# On the parametric description of US city size distribution: New empirical evidence

Ramos, Arturo and Sanz-Gracia, Fernando and González-Val, Rafael

Departamento de Análisis Económico. Universidad de Zaragoza, Institut d'Economia de Barcelona (IEB), Universitat de Barcelona

9 June 2014

# On the parametric description of US city size distribution: New empirical evidence

ARTURO RAMOS[*] FERNANDO SANZ-GRACIA[†]
RAFAEL GONZÁLEZ-VAL[‡]

June 9, 2014

**Abstract**

We study US city size distribution using places data from the Census, without size restrictions, for the period 1900-2010, and the recently constructed US City Clustering Algorithm (CCA) data for 1991 and 2000.

We compare the lognormal and the double Pareto lognormal with two newly introduced distributions. The empirical results are overwhelming: one of the new distributions greatly outperforms any of the previously-used density functions for both types of data.

We also discuss the implications of these results for the possible existence of a class of stochastic processes broader than the standard geometric Brownian motion with drift with or without a Yule process, which might generate the new density functions.

**JEL:**C13, C16, R00.

**Keywords:** US city size distribution, population thresholds, lower and upper tail, new statistical distributions

[*]Department of Economic Analysis, Zaragoza (SPAIN) aramos@unizar.es
[†]Department of Economic Analysis, Zaragoza (SPAIN) fsanz@unizar.es
[‡]Department of Economic Analysis, Universidad de Zaragoza and Institut d'Economia de Barcelona (IEB), Universitat de Barcelona (SPAIN) rafaelg@unizar.es

# 1 Introduction

The study of city size distribution has a long tradition, a few examples being Black and Henderson (2003), Ioannides and Overman (2003), Soo (2005), Anderson and Ge (2005), Bosker et al. (2008) and the more recent ones of Giesen et al. (2010) and Ioannides and Skouras (2013).

Over the years, the Pareto distribution (Pareto, 1896) (for the upper tail, subindex "ut") has generated a huge amount of research and received widespread acceptance. The normalized density function for this distribution reads

$$f_{\mathrm{ut}}(x, x_m, \zeta) = \frac{\zeta}{x}\left(\frac{x_m}{x}\right)^{\zeta}, \quad x > x_m ,$$

where $x > x_m$ is the population of urban centers, $x_m$ is the minimum threshold size and $\zeta > 0$ is the *Pareto exponent*. [1]

In an influential paper on city size distribution, Eeckhout (2004) essentially proposes the lognormal (abbreviated in this work as "lgn") to describe it, using US Census data for the year 2000 of all unincorporated and incorporated places in his analysis. Lognormal distributions had previously been proposed by Parr and Suzuki (1973), but one of the main points in Eeckhout (2004) is that one should take into account the whole set of cities when studying their distribution. Later, Levy (2009) argued that the upper tail of the city size distribution and, thus, most of the population (for the US places), followed a Pareto distribution, not a lognormal one.

In this line of research, the contribution of Ioannides and Skouras (2013) has appeared; it aims to reconcile the two views by means of the proposal of two distributions which have a lognormal body and, above an explicit threshold, a Pareto power law or a linear combination of Pareto and lognormal in the upper tail.

In parallel to the appearance of these works, a distribution has been proposed which has a lognormal body and power laws in the tails, but without clearly delineating between the three behaviors, called the double Pareto lognormal (dPln); see, e.g., Reed (2002, 2003) and Reed and Jorgensen (2004). The fit of this distribution is remarkably good for a number of countries (see Giesen et al. (2010), for eight countries, and the recent contribution of González-Val et al. (2013b) for a more comprehensive data set).

In what follows of this Introduction, we will try to motivate the appropriateness of our approach (see Section 3 for details).

[1] The cumulative distribution function is

$$\mathrm{cdf}_{\mathrm{ut}}(x, x_m, \zeta) = 1 - \left(\frac{x_m}{x}\right)^{\zeta}, \quad x > x_m$$

so that

$$1 - \mathrm{cdf}_{\mathrm{ut}}(x, x_m, \zeta) = \left(\frac{x_m}{x}\right)^{\zeta}$$

and

$$\ln(1 - \mathrm{cdf}_{\mathrm{ut}}(x, x_m, \zeta)) = \zeta \ln x_m - \zeta \ln x$$

Thus, for a Pareto distribution, the quantity $\ln(1 - \mathrm{cdf})$ is linear in $\ln x$ with a negative slope of absolute value $\zeta$. The case of $\zeta = 1$ corresponds to the well-known *Zipf's law* (Zipf, 1949); see the surveys on this subject by Cheshire (1999) and Gabaix and Ioannides (2004). This is the basis of the well-known *Zipf plots*.

Nowadays, there is a certain consensus in the study of city size distribution that a combination of Pareto and lognormal provides the best fit, the distributions of Ioannides and Skouras (2013) having a component of Pareto only in the upper tail and dPln having components of Pareto in the upper and lower tails. We build on this relevant strand of the literature and go further in two ways. First, by proposing two new distributions that systematically outperform the lognormal and dPln.[2] Second, pointing out the implications of these empirical conclusions for the existence of an underlying stochastic process more general than the geometric Brownian motion with drift with or without an associated Yule process, which in principle might be able to generate the two newly introduced parametric densities.

For the lower tails (subindex "lt") of city size distributions, Reed (2001, 2002) observes that they indeed follow a power law, using the smallest 5,000 settlements for the US in 1998. He plots the natural logarithm of cumulative frequencies against that of population and observes indeed a linear behavior.[3] This fact seems to be overlooked in the literature and, as we will see below, is one of the important points one should take into account in order to obtain an excellent overall fit.

Against this background, we have decided to compare in detail the lognormal and dPln distributions with two new ones which contain the essence of the views of the dPln and the distributions of Ioannides and Skouras (2013) and take a step forward. They are:

- The "threshold double Pareto Singh–Maddala" (tdPSM), which is a distribution with a Singh–Maddala one (Singh and Maddala, 1976) in the body and with both tails following a power law, but with two thresholds which exactly delineate the switch between the different behaviors. It is like the first distribution of Ioannides and Skouras (2013) but with the lower tail modeled as a pure power law and the body being Singh–Maddala instead of lognormal. As far as we know, the tdPSM is a completely new distribution.

- The "double mixture Pareto Champernowne Pareto" (dm PChP), which is a distribution with a Champernowne distribution (Champernowne, 1952) body and with a linear combination of Champernowne and Pareto in both tails, also with two population thresholds which exactly delineate the switch between the different behaviors. It is like the second distribution of Ioannides and Skouras (2013) but with the lower tail modeled as a mixture of Champernowne and power law,

[2] And also systematically outperform the distributions of Ioannides and Skouras (2013), more details avalaible from the authors upon request.

[3] For the lower tail, we can define the Pareto density function

$$f_{\mathrm{lt}}(x, x_M, \rho) = \frac{\rho}{x}\left(\frac{x}{x_M}\right)^{\rho}, \quad 0 < x < x_M ,$$

where $x_M$ is now the maximum size threshold and $\rho > 1$ is the Pareto exponent. The cumulative distribution function is then

$$\mathrm{cdf}_{\mathrm{lt}}(x, x_M, \rho) = \left(\frac{x}{x_M}\right)^{\rho}, \quad 0 < x < x_M ,$$

and, therefore, $\ln(\mathrm{cdf}_{\mathrm{lt}}(x, x_M, \rho)) = \rho \ln x - \rho \ln x_M$. So, we have that, for a lower tail Pareto distribution, the natural logarithm of $\mathrm{cdf}$ gives a straight line in $\ln x$ with a positive slope $\rho$. We will plot the $\ln(\mathrm{cdf})$'s in the left-hand panels of Figures 1 and 2.

and the lognormal substituted by a Champernowne in general. This is, to the best of our knowledge, also a new distribution.[4]

These distributions yield extremely good, strong and encouraging results, and they are based on the following important improvements:

- The extremely important need to specifically model the lower tail as a power law in order to get an overall good fit, as mentioned above.
- The mixtures in the tails become very important when considering some of our data; this is due to the fact that the tails of these samples are slightly curved on a log-log plot and so the Pareto needs to be combined with another distribution in order to improve the fit notably.
- The use of the Singh–Maddala and Champernowne distributions instead of the lognormal all lead to a very important improvement. This means that the standard theories generating the lognormal (Eeckhout, 2004) or the double Pareto lognormal (Reed, 2002, 2003; Reed and Jorgensen, 2004) can be enhanced notably.

The article is organised as follows. Section 2 describes the databases used. Section 3 motivates the need to search for new and better distributions. Section 4 shows the definitions and main properties of the distributions studied. Section 5 shows the detailed results. In Section 6, we offer a discussion. Finally, Section 7 concludes.

# 2 The databases

In this article, we use data about US urban centers from three sources. The first is the decennial data of the US Census Bureau of "incorporated places" without any size restriction, in the period 1900-2000. These include governmental units classified under state laws as cities, towns, boroughs or villages. Alaska, Hawaii and Puerto Rico have not been considered due to data limitations. The data have been collected from the original documents of the annual census published by the US Census Bureau.[5] This data was first introduced in González-Val (2010), see therein for details, and later used in other works like González-Val et al. (2013b).

The second source consists of all US urban places, unincorporated and incorporated, and without size restrictions, also provided by the US Census Bureau for the

[4]These two distributions are the outcome of a research process in which we have tried different ones. We started with the lognormal for the body as it is used in the distributions of Ioannides and Skouras (2013). But we realized that a much better performance could be obtained with the Fisk ("Fi") distribution (Fisk, 1961) for the body and (the mixtures at) the Pareto tails. Both the Singh–Maddala and Champernowne distributions generalize that of Fisk (and have one parameter more) so we tried them as well. For the sake of brevity, we present only *the best results obtained*, corresponding to the new distributions mentioned. We have also worked with (with obvious notation) tdPln, tdPFi, tdPCh, dm PlnP, dm PFiP, dm PSMP that, although all provide better results than the lognormal, dPln, and the distributions of Ioannides and Skouras (2013), perform worse than the ones finally presented here.

