Available online at www.sciencedirect.com

ScienceDirect

Physica A 00 (2022) 1–??

Physica A

Comment on “A statistical evidence of power law distribution in the upper tail of world billionaires’ data 2010–20” [Physica A 581 (2021) 126198]

Arturo Ramos[a,*]

[a]Departmento de Análisis Económico, Universidad de Zaragoza, Zaragoza, Spain

Abstract

In a recently published paper in Physica A it is claimed the statistical evidence of having a power law distribution in the upper tail of the world billionaires’ data in the period 2010-20. We address (almost) the same data sets, with the same truncation points in the data, and obtain two main conclusions. Firstly, both the truncated lognormal and the power law or Pareto distributions might be improved in principle to describe the upper tails’ data to the level of quality of recent studies for other types of data, as the Anderson–Darling test yields rejection of both models almost always. Secondly, the truncated lognormal is statistically comparable or even better to the Pareto by different criteria, formal and graphical.

Keywords: truncated lognormal distribution, power law, Pareto distribution, Anderson-Darling test, world’s billionaires

1. Introduction

In the recently published paper [1] it is claimed that there is statistical evidence about that the upper tails of the data sets of world billionaires is a power law. The topic of discerning whether there is a power law for some kind of data (firm sizes, city sizes, size of earthquakes and many other types of data) is a recurring theme in the literature, let us cite a few: [28, 4, 19, 5, 15, 3, 6]. As an example, in the recent paper [12] it is found that the upper tail of the distribution of strike sizes can be described almost equally by a truncated lognormal or by a power law or Pareto distribution. The authors in [1] seem to not have noticed that when fitting the truncated upper tail to something related with the lognormal distribution, one should use a *truncated* lognormal [14] and therefore should not use the usual maximum likelihood estimators of the lognormal distribution with support $(0, \infty)$, which as it is well-known are the mean and the standard deviation of the natural logarithm of the data.

Thus our purpose in this Comment is just to compare the power law or Pareto distribution to the truncated lognormal distribution for (most of) the truncated data sets used in [1], and we obtain mainly two conclusions that are as follows:

- Both models, power law or Pareto distribution and truncated lognormal distribution, might be outperformed by another statistical law so as to achieve that the standard Anderson–Darling test yields non-rejections. Let us recall that when assessing the adequacy at the tails, one should rely on such test [13].

*Corresponding author
E-mail address: aramos@unizar.es

- The truncated lognormal distribution is statistically comparable or even better to the power law or Pareto distribution for the cited datasets.

The rest of the Comment is organized as follows. The next Section describes briefly the distributions used. Section 3 presents the data. Section 4 is a brief account of our results. The last Section concludes.

2. The distributions

Let us denote the size of the wealth of the top billionaires in the world by the variable x, measured in US billions of US dollars. When dealing with wealth variable, it is not necessarily positive, as many people are heavily indebted and therefore their wealth is negative. In any case we are considering here the very top of the wealth distribution data in the world, so it would be necessary to consider the conditional distribution for the right tail as the full wealth distribution is not known, but only the very extreme part of the right tail is available.

The first alternative we take for the distribution of top billionaires, advocated in [1], is the power law or Pareto distribution (in [1] there is a slight missprint: their $\alpha - 1$ in the numerator should be only α, so it should be in their paper for example $\bar{f}_P(x;\alpha,x_{\min}) = \frac{\alpha}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-(\alpha+1)}$, but it is important: such a change affects the form of the maximum likelihood (ML) estimator of the Pareto distribution by a unit of difference), with probability density function (PDF) given by (see, e.g., [14, 26, 27, 7, 11, 12]):

$$f_P(x;\alpha,x_{\min}) = \frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha} \tag{1}$$

where $\alpha > 1$ is the power law exponent and $x_{\min}$ is the lower bound on power law behavior, that is, $0 < x_{\min} \leq x$. The corresponding cumulative distribution function (CDF) is simply

$$\mathrm{cdf}_P(x;\alpha,x_{\min}) = 1 - \left(\frac{x}{x_{\min}}\right)^{1-\alpha}, \quad 0 < x_{\min} \leq x$$

However, one issue that has been highlighted in the research literature is that it can be difficult to distinguish a power law with the upper tail of other heavy tailed distributions [14, 23, 4, 7, 3, 5, 6, 11, 25, 33]. This means that there could be other distributions that are a plausible fit to the upper tail of the data. An implication of this observation is that an empirical analysis should also consider the fit of at least one alternative distribution to the upper tail [14, 7]. The authors of [1] use the lognormal and the exponential distribution.

