law. This result coincides with the traditional interpretation of employment in the agricultural sector in theoretical models as a force for dispersion of economic activity, since the pioneering paper of Krugman (1991).

The other employment sector identified as a dispersing force is construction, as it is closely related to housing space[18]. The results show that the variable "construction" has a similar effect. Figure 6 shows category 4 (positive deviation) as the most likely outcome. Thus, the larger the percentage of labour employed in construction, the greater the probability that city size will be less than would correspond to a Pareto exponent equal to one. Although in Figure 6, categories 3 and 4 are joined by a line, indicating that an increase of 1% in “construction” makes outcome 4 more likely than categories 1 and 2, regarding category 3 (weak negative deviation) the effect is not significant.

However, in the case of employment in manufacturing, a sector which usually presents economies of scale, Figure 6 shows category 3 as the most likely outcome, a weak negative deviation (although the effect on category 4 is not significant). Thus, an increase in industrial employment increases the probability of the city being larger than would correspond with a Pareto exponent equal to one, although the deviation from Zipf’s Law will be small.

In services, we can see differentiated behaviour. Increases in the percentage of employment dedicated to “finance; insurance; real estate, rental and leasing” and “wholesale and retail trade” increase the probability of the city size being much larger than the size fulfilling Zipf’s Law (strong negative deviation), while if employment increases in “educational, health and social services” or in “public administration” the most likely outcome is a weak negative deviation (category 3). Again, this result must be seen in relation to that obtained for per capita income, as the activities “finance; insurance; real estate, rental and leasing” and “wholesale and retail trade” depend directly on the size of the local market, so that the percentage of employment in these activities will be higher in cities with higher per capita income. In contrast, cities with lower per capita income will have a higher employment percentage in social services.

5. Conclusions

City size distribution has been the subject of numerous empirical investigations by urban economists, statistical physicists, and urban geographers. Zipf's Law, an empirical regularity related to the shape of the city size distribution, has also attracted much attention from social science and physics researchers. Zipf's Law states that the exponent of the Pareto distribution is equal to unity and means that, ordered from largest to smallest, the size of the second city is half that of the first, the size of the third is a third of the first, and so on.

Recently, Eeckhout (2004) has demonstrated that, when the entire sample is considered, the size distribution of US cities in the year 2000 follows a lognormal and not a Pareto distribution. In this work, we present a simple method for calculating deviations city by city in relation to their size and the size which would correspond to a Pareto exponent equal to one (Zipf's Law). Our objective is to analyse the distribution element by element and to explain the deviation from Zipf's Law using data for each city on per capita income, distribution of employment among sectors, level of education of individuals, etc.; variables which try to capture the influence of local externalities. For this purpose, a Multinomial Logit Model is used, enabling us to discover the influence of each of these variables in terms of probability.

The results show two differentiated behaviours. Of the 23,519 cities of the sample, 18,874 present a negative deviation (80.25%), meaning they present a greater size than that fulfilling Zipf's Law. The variables which increase the probability that cities present this type of deviation are: per capita income in 1999; percent owner-occupied housing units with a mortgage, contract to purchase or similar debt (which we use as a proxy for the cost of urban congestion through housing cost); higher levels of human capital (some college or higher degree); and employment in certain services (finance; insurance; real estate, rental and leasing; and wholesale and retail trade).

Meanwhile, the size of the remaining 4,645 cities (19.75% of the sample) is lower than would fulfil Zipf's Law (which we define as a positive deviation). In this case, the variables which increase the probability of presenting a positive deviation are: variables measuring the age of the city (whether the median age of inhabitants or physical age of buildings, determined as percent housing units built 1939 or earlier); the percentage of the population educated from a wider human capital point of view (high school graduate or higher degree); and employment in productive sectors with constant or decreasing returns to scale (agriculture; forestry; fishing and hunting; mining; and construction).

Acknowledgments

The author has benefited from the helpful comments of Luis Lanaspa, Fernando Sanz, and Marcos Sanso-Navarro on an earlier version of this paper. The assistance of Miriam Marcén with the Multinomial Logit Model is greatly appreciated. The author acknowledges financial support from the Spanish Ministerio de Educación y Ciencia (SEJ2006-04893/ECON and ECO2009-09332 projects, and AP2005-0168 grant from FPU programme), the Regional Government of Aragón (ADETRE research group) and FEDER.

Endnotes

[1] However, the values of the Pareto exponent vary greatly between countries (Rosen and Resnick, 1980; Soo, 2005). And recent works demonstrate its sensitivity to the geographical unit chosen and the sample size (Eeckhout, 2004).

[2] Although Auerbach had previously observed in 1913 the Pareto pattern of city size distribution.

