Spurious Zipf's Law[¤]

Juan Carlos Candeal, Antonio Montañés[y] and Irene Olloqui
University of Zaragoza, Gran Vía, 2. 50005 Zaragoza (Spain)

29 November, 2001

Abstract

This paper shows that the acceptance of the Zipf's Law may sometimes be the result of a spurious artifact. By way of some Monte Carlo exercises we provide evidence in favour of the fact that the Zipf's law can be spuriously accepted when the variable being studied is generated by a random distribution. This result is explained by taking account the, so-called, spurious detrending problem.

Keywords: Zipf Law; Spurious Detrending; Time Series;

JEL Classi…cation: C22

[¤] The authors whish to express their gratitude for the DGES for …nantial support under project PB 98-1614

[y] Corresponding author. Address: Gran Vía, 2, 50005 Zaragoza (Spain). Phone +34 +976 76 22 21. Fax + 34 +976 76 19 96. email: amontane@posta.unizar.es.

1 Introduction

Zipf's law, named after the Harvard Professor of Linguistics, George Kingsley Zipf (1902-1950), is the observation that the frequency of occurrence of some event (x_i), as a function of the rank (r_i), when that rank is determined by the above frequency of occurrence, is a power-law function stated as $r_i x_i = A$, where A is a constant. This law can be generalized to $r_i x_i^a = A$, where a is now a positive constant. This more general version is often referred to as the rank size rule. For the sake of simplicity, we will refer to this parameter a as the Zipf parameter. Thus, we will consider that Zipf's law holds whenever this parameter is equal to 1.

Zipf's law has been widely used in a signi…cant number of academic disciplines when seeking to explain the evolution of a variety of variables. In this regard, the most famous example of its application can be found in the analysis of the frequency with which English words appear in a text. We can …nd other examples in a number of di¤erent disciplines, such as physics (Marsilli and Zhang, 1998), bibliometrics (Silagadaze, 2000) or business studies (Ramsden and Kiss-Haypdl, 2000). However, arguably the most popular use of Zipf's law is in the …eld of urban economics[1], where it has been used to explain the evolution of the population of metropolitan areas in developed countries. Whilst the literature contains a number of papers devoted to this issue, most of them pay particular attention to the US case. In this regard, we can cite the papers of Krugman (1996), Gabaix (1999a, 1999b), Dobkins and Ioannides (2000) or Ioannides and Overman (2000), which all present some evidence in favor of the acceptance of this law. In addition to this empirical support, Gabaix (1999b) has recently derived a statistical explanation of Zipf's law for cities or metropolitan areas, one that is based on Gibrat's law. Nevertheless, despite this body of evidence in favor of Zipf's law, some problems regarding the veri…cation of this hypothesis still remain, as indicated in Ioannides and Overman (2000, pg 2).

Against this background, in this paper we raise the possibility of a further potential problem, namely that the results obtained from applying Zipf's law may be the result of a spurious statistical artifact. Here, we should note that the presence of a spurious relationship has been the subject of extensive study in econometrics and is a concept that is intimately related to time series analysis. Thus, ever since the work of Yule (1929), and up to the more recent contributions of Granger and Newbold (1974) or Phillips (1986), it has become well established that the existence of a high degree of correla-

[1] Although we will frequently use the Zipf's law for cities as the benchmark, our discussion is not exclucisvely directed at this case, but it also valid for the rest of the cases where the Zipf's law is employed.

tion between two economic variables may not necessarily be induced by the presence of some kind of economic relationship, but rather by the statistical properties of the series under analysis. In this latter case, the relationship has no economic foundation, and consequently the relationship between the variables should be considered as spurious. In this regard, we can cite Engle and Granger (1986) or Banerjee et al (1990) which review the methods offered in the literature for determining whether or not a relationship is indeed spurious.

Thus, the central aim of this paper is to show that it is possible to find spurious relationships when studying the application of Zipf's law. A priori, this might appear to be a somewhat surprising hypothesis in that, as we have mentioned earlier, the presence of a spurious relationship is associated to the use of time series data, whilst the studies that have analyzed Zipf's law have, in their majority, used cross-section data. However, and as we will see later, account should be taken of the fact that, in such cases, the observed variable is transformed in such a way that its behavior is imitating that of a non-stationary time series variable. Consequently, the method for estimating the Zipf's law parameter is closely related to the spurious detrending case studied in Durlauf and Phillips (1988), amongst others. The importance of our hypothesis is that, if it can be shown to be true and, therefore, if it is confirmed that most of the evidence in favor of Zipf's law has a spurious nature, then this would imply that any conclusion in its favor would not be accurate. At this point, we should place emphasis on the fact that we are not challenging the law in itself, but rather questioning the methods that are commonly used to determine whether it holds.

The rest of the paper is organized as follows. In Section 2 we interpret Zipf's law from a time series point of view. Section 3 is devoted to an analysis of the results presented in Section 2 by way of some Monte Carlo simulations. Here, we show that the methods commonly used to study Zipf's law may induce its erroneous acceptance in the circumstances where the variable being studied is randomly generated. Section 4 illustrates the results from an empirical point of view by considering the case of the US metropolitan areas in 1998. The paper closes with a review of the most important conclusions.

2 Zipf's Law and Spurious Regressions

The aim of this Section is to discuss, from a time series perspective, the methods that have been commonly used to determine whether Zipf's law holds for a given variable. Even at the risk of repetition, we should again emphasize that we are not questioning the appropriateness of the law in itself, which can be considered as an useful tool in many scenarios; rather, we are focusing on the way it is habitually tested for in empirical analyses. We are particularly interested in showing that the most commonly used methods may sometimes be inappropriate and, therefore, that the results based on such analyses are not accurate. Finally, although the analysis reported below is not restricted to any particular scientific area, we will often particularize our results to the discussion of Zipf's law as it is applied to metropolitan areas, given that this is an area in which it has received a significant amount of attention.