As in [14, Table I], for the lognormal one should use a truncated lognormal for the upper tail. The other alternative, the exponential distribution, is not appropriate if one expects that the data follows a power law. It is the natural logarithm of the data which follows an exponential (see, e.g., [12]), and comparing the power law *and* the exponential distribution *for the same data* is a very hard endeavour (see, e.g. [17] for some effects of taking logarithms of the data on the underlying distribution).

So let us introduce the well-known lognormal PDF

$$f_{LN}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right), \quad x \in (0,\infty) \tag{2}$$

where $\mu = E[\ln(x)]$ is the mean of $\ln(x)$ and $\sigma > 0$ is its standard deviation according to this distribution. Lognormal distributions have been found to fit many economic and physical phenomena with heavy tails [24].

As mentioned, when dealing only with the upper tail using the lognormal distribution, one should censor it in the usual way. Let us recall the general procedure [20]. Suppose we have a random variable X that is distributed according to some PDF $f(x)$, with CDF $\mathrm{cdf}(x)$, both of which have support equal to $(-\infty,\infty)$. Suppose we wish to know the PDF of the random variable after restricting the support to $[a,b)$ with $0 < a < b \leq \infty$ (because, as said, only the very extreme part of the right tail is available). Then

$$f(x|a \leq X < b) = \frac{g(x)}{\mathrm{cdf}(b) - \mathrm{cdf}(a)} \tag{3}$$

where $g(x) = f(x)I(\{a \le x < b\})$ and I is the indicator function.

Now, returning to the case of the lognormal distribution, $a = x_{\min}$, $b = \infty$, $\mathrm{cdf}(b) = \mathrm{cdf}_{\mathrm{LN}}(\infty) = 1$ and the lower truncated lognormal distribution with minimum value $x_{\min}$ (see, e.g., [7, 11]) has PDF

$$f_{\mathrm{LNt}}(x;\mu,\sigma,x_{\min}) = \frac{f_{\mathrm{LN}}(x;\mu,\sigma)}{1-\mathrm{cdf}_{\mathrm{LN}}(x_{\min};\mu,\sigma)} \tag{4}$$

where $x \in [x_{\min}, \infty)$ and

$$\mathrm{cdf}_{\mathrm{LN}}(x;\mu,\sigma) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\ln(x)-\mu}{\sqrt{2}\sigma}\right)$$

is the CDF of the lognormal distribution with PDF given by (2) (erf is the error function associated to the standard normal distribution). The truncated lognormal distribution has been considered as an alternative to a power law in the upper tail of wealth [7, 11] and the size of strikes [12].

If one wants to apply the construction of (3) for the Pareto case with $a = x_{\min}$, $b = \infty$, one would have

$$f(x|x_{\min} \le X < \infty) = \frac{\frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha}}{1-1+\left(\frac{x_{\min}}{x_{\min}}\right)^{1-\alpha}} = \frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha}$$

which is again the expression of the right hand side (RHS) of (1). This is to be expected, as this distribution has support $[x_{\min}, \infty)$ by construction.

Let us remark that for the lower truncated lognormal (4) the parameter μ is no longer the expected value of $\ln(x)$, $E[\ln(x)]$. In fact, we have

$$E[\ln(x)] = \int_{x_{\min}}^{\infty} \ln(x) f_{\mathrm{LNt}}(x;\mu,\sigma,x_{\min})\,dx = \mu + \frac{\exp\left(-\frac{\mu^2+(\ln(x_{\min}))^2}{2\sigma^2}\right)\sqrt{\frac{2}{\pi}}\sigma x_{\min}^{\mu/\sigma^2}}{1+\mathrm{erf}\left(\frac{\mu-\ln(x_{\min})}{\sqrt{2}\sigma}\right)}$$

and therefore

$$\mu = E[\ln(x)] - \frac{\exp\left(-\frac{\mu^2+(\ln(x_{\min}))^2}{2\sigma^2}\right)\sqrt{\frac{2}{\pi}}\sigma x_{\min}^{\mu/\sigma^2}}{1+\mathrm{erf}\left(\frac{\mu-\ln(x_{\min})}{\sqrt{2}\sigma}\right)} \tag{5}$$

In this expression, if $x_{\min} \to 0$, then $\mu \to E[\ln(x)]$ as expected. Since in all of our samples $\ln(x_{\min}) > 0$, we have that μ in (5) is the difference of two positive quantities and may be negative. This is observed indeed later for the estimated truncated lognormal distribution for the samples in this paper.

3. The data

The data used in this Comment is obtained from the web page https://stats.areppim.com/stats/links_billionairexlists.htm and they are also used in [1]. This web page does not contain the full data of world billionaires for the years 2019 and 2020, and the data for these two years cannot be obtained in a simple or free way, making them to be not available. So we will use the above data for the period 2010-2018 on a yearly basis. This does not invalidate our analysis.