[3] Gibrat (1931) observed that the size distribution (measured by sales or number of employees) of firms tends to be lognormal, and his explanation was that the growth process of firms could be multiplicative and independent of the size of the firm. In the field of urban economics, Gibrat's Law, especially since the 1990s, has given rise to numerous empirical studies contrasting its validity for city size distributions, arriving at a majority consensus, though not absolute, that it holds in the long term.

[4] Ades and Glaeser (1995) find that the main city will tend to be more dominant the more political instability there is in a country and the more authoritarian is its regime. Main cities tend to be 41% larger if they are also capitals.

[5] Gabaix and Ioannides (2004) show that the Hill (Maximum Likelihood) Estimator is more efficient if the underlying stochastic process is really a Pareto distribution. This is not the distribution that the data follow, and so we use the OLS estimator because, when the size distribution of cities does not follow a

Pareto distribution, the Hill estimator may be biased (Soo, 2005). At the same time, the OLS estimate also presents some problems, see Goldstein et al. (2004) and Nishiyama et al. (2008).

[6] This value coincides with that obtained by Eeckhout (2004).

[7] Wilcoxon's test (rank-sum test) is a non-parametric test for assessing whether two samples of observations come from the same distribution. The null hypothesis is that the two samples are drawn from a single population and that their probability distributions are therefore equal, in our case, the lognormal distribution. Wilcoxon's test has the advantage of being appropriate for any sample size.

[8] The estimated Pareto exponent is close to the value 1, but does not equal unity. For the 100 largest cities, $\hat{a} = 1.32$.

[9] Malevergne et al. (2009) perform the uniformly most powerful unbiased test for the null hypothesis of the Pareto distribution against the lognormal.

[10] In the sample, this point corresponds to cities with 310 inhabitants.

[11] Probably the best units to describe the spatial distributions of the populations are the metropolitan areas constructed using the City Clustering Algorithm developed by Rozenfeld et al. (2008).

[12] A third option, intermediate between places and Metropolitan Areas, involves taking the urbanized areas (Garmestani et al., 2008). An urbanized area, according to the Census Bureau, consists of a central place(s) and adjacent territory with a general population density of at least 1,000 people per square mile of land area that together have a minimum residential population of at least 50,000 people. Therefore, data from urbanized areas also impose an implicit truncation point.

[13] The US Census Bureau offers information on a wide range of variables for different geographical levels, available through its website: www.census.gov/main/www/cen2000.html.

[14] This involves possible endogeneity and simultaneity problems for any regression we might attempt. On the other hand, a static analysis has the advantage of avoiding a spurious effect due to a change in administrative boundaries.

[15] The MNLM makes the assumption known as the independence of irrelevant alternatives (IIA). In this model: $\ln \frac{\Pr(K = m|\mathbf{x})}{\Pr(K = n|\mathbf{x})} = e^{\mathbf{x}'(\beta_{m|\mathbf{b}} - \beta_{n|\mathbf{b}})}$, where the odds between each pair of alternatives do not depend on other available alternatives. Thus, adding or deleting alternatives does not affect the odds between the remaining alternatives. The assumption of independence follows from the initial assumptions that the disturbances are independent and homoscedastic. We have considered one of the most common tests developed for testing the validity of the assumption, the Small-Hsiao test (1985), and we could not reject the null hypothesis, that is, the odds are independent of other alternatives, indicating that the MNLM is appropriate.

[16] The author thanks one anonymous referee for this argument.

[17] These results coincide with those of other studies analysing the influence of education in city growth. Glaeser and Shapiro (2003) also find that workers have a different impact depending on their education level (high school or college). In their sample of cities the different effect is entirely due to the impact of California.

[18] Helpman (1998) and Tabuchi (1998) introduce housing as a dispersion force into Krugman's model.

References

[1] Ades, A. F., and E. L. Glaeser, (1995). Trade and circuses: explaining urban giants. The Quarterly Journal of Economics, 110: 195–227.

[2] Batty, M., and L. March, (1976). The method of residues in urban modelling. Environment and Planning A, 8: 189–214.

[3] Beeson, P.E., D. N. DeJong, and W. Troesken, (2001). Population Growth in US Counties, 1840–1990. Regional Science and Urban Economics, 31: 669–699.

[4] Benguigui, L., and E. Blumenfeld-Lieberthal, (2007). A dynamic model for city size distribution beyond Zipf's law. Physica A: Statistical Mechanics and its Applications, 384(2): 613–627.

[5] Black, D., and V. Henderson, (1998). Urban evolution in the USA. Brown University Working Paper No. 98-21.