Let us consider that we dispose of an N-dimensional sample size of a variable x, with this variable measuring a number of activities, for example, the size of a city. If we order this sequence, we have that $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(r)} \geq \cdots \geq x_{(N)}$, where r is the rank and $x_{(r)}$ is the size of the variable. Zipf's law asserts that a graph of the rank against the size would render a perfect rectangular hyperbola. This implies that, for some given c and all r, it holds that:

$$r\,x_{(r)} = c \tag{1}$$

Since the seminal paper of Zipf (1949), a method that is frequently applied to determine whether this result holds is simply to plot the natural logarithm of the rank against the natural logarithm of the size. If this graph shows a straight line with a slope equal to -1, then it is considered that Zipf's law holds. However, the simple visual inspection of this graph does not seem to be an accurate and robust method for determining the appropriateness of this law, and thus it seems advisable take a second route. Thus, if we take natural logarithms in (1), it is simple to show that:

$$\ln r = \beta_1 + \beta_2 \ln x_{(r)} + e_r; \qquad r = 1; 2; \ldots; N \tag{2}$$

or, similarly, that

$$\ln r = \beta_1 + \beta_2 z_r + e_r; \qquad r = 1; 2; \ldots; N \tag{3}$$

where, for the sake of simplicity, we refer to the natural logarithm of $x_{(r)}$ as z_r. As a consequence, the closer the estimation of β_2 to -1, the greater the evidence in favor of Zipf's law. This is the route adopted by Krugman (1996), Gabaix (1999a,1999b), Dobkins and Ioannides (2000) and Ioannides and Overmann (2000), amongst others, to verify whether Zipf's law holds for cities. Whilst there are some alternative procedures, such as those of Alperovitch (1989) or Urzúa (2000), we should recognize that most of the analyses are simply based on a comparison of $\hat{\beta}_2$ with respect to -1 and, therefore, in what follows we will focus exclusively on this method.

The first aspect of this standard approach that we wish to question is that the use of the regression (3) implies a change in the type of data being considered. As mentioned earlier, we originally dispose of a cross-section of variable x which, of course, should not show any type of autocorrelation pattern. The inverse ranking of the variable x implictly implies that we are indeed changing the properties of the variable, in the sense that its behavior may now imitate that of a time series. In order to give an intuitive explanation of this change, let us consider that the variable x is generated accordingly to model (1). Thus, it is straightforward to show that:

$$z_r = c - \ln r; \quad r = 1, 2, \ldots, N \tag{4}$$

We should now note that r can be interpreted as being a deterministic trend. Consequently, it is clear that z_r is indeed indexed with a time trend and, therefore, is exhibiting a time series behavior.. Even more precisely, and following the terminology proposed in Nelson and Plosser (1982), z_r can be characterized as a Trend Stationary variable (TS). This implies that we should remove the deterministic trend component in order to achieve stationarity; that is to say, z_r is stationary around $\ln r$. Thus, given this first result, we can state that if it is not possible to characterize the variable $\ln x_{(r)}$ as being TS, then we should conclude that we have found evidence against Zipf's law. Nevertheless, we should recall that not all the shapes of trend stationarity imply Zipf's law. This is the case, for example, in the model $z_r = c - r$. Whilst it is obvious that z_r and r are related and that z_r is TS, this model is not equivalent to model (1) and, therefore, Zipf's law does not hold in this case.

This first analysis has shown that although the sample used to analyze Zipf's law in the variable x_t is, at least in spirit, a cross-section, we should properly treat it by following a time series approach. Consequently, we should take into account the problems which commonly appear in time series analysis. As we have already indicated, time series researchers are well aware of a problem that commonly appears in time series regressions, namely the possible existence of spurious relationships. In order to illustrate this, let us consider that a variable w_t is generated by the following model:

$$w_t = \mu + w_{t-1} + u_t \tag{5}$$

where μ is a parameter and u_t is an innovation that satisfies the conditions stated in Phillips (1986), for example, which imply that the innovation of the model may be generated by any stationary and invertible ARMA(p,q) model. Thus, w_t is an integrated or Difference Stationary (DS) variable, again adopting the Nelson-Plosser terminology. Let us estimate the following model:

$$w_t = b_1 + b_2 t + e_t \tag{6}$$

where t is a deterministic trend. This model implies that we are removing a deterministic trend from variable z_t. Given that the DGP of this variable does not apparently reject any trend, we should expect the estimation of (6) to exhibit a low degree of correlation and $\hat{b}_2$ to go towards 0. However, the empirical results are quite different, in that we can easily show that the estimation of model (6), when the variable being studied is generated by (5), leads the researcher to find that the determination coefficient is close to 1.0, which implies that w_t is highly correlated with the deterministic trend. As a consequence, we should conclude that the parameter b_2 is highly significant when explaining the evolution of w_t. These contra-intuitive results are commonly known as spurious detrending and have been analyzed from an analytical point of view in Durlauf and Phillips (1988). In order to better understand the nature of this phenomenon, we should consider that the variable w_t can be represented as follows:

$$w_t = \mu t + S_t \tag{7}$$

where $S_t = \sum_{i=1} u_i$. Thus, w_t shows a trend behavior.. However, its nature is not deterministic, as in model (6), but rather stochastic. Consequently, whilst the estimation of the model (6) has apparently good statistical properties, it is clearly misspecified. Fortunately, we can detect this misspecification problem by simply considering that the Durbin-Watson statistic goes towards 0 in this kind of model, as Durlauf and Phillips (1988) show.

We should now note that similar problems would appear if, instead of estimating model (6), the following model is estimated:

$$t = b_1' + b_2' w_t + e_t' \tag{8}$$

We can easily show that the estimation of this new model will suffer from the same problems as those mentioned earlier whenever w_t is generated by (5). Thus, given the direct similarity between models (8) and (3), it is also straightforward to assume that whenever $\ln x_{(r)}$ follows a drifted random walk, there will necessarily be a high degree of correlation between $\ln x_{(r)}$ and $\ln r$, with this being the consequence of a statistical artifact. In these cases, it would be totally erroneous to conclude that there is evidence in favor of any type of rank size rule. Therefore, it is necessary to distinguish between a spurious Zipf's law and those cases where this law truly holds. We can draw such a distinction by examining the time properties of the residuals. Thus, if we cannot conclude that the values of the Durbin-Watson statistic are statistically different from 0, we should interpret this result as evidence against any type of rank size rule. By contrast, if we can admit the Durbin-Watson statistic to be statistically different from 0, then it is possible to conclude in favor of either the rank size rule or Zipf's law.

Up to this point, we have questioned the results of Zipf's law to the extent that they may be founded on spurious regressions. In order to verify whether or not our hypothesis is correct, let us now carry out a number of Monte Carlo studies.

3 Spurious Zipf's Law: A Monte Carlo analysis.

In this Section we will use a number of Monte Carlo exercises to examine whether the acceptance of Zipf's law may sometimes be spurious in nature. This would be the case, for example, in the circumstances where the observations of the variable being studied have been generated by a model which does not show any connection to this law, but where the estimation of the parameter β_2 in model (3) is very close to -1.