Moreover, for the sake of brevity we will impose the same cut-offs to the obtained data sets as in [1] according to their Table 2. Thus, our data is as described in our Table 1.

4. Results

We have estimated the parameters of the power law or Pareto distribution and truncated lognormal distribution by ML estimation. For the index parameter of the Pareto distribution taking into account our specification (1), we can use the ML estimator closed formula [14]

$$\alpha = 1 + n\left(\sum_{i=1}^{n}\ln\left(\frac{x_i}{x_{\min}}\right)\right)^{-1} \tag{6}$$

Sample	Obs	Mean	SD	Skewness	Kurtosis	Mean (Log scale)	SD (Log scale)	Skewness (Log scale)	Kurtosis (Log scale)	Min	Max
2010	362	6.99585	6.50749	3.58729	20.87101	1.71189	0.61028	1.18659	3.91334	2.8	53.5
2011	563	6.28792	6.74985	4.31885	31.25266	1.55118	0.67441	1.11308	3.80102	2.2	74
2012	600	6.08716	6.50868	4.25306	30.24849	1.52177	0.66753	1.17596	3.88796	2.2	69
2013	611	6.82626	7.16804	4.34934	30.61189	1.64701	0.65795	1.12895	3.85878	2.5	73
2014	686	7.16574	7.82852	4.20985	27.53067	1.68338	0.66403	1.22682	4.13489	2.6	76
2015	511	9.10293	9.25298	3.90594	23.29646	1.95171	0.63150	1.24873	4.22177	3.5	79.2
2016	391	9.63375	9.09206	3.48100	18.42967	2.03326	0.60104	1.30657	4.30857	4	75
2017	716	7.63449	8.78315	4.50319	29.43380	1.74731	0.64720	1.43549	4.90181	2.9	86
2018	605	9.83570	11.30441	4.37670	27.58895	2.00116	0.64470	1.47332	4.98941	3.8	112

Table 1: Descriptive statistics of the truncated upper tail samples of world billionaires according to [1].

where x_i, $i = 1, \ldots, n$, are the data of wealth and n is the sample size of each data set. Note that the authors of [1] use instead

$$\alpha_{[1]} = n\left(\sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{\min}}\right)\right)^{-1} \tag{7}$$

which would be the correct expression had they used as well for example another correct specification like $\bar{f}_{\mathrm{P}}(x; \alpha, x_{\min}) = \frac{\alpha}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-(\alpha+1)}$. Thus, the ML estimator results for $\alpha_{[1]}$ of [1] should be statistically equivalent to those obtained using (6) minus one unit.

For the truncated lognormal distribution, we have used the command `mle` of the software MATLAB®.

Our results are presented in Tables 2 and 3. We observe that the difference of the corresponding (comparable) results for the parameter α obtained by using (6) and those of using (7) as in [1] differ in quantities statistically non significant, using the standard errors (SE) reported by [1] for these estimations.

Sample	α by (6)	$\alpha_{[1]}$ by (7)	$\alpha_{[1]} + 1$ by (7)	$\lvert\Delta\rvert$	SE of $\alpha_{[1]}$ given by [1]
2010	2.46567	1.545	2.545	0.07932	0.082
2011	2.31108	1.311	2.311	8.75726E-05	0.055
2012	2.36366	1.364	2.364	0.00033982	0.056
2013	2.36851	1.369	2.369	0.000489018	0.052
2014	2.37385	1.374	2.374	0.000144011	0.052
2015	2.43072	1.431	2.431	0.000278756	0.063
2016	2.54565	1.546	2.546	0.000343464	0.078
2017	2.46496	1.465	2.465	3.28206E-05	0.055
2018	2.50113	1.501	2.501	0.000135689	0.061

Table 2: ML estimators of the power law or Pareto distribution by (6), by (7) and their comparison. The $x_{\min}$ are those described in Table 1.

Sample	μ	σ
2010	-1.85977	1.67579
2011	-1.91105	1.75919
2012	-3.43608	2.02004
2013	-2.57208	1.87488
2014	-3.41233	2.03699
2015	-2.27454	1.83083
2016	-4.02569	2.06889
2017	-8.02712	2.66278
2018	-14.8595	3.41277

Table 3: ML estimators of the truncated lognormal distribution for our samples. The $x_{\min}$ are those described in Table 1.