[6] Black, D., and V. Henderson, (2003). Urban evolution in the USA. Journal of Economic Geography, 3: 343–372.

[7] Cheshire, P., (1999). Trends in sizes and structure of urban areas. In: Handbook of Regional and Urban Economics, Vol. 3, P. Cheshire and E. S. Mills, eds. Amsterdam: Elsevier Science, Chapter 35, pp.1339–1373.

[8] Coleman, J. S., (1964). Introduction to Mathematical Sociology. Free Press of Glencoe, New York.

[9] Córdoba, J. C., (2008). A Generalized Gibrat's Law. International Economic Review, 49(4): 1463–1468.

[10] Duranton, G., (2007). Urban Evolutions: The Fast, the Slow, and the Still. American Economic Review, 97(1): 197–221.

[11] Eeckhout, J., (2004). Gibrat's Law for (All) Cities. American Economic Review, 94(5): 1429–1451.

[12] Eeckhout, J., (2009). Gibrat's Law for (All) Cities: Reply. American Economic Review, 99(4): 1676–1683.

[13] Gabaix, X., (1999). Zipf's Law for cities: An explanation. Quarterly Journal of Economics, 114(3): 739–767.

[14] Gabaix, X., and Y. M. Ioannides, (2004). The evolution of city size distributions. In: Handbook of urban and regional economics, Vol. 4, J. V. Henderson and J. F. Thisse, eds. Amsterdam: Elsevier Science, North-Holland, pp.2341–2378.

[15] Garmestani, A. S., C. R. Allen, and C. M. Gallagher, (2008). Power laws, discontinuities and regional city size distributions. Journal of Economic Behavior & Organization, 68: 209–216.

[16] Garmestani, A. S., C. R. Allen, C. M. Gallagher, and J. D. Mittelstaedt, (2007). Departures from Gibrat's Law, Discontinuities and City Size Distributions. Urban Studies, 44(10): 1997–2007.

[17] Glaeser, E. L., J. Gyourko, and R. E. Saks, (2006). Urban growth and housing supply. Journal of Economic Geography, 6: 71–89.

[18] Glaeser, E. L., and M. E. Kahn, (2004). Sprawl and Urban Growth. In: Handbook of urban and regional economics, Vol. 4, J. V. Henderson and J. F. Thisse, eds. Amsterdam: Elsevier Science, North-Holland, Chapter 56, pp.2482–2527.

[19] Glaeser, E. L., and A. Saiz, (2003). The Rise of the Skilled City. Harvard Institute of Economic Research, Discussion Paper number 2025.

[20] Glaeser, E. L., J. A. Scheinkman, and A. Shleifer, (1995). Economic growth in a cross-section of cities. Journal of Monetary Economics, 36: 117–143.

[21] Glaeser, E. L., and J. Shapiro, (2003). Urban Growth in the 1990s: Is city living back? Journal of Regional Science, 43(1): 139–165.

[22] Gibrat, R., (1931). Les inégalités économiques. Paris: Librairie du Recueil Sirey.

[23] Giesen, K., A. Zimmermann, and J. Suedekum, (2009). The size distribution across all cities - Zipf, lognormal, double Pareto, or what? Paper presented at: 56th Annual North American Meetings of the Regional Science Association International, San Francisco (USA).

[24] Goldstein, M. L., S. A. Morris and G. G. Yen, (2004). Problems with fitting to the Power-law distribution. The European Physical Journal B - Condensed Matter, 41(2): 255–258.

[25] González-Val, R., (2010). The Evolution of US City Size Distribution from a Long Term Perspective (1900–2000). Journal of Regional Science, forthcoming.

[26] Guérin-Pace, F., (1995). Rank-Size Distribution and the Process of Urban Growth. Urban Studies, 32(3): 551–562.

[27] Helpman, E., (1998). The size of regions. In: Topics in Public Economics, D. Pines, E. Sadka, and I. Zilcha, eds. Cambridge University Press.

[28] Ioannides, Y. M., and H. G. Overman, (2003). Zipf's Law for cities: an empirical examination. Regional Science and Urban Economics, 33: 127–137.

[29] Ioannides, Y. M., and S. Skouras, (2009). Gibrat's Law for (All) Cities: A Rejoinder. Discussion Papers Series, Department of Economics, Tufts University 0740, Department of Economics, Tufts University.

[30] Krugman, P., (1991). Increasing returns and economic geography. Journal of Political Economy, 99: 483–499.

[31] Krugman, P., (1996). The Self-organizing economy. Cambridge: Blackwell.

[32] Long, J. S., and J. Freese, (2006). Regression Models for Categorical Dependent Variables Using Stata. Stata Press Publication, 2nd ed.