Let us start by considering the existence of a variable u_t, which has been generated by a non-correlated random distribution. We could have considered a number of alternative distributions in order to generate the values of this variable u_t. However, if we take into account that the estimation of model (3) implies the use of natural logarithms, it seems advisable to impose the restriction that u_t can only take values greater than 0. Given this restriction, we have opted to generate the values of u_t by way of a (positive) truncated $niid(0;1)$. Nevertheless, we should note that qualitatively similar results were obtained when a uniform $iid(0;1)$ distribution was used[2].

As regards the DGP of the variable, we should expect the estimation of model (3) to allow us to conclude that Zipf's law does not hold. Further, we should observe a value of the determination coefficient very close to 0 and, therefore, the Zipf parameter β_2 should go towards 0. However, we will verify that these results do not appear in the case under study. Throughout this section, z_t denotes the natural logarithm of the variable u_t, once this variable is inversely ranked.

[2] These results are not reported here for the sake of brevity, but are available from the authors upon request.

At this point we should recall that, given the DGP of the variable u_t, it would seem reasonable to accept that z_t should not exhibit any kind of autocorrelation pattern and that the coefficients of the autocorrelation function of this variable should go towards 0. Nevertheless, we find that this is not true. Table 1 reports the simulated values of the first 9 coefficients of the autocorrelation function of variable z_t. These values have been obtained by considering the mean and the standard deviation of the different autocorrelation coefficients of z_t. We have considered different sample sizes and have carried out $50;000$ replications for each of them. As we can see, the autocorrelation function denotes the presence of a high degree of autocorrelation, which is a contra-intuitive result if we bear in mind that the observations of u_t have been generated independently. Moreover, the simulated values of the autocorrelation coefficients take values close to 1, which allows us to conclude that they are exhibiting the typical performance of a non-stationary variable.

It is possible to offer an intuitive explanation of why this phenomenon is occurring. First, we should remember that the ranking of the variable u_t indeed implies that z_i is z_{i-1} plus a quantity. This is equivalent to saying that z_t can be very well reflected by model (5). From this perspective, the results reported in Table 1 can be perfectly understood. We have simply to note that z_t, the log of the inversely ranked values of the original variable x, is imitating the performance of a non-stationary variable, although the original values x have been generated by a process which has no connection to time.

This first result is quite relevant. It would appear to confirm our hypothesis as stated in the previous Section, in the sense that the ranking process of the original variable u_t implies a change in its time properties and, thus, we can regard this variable z_t as being a time series.

This result opens the door to the possibility of considering the results of Zipf's law from a time series perspective. In particular, we should be aware of the possible appearance of spurious evidence in its favor.. In order to verify this possibility, we have carried a number of additional Monte Carlo simulations, where the variable u_t has again been generated by a (positive) truncated $niid(0;1)$, with the sample sizes being $T = \{25; 50; 75; 100; 150; 200; 250; 300; 500; 1000\}$. We have then estimated model (3), with the main results being reported in Table 2.

The results set out in this Table show that the estimation of the parameter β_2 goes towards a value of around $-2/3$. This is a very surprising result, in that we could expect this estimator to go towards 0, given that the DGP of u_t does not show any deterministic trend pattern. Furthermore, the mean of the determination coefficient is very close to 1, which allows us to conclude in favor of the existence of a high degree of correlation between the variable

z_t and the log of the rank. This is another surprising result, in that we should also expect this value to go towards 0. However, if we bear in mind the discussion presented in the previous Section, we can easily explain these results. In essence, the key to understanding this puzzle lies in the fact that we are facing a spurious relationship. In order to overcome this problem, we should simply recall that the Durbin-Watson statistic goes towards 0. Consequently, and as is also the case in time series analysis, this statistic should always accompany the results of Zipf's law. Thus, if we ignore the value of the Durbin-Watson statistic, the possibility of falling into the trap of a spurious relationship is extremely high.

In order to provide an analytical explanation to the results obtained in our simulation, we have derived the limit of the estimation of β_2 in (3) when the variable x is generated by a drifted random walk. Although is not possible to offer an exact value for this limit, by taking approximation techniques, we have been able to prove that this value is close to -2/3. Thus, this confirms our intuition, showing that the results obtained in the simulations are closely related to those of a spurious detrending phenomenon.

The results presented up to this point alert us to the possible existence of spurious relationships in the analysis of Zipf's law. Fortunately, we have observed that the estimation of the parameter β_2 in model (3) is always so far away from -1 that we are never going to conclude that Zipf's law holds, even in the circumstances where the researcher is not aware of the existence of a spurious relationship. However, in other cases we may not be so fortunate, and it is possible to find some situations where the value of the Zipf parameter is very close to -1, although the variable being studied clearly does not satisfy Zipf's law. It is precisely in these cases where we can appreciate the full extent of the damage that can be caused by the existence of a spurious relationship, in that we could conclude in favor of this law when, in reality, it is entirely false. In order to illustrate these cases, let us consider that the variable being studied is again generated by a (positive) truncated $niid(0;1)$ distribution. The difference with respect to the previous analysis is that now the model is not being estimated using all the available information, but rather with some observations being excluded. This is a particularly common situation in the empirical analysis of Zipf's law. For example, Gabaix (1999a, 1999b) and Krugman (1996) find evidence in favor of Zipf's law for the USA when only the 135 largest metropolitan areas are taken into consideration: however, they cannot accept it when all the US metropolitan areas are included.

Given that the habitual procedure is to consider only one part of the sample, it seems to be appropriate to study the behavior of the estimator of the parameter β_2 when model (3) is recursively estimated. This implies that we first estimate this model for observations 1 to t_0 and, subsequently, add a new observation in each iteration. The procedure stops when t_0 coincides with the total number of observations. In this latter case, $t_0 = T$, the results would be those presented in Table 2.

Table 3 presents the results obtained from the application of the above-mentioned procedure. This Table reflects the number of times (in %) that the difference between the estimator of the parameter β_2 and -1 is lower than a predetermined value ξ. The results are quite interesting, in that we can conclude that it is always possible to find a sample size which guarantees an approximation of $\hat{\beta}_2$ towards a value as near to -1 as we wish. For example, we can see that if we dispose of a total of 25 observations, the parameter ξ is lower than 0.01 in 17.6% of the cases. This percentage increases to 76.1% when $T = 150$ and to 100% when $T = 1,000$. This might explain why the use of one particular sample size could imply the acceptance of Zipf's law, whilst the increase (or decrease) of this sample may induce us to reject it.