In Table 3 they are shown the ML estimations of the parameters μ, σ of the truncated lognormal distribution (4) for the employed data sets. It is observed that the μ's are always negative, something that it is allowed by virtue of (5). In Table 4 this relation is independently verified computing separately the two terms on the RHS of (5) either by sample or density expectations, subtracting and comparing with the corresponding μ.

Sample	μ	RHS of (5), sample	RHS of (5), density
2010	-1.85977	-1.85977	-1.85977
2011	-1.91105	-1.91105	-1.91105
2012	-3.43608	-3.43608	-3.43608
2013	-2.57208	-2.57208	-2.57208
2014	-3.41233	-3.41233	-3.41233
2015	-2.27454	-2.27454	-2.27454
2016	-4.02569	-4.02569	-4.02569
2017	-8.02712	-8.02712	-8.02712
2018	-14.8595	-14.8595	-14.8595

Table 4: Computation of the right hand side (RHS) of (5) using the sample mean of $\log(x_i)$ and using the mean of $\log(x)$ of the density (4) and its comparison with the estimated ML μ for each sample. The $x_{\min}$ are those described in Table 1. It is verified (5) in all cases.

Afterwards, we have performed standard Kolmogorov–Smirnov (KS), Crámer–von Mises (CM) and Anderson–Darling (AD) tests to assess the adequacy of the fits. Let us recall that the AD test is very convenient when assessing the fit at the tails [13]. The results are shown in Table 5. From this table, we observe that the KS and CM test yield non-rejection almost always (the only exception is the 2013 sample for the LNt and by a very small margin), and on the opposite side the AD test yields almost always rejection (the only exceptions are the samples of 2010 and 2016 for the Pareto and the 2010 for the truncated lognormal distribution). This means that the fits at the tails, and we are working at the very upper tail only, might be improved by another different distribution to those of Pareto and truncated lognormal. Let us remark that there exist studies that offer a very good fit in terms of AD test even with a much higher sample size than in the current samples, see, e.g., [29, 30]. So in much more demanding situations it is possible to obtain a better fit at the tails. Let us recall that the Anderson–Darling test has increasing power with sample size [31].

So the first conclusion is that both Pareto and truncated lognormal distribution might be outperformed by another different distribution.

	Pareto			LNt		
Sample	KS	CM	AD	KS	CM	AD
2010	**0.404 (0.047)**	**0.489 (0.122)**	**0.085 (2.06)**	**0.392 (0.047)**	**0.531 (0.112)**	**0.104 (1.900)**
2011	**0.261 (0.042)**	**0.196 (0.244)**	0.000665 (6.350)	**0.127 (0.049)**	**0.311 (0.179)**	0.000819 (6.160)
2012	**0.345 (0.038)**	**0.354 (0.162)**	0.00211 (5.280)	**0.089 (0.051)**	**0.236 (0.217)**	0.00155 (5.570)
2013	**0.201 (0.043)**	**0.193 (0.246)**	0.00116 (5.830)	0.0486 (0.055)	**0.127 (0.310)**	0.000639 (6.390)
2014	**0.375 (0.035)**	**0.315 (0.178)**	0.0029 (4.990)	**0.102 (0.046)**	**0.203 (0.239)**	0.00236 (5.180)
2015	**0.654 (0.032)**	**0.507 (0.117)**	0.030 (2.920)	**0.721 (0.030)**	**0.574 (0.102)**	0.034 (2.810)
2016	**0.763 (0.033)**	**0.597 (0.097)**	**0.051 (2.470)**	**0.302 (0.049)**	**0.403 (0.146)**	0.043 (2.620)
2017	**0.401 (0.033)**	**0.633 (0.090)**	0.00589 (4.350)	**0.401 (0.033)**	**0.47 (0.126)**	0.00488 (4.520)
2018	**0.540 (0.032)**	**0.447 (0.133)**	0.018 (3.320)	**0.344 (0.038)**	**0.291 (0.188)**	0.014 (3.570)

Table 5: Results of the KS, CM and AD tests for the hypothesized distributions and the samples used. The format is p-value (test statistic). Non-rejections at the 5% level are marked in bold.

In order to select amongst the two models we use three well-known information criteria. They are:

- The Akaike Information Criterion (AIC) [2, 8, 9], defined as

$$AIC = 2k - 2\ln L^*$$

where k is the number of parameters of the distribution and $\ln L^*$ is the corresponding (maximum) log-likelihood. The minimum value of AIC corresponds (asymptotically) to the minimum value of the Kullback–Leibler divergence, so a model with the lowest AIC is selected from among the competitors.

- The Bayesian or Schwarz Information Criterion (BIC) [8, 9, 32], defined as

$$BIC = k\ln(n) - 2\ln L^*$$

where k is the number of parameters of the distribution, n the sample size and $\ln L^*$ is as before. The BIC penalizes more heavily the number of parameters used than does the AIC. The model with the lowest BIC is selected according to this criterion.