[33] Malacarne, L. C., R. S. Mendes and E. K. Lenzi, (2001). q-exponential distribution in urban agglomeration. Physical Review E, 65, 017106.

[34] Malevergne, Y., V. Pisarenko, and D. Sornette, (2009). Gibrat's law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal. arXiv:0909.1281v1 [physics.data-an].

[35] Michaels, G., F. Rauch, and S. Redding, (2009). Rethinking urbanisation. VoxEU Jan 3, 2009.

[36] Newman, M. E. J., (2006). Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 46: 323–351.

[37] Nishiyama, Y., S. Osada and Y. Sato, (2008). OLS estimation and the t test revisited in rank-size rule regression. Journal of Regional Science, 48(4): 691–715.

[38] Parr, J. B. and K. Suzuki, (1973). Settlement populations and the lognormal distribution. Urban Studies, 10: 335–352.

[39] Petrakos, G., P. Mardakis, and H. Caraveli, (2000). Recent Developments in the Greek System of Urban Centres. Environment and Planning B: Planning and Design, 27(2): 169–181.

[40] Rosen, K. T., and M. Resnick, (1980). The Size Distribution of Cities: An Examination of the Pareto Law and Primacy. Journal of Urban Economics, 8: 165–186.

[41] Rossi-Hansberg, E., and M. L. J. Wright, (2007). Urban structure and growth. Review of Economic Studies 74, 597–624.

[42] Rozenfeld, H. D., D. Rybski, J. S. Andrade, Jr., M. Batty, H. E. Stanley, and H. A. Maksea, (2008). Laws of population growth. Proceedings of the National Academy of Sciences, 105(48): 18702–18707.

[43] Semboloni, F., (2008). Hierarchy, cities size distribution and Zipf's law. The European Physical Journal B, 63: 295–301.

[44] Simon, C. J., and C. Nardinelli, (2002). Human capital and the rise of American cities, 1900–1990. Regional Science and Urban Economics, 32: 59–96.

[45] Small, K.A., and C. Hsiao, (1985). Multinomial logit specification tests. International Economic Review, 26: 619–627.

[46] Song, S., and K. H. Zhang, (2002). Urbanisation and City Size Distribution in China. Urban Studies, 39(12): 2317–2327.

[47] Soo, K. T., (2005). Zipf’s Law for cities: a cross-country investigation. Regional Science and Urban Economics, 35: 239–263.

[48] Soo, K. T., (2007). Zipf's Law and Urban Growth in Malaysia. Urban Studies, Vol. 44(1): 1–14.

[49] Tabuchi, T., (1998). Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman. Journal of Urban Economics, 44: 333–351.

[50] Wheaton, W. C., and H. Shishado, (1981). Urban concentration, agglomeration economies and the level of economic development. Economic Development and Cultural Change, 30:17–30.

[51] Zipf, G., (1949). Human Behaviour and the Principle of Least Effort. Cambridge, MA: Addison-Wesley.

Tables

Table 1. Deviations for the ten biggest and ten smallest cities

Ranking	City	S	$\left(S^{z}/S\right)$
1	New York City	8,008,278	0.787
2	Los Angeles	3,694,820	0.853
3	Chicago	2,896,016	0.726
4	Houston	1,953,631	0.807
5	Philadelphia	1,517,550	0.831
6	Phoenix	1,321,045	0.795
7	San Diego	1,223,400	0.736
8	Dallas	1,188,580	0.663
9	San Antonio	1,144,646	0.612
10	Detroit	951,270	0.663
24,991	Stotesbury city	43	5.870
24,992	Antelope CDP	43	5.870
24,993	Saltaire village	43	5.870
24,994	Braddock city	43	5.869
24,995	Regan city	43	5.869
24,996	Atlantic CDP	43	5.869
24,997	Hetland city	43	5.869
24,998	Washam CDP	43	5.868
24,999	McCarthy CDP	42	6.008
25,000	Montezuma city	42	6.008

Note:

S : City Population in 2000 (Source: US Census Bureau),

S^{z} : Population which would correspond to a Pareto exponent equal to 1.