Therefore, we can appreciate that Zipf's law might be wrongly accepted in a large number of situations. This acceptance, which is spurious in nature, depends on the total number of observations. The larger the number of observations available, the easier it is to approximate the Zipf parameter towards a value as close to -1 as we wish. This finding offers the possibility a very rich interpretation from the point of view of the analysis of Zipf's law when applied to cities. Thus, if we admit, as seems quite reasonable, that the number of cities or metropolitan areas increases with the surface area of a country, we could conclude that it is easier to approximate $\hat{\beta}_2$ to -1 in a big country (for example, the USA) than in a relatively small one (any of the countries of the European Union, for instance). Accordingly, this approach allows us to understand why the evidence in favor of Zipf's law for cities seems to be stronger for the USA than for the European countries. However, in our view, the results presented in previous papers devoted to this particular theme should be reworked in the light of our Monte Carlo results in order to distinguish between the spurious and true evidence in favor of Zipf's law.

As we have mentioned earlier, there is a tool available to us, namely the Durbin-Watson statistic, that we can use to overcome this problem. From the values of this statistic, which are reported in the last column of Table 3, we can observe that the minimization of the parameter ξ, implies that the Durbin-Watson goes towards 0, as occurs in the spurious relationship cases. Thus, the value of the Durbin-Watson statistic can help us to discriminate between spurious and true Zipf's law. Thus, if this statistic goes to 0, we should conclude that we are facing a spurious regression; by contrast, if it takes a value that is statistically different from 0, then the relationship will not be spurious. If, furthermore, $\hat{\beta}_2$ is close enough to -1, we can consider this as being evidence in favor of Zipf's law.

4 An empirical example: The US metropolitan areas

In order to illustrate the problematic that we have discussed in the previous Sections, let us study whether Zipf's law is a valid instrument to explain the evolution of the US metropolitan areas. At this point, we should recall that a number of papers have previously studied this issue, with most of them finding evidence in favor of Zipf's law. However, and taking into account the results of our Monte Carlo exercise, it is our hypothesis that these results should be interpreted with a degree of caution, given the possible presence of some spurious relationships.

In our example, we have information on the population of the 276 largest agglomerations in the USA. The largest is the metropolitan area of New York, with a population of 20,156,150 inhabitants. This is followed by the metropolitan areas of Los Angeles and Chicago with 15,781,273 and 8,809,846, respectively. By contrast, the smallest areas are Pocatello (ID), Casper (WY) and Enid (OK) with 74,866, 63,341 and 56,859 inhabitants, respectively. All the data have been obtained from the US Census Bureau and are related to the estimations for the year 1998.

In order to determine whether Zipf's law holds, we have estimated the model (2) for the whole sample, obtaining the following results:

$$\ln \hat{t} = 15.48 - \underset{(0.008)}{0.8536} \ln \widehat{POP} \qquad R^2 = 0.97 \qquad DW = 0.06 \tag{9}$$

where t is a deterministic trend, which represents the rank, and $\ln \widehat{POP}$ is the natural logarithm of the population of the different metropolitan areas, once these have been inversely ranked. If we do not take into account the value of the Durbin-Watson statistic, as has commonly been the cases in all the previous studies related to Zipf's Law, we would draw the conclusion that the estimation presented in (9) clearly fails to support it. This is due to the fact that the Zipf parameter is far from -1. Nevertheless, we can admit the existence of a high degree of correlation between the population and the rank, in that the R^2 takes a value close to 1.

However, we should recall that these earlier studies have not included all the available information on the metropolitan areas. Rather, it is common to find that the strongest evidence in favor of Zipf's law is adduced when only a part of the sample is used. To analyze what occurs in such circumstances, we have estimated the model (2) recursively. The initial sample size was $T = 28$. Table 4 summarizes the main results obtained when we consider different sample sizes. This table reflects the value of the estimation of parameter β_2, the determination coefficient and the Durbin-Watson statistic for each estimation of the model. For the sake of simplicity, we have chosen not to present all the results obtained and, whilst the omitted results are available upon request, their consideration does not change the conclusion that can be drawn from the analysis of Table 4.

The results reported in this table invite us to conclude that the estimation of the parameter β_2 can take values which are very close to -1. The minimum distance of this estimator to -1 is attained when the sample size considers the 138 largest metropolitan areas (the last one included is Anchorage, AK). In this case, $\hat{\beta}_2$ is equal to -0.9996. Here, attention should be drawn to the fact that the sample size which minimizes the distance of the estimation of the Zipf parameter to 1 is qualitatively similar to that used in previous papers that analyze whether Zipf's Law holds. Thus, for example, Krugman (1996) and Gabaix (1998), who used the values of the population of 1990, consider the 135 largest US metropolitan areas.

Table 4 offers other interesting insights. Thus, we can observe that the R^2 is always greater than 0.98. At the same time, we can also see that the value of the estimator $\hat{\beta}_2$ continuously decays from the initial value of -1.2379 when only the 28 largest metropolitan areas are used, to the value of -0.8536 for the case where all the 256 metropolitan areas are used in the estimation. Thus, a second conclusion would be that whilst Zipf's law holds for a certain sample size, it does not when all the observations are included.

Up to this point, we have followed the approach that has been previously adopted in the literature, without giving any consideration to the results presented in Sections 2 and 3. However, if we approach the analysis from a time series perspective, the interpretation of the results of Table 4 changes radically. Thus, if we first consider the last column of this Table, we can see that the values of the Durbin-Watson statistic are always very low, denoting the inappropriate nature of the estimated model. If we combine this result with the fact that the R^2 is very high, and further take into account the results of the previous Section, we could explain this puzzle by admitting the possibility that the relation we are estimating is spurious. In order to provide support for this hypothesis, we have tested for the presence of a unit root in the variable $\ln POP$. To that end, we have estimated the following model

$$y_t = \mu + \beta t + \rho y_{t-1} + \sum_{i=1}^{\ell} \phi_i \Delta y_{t-i} + e_t \qquad (10)$$

and have then obtained the pseudo t-ratio for testing whether the autoregressive parameter is 1 (in this regard, see Dickey and Fuller, 1979). The parameter ℓ is commonly known as the lag truncation parameter and its aim is to control the number of lags of Δy_t that it is necessary to include in the model specification so as to guarantee that there are no autocorrelation problems. Although a number of methods for the selection of the lag truncation parameter are available, we have chosen to use the k(t) procedure recommended in Ng and Perron (1995). This method involves a general-to-specific strategy, starting with a predetermined value of the lag truncation parameter (kmax) and then testing the significance of the single coefficient associated with the last lag until a significant statistic is encountered. The single significance of the lags is analyzed by comparing their t-ratios with the value 1.65. When following this procedure, we have found that the value of the ADF test when all 256 observations are used is -0.63. Thus, we should accept the presence of a unit root in the variable $\ln \widehat{POP}$.