[5]`http://www.census.gov/prod/www/decennial.html` Last accessed: June 9$^{\text{th}}$, 2014.

years 2000 and 2010. The data for the year 2000 was first used in Eeckhout (2004) and later in Levy (2009), Eeckhout (2009), Giesen et al. (2010), Ioannides and Skouras (2013) and Giesen and Suedekum (2013). The two samples were also used in González-Val et al. (2013b).

The third comes from a different and recent approach to defining city centers, described in detail in Rozenfeld et al. (2008, 2011). They use a so called "City Clustering Algorithm" (CCA) to get "an automated and systematic way of building population clusters based on the geographical location of people." (*op. cit.*) We use their US clusters data based on the radii of 2, 3, 4, 5 km. and for the years 1991 and 2000. This data was used in Ioannides and Skouras (2013) and Giesen and Suedekum (2013).

[Table 1 near here]

The descriptive statistics of the data can be seen in Table 1. As Giesen and Suedekum (2013) indicate, the CCA data comprises a higher percentage of the whole population than the Census data.

# 3 Motivation of our approach

As a preliminary analysis, we take the sample of all US places in 2010, in order to see whether the previous lognormal and dPln provide a good fit. For these density functions, we use some of the estimation results in Table 2. In Figure 1, we show, in the left-hand panel, the empirical and estimated (by maximum likelihood, ML) $\ln(\mathrm{cdf})$ against $\ln x$ for the lower tail and in the right-hand panel, the analogous quantities $\ln(1-\mathrm{cdf})$ against $\ln x$ for the upper tail.[6] In the center panel, we show the usual empirical density functions (obtained through an adaptive Gaussian kernel) compared to the estimated density functions, all three for the cases of the lognormal and dPln.

[Figure 1 near here]

We see, in the left-hand panel of Figure 1, that the $\ln(\mathrm{cdf})$'s of both the lognormal and dPln (in red) are not so linear as the empirical ones, the fit of the dPln being slightly better than that of the lognormal. In the middle panel, we observe that the empirical and estimated densities differ clearly in the body and also in the tails. In the right-hand panel, corresponding to the upper tails, we see that the fit is also not so good for the lognormal and dPln (serious discrepancies starting at $\ln x > 11$, i.e., $x > 59,874$ inhabitants).[7] Advancing some results of Table 4, we will see that both of two standard but demanding tests, given the high sample size, (Kolmogorov–Smirnov (KS) and Cramér-Von Mises (CM)) clearly reject the cited models.[8]

[6]The difference between the empirical and the estimated quantities are amplified because we take the natural logarithms of cdf or $(1-\mathrm{cdf})$ for the lower and upper tails, respectively (González-Val et al., 2013a).

[7]A linear OLS estimation has been calculated and shown in green, only for reference purposes, for the lower and upper tails. If one wanted to obtain accurate numerical results by this method, techniques inspired in Gabaix and Ibragimov (2011) might be appropriate for both tails. However, our formal estimations are performed by the standard maximum likelihood (ML).

[8]When performing the tests, we take the whole studied sample, and not subsamples, in order to achieve the maximum power of the KS and CM tests (compare with Giesen and Suedekum (2013)).

Therefore, it makes sense to look for one or a number of new distributions that cannot be rejected in the majority of cases and that offer a better fit to the data. We will see that this can be achieved by introducing some simple but significant changes into the distributions of Ioannides and Skouras (2013), which act as our baseline distributions.

# 4 Description of the distributions used

In this section, we will introduce the distributions used in the paper. Firstly, we recall the lognormal and define some basic functions which are employed by our new distributions.

We thus set

$$f_{\mathrm{lgn}}(x,\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right) \tag{1}$$

$$f_{\mathrm{SM}}(x,\mu,\sigma,\alpha) = \frac{\alpha\,(\mathrm{e}^{-\mu}x)^{1/\sigma}}{x\sigma(1+(\mathrm{e}^{-\mu}x)^{1/\sigma})^{1+\alpha}} \tag{2}$$

$$f_{\mathrm{Ch}}(x,\mu,\sigma,\beta) = \frac{\sin\beta}{x\beta\sigma((\mathrm{e}^{-\mu}x)^{-1/\sigma}+(\mathrm{e}^{-\mu}x)^{1/\sigma}+2\cos\beta)} \tag{3}$$

$$u(x,\zeta) = \frac{1}{x^{1+\zeta}} \tag{4}$$

$$l(x,\rho) = x^{\rho-1} \tag{5}$$

where $\mu,\sigma>0$ are, respectively, the mean and the standard deviation of $\ln x$ for the lognormal density $f_{\mathrm{lgn}}$. For the $f_{\mathrm{SM}}$, $f_{\mathrm{Ch}}$ distributions, the corresponding $\mu,\sigma>0$ are also related to the mean and standard deviation of $\ln x$ (Singh and Maddala, 1976; Champernowne, 1952).[9] The function $u(x,\zeta)$ will model the Pareto part of the upper tail of our distributions, $\zeta>0$ is the Pareto exponent, and $l(x,\rho)$ corresponds to the Pareto lower tail, $\rho>1$ being the power law exponent. The functions $u, l$ are not normalized at this stage in accordance with the practice of Ioannides and Skouras (2013).

## 4.1 The double Pareto lognormal distribution (dPln)

The probability density function of the double Pareto lognormal distribution is (Reed, 2002, 2003; Reed and Jorgensen, 2004):

$$\begin{aligned} f_1(x,\alpha,\beta,\mu,\sigma) = {} & \frac{\alpha\beta}{2x(\alpha+\beta)}\exp\left(\alpha\mu+\frac{\alpha^2\sigma^2}{2}\right)x^{-\alpha}\left(1+\mathrm{erf}\left(\frac{\ln x-\mu-\alpha\sigma^2}{\sqrt{2}\sigma}\right)\right) \\ & -\frac{\alpha\beta}{2x(\alpha+\beta)}\exp\left(-\beta\mu+\frac{\beta^2\sigma^2}{2}\right)x^{\beta}\left(\mathrm{erf}\left(\frac{\ln x-\mu+\beta\sigma^2}{\sqrt{2}\sigma}\right)-1\right) \end{aligned} \tag{6}$$

[9] We have taken the Champernowne density (2.4) in Champernowne (1952) with $\lambda=\cos\beta$ since this particular specification covers all the cases estimated in this paper. Also, the $f_{\mathrm{SM}}$ is directly related to the Burr Type XII distribution (Burr, 1942). See also Kleiber and Kotz (2003).

where $\alpha, \beta, \mu, \sigma > 0$ are the four distribution parameters to be estimated. The dPln distribution has the property that it approximates different power laws at its two tails, namely $f_1(x) \approx x^{-\alpha-1}$ when $x \to \infty$ and $f_1(x) \approx x^{\beta-1}$ when $x \to 0$, hence the name double Pareto. The central part of the distribution is approximately lognormal, although it is not possible to delineate the lognormal body part and the Pareto tails exactly (Giesen et al., 2010).

The dPln distribution is the steady-state distribution of an evolutionary process of a simple stochastic model of settlement formation and growth based on a geometric Brownian motion with drift and a Yule process. Mathematically, the dPln is the log version of the convolution of the normal distribution and the (asymmetric) double Laplace distribution, see Reed (2002, 2003); Reed and Jorgensen (2004) and references therein for details.

For more recent work on an economic model which incorporates the stochastic derivation of Reed (2002, 2003), see Giesen and Suedekum (2012, 2013). The key in these latter models is the endogenous creation of cities and the resulting age heterogeneity in cities within the distribution. Giesen and Suedekum (2012, 2013) argue that Eeckhout (2004)'s theoretical framework and the lognormal distribution represent a particular scenario of their model, the case in which there is no city creation and all cities have the same age.

## 4.2 The threshold double Pareto Singh–Maddala (tdPSM)

We introduce here the first of our distributions. It is a variant of the first distribution of Ioannides and Skouras (2013) in which we model the lower tail as a Pareto power law and the body as Singh–Maddala instead of lognormal. Thus, the tdPSM has a Singh–Maddala body and Pareto tails, the three regions exactly delineated by two thresholds: $\epsilon > 0$ separates the Pareto power law in the lower tail from the Singh–Maddala body, and $\tau > \epsilon$ separates the body from the Pareto power law in the upper tail. We impose continuity of the density function on the two threshold points and normalization of the former to unity. The resulting density reads

$$f_2(x, \rho, \epsilon, \mu, \sigma, \alpha, \tau, \zeta) = \begin{cases} b_2\, e_2\, l(x, \rho) & 0 < x < \epsilon \\ b_2\, f_{\mathrm{SM}}(x, \mu, \sigma, \alpha) & \epsilon \leq x \leq \tau \\ b_2\, a_2\, u(x, \zeta) & \tau < x \end{cases} \tag{7}$$

where now

$$e_2 = \frac{f_{\mathrm{SM}}(\epsilon, \mu, \sigma, \alpha)}{l(\epsilon, \rho)} \tag{8}$$

$$a_2 = \frac{f_{\mathrm{SM}}(\tau, \mu, \sigma, \alpha)}{u(\tau, \zeta)} \tag{9}$$

$$b_2^{-1} = e_2 \frac{\epsilon^\rho}{\rho} + \mathrm{e}^{\mu\alpha/\sigma}((\mathrm{e}^{\mu/\sigma} + \epsilon^{1/\sigma})^{-\alpha} - (\mathrm{e}^{\mu/\sigma} + \tau^{1/\sigma})^{-\alpha}) + \frac{a_2}{\zeta\, \tau^\zeta} \tag{10}$$

This distribution depends on seven parameters $(\rho, \epsilon, \mu, \sigma, \alpha, \tau, \zeta)$ to be estimated.