Table 2. Descriptive statistics

Variables	Average	Stand. dev.	Minimum	Maximum
City characteristics and External local effects variables				
Per capita income in 1999	18947.70	9713.34	1539	200087
Total population: Median age	37.32	6.60	10.80	79.20
Percent housing units: Built 1939 or earlier	22.42	19.00	0	97.88
Percent owner-occupied housing units with a mortgage; contract to purchase; or similar debt	47.96	17.39	0	100
Workers 16 years and over who did not work at home: Median travel time to work (in minutes)	24,45	6,94	2,59	109,05
Percent workers 16 years and over: Car; truck; or van; Drove alone	76.93	10.37	0	100
Percent workers 16 years and over: Public transportation	1.36	3.25	0	57.16
Human capital variables				
Percent population 25 years and over: Some college or higher degree	43.79	16.89	0	99.57
Percent population 25 years and over: High school graduate (includes equivalency) or higher degree	78.30	12.32	5.11	100
Productive structure variables				
Percent employed civilian population 16 years and over:				
Agriculture; forestry; fishing and hunting; and mining	3.52	5.40	0	72.75
Construction	7.62	4.15	0	40.32
Manufacturing	16.31	10.32	0	70.63
Wholesale and Retail trade	15.27	4.69	0	67.86
Finance; insurance; real estate and rental and leasing	5.16	3.54	0	46.67
Educational; health; and social services	20.32	7.20	0	87.18
Public administration	5.21	4.19	0	60.71

Note: All the variables correspond to 2000, except for the per capita income in 1999. Source: US Census Bureau.

Table 3.- Multinomial Logit coefficients relative to Category 1

	Categories		
	2	3	4
City characteristics and External local effects variables			
Log (Per capita income in 1999)	-0.4305***	-0.4956***	-0.6238***
Total population: Median age	0.0386***	0.0374***	0.0459***
Percent housing units: Built 1939 or earlier	0.0163***	0.0331***	0.0548***
Percent owner-occupied housing units with a mortgage; contract to purchase; or similar debt	-0.0280***	-0.0550***	-0.0860***
Workers 16 years and over who did not work at home: Median travel time to work (in minutes)	0.0390***	0.0734***	0.0922***
Percent workers 16 years and over: Car; truck; or van; Drove alone	0.0055**	-0.0084***	-0.0125***
Percent workers 16 years and over: Public transportation	0.0274***	-0.0063	-0.0755***
Human capital variables			
Percent population 25 years and over: Some college or higher degree	-0.0137***	-0.0367***	-0.0434***
Percent population 25 years and over: High school graduate (includes equivalency) or higher degree	0.0379***	0.0856***	0.1049***
Productive structure variables			
Percent employed civilian population 16 years and over:			
Agriculture; forestry; fishing and hunting; and mining	0.1190***	0.1709***	0.1939***
Construction	0.0942***	0.1286***	0.1297***
Manufacturing	0.0358***	0.0539***	0.0498***
Wholesale and Retail trade	-0.0157***	-0.0140**	-0.0359***
Finance; insurance; real estate and rental and leasing	-0.0188**	-0.0347***	-0.0790***
Educational; health; and social services	0.0235***	0.0280***	-0.0080*
Public administration	0.0306***	0.0404***	0.0367***

1 is the base outcome. ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level

Table 4. Marginal effects for each category and the average absolute change in the probability

	Categories				
	1	2	3	4	Total average
City characteristics and External local effects variables					
Log (Per capita income in 1999)	0.0856***	-0.0231***	-0.0301***	-0.0324***	0.0428***
Total population: Median age	-0.0070***	0.0033***	0.0017***	0.0020***	0.0035***
Percent housing units: Built 1939 or earlier	-0.0050***	-0.0021***	0.0028***	0.0043***	0.0035***
Percent owner-occupied housing units with a mortgage; contract to purchase; or similar debt	0.0082***	0.0030***	-0.0047***	-0.0065***	0.0056***
Workers 16 years and over who did not work at home: Median travel time to work (in minutes)	-0.0104***	-0.0025***	0.0069***	0.0061***	0.0065***
Percent workers 16 years and over: Car; truck; or van; Drove alone	0.0003	0.0028**	-0.0017***	-0.0014***	0.0016***
Percent workers 16 years and over: Public transportation	0.0001	0.0110***	-0.0014	-0.0097***	0.0056***
Human capital variables					
Percent population 25 years and over: Some college or higher degree	0.0046***	0.0025***	-0.0041***	-0.0030***	0.0035***
Percent population 25 years and over: High school graduate (includes equivalency) or higher degree	-0.0114***	-0.0046***	0.0088***	0.0071***	0.0080***
Productive structure variables					
Percent employed civilian population 16 years and over:					
Agriculture; forestry; fishing and hunting; and mining	-0.0262***	0.0020***	0.0139***	0.0103***	0.1310***
Construction	-0.0197***	0.0034***	0.0105***	0.0057***	0.0098***
Manufacturing	-0.0078***	0.0008***	0.0049***	0.0021***	0.0039***
Wholesale and Retail trade	0.0033***	-0.0006***	0.0001**	-0.0028***	0.0017***
Finance; insurance; real estate and rental and leasing	0.0060***	0.0029**	-0.0021***	-0.0068***	0.0045***
Educational; health; and social services	-0.0035***	0.0033***	0.0031***	-0.0030*	0.0032***
Public administration	-0.0061***	0.0015***	0.0033***	0.0013***	0.0031***