As we have already discussed in Section 2, the presence of a unit root in the variable under analysis is totally incompatible with Zipf's Law and, if we combine the presence of a unit root in the variable $\ln \widehat{POP}$ with the proximity of the Durbin-Watson to 0, then we can confirm the spurious nature of the estimation reported in (9). Consequently, we can refute the conclusion which maintains that Zipf's Law holds for the USA, at least in the circumstances where this law is analyzed using the 256 largest metropolitan areas.

5 Conclusions

In this paper we have provided some evidence in favor of the possible presence of spurious relationships when studying Zipf's law. First, we have shown that it would be advisable for studies devoted to this issue to adopt a time series approach, even though the original sample size has a cross-section nature. Furthermore, if Zipf's law really holds, then once the variable being studied has been inversely ranked, it must be characterized as trend stationary. As a consequence, a necessary condition in order for Zipf's law to hold is that the transformed variable shows a trend stationary pattern. The time series literature on unit root tests provides us with an excellent framework within which to carry out this analysis. On this basis, if we cannot reject the unit root null hypothesis against the trend stationary alternative, we should never conclude that Zipf's law holds, even in those case where the estimation of the Zipf parameter is virtually ¡1.

Secondly, and as consequence of the imitation of time series behavior on the part of the transformed variable, we have proved that the studies on Zipf's law may su¤er from the presence of spurious relationships. Moreover, we have provided evidence, by way of some Monte Carlo analyses, that a variable which has been generated randomly may lead us to wrongly accept Zipf's law. The probability of this phenomenon occurring depends mainly on the total number of observations available. This result is clearly related to the spurious detrending problems that are very well known in the time series literature. The method to combat this problem is to use the Durbin-Watson statistic. If Zipf's law really holds, then this statistic should not show any autocorrelation pattern; however, if this law has a spurious nature, then the Durbin-Watson goes towards 0. Thus, we should never conclude in favor of Zipf's law when this statistic takes a value close to 0.

Finally, as an empirical illustration we have analyzed the case of the US metropolitan areas, using the data projections for 1998. We have found that whilst an appropriate selection of the sample size allows us to obtain an estimation of the slope parameter very close to ¡1, the Durbin-Watson statistic takes a value very close to 0: Furthermore, the unit root null hypothesis is not reject for the inversely ranked population variable. Under these circumstances, the conclusion should be drawn that., in this case, Zipf's law does not hold, a …nding which challenges the dominant line of results habitually presented in the literature.

6 References

Alperovitch, G. (1989). "The distribution of city size: A sensitivity analysis", Journal of Urban Economics, 25, 93-102.

Banerjee, A., J.J. Dolado, J.W. Galbraith and D.Hendry (1990). Co-integration, Error Correction, and the Econometric of Non-Stationary Data, Oxford University Press.

Brakma, S., H. Garretesen, C. van Marrewijk and M. van der Berg (1999). "The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution", Journal of Regional Science, 39, 183-213.

Dickey, D. A., and W.A. Fuller, (1979). "Distribution of the estimators for autoregressive time series with a unit root", Journal of the American Statistical Association 74, 427-31.

Dobkins, L. H. and Y.N. Ioannides (2000), "Dynamic evolution of the US city size distribution", 217-260, J. M. Huriot and J. F. Thisse, eds., Economics of cities, Cambridge University Press, New York.

Durlauf, S. and P.C.B. Phillips (1988). "Trends versus Random Walks in Time Series Analysis", Econometrica, 56, 1333-1354.

Eaton, J. and Z. Eckstein (1997). "Cities and Growth: Theory and Evidence from France and Japan", Journal of Monetary Economics, 27, 443-474.

Engle, R. and C. W. J. Granger (1987), 'Cointegration and Error Correction, Representation, Estimation and Testing', Econometrica, 55, 251-76.

Gabaix, X. (1999a). "Zipf's Law and the Growth of Cities", American Economic Review, 89, 129-132.

Gabaix, X. (1999b). "Zipf's Law for Cities: An Explanation", Quarterly Journal of Economics, 114, 739-767.

Granger, C. W. and P. Newbold (1974). "Spurious Regressions in Econometrics", Journal of Econometrics, 2, 111-120.

Ioannides, Y. M. and H.G. Overman (2000), "Zipf's law for cities: An empirical examination", Working Paper.

Krugman, P. (1996). The Self-Organizinf economy, Oxford, Blackwell Publishers.

Marsili, M. and Y. C. Zhang (1998), "Interacting individuals leading to Zipf's law", Physical Review Letters, 80, 2741-2744.

Nelson, C. and Plosser (1982). "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", Journal of Monetary Economics, 10, 139-162.

Ng, S. and P. Perron (1995), "Unit Root Test in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag", Journal of the American Statistical Association, vol. 90, 268-281.

Phillips, P.C.B. (1986). "Understanding Spurious Regressions", Journal of Econometrics, 33, 1986, 311-340.

Ramsden, J.J. and G. Kiss-Haypdl (2000), "Company size distribution in di¤erent countries", Physica A, 277, 220-227.

Silagadze, Z.K. (2000), "Citations and the Zipf-Mandelbrot law", Complex Systems, 11.

Urzúa, C. M. (2000). "A Simple and E¢cient Test for Zipf's Law", Economics Letters, 66, 257-260.

Yule, G. U. (1926). "Why do we sometimes get nonsense correlations between time series- a study in sampling and the nature of time series", Journal of the Royal Statistical Society, 89, 1-64.

7 Appendix

In this Appendix we provide a theoretical argument which leads us to offer a realistic approximation to the estimation of the parameter β_2 in 3. To that end, let us consider that r is a deterministic trend and that x_t is generated by a drifted random walk.