### 4.3 The double mixture Pareto Champernowne Pareto (dm PChP)

The second distribution we introduce is a variant of the second distribution of Ioannides and Skouras (2013) in the sense that we now consider linear combinations of the Champernowne and Pareto distributions in the two tails, while maintaining a Champernowne body. The tails and the body are separated by two exact thresholds $\epsilon$ and $\tau$ with similar meaning to those of the tdPSM. For the lower tail, the combining coefficient will be denoted by $\nu$, and $\theta$ for the upper tail as before. We require, as usual, continuity of the density function at the threshold points and overall normalization to one. The following conditions are also imposed:

$$a_3 \int_\tau^\infty u(x,\zeta)\,dx = c_3 \int_\tau^\infty f_{\mathrm{Ch}}(x,\mu,\sigma,\beta)\,dx$$
$$e_3 \int_0^\epsilon l(x,\rho)\,dx = d_3 \int_0^\epsilon f_{\mathrm{Ch}}(x,\mu,\sigma,\beta)\,dx$$

so that the parameters $\theta$, $\nu$ control the proportion of the density in the combination in the upper (resp. lower) tail. The resulting composite density is given by:

$$\begin{aligned} & f_3(x,\rho,\epsilon,\nu,\mu,\sigma,\beta,\tau,\zeta,\theta) \\ & = \begin{cases} b_3\ [(1-\nu)\,d_3\,f_{\mathrm{Ch}}(x,\mu,\sigma,\beta) + \nu\,e_3\,l(x,\rho)] & 0 < x < \epsilon \\ b_3\,f_{\mathrm{Ch}}(x,\mu,\sigma,\beta) & \epsilon \le x \le \tau \\ b_3\ [(1-\theta)\,c_3\,f_{\mathrm{Ch}}(x,\mu,\sigma,\beta) + \theta\,a_3\,u(x,\zeta)] & \tau < x \end{cases} \end{aligned} \tag{11}$$

where the constants are now given as follows:

$$d_3^{-1} = 1 - \nu + \frac{\nu\rho(\beta - \operatorname{arccot}[\cot\beta + (\mathrm{e}^{-\mu}\epsilon)^{1/\sigma}\csc\beta])l(\epsilon,\rho)}{\epsilon^\rho\beta f_{\mathrm{Ch}}(\epsilon,\mu,\sigma,\beta)} \tag{12}$$

$$e_3^{-1} = \frac{\beta\epsilon^\rho(1-\nu)}{\rho(\beta - \operatorname{arccot}[\cot\beta + (\mathrm{e}^{-\mu}\epsilon)^{1/\sigma}\csc\beta])} + \frac{\nu\,l(\epsilon,\rho)}{f_{\mathrm{Ch}}(\epsilon,\mu,\sigma,\beta)} \tag{13}$$

$$c_3^{-1} = 1 - \theta + \frac{\theta\zeta\tau^\zeta \operatorname{arccot}[\cot\beta + (\mathrm{e}^{-\mu}\tau)^{1/\sigma}\csc\beta]u(\tau,\zeta)}{\beta f_{\mathrm{Ch}}(\tau,\mu,\sigma,\beta)} \tag{14}$$

$$a_3^{-1} = \frac{\beta(1-\theta)}{\zeta\tau^\zeta \operatorname{arccot}[\cot\beta + (\mathrm{e}^{-\mu}\tau)^{1/\sigma}\csc\beta]} + \frac{\theta\,u(\tau,\zeta)}{f_{\mathrm{Ch}}(\tau,\mu,\sigma,\beta)} \tag{15}$$

$$\begin{aligned} b_3^{-1} = {} & e_3\frac{\epsilon^\rho}{\rho} + \frac{1}{\beta}\arctan\left(\frac{\sin\beta}{(\mathrm{e}^{-\mu}\epsilon)^{1/\sigma} + \cos\beta}\right) \\ & -\frac{1}{\beta}\arctan\left(\frac{\sin\beta}{(\mathrm{e}^{-\mu}\tau)^{1/\sigma} + \cos\beta}\right) + \frac{a_3}{\zeta\,\tau^\zeta} \end{aligned} \tag{16}$$

This distribution depends on nine parameters $(\rho,\epsilon,\nu,\mu,\sigma,\beta,\tau,\zeta,\theta)$ to be estimated.

# 5 Results

## 5.1 Estimation of the distributions

Maximum likelihood (ML) is a standard technique which allows the estimation of the parameters of a distribution given a sample of data. For the case of the lognormal density function, the corresponding ML estimators can be found easily in an exact closed form (the $\mu$ and $\sigma$ are then the mean and the standard deviation (SD) of the natural logarithm of the data). However, for the other distributions $f_1,\ f_2,\ f_3$ used in this article, one must resort to numerical optimization methods in order to find the ML estimators.[10] It is worth noting that the threshold population parameters $\epsilon$ and $\tau$ present in the cited density functions are to be estimated endogenously by ML, letting the data "decide" the optimum threshold values which maximize the log-likelihood.

Previous work on similar matters includes that of Bee (2012), which deals with a distribution similar to the first one of Ioannides and Skouras (2013) with ML. The log-likelihood function of the dPln is also found in Reed and Jorgensen (2004). The other cases of our paper can be dealt with in a similar fashion.[11]

When performing the estimations, not all density functions can always be treated by our numerical procedure because it seems that, in the corresponding cases, the estimators simply do not exist. This may happen when dealing with composite densities, see, e.g., Bee (2012) for a theoretical discussion in a related sample situation. Specifically, for the US places data, the dm PChP cannot be estimated so, for the sake of brevity, we include only the results of the new distributions which can be estimated for each type of data (US places and CCA clusters, separately) and for all periods and which provide the best performance in each case.

We present the results of the estimation procedure for the US places data in Table 2. For the sample of the US (2000, all places) we essentially replicate the results of Giesen et al. (2010) and Giesen and Suedekum (2013) for the dPln, apart from slight non-essential and very small numerical discrepancies. We have found that the log-likelihood function is smooth near its maximum in all of the estimated cases, see also Bee (2012).

[Table 2 near here]

We observe in these results that the tdPSM offers quite stable, or with a soft trend, estimates. Its lower ($\epsilon$) threshold vary between 99 and 178, and the upper ($\tau$) threshold vary between 3,405 and 54,144. This is an observed first good feature of the tdPSM.

Next, we show the estimation results for the US CCA samples in Table 3. For these data, we also replicate essentially the results of Giesen and Suedekum (2013). The estimation process is smoother than for the places data, and the distribution dm PChP can be estimated for all of these samples. This is a remarkable feature of the cluster data: the City Clustering Algorithm considers an actual agglomeration of people within a prescribed radius as an urban center, irrespectively of legally-established borders, giving an economic and physical entity to the clusters considered. This fact seems to

[10] We have used MATLAB in order to perform the ML estimations.

[11] More details are available from the authors upon request.

be reflected in the data obtained, which allows the estimation of more density functions and, in general, with narrower confidence intervals. For the dm PChP, $\epsilon$ varies between 1,118 and 2,671 and $\tau$ between 14,253 and 20,381.

[Table 3 near here]

We have used the graphical tools in Section 3 to introduce the need of continuing to search for distributions with better fit. But, when performing a high precision exercise, these graphical tools can be misleading in assessing the quality of fit, see González-Val et al. (2013a). So, we resort to standard statistical tests and information criteria to see when the hypothesized distributions offer a good fit and which model is selected from amongst the ones studied. This is done in the following subsections.

## 5.2 Standard statistical tests

In this subsection, we provide independent tests to verify the goodness of fit in all of the cases studied. As in González-Val et al. (2013b), we have chosen the Kolmogorov–Smirnov (KS) test, which is also mentioned in Giesen et al. (2010), Giesen and Suedekum (2012, 2013) and is standard in the literature. We also use the Cramér-von Mises (CM) test, cited in Ioannides and Skouras (2013).

The KS and CM tests have similar power, quite low for small sample sizes but very high for large sample sizes (Razali and Wah, 2011). Both tests are extremely precise for large and very large sample sizes like the ones used in this paper, for which non rejections only occur if the deviations (statistics) are extremely small. The significance level chosen is always 5%. Non rejections are indicated in bold.

[Tables 4 and 5 near here]

In Table 4, we show the results for the samples of US places. We offer the $p$-values of the tests together with the values of the statistics (in parentheses). A first observation is that the lognormal model is very strongly rejected for all samples. The dPln is also rejected in almost all cases (except two). Moreover, a big jump in performance is obtained with the tdPSM. Indeed, this distribution is not rejected in 100% of the cases, with much lower values of the tests' statistics. Thus, modeling both tails as a pure Pareto and the body as the Singh–Maddala distribution leads to a strikingly better improvement. Thus, the tdPSM reveals itself as an excellent and robust specification for the US places size distribution.

We move on to the results of the tests for the US CCA clusters in Table 5. Again, we show the $p$-values and the tests' statistics in parentheses. Here, the lognormal is again always strongly rejected. The dPln is always rejected as well (with lower values of the tests' statistics). Again, a wide jump is obtained when considering the dm PChP, which is not rejected in 100% of the cases, with considerably lower values of the tests' statistics. This means that modeling the two tails as a Pareto-Champernowne mix and the body as Champernowne leads to an excellent fit. These final results are robust to the different radii the clusters are constructed with (2, 3, 4 and 5 km.), and to the years studied (1991 and 2000). In this way, we obtain an excellent model for the US CCA clusters size distribution, the dm PChP.

In the next subsection, we study the distributions with the information criteria.

