Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrie & Operationele Research

Analyzing The Distribution Of Economic Aspects: Top Wealth And U.S. City Size

Joey van der Eijk (583253)

Supervisor: Phyllis Wan
Second assesor: Jens Klooster
Date: 1st July 2024

Abstract

This paper studies the distribution of Top Wealth using the *Forbes Billionaires Evolution List*. Essentially, it extends the paper of Teulings & Toussaint (2023). We use their proposed test statistic to propose a more rigorous test. Next to this, another class of methods is proposed. These are the Anderson-Darling test and the Cramér-von Mises test. These statistical tests can be used to assess whether a sample of data follows a specific distribution. All tests reject the Pareto distribution while the Weibull distribution is not rejected for the years 2001-2012. This methodology is also applied to the U.S. City Size, specifically the population of metropolitan statistical areas (MSAs). Again, the Pareto distribution is rejected. However, the Weibull distribution is not rejected based on the results of the statistical tests. Ultimately, the Pickands-Balkema-De Haan Theorem, also called the second extreme value theorem, motivates the consideration of the Generalized Pareto distribution (GPD) as a more justified alternative to Pareto or Weibull for the exceedances above a certain threshold. All tests do not reject the GPD for the exceedances above the 90th quantile of Top Wealth.

1 Introduction

Many features of economic life exhibit right-skewed distributions (Gabaix (2009), Gabaix (2016)). For example, income and wealth distributions are often right-skewed (Vermeulen, 2018). The same holds for city populations (Gabaix, 1999), and also firm size often follows a right-skewed distribution (Luttmer, 2011). Other examples are housing prices within a region and salaries within a company, which also tend to follow a right-skewed distribution.

Such data is often assumed to follow a Pareto distribution. For example, Krugman (1996) and Gabaix (1999) assumed the default of Pareto for the 135 largest metropolitan statistical areas (MSAs). Also, Eeckhout (2004) has already challenged the Pareto assumption of city size for log normal. Despite this, empirical and also theoretical evidence is increasing from which it is shown that the Pareto distribution actually is a poor fit in many of these occasions. For example, Blanchet et al. (2022) already show that the Pareto distribution is not a good fit regarding predictions of mean wealth and income.

The Pareto distribution only considers one parameter, α. This parameter determines the heaviness of the tail of the distribution. However, this simplicity may not be sufficient to model these economic features. This is why we explore other distributions to analyze these phenomena. For example, research has already been conducted considering the Weibull distribution for economic features (Teulings & Toussaint, 2023). They conclude that Top Wealth actually follows a Weibull distribution, not a Pareto distribution.

We would also like to explore other distributions besides Pareto for modelling various economic features. Hence, we consider two other distributions, the Weibull distribution and the Generalized Pareto distribution (GPD). Compared to the Pareto distribution, the Weibull distribution has two parameters, α and γ, the scale and shape parameter respectively. This makes the Weibull distribution more flexible.

However, to understand the practical difference, one must consider the logs of both Pareto and Weibull, which are Exponential and Gompertz respectively. Teulings & Toussaint (2023) show that the Gompertz distribution has an increasing hazard rate ($\alpha^{-1} \exp(\omega\gamma)$) while the

Exponential distribution exhibits a constant hazard rate (α^{-1}). The constant hazard rate of the Exponential distribution implies that the amount of people that have more wealth than a particular lower bound decreases by a constant percentage when this lower bound increases by one unit. The increasing hazard rate of the Gompertz distribution, where $\gamma > 0$ and ω is the lower bound, implies that the probability of being more wealthier increases when one has more wealth. They also show that the increasing hazard of the Gompertz distribution fits their data (*Forbes List of Billionaires*) well while the constant rate of the Exponential distribution is strongly at odds with this data.

The GPD is considered because of the Pickands-Balkema-De Haan theorem which states that, above a certain threshold, the exceedances converge to the GPD. For this theorem to hold, the threshold should be sufficiently high and the data needs to have enough extreme values which means that the distribution should not decrease to zero too fast. This provides motivation to consider GPD as a more justified alternative to Pareto or Weibull.