Let us consider that we dispose of a sample of dimension n and we want to analyze the relationship between a time trend, a proxy of the rank, and the inverse of a drifted random walk. These two variables can be represented by $\{x_m; y_m\}_{m=1}^n = \{\ln m; \ln(n+1-m)\}_{m=1}^n$, where x_m is a deterministic trend, that is to say, the rank, whilst y_m is an asymptotic approximation to the behavior of a drifted random walk. To see this, we should first consider that this drifted random walk is defined as follows $y_m = y_{m-1} + \mu + u_m$, where $\{u_m\}_{m=1}^n$ is an innovation that satisfies the conditions stated in Phillips (1986), for example. This leads us to $y_m = \mu m + S_m$, where $S_m = \sum_{i=1}^m u_i$. If we take into consideration the asymptotic results of Phillips (1986), it is straightforward to see that the y_m is dominated by the trend, m. Thus, the effect of S_m is asymptotically negligible. Therefore, we have omitted this term throughout this Section, given that its inclusion would not modified the results presented here. Furthermore, given that $\ln \mu m = \ln m + \ln \mu$, we will omit $\ln \mu$, focusing on the first summand. Thus, if we inversely rank the variable y_m, we can approximate to its asymptotic behavior by way of $\{\ln(n+1-m)\}_{m=1}^n$. Finally, for reasons of simplicity in the derivation of the limit values, we have considered the inverse regression to that presented in 3. Once again, this does not involve any change in the limit values that we will obtained, given the symmetry of the limit value.

Against this background, a direct application of the least squares principle leads to:

$$\hat{\beta}_2(n) = \frac{n \sum_{m=1} \ln(n+1-m)\ln(m) - \sum_{m=1} \ln(n+1-m) \sum_{m=1} \ln(m)}{n \sum_{m=1} \ln^2 m - \left(\sum_{m=1} \ln(m)\right)^2}$$

Let us denote $a_n = \sum_{m=1} \ln(n+1-m)\ln(m)$, $b_n = \left(\sum_{m=1} \ln(m)\right)^2 = (\ln n!)^2$ and $c_n = \sum_{m=1} \ln^2 m$. If we additionally consider that $\sum_{m=1} \ln(n+1-m) = \sum_{m=1} \ln(m)$, then the previous equation turns into:

$$\hat{\beta}_2(n) = \frac{n a_n - b_n}{n c_n - b_n}$$

Our aim is to calculate the $\lim_{n\to\infty} \psi_2(n)$. To that end, our strategy is based on the following steps.

Step 1. The sequence $D(n) = nc_n - b_n$ is monotone increasing and $\lim_{n\to\infty} D(n) = \infty$. In particular, $D(n) > 0; \forall n > 1$.

Proof . It is enough to compute $D(n+1) - D(n)$ for a given $n \in \mathbb{N}$.

$$
\begin{aligned}
D(n+1) - D(n) &= n\,(c_{n+1} - c_n) + c_{n+1} + (b_n - b_{n+1}) \\
&= n \ln^2(n+1) + \sum_{m=1} \ln^2 m - \ln(n+1) \ln\left[(n+1)\,(n!)^2\right] \\
&= n \ln^2(n+1) + \sum_{m=1} \ln^2 m + \ln^2(n+1) - \ln^2(n+1) - \\
2\ln(n+1)\ln n! &= n \ln^2(n+1) + \sum_{m=1} \ln^2 m - \sum_{m=1} 2\ln(n+1)\ln m \\
&= \sum_{m=1} \left[\ln^2(n+1) + \ln^2 m - 2\ln(n+1)\ln m\right] \\
&= \sum_{m=1} [\ln(n+1) - \ln m]^2 = \sum_{m=1} \ln^2\left(\frac{n+1}{m}\right)
\end{aligned}
$$

Thus, this shows that $D(n+1) - D(n) > 0; \forall n \geq 1$ and, therefore, $D(n)$ is a monotone increasing sequence. The fact that $\lim_{n\to\infty} D(n) = \infty$ is obvious given that $\lim_{n\to\infty} [D(n+1) - D(n)] \geq \lim_{n\to\infty} \ln^2(n+1) = \infty$.

Step 2. $\psi_2(n)$ is a bounded sequence. In particular, it holds that $-1 \leq \psi_2(n) \leq 1, \forall n \in \mathbb{N}$

Proof.

In order to show that $\psi_2(n) \leq 1$ we should consider that this is the same as $\frac{na_n - b_n}{nc_n - b_n} \leq 1$ and, by step 1, this last inequality is equivalent to $na_n - b_n \leq nc_n - b_n$. Hence, $\psi_2(n) \leq 1$ holds if and only if $n\,(c_n - a_n) \geq 0; \forall n \in \mathbb{N}$, with the latter being true if and only if $c_n - a_n \geq 0, \forall n \in \mathbb{N}$. Somewhat tedious calculations show that

$$c_n - a_n = \sum_{m=1}^{n} \ln m \ln\left(\frac{m}{n+1-m}\right)$$

$$= \begin{cases} \sum_{m=1}^{n/2} \ln^2\left(\frac{n+1-m}{m}\right) & \text{for even } n \\ \sum_{m=1}^{(n+1)/2} \ln^2\left(\frac{n+1-m}{m}\right) & \text{for odd } n \end{cases}$$

Therefore, $c_n - a_n \geq 0, \forall n \in N$ and $\chi_2(n) \leq 1$. Moreover, $c_n - a_n > 0, \forall n > 1$.

In order to prove the second inequality, $-1 \leq \chi_2(n), \forall n \in N$ we can proceed in a similar way. We omit the details for the sake of brevity.

Step 3. Let us consider the function $(r;n) \in \Re \times N \mapsto F(r;n) = \sum_{m=1}^{n} \ln m \ln[(n+1-m)\, m^r] - (1+r)\frac{(\ln n!)^2}{n} \in \Re$. Then F is a continuous function on $\Re \times N$ which is strictly increasing as a function of "r", for a fixed $n \in N$. Moreover, if for a fixed $r \in \Re$, $F(r;n) > 0$, then it holds that $F(r;m) > 0$, $\forall m > n$.

Proof.

It should be noted that F can be written as $F(r;n) = a_n + r c_n - (1+r)\frac{b_n}{n} = \left(\frac{n c_n - b_n}{n}\right) r + \frac{n a_n - b_n}{n}$. Thus, F is obviously continuous. Moreover, the second assertion follows, in that, for a fixed $n \in N$, F is an affine function of "r" and it holds that the term $\frac{n c_n - b_n}{n} \geq 0$, by step 1. This last assertion can be shown inductively.

Step 4. Let us define the following set $A = \{r \in \Re; F(r;n) < 0, \forall n > 1\}$. Then, it holds that

a) A is non-empty

b) A is bounded above.