## 5.3 Information criteria

To select a distribution from among those studied, we compute two information criteria very well-suited to the maximum likelihood method which we have used to estimate the parameters of the distributions, namely, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (see, e.g., Burnham and Anderson (2002, 2004); Giesen et al. (2010) and references therein).

[Tables 6 and 7 near here]

In Table 6, we show the results for the US places samples and the distributions presented. We obtain a similar result to those of the KS and CM tests: choosing an ordering of ascending values of the AIC for each sample (the results with the BIC are the same), we deduce a robust ordering of the distributions (the lower the value of AIC, the better the distribution). For the US incorporated places and all places samples in the period 1900-2010 we have

$$\mathrm{AIC}_{\mathrm{tdPSM}} < \mathrm{AIC}_{\mathrm{dPln}} < \mathrm{AIC}_{\mathrm{lgn}}$$

Therefore, the selected model is the tdPSM in 100% of the samples. This, together with the outcomes of the KS and CM tests, yields a new and strong result: the US city size distribution (incorporated places and all places) can be safely taken as the new tdPSM.

For the US CCA cluster samples, we refer to Table 7. We again have strong regularities. The ordering of the distributions by ascending values of the AIC is (the ordering by BIC is the same)

$$\mathrm{AIC}_{\mathrm{dm\ PChP}} < \mathrm{AIC}_{\mathrm{dPln}} < \mathrm{AIC}_{\mathrm{lgn}}$$

It is striking that our new distribution dm PChP is systematically preferred to others known up to now in the literature. In short, we have that the selected distribution (amongst those studied here and others not shown for the sake of brevity) is the dm PChP in 100% of the cases, with values of the AIC and BIC much lower than for the other previously-known distributions. This, together with the results of the KS and CM tests, yields a second strong and new result: the US city size distribution (CCA clusters) can be safely taken as the new dm PChP.

In both the US places and CCA clusters samples, we have another result. To achieve an exceptional performance, it seems to be essential to model both tails as a Pareto distribution, in a pure form, with a Singh–Maddala body (places), or as part of a linear combination mixture with the Champernowne distribution, and a Champernowne body (clusters).[12]

As a complement to the KS, CM, AIC and BIC results, in Figure 2, we show an informal graphical approximation of the fits obtained in two different cases. The first row for the sample of all US places (2010) and the tdPSM, and the second for the sample of US CCA clusters (2000, 2km.) and the dm PChP. We see that the lower

[12]It is worth mentioning that both of the AIC and BIC information criteria penalize the number of parameters of the compared distributions, so increasing the number of parameters of an hypothesized distribution does not necessarily yield a better fit nor a lower AIC or BIC. Thus, the fact that the selected distributions have a high number of parameters means that the fit is really good. In the same way, the fact that the worst distribution (out of the ones compared according to these criteria) has only two parameters means that the fit it provides, compared with the others studied, is quite poor.

tail of the first sample fits nicely (the empirical $\ln(\text{cdf})$ of that of clusters is not so linear), for the upper tails the fit is quite remarkable in the two cases and, for the middle panel, it is very hard to see discrepancies between the empirical and estimated density functions, compare with Figure 1. In particular, visually the improvement of the tdPSM with respect to the dPln is greater than the improvement of the dPln with respect to the lognormal. The formal tests performed in Subsection 5.2 agree with this visual appreciation.

[Figure 2 near here]

[Table 8 near here]

We also show, in Table 8, the percentages of population and urban units in the tails and the body of the selected distributions for each type of data (places and clusters). As an approximation, we classify the urban units in the lower tail as those having a population less than the value of the $\epsilon$ threshold, those in the upper tail having a population greater than the $\tau$ threshold, and the body is formed by urban units with a population between $\epsilon$ and $\tau$. The values of these thresholds for places are those of Table 2 and for clusters those of Table 3. It can be observed that, although the percentages of population in the lower tails are generally quite low, the percentages of urban units in the lower tail are comparable to or even higher than those in the upper tail. This fact explains the need to take into account the appropriate modeling of the lower tail in order to obtain an excellent overall fit.

# 6 Discussion

We have seen that two new density functions perform better than some previously known ones including the lognormal used by Eeckhout (2004) and others, and the dPln of Reed (2002, 2003); Reed and Jorgensen (2004); Giesen et al. (2010) and others, when fitting US city data. More precisely, the tdPSM is the preferred model for US incorporated and all places data and the dm PChP is the preferred density function for the US CCA clusters of Rozenfeld et al. (2008, 2011).

In our current study we have considered the lognormal and double Pareto lognormal specifications, which are generated by well known stochastic processes. Indeed, the first is obtained when considering a process where the log-increments are independent on the initial log-sizes and stationarily distributed with finite variance, which is also considered sometimes as a direct implementation of Gibrat's law (Sutton, 1997; Eeckhout, 2004). The second is generated by a geometric Brownian motion with drift and a Yule process (Reed, 2002, 2003; Reed and Jorgensen, 2004) which accounts for the endogenous creation of cities and age heterogeneity within the urban system. In addition, the Pareto distribution for the whole range of the population variable can be obtained as well from an underlying geometric Brownian motion with drift plus a reflecting barrier at very low values of the population (Gabaix, 1999, 2009). These three parametric models have associated economic foundations (*op. cit.* and references therein).

However, we have seen that both the lognormal and the double Pareto lognormal are rejected by our empirical analysis in almost all cases, by means of standard statistical tests.[13]

In turn, by the same methods and techniques, we obtain that the tdPSM is never rejected for US places and that the dm PChP is never rejected for US CCA clusters, and with remarkable improvements in the values of the statistics of the KS and CM tests. In addition, the values of the AIC and BIC of the new distributions are greatly improved with respect to those of the other distributions studied, despite of the new density functions having more parameters.

This means that, probably, there is something substantial in the newly introduced distributions, and that one should study further the stochastic processes that would occasionally generate the tdPSM and the dm PChP. These processes should be more general and/or different than the cited geometric Brownian motion with drift with or without a Yule process.

In any case, the cited and still not characterized stochastic process should have certain features already known in the literature of city size distributions.

The first is the stability and persistence of the population and hierarchical structure of cities over time, at least in the short term, as the empirical evidence clearly shows (Black and Henderson, 1999; Kim, 2000; Beeson et al., 2001; Sharma, 2003). Furthermore, this stability or persistence is corroborated even when the cities suffer strong temporal shocks, like the US Civil War (Sanso-Navarro et al., 2013), the WWII atomic bombing in Japan (Davis and Weinstein, 2002), the WWII bombing in Germany (Brakman et al., 2004; Bosker et al., 2008), the US bombing in Vietnam (Miguel and Roland, 2011) and urban terrorism (Glaeser and Shapiro, 2002).[14]

Another feature of the desired stochastic process is that it should generate pure Pareto power laws at the tails, mixed or not with the distribution present at the body, which would be preferably a Singh–Maddala or Champernowne distribution. It is striking that Gibrat's law, understood as an implementation of a geometric Brownian motion with drift and a low value reflecting barrier, is a sufficient condition for the generation of the Pareto upper tail (Gabaix, 1999, 2009), but it seems to be not necessary. For the upper tail, this fact is known in the firm size distribution literature, see, e.g., Fujiwara et al. (2004). However, it could still be the case that Gibrat's law, understood as the statistical independence of the log-growth rates on the initial log-sizes (see, e.g., Sutton (1997) and references therein), may hold even when the underlying stochastic process is not a geometric Brownian motion with drift. This also points out to the need of studying alternative stochastic processes which could generate the new distributions.

In short, we have achieved to describe two new parametric probability density functions which can never be rejected empirically for each type of US city data (places and CCA clusters), have pure Pareto tails mixed or not with the body by means of a linear combination, and that are presumably generated by a process that is more general and/or different than the geometric Brownian motion with drift, with or without a Yule

[13] The Pareto distribution for the whole range of the population variable is always rejected as well by both the KS and CM tests. More details are available from the authors upon request.

[14] In the long term, things become different: in the extreme situation, we have the contribution of Batty (2006), which defends that the changes in the internal hierarchy of cities can be very important, although the aggregate distribution appears to be quite stable. This is not incompatible with the short term persistence literature, because Batty's temporal horizon is very large (world data from 430 BC.)

process.

The further study of these, maybe new, stochastic processes is a subject of current research.

# 7 Conclusions

Since the work of Eeckhout (2004), the risks of considering only the largest cities, that is, only the upper tail, have been demonstrated. One of the main lessons of this work is that, when possible, one should use city data without minimum size restrictions.[15] In turn, if the availability of data allows it, the analysis of city size distribution should be done in the long term. With both considerations as premises, this article uses US Census data for the period 1900-2010, incorporated places from 1900 to 2000, in decades, and all places for 2000 and 2010. We also use the US City Clustering Algorithm (CCA) clusters data of Rozenfeld et al. (2008, 2011) for the years 1991 and 2000 and radii of the clusters of 2, 3, 4 and 5 km.

This work has minutely examined four density functions. As well as the lognormal and dPln, known in the field of city size distributions, we have explicitly introduced into Section 4 two new density functions, which we call tdPSM and dm PChP. The essential point of the new functions is the modeling of *both* tails as a Pareto distribution with or without mixing (by means of a linear combination) with the Champernowne or Singh–Maddala distributions, which conform the body.

These two new distributions are associated with two "philosophical" principles:

i) For the US, it seems to be necessary to pay attention to the lower tail of the distribution, despite it represents a small percentage of the population, in order to obtain an excellent overall fit. In a nutshell, *small nuclei do matter*.

ii) The body of the distribution is better described by a Singh–Maddala or Champernowne distribution than by a lognormal. This constitutes a relevant difference to the evidence accumulated so far.