***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level

Figures

Figure 1. Histograms for the year 2000

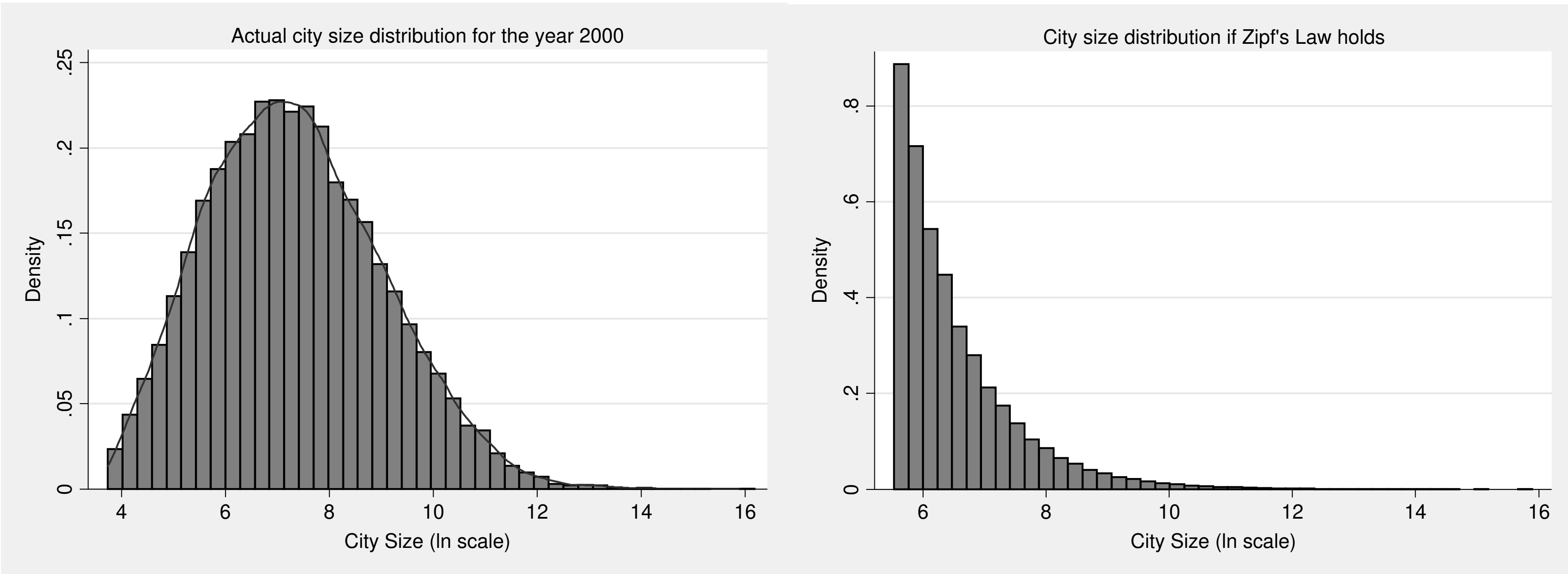


Figure 2. Empirical density functions for the year 2000

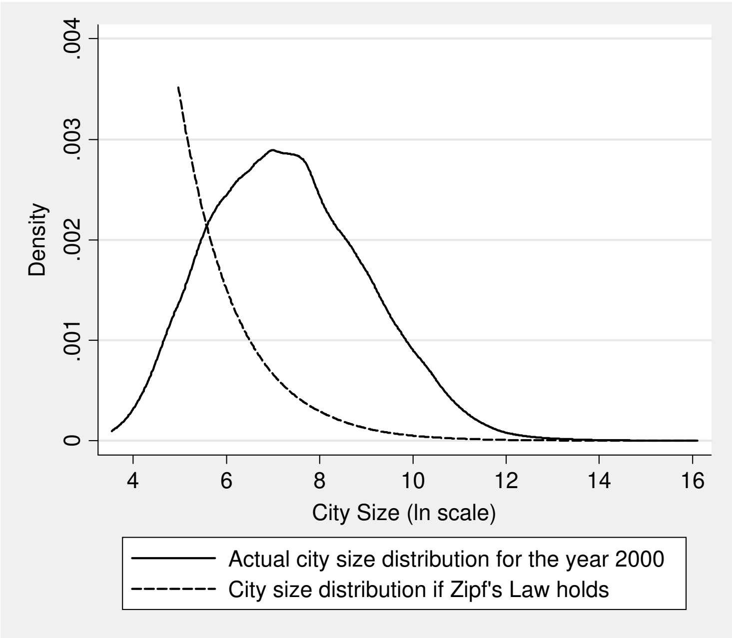