Teulings & Toussaint (2023) propose a statistic $\mathcal{R}_k$ to test for Pareto, but not a rigorous test. In this paper, two classes of methods are proposed for the evaluation of the distribution of economic features. The proposed statistics are used for the first part of the methodology. The statistic is simulated a various amount of times using simulated data from the null distribution where the parameters are estimated from the data. This results in the distribution of the test statistic. With this information, we can test for a distribution on a specified significance level. Second, two statistical tests are also used to assess if a sample follows a specified distribution. These tests are the Anderson-Darling test and the Cramér-von Mises test. With these tests, one can also assess if a sample follows a specified distribution. Next to this, with these tests, one can also assess whether two samples follow the same underlying distribution.

In this paper, Top Wealth and U.S. City Size are considered. For both these sets of data, the Pareto distribution is rejected. However, the Weibull distribution is not rejected for U.S. City Size and its sub-samples. The same holds for Top Wealth, but only for the first 12/13 years of this century. After this, the evidence for a Weibull distribution is declining. This shows that the distribution of Top Wealth has been shifting towards another, unknown distribution. Finally, the exceedances above the top decile of Top Wealth conform the Generalized Pareto distribution. This provides evidence for the Pickands-Balkema-De Haan theorem.

This paper is organised as follows: In Section 2, the literature is discussed. In Section 3, the Pareto distribution is discussed. Section 4 shows the theory about the Weibull distribution. Section 5 discusses the theory about the Generalized Pareto distribution. In Section 6 the two different datasets are presented. In Section 7, the methodology from Teulings & Toussaint (2023) is used on these two sets of data. In Section 8, statistical tests are discussed and in Sections 9, 10 and 11, we test for Pareto, Weibull and Generalized Pareto respectively. Finally, in Section 12, the conclusions are drawn.

2 Literature Review

This section provides the literature about the concepts which are used in this research. More information regarding the distributions can be found in Sections 3, 4 and 5.

2.1 Cramér-von Mises test

The first statistical test which is used to assess the distribution of the economic features in this research is called the Cramér-von Mises Test. This test takes the overall difference between the two distributions into account. More information regarding this can be found in Section 8.2. Cramér (1928) first wrote a paper which essentially was the groundwork for theory under which the Cramér-von Mises criterion was formed. Mises (1931) later wrote a book in which the statistical criteria is introduced which later would be part of the basis of this test.

2.2 Anderson-Darling test

The second statistical test which is used to assess the distribution of the economic features in this research is called the Anderson-darling test. Anderson & Darling (1952) created this test for goodness-of-fit testing. This test is called the Anderson-Darling Test. Compared to other test, this test focuses more on the tails. Hence, this test is a modification of the Cramér-von Mises test. The Anderson-Darling test is more sensitive to deviations in the tail of the data.

2.3 Pickands–Balkema–De Haan theorem (second extreme value theorem)

The Pickands–Balkema–De Haan theorem, often referred to as the second extreme value theorem states that the exceedances of a sample can be approximated by the GDP above a certain threshold T. Pickands III (1975) was the first to introduce the concept about GPD. Balkema & De Haan (1974) on the other hand made their contribution by understanding this asymptotic behaviour better. They did research about the conditions under which the data can be distributed by the GDP. Altogether, this theorem was created. More information regarding this theorem can be found in Section 5.1.

3 Pareto: Theory

Following the methodology presented by Teulings & Toussaint (2023), a random variable $\underline{X} \geq \Omega$, with $\Omega > 0$ is considered. For instances where no lower bound is known, the lowest observation should be used as lower bound. In this research, Ω is assumed to be known.

The probability density function (PDF) of the Pareto distribution is shown below:

$$f(x) = \frac{\alpha\Omega^{\alpha}}{x^{\alpha+1}}, \qquad \text{for} \quad x \geq \Omega, \tag{1}$$

where $\Omega > 0$ is the scale parameter and $\alpha > 0$ is the shape parameter. For example, for the *Forbes Billionaires Evolution List*, $\Omega = 1$ billion USD. For U.S. City Size, $\Omega \approx 50.000$ people. The moments of the Pareto distribution are given by:

$$E[\underline{X}^k] = \frac{1}{1-\alpha k}\Omega^r, \quad 0 < k < \alpha^{-1}, \tag{2}$$

where α is strictly positive. Equation (2) is only identified in the case that $k \leq \alpha^{-1}$. Jones (2015) states that α^{-1} is usually called the tail index. Vermeulen (2018) shows that the tail index is approximately 1.5 for wealth distributions and Eeckhout (2004) states that the index is close to 1 for city sizes. With this information, it can be assumed that the index doesn't exceed 2. In this case, only the first moment exists. Actually, Teulings & Toussaint (2023) state that in many cases even the first moment doesn't exist.