After estimating the parameters of all of the distributions by maximum likelihood (ML), we have tested the fit provided by each distribution using the Kolmogorov–Smirnov (KS) and Cramér-von Mises (CM) tests. Afterwards, we have computed the AIC and BIC information criteria.

The results are extremely robust and regular. The two new density functions notably improve on the performance of the lognormal and dPln. The tdPSM is a new distribution that is not rejected in 100% of the cases by either the KS or the CM tests,

[15] In this work, we have not shown the results corresponding to the data of the so-called Metropolitan and Micropolitan areas (MMA), see, e.g., Ioannides and Skouras (2013) for their definition, because, in them, a not small minimum threshold size (about 13,000 inhabitants) is imposed. We simply mention that the KS and CM tests for a truncated version of all of the distributions used in this paper yield rejection, even though the sample sizes of MMA data are much lower than for US places or CCA clusters (less than 1,000 observations). This means that the modeling of the MMA size distribution is much more demanding than for the US places or CCA clusters, possibly due to the cut-off imposed on the data.

and is the model selected (of the distributions studied) by both the AIC and BIC for the whole period 1900-2010 of samples of US incorporated and all places. Likewise, the dm PChP is a new distribution that is not rejected in 100% of the cases of CCA clusters by either the KS or the CM tests, and is the model selected for all these samples by both the AIC and BIC.

In short, we find empirically that the US city size distribution for places can be safely taken as a *Singh–Maddala* body with pure Pareto tails, the three regions separated by two exact thresholds. For US CCA clusters, an analogous situation occurs but where the body is Champernowne and, in the tails, it is advantageous *to mix* the Pareto distributions with the Champernowne one.

Moreover, we have briefly discussed the theoretical implications of the empirical results on the possible existence of stochastic processes which would occasionally generate the best of the studied distributions for each type of data (places, CCA clusters). These processes should be more general and/or different than the geometric Brownian motion with drift, associated or not with a Yule process, but should still generate Pareto upper and lower tails of the city size distribution.

## Acknowledgements

This work is supported by Aragon Government, ADETRE Consolidated Group.

## References

Anderson, G. and Ge, Y. (2005). The size distribution of Chinese cities. *Regional Science and Urban Economics*, 35(6):756–776.

Batty, M. (2006). Rank clocks. *Nature*, 444(7119):592–596.

Bee, M. (2012). Statistical analysis of the lognormal-Pareto distribution using probability weighted moments and maximum likelihood. Technical report, Department of Economics, University of Trento, Italia.

Beeson, P., DeJong, D., and Troesken, W. (2001). Population growth in US counties, 1840–1990. *Regional Science and Urban Economics*, 31(6):669–699.

Black, D. and Henderson, V. (1999). Spatial evolution of population and industry in the United States. *American Economic Review*, 89(2):321–327.

Black, D. and Henderson, V. (2003). Urban evolution in the USA. *Journal of Economic Geography*, 3(4):343–372.

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2008). A century of shocks: The evolution of the German city size distribution 1925-1999. *Regional Science and Urban Economics*, 38(4):330–347.

Brakman, S., Garretsen, H., and Schramm, M. (2004). The strategic bombing of cities in Germany in World War II and its impact on city growth. *Journal of Economic Geography*, 4:201–218.

Burnham, K. and Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer-Verlag.

Burnham, K. and Anderson, D. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33:261–304.

Burr, I. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics*, 13:215–232.

Champernowne, D. (1952). The graduation of income distributions. *Econometrica*, 20(4):591–615.

Cheshire, P. (1999). Trends in sizes and structure of urban areas. In Cheshire, P. and Mills, E., editors, *Handbook of Regional and Urban Economics*, volume 3, chapter 35. Elsevier, Amsterdam.

Davis, D. and Weinstein, D. (2002). Bones, bombs and break points: The geography of economic activity. *American Economic Review*, 92:1269–1289.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. *American Economic Review*, 99:1676–1683.

Fisk, P. (1961). The graduation of income distributions. *Econometrica*, 29:171–185.

Fujiwara, Y., Di Guilmi, C., Aoyama, H., Gallegati, M., and Souma, W. (2004). Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A*, 335:197–216.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, 114:739–767.

Gabaix, X. (2009). Power laws in Economics and finance. *Annu. Rev. Econ.*, 2009:255–293.

Gabaix, X. and Ibragimov, R. (2011). Rank $-1/2$: A simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

Gabaix, X. and Ioannides, Y. (2004). The evolution of city size distributions. In Henderson, V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 53, pages 2341–2378. Elsevier.

Giesen, K. and Suedekum, J. (2012). The French overall city size distribution. *Région et Développement*, 36:107–126.

Giesen, K. and Suedekum, J. (2013). City age and city size. Conference paper, ECONSTOR.

Giesen, K., Zimmermann, A., and Suedekum, J. (2010). The size distribution across all cities-double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137.

Glaeser, E. and Shapiro, J. (2002). Cities and warfare: The impact of terrorism on urban form. *Journal of Urban Economics*, 51(2):205–224.

González-Val, R. (2010). The evolution of US city size distribution from a long term perspective (1900–2000). *Journal of Regional Science*, 50:952–972.

González-Val, R., Ramos, A., and Sanz-Gracia, F. (2013a). The accuracy of graphs to describe size distributions. *Applied Economics Letters*, 20(17):1580–1585.

González-Val, R., Ramos, A., Sanz-Gracia, F., and Vera-Cabello, M. (2013b). Size distribution for all cities: Which one is best? *Papers in Regional Science. Forthcoming. doi:10.1111/pirs.12037*.

Ioannides, Y. and Overman, H. (2003). Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics*, 33(2):127–137.

Ioannides, Y. and Skouras, S. (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics*, 73:18–29.

Kim, S. (2000). Urban development in the United States, 1690-1990. *Southern Economic Journal*, 66(4):855–880.

Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in Economics and actuarial sciences*. Wiley-Interscience.

Levy, M. (2009). Gibrat's law for (all) cities: Comment. *American Economic Review*, 99:1672–1675.

Miguel, E. and Roland, G. (2011). The long-run impact of bombing Vietnam. *Journal of Development Economics*, 96:1–15.

Pareto, V. (1896). *Cours d'economie politique*. Geneva: Droz.

Parr, J. and Suzuki, K. (1973). Settlement populations and the lognormal distribution. *Urban Studies*, 10(3):335–352.

Razali, N. and Wah, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2:21–33.

Reed, W. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, 74:15–19.

Reed, W. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, 42:1–17.

Reed, W. (2003). The Pareto law of incomes–an explanation and an extension. *Physica A*, 319:469–486.

Reed, W. and Jorgensen, M. (2004). The double Pareto-lognormal distribution–a new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 33(8):1733–1753.

Rozenfeld, H., Rybski, D., Andrade, J., Batty, M., Stanley, H., and Makse, H. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48):18702–18707.

Rozenfeld, H., Rybski, D., Gabaix, X., and Makse, H. (2011). The area and population of cities: new insights from a different perspective on cities. *American Economic Review*, 101:2205–2225.

Sanso-Navarro, M., Sanz-Gracia, F., and Vera-Cabello, M. (2013). The impact of the American Civil War on city growth. Mimeo.

Sharma, S. (2003). Persistence and stability in city growth. *Journal of Urban Economics*, 53(2):300–320.

Singh, S. and Maddala, G. (1976). A function for size distribution of incomes. *Econometrica*, 44(5):963–970.

Soo, K. (2005). Zipf's Law for cities: A cross-country investigation. *Regional Science and Urban Economics*, 35(3):239–263.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, 35:40–59.

Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge, Massachusetts: Addison-Wesley Press.

Table 1: Descriptive statistics of the US data samples used

| Sample | Obs. | % of US pop. | Mean | SD | Min. | Max. |
|---|---|---|---|---|---|---|
| Inc. Places 1900 | 10,596 | 46.99 | 3,376 | 42,324 | 7 | 3,437,202 |
| Inc. Places 1910 | 14,135 | 54.90 | 3,561 | 49,351 | 4 | 4,766,883 |
| Inc. Places 1920 | 15,481 | 58.62 | 4,015 | 56,782 | 3 | 5,620,048 |
| Inc. Places 1930 | 16,475 | 62.69 | 4,642 | 67,854 | 1 | 6,930,446 |
| Inc. Places 1940 | 16,729 | 63.75 | 4,976 | 71,299 | 1 | 7,454,995 |
| Inc. Places 1950 | 17,113 | 63.48 | 5,613 | 76,064 | 1 | 7,891,957 |
| Inc. Places 1960 | 18,051 | 64.51 | 6,409 | 74,738 | 1 | 7,781,984 |
| Inc. Places 1970 | 18,488 | 64.51 | 7,094 | 75,320 | 3 | 7,894,862 |
| Inc. Places 1980 | 18,923 | 61.78 | 7,396 | 69,170 | 2 | 7,071,639 |
| Inc. Places 1990 | 19,120 | 61.33 | 7,978 | 71,874 | 2 | 7,322,564 |
| Inc. Places 2000 | 19,296 | 61.49 | 8,968 | 78,015 | 1 | 8,008,278 |
| All places 2000 | 25,358 | 73.98 | 8,232 | 68,390 | 1 | 8,008,278 |
| All places 2010 | 29,461 | 74.31 | 7,826 | 65,494 | 1 | 8,175,133 |
| CCA 1991 (2000m) | 30,201 | 97.46 | 8,180 | 104,954 | 1 | 12,511,237 |
| CCA 1991 (3000m) | 23,499 | 97.46 | 10,513 | 147,360 | 1 | 15,191,634 |
| CCA 1991 (4000m) | 19,912 | 97.46 | 12,407 | 180,751 | 2 | 17,064,816 |
| CCA 1991 (5000m) | 17,569 | 97.46 | 14,062 | 212,084 | 2 | 19,439,862 |
| CCA 2000 (2000m) | 30,201 | 96.08 | 8,977 | 108,342 | 1 | 12,734,150 |
| CCA 2000 (3000m) | 23,499 | 96.08 | 11,537 | 154,157 | 1 | 15,594,627 |
| CCA 2000 (4000m) | 19,912 | 96.08 | 13,615 | 190,528 | 1 | 17,567,010 |
| CCA 2000 (5000m) | 17,569 | 96.08 | 15,431 | 223,825 | 1 | 19,952,762 |