Proof

a) Let us see that $-1 \in A$. Indeed, and taking into account the results of Step 2, it holds that $F(-1;n) = -\frac{n c_n - b_n}{n} + \frac{n a_n - b_n}{n} = \frac{1}{n}(n a_n - n c_n) = a_n - c_n < 0, \forall n > 1$.

b) Let us prove that 1 is an upper bound for A. First of all, we should consider that $F(1;n) = \frac{n c_n - b_n}{n} + \frac{n a_n - b_n}{n} = c_n + a_n - 2\frac{b_n}{n} \geq 0, \forall n \in N$, with this result coming from the direct consideration of the results of Step 2. Thus, it holds that $1 \notin A$. Further, by Step 3, it is true that $F(r;n) > F(1;n) > 0$, $\forall r > 1, \forall n > 1$. Thus, for every $r \in A$, it holds that $r < 1$ and, therefore, A is bounded above.

Step 5. Let B denote the following set of the reals,

$B = \{ r \in \mathbb{R}; \exists N(r) \in \mathbb{N}; \forall n \geq N(r), F(r;n) \geq 0 \}$ Then, it holds that:

a) B is non-empty

b) B is bounded below.

Proof

The Proof is similar to that offered for Step 4 and is therefore omitted.

Step 6. Given a subset of the reals Γ and a real sequence $(\alpha_n)_{n \geq 1}$, let us denote $\sup \Gamma$, $\inf \Gamma$, $\overline{\lim}_{n \to \infty} \alpha_n$ and $\underline{\lim}_{n \to \infty} \alpha_n$ as being the supremum of Γ, the infimum of Γ, the superior limit and the inferior limit of $(\alpha_n)_{n \geq 1}$, respectively. Then, it holds that:

a) $\sup A \leq \inf B$

b) $\sup A = \inf B$

c) $\sup A \leq \underline{\lim}_{n \to \infty} -\lambda_2(n)$

d) $\inf B \geq \overline{\lim}_{n \to \infty} -\lambda_2(n)$

e) $\exists \lim_{n \to \infty} \lambda_2(n)$

f) $\lim_{n \to \infty} \lambda_2(n) \simeq -\frac{2}{3}$

Proof

a) Let $r \in A$, $s \in B$ and suppose by contradiction that $s \leq r$. Then, by using the results of Step 3, $F(s;n) \leq F(r;n)$, $\forall n > 1$. In particular, it holds that $F(s;n) < 0$, $\forall n > 1$. However, this result contradicts the fact that $s \in B$. Therefore, we have that $r < s$ for each $r \in A$, $s \in B$. Thus, it holds that $\sup A \leq \inf B$:

b) Let us suppose, again by contradiction, that $\sup A < \inf B$. Then, if we take into account the definitions of A and B, and considering the last assertion of Step 3, if follows that there exist ℓ, ℓ', $\ell'' \in \mathbb{R}$ and $n \in \mathbb{N}$ such that $\sup A < \ell' < \ell < \ell'' < \inf B$ and $F(\ell', n) \geq 0; F(\ell'', n) < 0$. However, this again would lead to the following contradiction $0 \leq F(\ell', n) < F(\ell, n) <$

$F(\ell^{\infty}, n) < 0$, by simply considering Step 3. Thus, we should conclude that $\sup A = \inf B$.

c) Let $r \in A$. Then, $F(r; n) < 0$, $\forall n > 1$. This means that $\frac{n c_n - b_n}{n} r + \frac{n a_n - b_n}{n} < 0$, $\forall n > 1$ or, equivalently, $r < -\hat{\ }_2(n)$, $\forall n > 1$, what obviously implies that $r \leq \liminf_{n \to \infty} -\hat{\ }_2(n)$ and, therefore, it holds that $\sup A \leq \liminf_{n \to \infty} -\hat{\ }_2(n)$.

d) Let $s \in B$. Then, by definition, it is true that $F(s; n) \geq 0$, $\forall n \in N(s)$. This means that $s \geq -\hat{\ }_2(n)$, $\forall n \in N(s)$. Thus, $s \geq \limsup_{n \to \infty} \alpha_n - \hat{\ }_2(n)$ and, as a consequence, it holds that $\inf B \geq \limsup_{n \to \infty} -\hat{\ }_2(n)$.

e) Using the results of c) and d), we have that $\sup A \leq \liminf_{n \to \infty} -\hat{\ }_2(n) \leq \limsup_{n \to \infty} -\hat{\ }_2(n) \leq \inf B$. Furthermore, if we consider b), then if follows that $\liminf_{n \to \infty} -\hat{\ }_2(n) = \limsup_{n \to \infty} -\hat{\ }_2(n)$ and, therefore, there exists the limit of $\left\{ -\hat{\ }_2(n) \right\}_{n \geq 1}$. Obviously, it is also true that there exists the limit $\lim_{n \to \infty} \hat{\ }_2(n)$.

f) In order to provide a numerical value for $\lim_{n \to \infty} \hat{\ }_2(n)$ we have used numerical approximation. The results, which are not included here but are available from the authors upon request, reveal that the exact value of this limit would be close to $-2/3$, which helps us to understand the results based on our Monte Carlo simulations, confirming that these are produced by the so-called spurious detrending phenomenon.

Table 1. Autocorrelation Function

T	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
25	0:79 (0:10)	0:65 (0:11)	0:55 (0:11)	0:45 (0:10)	0:36 (0:09)	0:28 (0:07)	0:19 (0:06)	0:11 (0:04)	0:02 (0:03)
50	0:86 (0:07)	0:77 (0:08)	0:70 (0:09)	0:64 (0:09)	0:58 (0:09)	0:53 (0:08)	0:49 (0:08)	0:44 (0:08)	0:40 (0:07)
75	0:89 (0:06)	0:82 (0:07)	0:76 (0:07)	0:71 (0:07)	0:67 (0:08)	0:63 (0:08)	0:59 (0:07)	0:56 (0:07)	0:53 (0:07)
100	0:91 (0:05)	0:84 (0:06)	0:80 (0:06)	0:75 (0:07)	0:72 (0:07)	0:69 (0:07)	0:65 (0:07)	0:62 (0:07)	0:60 (0:07)
150	0:93 (0:04)	0:88 (0:04)	0:84 (0:04)	0:81 (0:05)	0:78 (0:05)	0:75 (0:05)	0:72 (0:06)	0:70 (0:06)	0:68 (0:06)
200	0:94 (0:03)	0:90 (0:04)	0:86 (0:04)	0:84 (0:04)	0:81 (0:05)	0:79 (0:05)	0:77 (0:05)	0:75 (0:05)	0:73 (0:05)
250	0:95 (0:02)	0:91 (0:03)	0:88 (0:04)	0:86 (0:04)	0:83 (0:04)	0:81 (0:04)	0:80 (0:04)	0:78 (0:04)	0:76 (0:04)
300	0:95 (0:02)	0:92 (0:03)	0:89 (0:03)	0:87 (0:03)	0:85 (0:04)	0:83 (0:04)	0:82 (0:04)	0:80 (0:04)	0:79 (0:04)
500	0:97 (0:01)	0:94 (0:02)	0:92 (0:02)	0:91 (0:02)	0:89 (0:03)	0:88 (0:03)	0:87 (0:03)	0:85 (0:03)	0:84 (0:03)
1000	0:98 (0:01)	0:97 (0:01)	0:95 (0:01)	0:94 (0:01)	0:93 (0:02)	0:92 (0:02)	0:91 (0:02)	0:91 (0:02)	0:90 (0:02)