- The Hannan–Quinn Information Criterion (HQC) [8, 9, 18], defined as

$$HQC = 2k\ln(\ln(n)) - 2\ln L^*$$

where k is the number of parameters of the distribution, n the sample size and $\ln L^*$ is as before. The HQC implements an intermediate penalization of the number of parameters when compared to the AIC and BIC. The model with the lowest HQC is selected according to this criterion.

The results in our case are displayed in Table 6. In it, it can be seen that the LNt is selected most of the time by both AIC and HQC, and only the BIC yields selection of the Pareto more often. The overall picture is that the truncated lognormal distribution performs not worse than the Pareto one.

Sample	Pareto log-likelihood	AIC	BIC	HQC	LNt log-likelihood	AIC	BIC	HQC	Min AIC	Sel AIC	Min BIC	Sel BIC	Min HQC	Sel HQC
2010	-849.54621	1701.09242	1704.98407	1702.63949	-846.958	1697.91599	1705.69928	1701.01013	1697.91599	LNt	1704.98407	Pareto	1701.01013	LNt
2011	-1300.459	2602.91805	2607.25133	2604.60968	-1295.5675	2595.13502	2603.80158	2598.51829	2595.13502	LNt	2603.80158	LNt	2598.51829	LNt
2012	-1342.9057	2687.81136	2692.20829	2689.523	-1339.9239	2683.84785	2692.64171	2687.27112	2683.84785	LNt	2692.20829	Pareto	2687.27112	LNt
2013	-1441.5818	2885.16352	2889.57862	2886.88083	-1437.7734	2879.54683	2888.37702	2882.98144	2879.54683	LNt	2888.37702	LNt	2882.98144	LNt
2014	-1639.5527	3281.1053	3285.63618	3282.85839	-1636.3129	3276.62585	3285.6876	3280.13202	3276.62585	LNt	3285.63618	Pareto	3280.13202	LNt
2015	-1335.6923	2673.38459	2677.62096	2675.04538	-1332.7093	2669.41864	2677.89137	2672.74023	2669.41864	LNt	2677.62096	Pareto	2672.74023	LNt
2016	-1024.0653	2050.13066	2054.09937	2051.70373	-1022.8734	2049.74685	2057.68427	2052.89297	2049.74685	LNt	2054.09937	Pareto	2051.70373	Pareto
2017	-1710.3243	3422.64852	3427.2222	3424.41466	-1709.215	3422.43001	3431.57737	3425.96231	3422.43001	LNt	3427.2222	Pareto	3424.41466	Pareto
2018	-1582.4162	3166.83241	3171.23764	3168.54664	-1582.061	3168.12203	3176.93248	3171.55048	3166.83241	Pareto	3171.23764	Pareto	3168.54664	Pareto

Table 6: Maximum log-likelihoods, AIC, BIC, HQC information criteria and the selected model according to each of them for each sample.

Sample	BIC LNt	BIC Pareto	Bayes factor	Conclusion
2010	1705.69928	1704.98407	1.42990397	Evidence in favor of Pareto
2011	2603.80158	2607.25133	0.17819539	Moderate evidence in favor of LNt
2012	2692.64171	2692.20829	1.24198121	Evidence in favor of Pareto
2013	2888.37702	2889.57862	0.54837335	Weak evidence in favor of LNt
2014	3285.6876	3285.63618	1.02604519	Evidence in favor of Pareto
2015	2677.89137	2677.62096	1.14477589	Evidence in favor of Pareto
2016	2057.68427	2054.09937	6.00413017	Evidence in favor of Pareto
2017	3431.57737	3427.2222	8.82499205	Evidence in favor of Pareto
2018	3176.93248	3171.23764	17.2432469	Evidence in favor of Pareto

Table 7: Computation of the approximate Bayes' factors $\mathrm{BF}^i \approx \exp\left(\frac{1}{2}(\mathrm{BIC}^i_{\mathrm{LNt}} - \mathrm{BIC}^i_{\mathrm{P}})\right)$ and conclusions according to Jeffrey's scale [21, 16, 10].

Let us perform Bayes factors' tests, for the distributions truncated lognormal and Pareto, as attempted in [1]. The Bayes factors can be approximated [21] by the expression $\mathrm{BF}^i \approx \exp\left(\frac{1}{2}(\mathrm{BIC}^i_{\mathrm{LNt}} - \mathrm{BIC}^i_{\mathrm{P}})\right)$ and the interpretation is as follows: if $\mathrm{BF}^i < 0.1$, we have strong support for model LNt, if $0.1 < \mathrm{BF}^i < 1/3$, then the support is moderate, while a Bayes factor greater than $1/3$ suggests a weak support for that model. There is a evidence in favour of Pareto model if $\mathrm{BF}^i > 1$. We observe in Table 7 that similar conclusions to the outcome of the BIC information criterion are achieved by the Bayes factors' tests.