Table 2: Estimators and 95% confidence intervals of the parameters of the dPln and tdPSM for the US (places) samples. The estimators for the lognormal are the mean and the standard deviation of the logarithm of population data

| Sample | lgn | | dPln | | | |
|---|---|---|---|---|---|---|
| | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ | $\mu$ | $\sigma$ |
| Inc. Places 1900 | 6.65 | 1.26 | 0.92±0.03 | 2.64±0.27 | 5.95±0.04 | 0.58±0.04 |
| Inc. Places 1910 | 6.65 | 1.29 | 0.89±0.03 | 2.96±0.35 | 5.86±0.04 | 0.61±0.04 |
| Inc. Places 1920 | 6.67 | 1.32 | 0.87±0.03 | 2.78±0.27 | 5.88±0.04 | 0.60±0.04 |
| Inc. Places 1930 | 6.69 | 1.40 | 0.80±0.02 | 2.21±0.14 | 5.89±0.04 | 0.57±0.04 |
| Inc. Places 1940 | 6.78 | 1.43 | 0.79±0.02 | 2.20±0.15 | 5.96±0.04 | 0.61±0.04 |
| Inc. Places 1950 | 6.84 | 1.50 | 0.80±0.03 | 2.15±0.17 | 6.06±0.05 | 0.78±0.04 |
| Inc. Places 1960 | 6.92 | 1.61 | 0.80±0.03 | 2.24±0.26 | 6.11±0.06 | 0.96±0.05 |
| Inc. Places 1970 | 7.00 | 1.67 | 0.83±0.03 | 2.62±0.22 | 6.18±0.05 | 1.13±0.04 |
| Inc. Places 1980 | 7.11 | 1.66 | 0.86±0.02 | 3.65±0.02 | 6.23±0.02 | 1.19±0.01 |
| Inc. Places 1990 | 7.10 | 1.74 | 0.87±0.02 | 3.59±0.01 | 6.23±0.01 | 1.31±0.003 |
| Inc. Places 2000 | 7.18 | 1.78 | 0.87±0.02 | 3.55±0.01 | 6.32±0.02 | 1.36±0.003 |
| All places 2000 | 7.28 | 1.75 | 1.22±0.03 | 3.15±0.005 | 6.78±0.01 | 1.52±0.002 |
| All places 2010 | 7.11 | 1.82 | 1.12±0.02 | 3.14±0.003 | 6.53±0.01 | 1.56±0.002 |

| Sample | tdPSM | | | | | | |
|---|---|---|---|---|---|---|---|
| | $\rho$ | $\epsilon$ | $\mu$ | $\sigma$ | $\alpha$ | $\tau$ | $\zeta$ |
| Inc. Places 1900 | 2.32±0.14 | 172±1 | 5.64±0.09 | 0.42±0.06 | 0.32±0.07 | 3,405±97 | 1.02±0.05 |
| Inc. Places 1910 | 2.48±0.15 | 147±1 | 5.62±0.06 | 0.44±0.04 | 0.34±0.05 | 8,190±308 | 1.09±0.06 |
| Inc. Places 1920 | 2.36±0.12 | 167±1 | 5.60±0.08 | 0.45±0.05 | 0.33±0.06 | 4,310±127 | 0.98±0.04 |
| Inc. Places 1930 | 2.06±0.09 | 178±1 | 5.52±0.06 | 0.45±0.05 | 0.31±0.05 | 8,465±222 | 1.00±0.05 |
| Inc. Places 1940 | 2.01±0.09 | 177±1 | 5.53±0.06 | 0.44±0.05 | 0.28±0.05 | 10,359±229 | 1.06±0.05 |
| Inc. Places 1950 | 1.89±0.09 | 150±1 | 5.62±0.08 | 0.54±0.06 | 0.34±0.05 | 11,741±382 | 1.06±0.05 |
| Inc. Places 1960 | 1.72±0.07 | 148±1 | 5.55±0.09 | 0.61±0.07 | 0.32±0.06 | 13,917±405 | 1.07±0.05 |
| Inc. Places 1970 | 1.60±0.07 | 141±1 | 5.71±0.10 | 0.69±0.07 | 0.38±0.06 | 25,937±682 | 1.18±0.07 |
| Inc. Places 1980 | 1.69±0.08 | 129±1 | 5.84±0.10 | 0.69±0.06 | 0.38±0.05 | 34,196±571 | 1.30±0.08 |
| Inc. Places 1990 | 1.51±0.06 | 140±1 | 5.91±0.14 | 0.85±0.08 | 0.48±0.08 | 41,945±1,003 | 1.31±0.08 |
| Inc. Places 2000 | 1.60±0.08 | 99±1 | 5.88±0.11 | 0.79±0.06 | 0.40±0.05 | 47,386±851 | 1.35±0.08 |
| All places 2000 | 1.46±0.06 | 127±1 | 6.80±0.24 | 1.14±0.08 | 0.82±0.14 | 36,081±746 | 1.33±0.07 |
| All places 2010 | 1.32±0.04 | 134±1 | 6.77±0.28 | 1.29±0.11 | 0.98±0.18 | 54,144±1,336 | 1.43±0.09 |

Table 3: Estimators and 95% confidence intervals of the parameters of the dPln and dm PChP for the US CCA clusters samples. The estimators for the lognormal are the mean and the standard deviation of the logarithm of population data

| Sample | lgn | | dPln | | | |
|---|---|---|---|---|---|---|
| | $\mu$ | $\sigma$ | $\alpha$ | $\beta$ | $\mu$ | $\sigma$ |
| CCA 1991 (2000m) | 8.33 | 0.85 | 1.95±0.04 | 1.85±0.03 | 8.36±0.01 | 0.14±0.02 |
| CCA 1991 (3000m) | 8.32 | 0.89 | 1.76±0.04 | 1.86±0.04 | 8.29±0.01 | 0.11±0.02 |
| CCA 1991 (4000m) | 8.32 | 0.92 | 1.64±0.03 | 1.88±0.04 | 8.25±0.01 | 0.10±0.02 |
| CCA 1991 (5000m) | 8.33 | 0.95 | 1.54±0.03 | 1.87±0.05 | 8.22±0.01 | 0.10±0.03 |
| CCA 2000 (2000m) | 8.44 | 0.87 | 1.86±0.04 | 1.82±0.03 | 8.45±0.01 | 0.18±0.02 |
| CCA 2000 (3000m) | 8.43 | 0.91 | 1.66±0.03 | 1.83±0.04 | 8.37±0.01 | 0.16±0.02 |
| CCA 2000 (4000m) | 8.42 | 0.94 | 1.55±0.03 | 1.84±0.05 | 8.32±0.02 | 0.15±0.03 |
| CCA 2000 (5000m) | 8.42 | 0.97 | 1.46±0.03 | 1.83±0.05 | 8.29±0.02 | 0.14±0.03 |

| Sample | dm PChP | | | | |
|---|---|---|---|---|---|
| | $\rho$ | $\epsilon$ | $\nu$ | $\mu$ | $\sigma$ |
| CCA 1991 (2000m) | 0.59±0.07 | 2,091±136 | 0.22±0.04 | 8.35±0.01 | 0.37±0.02 |
| CCA 1991 (3000m) | 0.63±0.09 | 2,134±161 | 0.19±0.05 | 8.31±0.01 | 0.37±0.02 |
| CCA 1991 (4000m) | 0.63±0.11 | 1,963±173 | 0.18±0.06 | 8.29±0.01 | 0.39±0.03 |
| CCA 1991 (5000m) | 0.57±0.12 | 2,671±314 | 0.09±0.03 | 8.27±0.01 | 0.42±0.03 |
| CCA 2000 (2000m) | 0.54±0.07 | 1,371±114 | 0.36±0.07 | 8.44±0.01 | 0.39±0.02 |
| CCA 2000 (3000m) | 0.56±0.09 | 1,323±134 | 0.32±0.08 | 8.40±0.01 | 0.40±0.02 |
| CCA 2000 (4000m) | 0.57±0.11 | 1,118±140 | 0.33±0.09 | 8.38±0.01 | 0.42±0.02 |
| CCA 2000 (5000m) | 0.58±0.12 | 1,279±166 | 0.26±0.09 | 8.35±0.01 | 0.42±0.03 |

| Sample | dm PChP | | | |
|---|---|---|---|---|
| | $\beta$ | $\tau$ | $\zeta$ | $\theta$ |
| CCA 1991 (2000m) | 1.29±0.22 | 17,171±898 | 0.96±0.11 | 0.78±0.10 |
| CCA 1991 (3000m) | 1.31±0.24 | 16,903±853 | 0.87±0.08 | 0.90±0.08 |
| CCA 1991 (4000m) | 1.45±0.24 | 16,495±864 | 0.83±0.08 | 0.92±0.08 |
| CCA 1991 (5000m) | 1.62±0.21 | 15,773±852 | 0.83±0.08 | 0.92±0.09 |
| CCA 2000 (2000m) | 1.13±0.24 | 20,381±1,231 | 0.95±0.12 | 0.69±0.11 |
| CCA 2000 (3000m) | 1.21±0.25 | 19,912±1,122 | 0.87±0.09 | 0.84±0.10 |
| CCA 2000 (4000m) | 1.36±0.24 | 20,083±1,173 | 0.84±0.09 | 0.89±0.10 |
| CCA 2000 (5000m) | 1.26±0.30 | 14,253±797 | 0.71±0.08 | 0.71±0.08 |