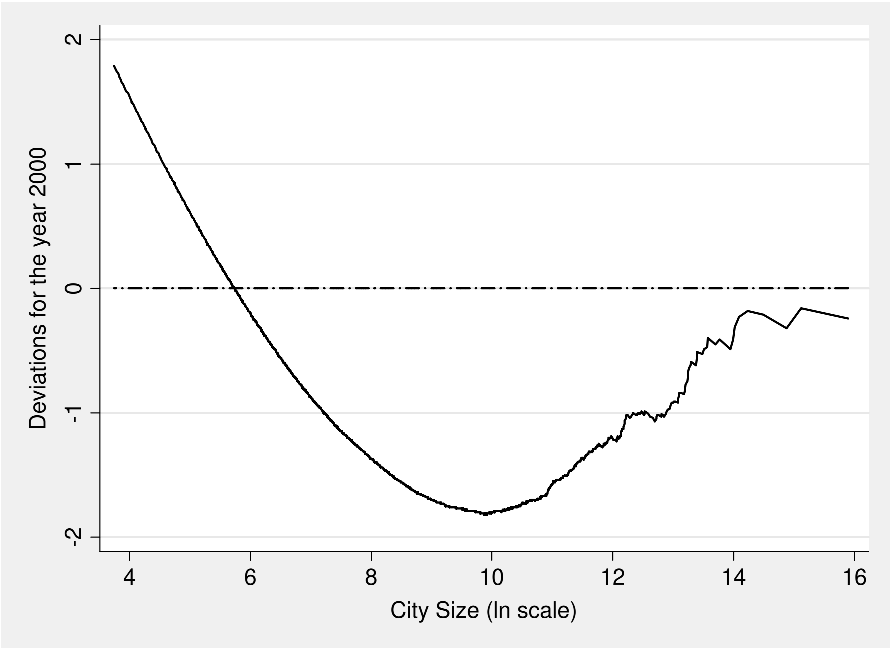
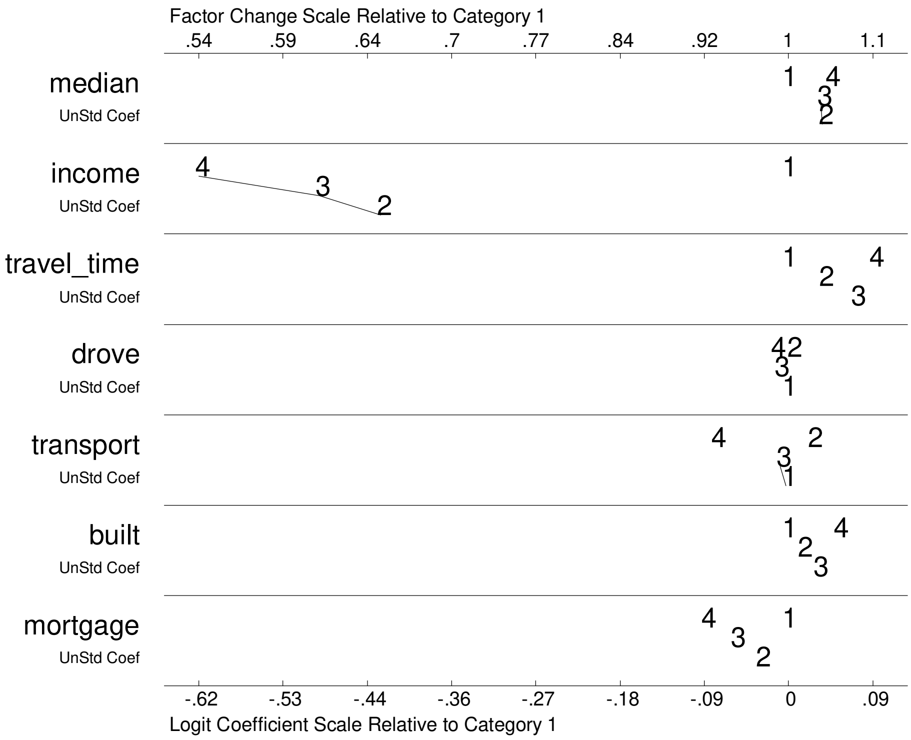


Figure 3. Deviations for the year 2000



Note: The figure represents the sample values of $\ln\left(S^{Z}/S\right)$, calculated by applying equation (3).