Therefore, similar to Teulings & Toussaint (2023), we work with the log-transformation of $\underline{X}$. First $\underline{W} := \underline{X}/\Omega$ is defined. Subsequently, $\underline{w} := ln\,\underline{W} = \underline{x} - \omega$ is defined, where $\underline{x} := ln\,\underline{X}$ and $\omega := ln\,\Omega$. The log transformation of a Pareto distributed variable follows an Exponential distribution. The moments of this distribution are shown below:

$$E[\underline{w}^k] = \alpha^k \Gamma(k+1) = \alpha^k k!, \tag{3}$$

where $\Gamma(\cdot)$ is the Gamma function. Because all moments do exist for this distribution, the Exponential distribution is more suitable for testing.

Teulings & Toussaint (2023) show that equation (3) can be used to construct simple test statistics to test whether $\underline{w}$ is Exponentially distributed. Their proposed test statistics look as follows:

$$\mathcal{R}_k := \frac{E[\underline{w}^k]}{k!E[\underline{w}]^k}, \quad \hat{\mathcal{R}}_k := \frac{\overline{w^k}}{k!\overline{w}^k}. \tag{4}$$

Under the null of $\underline{w}$ being Exponentially distributed, the expected value of $\hat{\mathcal{R}}_k$ is 1. The variance of $\hat{\mathcal{R}}_3$ is proven to be larger than the variance of $\hat{\mathcal{R}}_2$, thus making $\hat{\mathcal{R}}_2$ the more powerful statistic.

4 Weibull: Theory

Teulings & Toussaint (2023) compare the hazard rates of the Exponential distribution and the Gompertz distribution. These are the log transforms of the Pareto distribution and the Weibull distribution respectively. It appears that the hazard rate of the Exponential distribution does not fit their data (*Forbes List of Billionaires*) well, while the hazard rate of the Gompertz distribution does.

The PDF of the Weibull distribution is defined as follows:

$$f(x;\alpha,\gamma) = \begin{cases} \frac{\gamma}{\alpha}\left(\frac{x}{\alpha}\right)^{\gamma-1} e^{-(x/\alpha)^\gamma}, & \text{for} \quad x \geq 0, \\ 0, & \text{for} \quad x < 0, \end{cases} \tag{5}$$

where $\alpha > 0$ is the scale parameter and $\gamma > 0$ is the shape parameter.

The moments for Weibull are given by:

$$E[\underline{W}^k] = (\alpha\gamma)^{k/\gamma} e^{(\alpha\gamma)^{-1}} \Gamma(1 + k\gamma, (\alpha\gamma)^{-1}), \tag{6}$$

where $\Gamma(\cdot,\cdot)$ *is the upper incomplete Gamma function.*

Similar to the Pareto distribution, the log transformation is used. The log of a Weibull distributed variable follows a Gompertz distribution. The Gompertz distribution has 2 parameters, α & γ. Wingo (1989) estimates these parameters with method of moments, whereas Teulings & Toussaint (2023) use maximum likelihood to optimize the parameters, using the density of w. The derivation of their likelihood function can be found in Appendix C in their paper. This results in the following expressions for $\hat{\alpha}$ & $\hat{\gamma}$:

$$\hat{\alpha} = \gamma^{-1}(\overline{e^{\gamma w}} - 1), \tag{7}$$

$$N^{-1}\mathcal{L}(\gamma) = ln\gamma - ln(\overline{e^{\gamma w}} - 1) + \gamma\overline{w}, \tag{8}$$

$$\hat{\gamma} = \frac{\overline{e^{\hat{\gamma} w}} - 1}{\overline{e^{\hat{\gamma} w}}}(\hat{\gamma}^{-1} + \overline{w}). \tag{9}$$

Because of the relative simplicity of the expression for $\hat{\alpha}$, α itself can be eliminated in the likelihood. This is very practical for calculations. First, one $\hat{\gamma}$ is estimated over the whole sample where $N \geq 64$. Next, this common $\hat{\gamma}$ is used to separately calculate $\hat{\alpha}$. Hence, one $\hat{\gamma}$ is used for the calculation.