Table 4: $p$-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for the US places samples and the density functions used. Non rejections at the 5% significance level are in bold

| Sample | lgn | | dPln | |
|---|---|---|---|---|
| | KS | CM | KS | CM |
| Inc. Places 1900 | 0 (0.07) | 0 (17.22) | 0.03 (0.01) | **0.07 (0.42)** |
| Inc. Places 1910 | 0 (0.07) | 0 (21.81) | 0.001 (0.02) | 0.02 (0.66) |
| Inc. Places 1920 | 0 (0.07) | 0 (25.87) | 0.02 (0.013) | **0.09 (0.37)** |
| Inc. Places 1930 | 0 (0.07) | 0 (27.59) | 0 (0.017) | 0 (1.19) |
| Inc. Places 1940 | 0 (0.07) | 0 (25.59) | 0 (0.021) | 0 (1.60) |
| Inc. Places 1950 | 0 (0.06) | 0 (17.55) | 0 (0.021) | 0 (1.64) |
| Inc. Places 1960 | 0 (0.05) | 0 (14.26) | 0 (0.024) | 0 (2.02) |
| Inc. Places 1970 | 0 (0.05) | 0 (12.88) | 0 (0.021) | 0 (1.75) |
| Inc. Places 1980 | 0 (0.04) | 0 (11.36) | 0 (0.021) | 0 (1.99) |
| Inc. Places 1990 | 0 (0.04) | 0 (9.10) | 0 (0.021) | 0 (2.03) |
| Inc. Places 2000 | 0 (0.04) | 0 (9.35) | 0 (0.020) | 0 (2.28) |
| All places 2000 | 0 (0.02) | 0 (2.69) | 0 (0.01) | 0 (1.31) |
| All places 2010 | 0 (0.03) | 0 (4.86) | 0 (0.02) | 0 (2.01) |

| Sample | tdPSM | |
|---|---|---|
| | KS | CM |
| Inc. Places 1900 | **0.99 (0.005)** | **0.97 (0.03)** |
| Inc. Places 1910 | **0.62 (0.007)** | **0.84 (0.06)** |
| Inc. Places 1920 | **0.50 (0.007)** | **0.65 (0.09)** |
| Inc. Places 1930 | **0.96 (0.004)** | **0.97 (0.03)** |
| Inc. Places 1940 | **0.90 (0.005)** | **0.96 (0.03)** |
| Inc. Places 1950 | **0.87 (0.005)** | **0.78 (0.06)** |
| Inc. Places 1960 | **0.93 (0.004)** | **0.85 (0.05)** |
| Inc. Places 1970 | **0.94 (0.004)** | **0.96 (0.03)** |
| Inc. Places 1980 | **0.54 (0.006)** | **0.48 (0.12)** |
| Inc. Places 1990 | **0.71 (0.006)** | **0.75 (0.07)** |
| Inc. Places 2000 | **0.88 (0.005)** | **0.90 (0.05)** |
| All places 2000 | **0.65 (0.005)** | **0.47 (0.13)** |
| All places 2010 | **0.37 (0.006)** | **0.41 (0.14)** |

Table 5: $p$-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for the US CCA clusters samples and the density functions used. Non rejections at the 5% significance level are in bold

| Sample | lgn | | dPln | | dm PChP | |
|---|---|---|---|---|---|---|
| | KS | CM | KS | CM | KS | CM |
| CCA 1991 (2000m) | 0 (0.09) | 0 (92.70) | 0 (0.02) | 0 (1.84) | **0.86 (0.004)** | **0.82 (0.06)** |
| CCA 1991 (3000m) | 0 (0.10) | 0 (86.75) | 0 (0.02) | 0 (2.42) | **0.64 (0.005)** | **0.74 (0.07)** |
| CCA 1991 (4000m) | 0 (0.11) | 0 (78.08) | 0 (0.03) | 0 (2.46) | **0.86 (0.005)** | **0.69 (0.08)** |
| CCA 1991 (5000m) | 0 (0.11) | 0 (74.02) | 0 (0.03) | 0 (2.21) | **0.61 (0.006)** | **0.60 (0.10)** |
| CCA 2000 (2000m) | 0 (0.09) | 0 (73.26) | 0 (0.02) | 0.003 (1.09) | **0.58 (0.005)** | **0.73 (0.07)** |
| CCA 2000 (3000m) | 0 (0.09) | 0 (71.00) | 0 (0.02) | 0 (1.18) | **0.55 (0.006)** | **0.43 (0.14)** |
| CCA 2000 (4000m) | 0 (0.09) | 0 (62.27) | 0 (0.04) | 0 (1.79) | **0.36 (0.007)** | **0.28 (0.19)** |
| CCA 2000 (5000m) | 0 (0.10) | 0 (58.44) | 0 (0.05) | 0 (2.22) | **0.46 (0.007)** | **0.51 (0.12)** |

Table 6: Maximum log-likelihoods, AIC and BIC for the distributions used and the US places data. The lowest values of AIC and BIC for each sample are in bold

| Sample | lgn log-likelihood | AIC | BIC | dPln log-likelihood | AIC | BIC |
|---|---|---|---|---|---|---|
| Inc. Places 1900 | -87,943 | 175,891 | 175,905 | -87,254 | 174,516 | 174,545 |
| Inc. Places 1910 | -117,640 | 235,284 | 235,299 | -116,727 | 233,462 | 233,492 |
| Inc. Places 1920 | -129,580 | 259,164 | 259,179 | -128,521 | 257,050 | 257,081 |
| Inc. Places 1930 | -139,194 | 278,392 | 278,407 | -138,129 | 276,266 | 276,297 |
| Inc. Places 1940 | -143,097 | 286,198 | 286,213 | -142,179 | 284,366 | 284,397 |
| Inc. Places 1950 | -148,254 | 296,512 | 296,528 | -147,593 | 295,194 | 295,225 |
| Inc. Places 1960 | -159,142 | 318,288 | 318,304 | -158,679 | 317,366 | 317,397 |
| Inc. Places 1970 | -165,171 | 330,346 | 330,362 | -164,831 | 329,670 | 329,701 |
| Inc. Places 1980 | -171,088 | 342,180 | 342,196 | -170,777 | 341,562 | 341,593 |
| Inc. Places 1990 | -173,472 | 346,948 | 346,964 | -173,243 | 346,494 | 346,525 |
| Inc. Places 2000 | -177,127 | 354,258 | 354,274 | -176,931 | 353,870 | 353,901 |
| All places 2000 | -234,773 | 469,550 | 469,566 | -234,710 | 469,428 | 469,461 |
| All places 2010 | -268,748 | 537,499 | 537,516 | -268,657 | 537,323 | 537,356 |

| Sample | tdPSM log-likelihood | AIC | BIC |
|---|---|---|---|
| Inc. Places 1900 | -87,232 | **174,478** | **174,529** |
| Inc. Places 1910 | -116,690 | **233,393** | **233,446** |
| Inc. Places 1920 | -128,485 | **256,983** | **257,037** |
| Inc. Places 1930 | -138,060 | **276,134** | **276,188** |
| Inc. Places 1940 | -142,074 | **284,162** | **284,216** |
| Inc. Places 1950 | -147,486 | **294,986** | **295,040** |
| Inc. Places 1960 | -158,530 | **317,073** | **317,128** |
| Inc. Places 1970 | -164,680 | **329,375** | **329,430** |
| Inc. Places 1980 | -170,625 | **341,265** | **341,320** |
| Inc. Places 1990 | -173,106 | **346,226** | **346,281** |
| Inc. Places 2000 | -176,775 | **353,563** | **353,618** |
| All places 2000 | -234,633 | **469,280** | **469,337** |
| All places 2010 | -268,524 | **537,062** | **537,120** |

Table 7: Maximum log-likelihoods, AIC and BIC for the distributions used and the US CCA clusters data. The lowest values of AIC and BIC for each sample are in bold

| Sample | lgn log-likelihood | AIC | BIC | dPln log-likelihood | AIC | BIC |
|---|---|---|---|---|---|---|
| CCA 1991 (2000m) | -289,460 | 578,923 | 578,940 | -284,288 | 568,584 | 568,617 |
| CCA 1991 (3000m) | -226,140 | 452,284 | 452,300 | -221,851 | 443,710 | 443,742 |
| CCA 1991 (4000m) | -192,249 | 384,502 | 384,518 | -188,584 | 377,177 | 377,209 |
| CCA 1991 (5000m) | -170,343 | 340,690 | 340,706 | -167,096 | 334,201 | 334,232 |
| CCA 2000 (2000m) | -293,311 | 586,627 | 586,643 | -288,879 | 577,765 | 577,798 |
| CCA 2000 (3000m) | -229,171 | 458,347 | 458,363 | -225,494 | 450,996 | 451,028 |
| CCA 2000 (4000m) | -194,701 | 389,406 | 389,422 | -191,552 | 383,112 | 383,143 |
| CCA 2000 (5000m) | -172,389 | 344,783 | 344,798 | -169,586 | 339,179 | 339,211 |

| Sample | dm PChP log-likelihood | AIC | BIC |
|---|---|---|---|
| CCA 1991 (2000m) | -283,584 | **567,186** | **567,261** |
| CCA 1991 (3000m) | -221,218 | **442,454** | **442,526** |
| CCA 1991 (4000m) | -188,065 | **376,148** | **376,219** |
| CCA 1991 (5000m) | -166,669 | **333,356** | **333,426** |
| CCA 2000 (2000m) | -288,309 | **576,635** | **576,710** |
| CCA 2000 (3000m) | -225,020 | **450,057** | **450,130** |
| CCA 2000 (4000m) | -191,176 | **382,370** | **382,441** |
| CCA 2000 (5000m) | -169,277 | **338,572** | **338,642** |