We continue our brief analysis with a standard statistical test to assess whether two models are statistically equivalent or not, known as Vuong's test [34]. This test is used very often when assessing the statistical difference of non-nested models. However, when the models are nested or non-nested the usual log-likelihood ratio test (LRT) may be used [22], but then when the models are non-nested one should simulate the distribution of the log-likelihood ratio statistic as it is no longer distributed necessarily as a chi-square distribution. The authors of [1] attempt to consider log-likelihood ratio tests for the non-nested models Pareto, lognormal and exponential and say:

- That the tests' statistics are $2(\ln(L^i_{\text{PL}} - L^i_{\text{LN}}))$ and $2(\ln(L^i_{\text{PL}} - L^i_{\text{Exp}}))$ instead of the correct log-likelihood ratios (LRT) $2(\ln(L^i_{\text{PL}}) - \ln(L^i_{\text{LN}}))$ and $2(\ln(L^i_{\text{PL}}) - \ln(L^i_{\text{Exp}}))$, respectively.

- That the (in the best case, the correctly written) LRT statistics in the first point of this list are both distributed as the chi-square distribution with two degrees of freedom $\chi^2_{(2)}$, which, according to [22], need not to be the case since the compared models are non-nested.

For the sake of simplicity, we will use only Vuong's test [34], where it is proved that its test's statistic for non-nested models, namely the normalized difference of the log-likelihoods is distributed always as a standard *normal* variable. The authors of the paper [1] do not consider normalized difference of the log-likelihoods so actually they do not perform Vuong's tests.

Sample	Pareto vs LNt
2010	**0.132 (-1.504)**
2011	0.046 (-1.990)
2012	**0.107 (-1.610)**
2013	**0.079 (-1.755)**
2014	**0.098 (-1.653)**
2015	**0.113 (-1.581)**
2016	**0.299 (-1.036)**
2017	**0.326 (-0.980)**
2018	**0.569 (-0.568)**

Table 8: Results of Vuong's test for the null hypothesis that the Pareto and truncated lognormal are equivalent statistical models. The format is p-value (test statistic). Non-rejections at the 5% level are marked in bold.

Our results are shown in Table 8. In it we can see that the Pareto and truncated lognormal distributions are almost always non rejected to be statistically equivalent, with the exception of the year 2011 (rejection in favour of the LNt). But always in Table 8, the test's statistic of Vuong's test is negative, meaning a slight preference for the truncated lognormal distribution.

We complete the study by showing the Zipf's plots for the Pareto and truncated lognormal distributions in Figure 1. In these graphs, we observe visually that the fits of the truncated lognormal distribution (green) is as good or even better than those of the Pareto distribution (red). We observe as well that the actual data (blue) depart from linearity, with ups and downs, in a clear way, something that has been shown before in [26, 27].

So we have a second conclusion: the truncated lognormal distribution is statistically comparable or even better to the Pareto distribution for the top billionaires wealth data at least in the period 2010-2018.

5. Conclusions

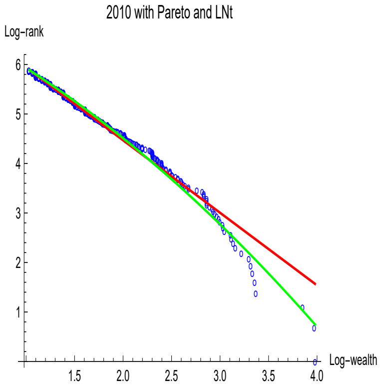
In this short Comment we have seen that the main conclusion of [1] is debatable: that there is statistical evidence that the upper tails of the world billionaires data sets (data measured in US billions of US dollars of wealth of individuals) follow a power law or Pareto distribution. This is not a new result, indeed in the previous references [26, 27] it is shown that the upper tail distribution of billionaires in Canada and in the world may be regarded as non-Pareto.