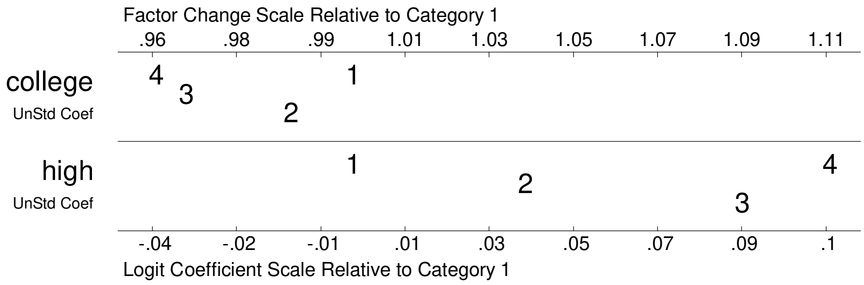
Figure 4. Odds-ratio plot of external local effects variables



Categories: 1: strong negative deviation (base category), 2: medium negative deviation, 3: weak negative deviation, 4: positive deviation.

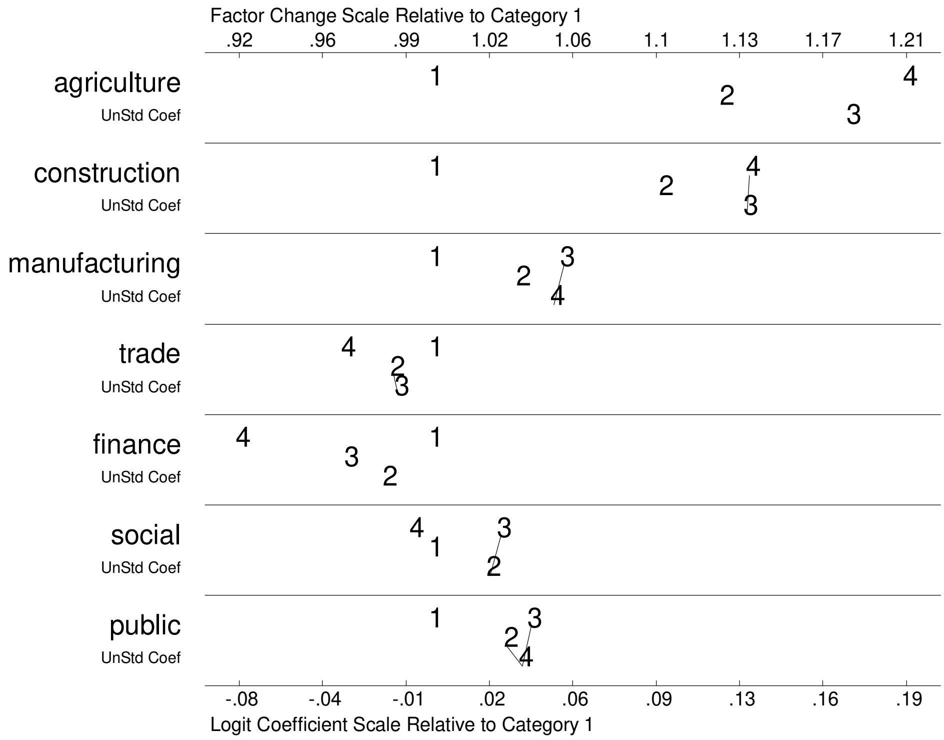
Variables: Median: "Total population: Median age", Income: "Log (Per capita income in 1999)", Travel_time: "Workers 16 years and over who did not work at home: Median travel time to work (in minutes)", Drove: "Percent workers 16 years and over: Car; truck; or van; Drove alone", Transport: "Percent workers 16 years and over: Public transportation", Built: "Percent housing units: Built 1939 or earlier", Mortgage: "Percent owner-occupied housing units with a mortgage; contract to purchase; or similar debt".

Figure 5. Odds-ratio plot of human capital variables



Categories: 1: strong negative deviation (base category), 2: medium negative deviation, 3: weak negative deviation, 4: positive deviation.
Variables: College: "Percent population 25 years and over: Some college or higher degree", High: "Percent population 25 years and over: High school graduate (includes equivalency) or higher degree".

Figure 6. Odds-ratio plot of productive structure variables



Categories: 1: strong negative deviation (base category), 2: medium negative deviation, 3: weak negative deviation, 4: positive deviation.

Variables: Percent employed civilian population 16 years and over: Agriculture: "Agriculture; forestry; fishing and hunting; and mining", Construction: "Construction", Manufacturing: "Manufacturing", Trade: "Wholesale and Retail trade", Finance: "Finance; insurance; real estate and rental and leasing", Social: "Educational; health; and social services", Public: "Public administration".