The values of this Table have been obtained as follows. First, we have generated a T-dimension vector of observations of the variable u_t by way of a positive truncated $niid(0; 1)$. Then, we have inversely sorted these values and, subsequently, have calculated the coe¢cients of the autocorrelation function. This procedure has been repeated 50; 000 times and the Table reports the mean value of these autocorrelation coe¢cients, with the standard deviation reported in parenthesis.

Table 2. Spurious result in the analysis of Zipf's law

T	$\hat{\beta}_2$	R^2	DW	$\hat{\rho}$
25	-0.66 (0.19)	0.936 (0.02)	0.227 (0.08)	1.261 (0.61)
50	-0.65 (0.14)	0.954 (0.01)	0.103 (0.04)	1.320 (0.34)
75	-0.65 (0.12)	0.961 (0.01)	0.067 (0.03)	1.277 (0.24)
100	-0.65 (0.11)	0.966 (0.01)	0.049 (0.02)	1.238 (0.18)
125	-0.65 (0.09)	0.971 (0.00)	0.032 (0.02)	1.191 (0.12)
200	-0.65 (0.08)	0.974 (0.00)	0.024 (0.02)	1.161 (0.10)
250	-0.65 (0.07)	0.976 (0.00)	0.019 (0.01)	1.141 (0.08)
300	-0.65 (0.06)	0.978 (0.02)	0.016 (0.01)	1.127 (0.07)
500	-0.65 (0.05)	0.981 (0.00)	0.010 (0.01)	1.092 (0.04)
1000	-0.66 (0.04)	0.985 (0.00)	0.005 (0.00)	1.059 (0.03)

This Table considers the presence of possible spurious relationships when studying Zipf's law for variable u_t, with this variable having been generated by a (positive) truncated niid(0,1). These values are inversely ordered and transformed into their natural logarithms, and model (2) is then estimated. The first column of this Table reflects the different sample sizes considered. The second presents the mean and the standard deviation (in parenthesis) of the estimated values of the Zipf parameter β_2 in (2). The third and the fourth report the mean of the determination coefficient and the mean of the Durbin-Watson statistic, respectively. All values were obtained as a result of 50,000 replications.

Table 3. Spurious Zipf's law

T	γ 10^{-4}	5×10^{-4}	10^{-3}	0.005	0.01	0.015	0.02	0.025	0.03	0.05	DW
25	0:2	0:9	1:7	8:8	17:6	26:0	34:0	41:5	48:0	67:7	0:29 (0:11)
50	0:3	1:6	3:2	15:1	30:3	45:0	57:8	68:0	75:7	91:2	0:11 (0:03)
75	0:4	2:1	4:3	21:9	43:0	61:9	76:1	85:2	90:6	97:9	0:06 (0:02)
100	0:5	2:6	5:5	27:4	54:3	75:6	88:1	94:0	96:9	99:5	0:04 (0:01)
150	0:8	4:0	8:0	39:4	76:1	93:6	98:2	99:4	99:8	100	0:02 (0:00)
200	1:1	5:3	10:5	52:0	90:6	98:8	99:8	100	100	100	0:02 (0:00)
250	1:3	6:5	12:9	64:3	97:4	99:9	100	100	100	100	0:01 (0:00)
300	1:6	7:7	15:3	75:5	99:4	100	100	100	100	100	0:01 (0:00)
500	2:4	12:4	24:9	98:8	100	100	100	100	100	100	0:01 (0:00)
1000	4:9	24:7	49:3	100	100	100	100	100	100	100	0:00 (0:00)

This Table reports the number of times (in %) that the estimation of the parameter β_2 in model (2) is lower than a value γ, when the variable being studied has been generated by a (positive) truncated niid(0,1) distribution. The first row includes the different values of γ, whilst the first column reports the different sample sizes considered. $50,000$ replications were carried out for each of the previous combinations of values.

t	$\hat{\ }_2$	R^2	DW
28	−1.2379	0.98	0.93
38	−1.1822	0.98	0.71
48	−1.1710	0.99	0.69
58	−1.1606	0.99	0.67
68	−1.1116	0.98	0.43
78	−1.0843	0.98	0.33
88	−1.0575	0.98	0.27
98	−1.0425	0.98	0.25
108	−1.0309	0.98	0.24
118	−1.0214	0.98	0.23
128	−1.0115	0.99	0.21
138	−0.9996	0.99	0.20
148	−0.9851	0.98	0.17
158	−0.9707	0.98	0.15
168	−0.9572	0.98	0.13
178	−0.9426	0.98	0.12
188	−0.9311	0.98	0.11
198	−0.9204	0.98	0.10
208	−0.9114	0.98	0.09
218	−0.9038	0.98	0.09
228	−0.8963	0.98	0.08
238	−0.8902	0.98	0.08
248	−0.8838	0.98	0.08
258	−0.8759	0.98	0.07
268	−0.8659	0.98	0.07

Table 4. Zipf's law for the US metropolitan areas

This Table reports the results related to the analysis of the Zipf's law for the 256 biggest agglomeration areas of the USA. These results are obtained by recursively estimating model (2). In each iteration, the sample size includes the t-th biggest US agglomeration areas. This Table then presents the values of the estimation of the Zipf parameter ($\hat{\ }_2$), the coefficient of determination (R^2) and the Durbin-Watson statistic (DW) of each of these estimations.