On the contrary, we have shown that there is statistical evidence on two points that are opposite to the former thesis. Namely:

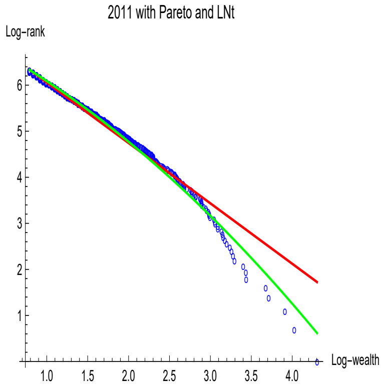
- The power law or Pareto and truncated lognormal distributions might be improved in principle by a different distribution to achieve a good description of the upper tail, in the sense of non-rejection by Anderson–Darling tests, for the data used in [1]. This is in line with [7].

- The truncated lognormal distribution is a comparable or even better model to the power law or Pareto distribution in terms of goodness-of-fit, statistical and graphical criteria, all for the upper tail data used in [1]. This result is also in line with the results of [7, 11].

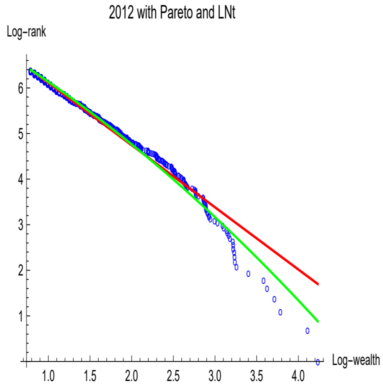
Thus the conclusions obtained in [1] should be modified.



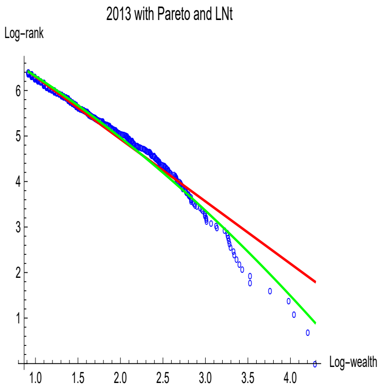
(a)



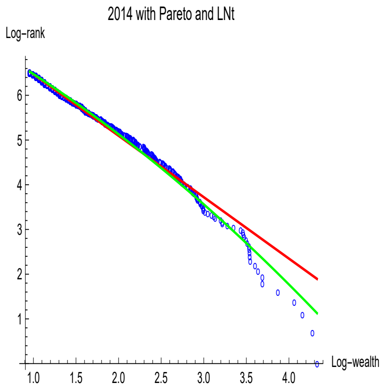
(b)



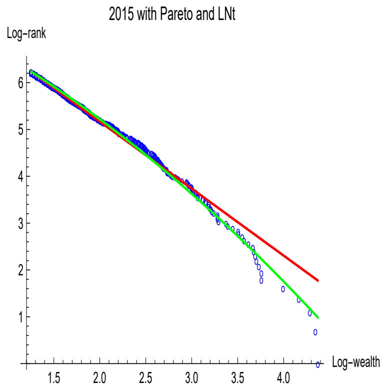
(c)



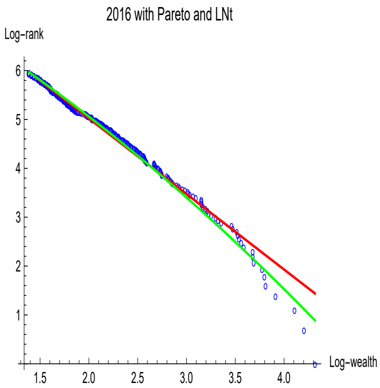
(d)



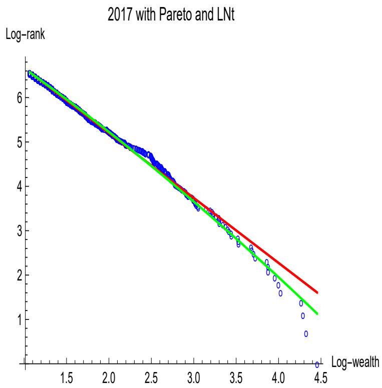
(e)



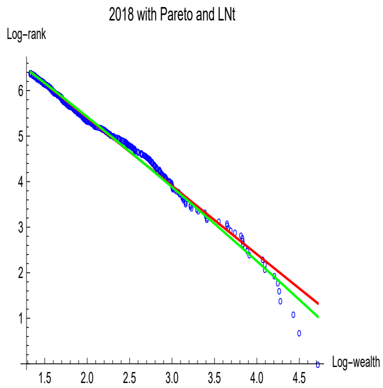
(f)



(g)



(h)



(i)

Figure 1: Log-rank plots for the truncated upper tails, using the power law or Pareto distribution (red), the truncated lognormal distribution LNt (green) and the empirical data (blue).

Author contributions

Arturo Ramos: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, resources, software, validation, visualization, writing-original draft, writing-review & editing.

Competing interests statement

The author declares to have no competing interests concerning the research carried out in this article.