Table 8: Percentages of population and urban units (places, clusters) in the tails and the body of the tdPSM for places and the dm PChP for clusters. For the definition of tails and body we use, in each case, the corresponding thresholds $\epsilon$ and $\tau$ of Table 2 for places and Table 3 for clusters

| | Population | | | Units | | |
|---|---|---|---|---|---|---|
| | Lower tail | Body | Upper tail | Lower tail | Body | Upper tail |
| Inc. Places 1900 | 0.3% | 20.8% | 78.9% | 7.4% | 81% | 11.6% |
| Inc. Places 1910 | 0.2% | 29.5% | 70.3% | 5.7% | 89% | 5.3% |
| Inc. Places 1920 | 0.2% | 19.6% | 80.2% | 7.5% | 82.2% | 10.3% |
| Inc. Places 1930 | 0.3% | 23.5% | 76.2% | 9.9% | 83.7% | 6.4% |
| Inc. Places 1940 | 0.2% | 25.6% | 74.2% | 9.2% | 84.8% | 6% |
| Inc. Places 1950 | 0.1% | 25.4% | 74.5% | 7.7% | 86.1% | 6.2% |
| Inc. Places 1960 | 0.1% | 26% | 73.9% | 8.5% | 84.9% | 6.6% |
| Inc. Places 1970 | 0.1% | 33.8% | 66.1% | 8.2% | 87.5% | 4.3% |
| Inc. Places 1980 | 0.1% | 39.5% | 60.4% | 6.2% | 90.2% | 3.6% |
| Inc. Places 1990 | 0.1% | 41.2% | 58.7% | 8.6% | 88.2% | 3.2% |
| Inc. Places 2000 | 0% | 41.4% | 58.6% | 5.2% | 91.5% | 3.3% |
| All places 2000 | 0.1% | 42.9% | 57% | 7.1% | 89% | 3.9% |
| All places 2010 | 0.1% | 49.3% | 50.6% | 10.1% | 87.5% | 2.4% |
| CCA 1991 (2000m) | 2% | 53.2% | 44.8% | 12.3% | 84.5% | 3.2% |
| CCA 1991 (3000m) | 1.8% | 39.3% | 58.9% | 13.9% | 82.2% | 3.9% |
| CCA 1991 (4000m) | 1.3% | 32.7% | 66% | 12.3% | 83.2% | 4.5% |
| CCA 1991 (5000m) | 3.1% | 26.2% | 70.7% | 24.6% | 70% | 5.4% |
| CCA 2000 (2000m) | 0.4% | 56.7% | 42.9% | 4.7% | 92.2% | 3.1% |
| CCA 2000 (3000m) | 0.3% | 42.2% | 57.5% | 4.8% | 91.3% | 3.9% |
| CCA 2000 (4000m) | 0.2% | 35.1% | 64.7% | 3.7% | 92% | 4.3% |
| CCA 2000 (5000m) | 0.3% | 27.7% | 72% | 5% | 87.6% | 7.4% |

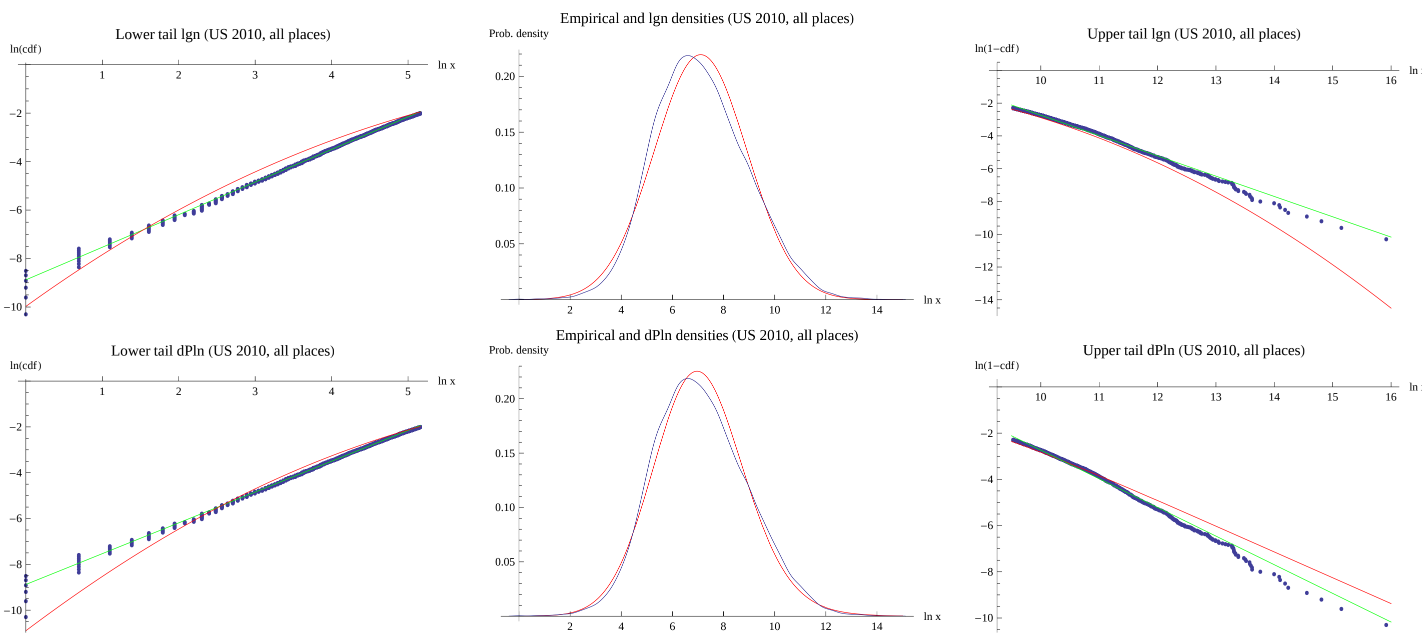


Figure 1: Left-hand column: Empirical, estimated and OLS linear fit lgn and dPln $\ln(\mathrm{cdf})$ for the lower tail. Center column: Empirical (Gaussian adaptive kernel density) and estimated lgn and dPln density functions. Right-hand column: Empirical, estimated and OLS linear fit lgn and dPln $\ln(1 - \mathrm{cdf})$ for the upper tail. The online version of this figure is in color.

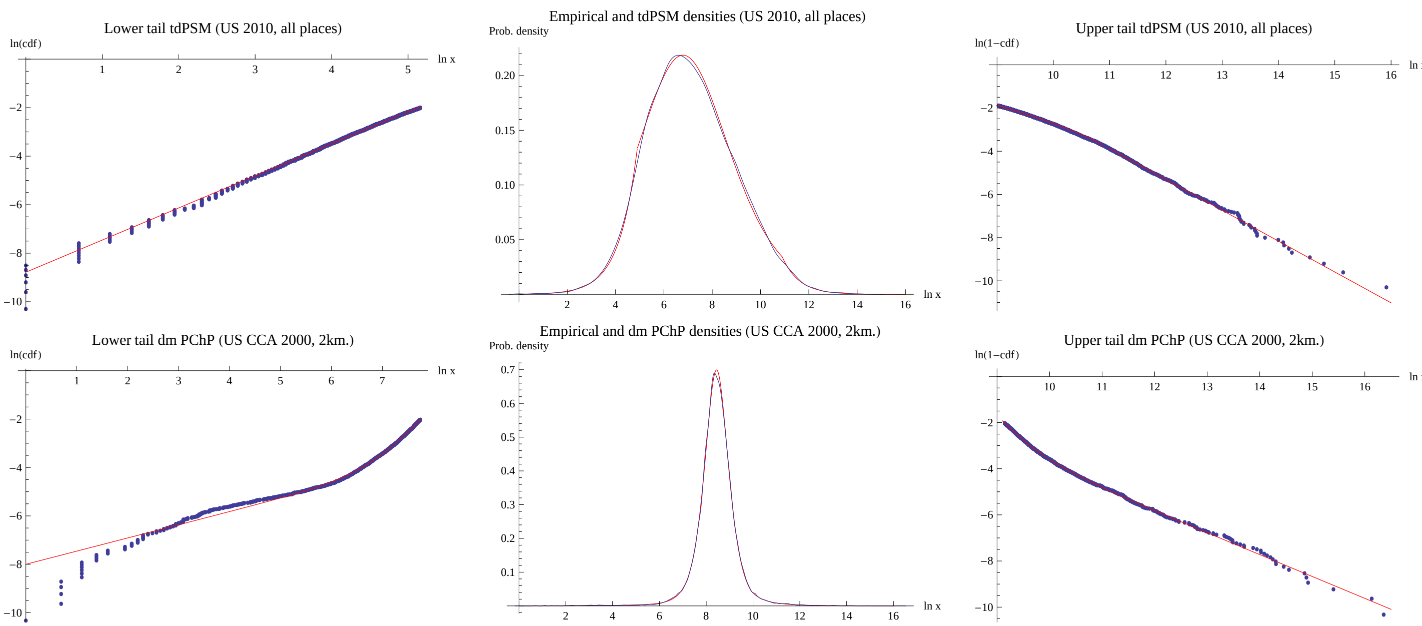


Figure 2: Left-hand column: Empirical and estimated tdPSM and dm PChP $\ln(\mathrm{cdf})$ for the lower tail. Center column: Empirical (Gaussian adaptive kernel density) and estimated tdPSM and dm PChP density functions. Right-hand column: Empirical and estimated tdPSM and dm PChP $\ln(1-\mathrm{cdf})$ for the upper tail. First row: US all places (2010). Second row: US CCA clusters (2000, 2km.). The online version of this figure is in color.