Acknowledgments

We acknowledge very useful and constructive comments from three anonymous reviewers and the Main Editor of Physica A, Dr. Tiziana Di Matteo, that have allowed a substantial improvement of previous versions of this Comment. This study was supported by PID2020-112773GB-I00 of the Spanish Ministerio de Ciencia e Innovación and ADETRE Reference Group S39_20R of Gobierno de Aragón.

References

[1] Asif, M., Hussain, Z., Asghar, Z., Hussain, M. I., Raftab, M., Shah, S. F., and Khan, A. A. (2021). A statistical evidence of power law distribution in the upper tail of world billionaires' data 2010–20. *Physica A: Statistical Mechanics and its Applications*, 581:126198.

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

[3] Bee, M. (2015). Estimation of the lognormal-Pareto distribution using probability weighted moments and maximum likelihood. *Communications in Statistics–Simulation and computation*, 44(8):2040–2060.

[4] Bee, M., Riccaboni, M., and Schiavo, S. (2011). Pareto versus lognormal: a maximum entropy test. *Physical Review E*, 84:026104.

[5] Bee, M., Riccaboni, M., and Schiavo, S. (2013). The size distribution of US cities: not Pareto, even in the tail. *Economics Letters*, 120:232–237.

[6] Bee, M., Riccaboni, M., and Schiavo, S. (2017). Where Gibrat meets Zipf: Scale and scope of French firms. *Physica A: Statistical Mechanics and its Applications*, 481:265–275.

[7] Brzezinski, M. (2014). Do wealth distributions follow power laws? Evidence from 'rich lists'. *Physica A: Statistical Mechanics and its Applications*, 406:155–162.

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

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

[10] Băncescu, I., Chivu, L., Preda, V., Puente-Ajovín, M., and Ramos, A. (2019). Comparisons of log-normal mixture and pareto tails, GB2 or log-normal body of Romania's all cities size distribution. *Physica A: Statistical Mechanics and its Applications*, 526:121017.

[11] Campolieti, M. (2018). Heavy-tailed distributions and the distribution of wealth: Evidence from rich lists in Canada, 1999-2017. *Physica A: Statistical Mechanics and its Applications*, 503:263–272.

[12] Campolieti, M. and Ramos, A. (2021). The distribution of strike size: Empirical evidence from Europe and North America in the 19th and 20th centuries. *Physica A: Statistical Mechanics and its Applications*, 563:125424.

[13] Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications*, 392:5947–5962.

[14] Clauset, A., Shalizi, C. R., and Newman, E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

[15] Fazio, G. and Modica, M. (2015). Pareto or log-normal? Best fit and truncation in the distribution of all cities. *Journal of Regional Science*, 55(5):736–756.

[16] 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.

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

[18] Hannan, E. J. and Quinn, B. G. (1979). The Determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, 41:190–195.

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

[20] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions. Volume 1*. John Wiley & Sons.

[21] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

[22] Lewis, F., Butler, A., and Gilbert, L. (2011). A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, 2:155–162.

[23] Malevergne, Y., Pisarenko, V., and Sornette, D. (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, 83:1–11.

[24] Mitzenmacher, M. (2004). A brief history of generative models for power law and log normal distributions. *Internet Mathematics*, 1(2):226–251.

[25] Montebruno, P., Bennett, R. J., van Lieshout, C., and Smith, H. (2019). A tale of two tails: Do power law and lognormal models fit firm-size distributions in the mid-Victorian era? *Physica A: Statistical Mechanics and its Applications*, 523:858–875.

[26] Ogwang, T. (2011). Power laws in top wealth distributions: evidence from Canada. *Empirical Economics*, 41:473–486.

[27] Ogwang, T. (2013). Is the wealth of the world's billionaires Paretian? *Physica A: Statistical Mechanics and its Applications*, 392:757–762.

[28] Perline, R. (2005). Strong, weak and false inverse power laws. *Statistical Science*, 20(1):68–88.

[29] Puente-Ajovín, M., Ramos, A., and Sanz-Gracia, F. (2020a). Is there a universal parametric city size distribution? Empirical evidence for 70 countries. *The Annals of Regional Science*, 65:727–741.

[30] Puente-Ajovín, M., Ramos, A., Sanz-Gracia, F., and Arribas-Bel, D. (2020b). How sensitive is city size distribution to the definition of city? The case of Spain. *Economics Letters*, 197:109643.

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

[32] Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

[33] Tomaschitz, R. (2020). Multiply broken power-law densities as survival functions: An alternative to Pareto and lognormal fits. *Physica A: Statistical Mechanics and its Applications*, 541:123188.

[34] Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.