Teulings & Toussaint (2023) show that equation 6 together with the moments of Gompertz can be used for the calculation of the asymptotic expectation of the test statistic:

$$\mathcal{R}_k(\alpha\gamma) = \frac{h(\alpha\gamma, k)}{k!\, h(\alpha\gamma, 1)^k} = \frac{\int_0^\infty q^{k-1}\, q\, \exp\left(-(\alpha\gamma)^{-1}e^{-q}\right)\, dq}{(k-1)!\, e^{(\alpha\gamma)^{k-1}}\mathrm{Ei}((\alpha\gamma)^{-1})^k}. \tag{10}$$

This equation is used by them to test for Weibull. This is discussed more clear at Section 7.2.

5 Generalized Pareto: Theory

The Generalized Pareto distribution (GPD) is often used to model the tails of other distributions. Compared to the Pareto and Weibull distribution, the GPD has more parameters. The motivation for using the GPD can be found in Section 5.1.

The PDF is:

$$f(x; T, \sigma, \xi) = \begin{cases} \frac{1}{\sigma}\left(1 + \frac{\xi(x-T)}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \text{if } \xi \neq 0, \\ \frac{1}{\sigma}\exp\left(-\frac{x-T}{\sigma}\right), & \text{if } \xi = 0, \end{cases} \tag{11}$$

where T is the threshold, σ >0 is the scale parameter and ξ is the shape parameter.

5.1 Pickands–Balkema–De Haan theorem

When the distribution of a random variable is not known, the Pickands–Balkema–De Haan theorem provides the asymptotic tail distribution of this variable. The theorem states that, above a certain threshold T, data can be approximated by the Generalized Pareto distribution. It is often also referred to as the second extreme value Theorem.

Consider a sequence of random variables $X_1, ..., X_n$ which are independently and identically distributed (i.i.d) together with a high enough threshold T. According to the Pickands–Balkema–De Haan theorem, the distribution of the exceedances ($Y_i = X_i - T$) of this threshold can be approximated by the GPD. For this theorem to hold, T should be high enough.

For a distribution function F of variable X, this theorem describes the conditional distribution of the variable above a certain threshold. This is called the conditional excess distribution function:

$$F_T(y) = P(X - T \leq y | X > T) = \frac{F(T+y) - F(T)}{1 - F(T)}. \tag{12}$$

This function thus describes the distribution of all excess values above the threshold. For various distributions F, F_T is approximated well by the GPD.

The theorem states that if there exist functions $a(T)$ and $b(T)$ such that $F_T(a(T)y + b(T))$ converge to a non-degenerate distribution. In this case, such limit is equal to the GPD:

$$F_T(a(T)y + b(T)) \to G_{\xi,\sigma}(y), \text{ as } T \to \infty, \tag{13}$$

where $G_{\xi,\sigma}(y)$ is the cumulative density function (CDF) of GPD. This CDF is defined as follows:

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - e^{-y/\sigma}, & \text{if } \xi = 0, \end{cases} \tag{14}$$

where y_i are the observed exceedances above a threshold T defined as: $y_i = x_i - T$.

6 Data

The *Forbes Billionaires Evolution List*[1] from 2001-2021 is used for this research. This set closely resembles the dataset used by Teulings & Toussaint (2023) (*Forbes List of Billionaires*). Similarly, the billionaires are categorized based on their citizenship.

To start with, the countries are clustered into regions. Next, some regions are split into sub-regions. Some of these sub-regions also consist out of multiple countries. The aim is to form regions out of countries which comply with the following three criteria: 1) the countries must be close to each other; 2) the countries must be close to each other in terms of GDP and 3) the amount of billionaires should exceed the rough threshold of 40. The classification of the regions can be found in Table 1.

Table 18 presents the mean summary statistics of all different regions for the period 2001-2021. Hence, N is the average amount of billionaires between 2001-2021. Similarly, $\hat{\mathcal{R}}_2$, $\hat{\mathcal{R}}_3$ and $\overline{w}$ also denote averages over the period 2001-2021.

Similar to Teulings & Toussaint (2023), all $\hat{\mathcal{R}}_2$ & $\hat{\mathcal{R}}_3$ of the (sub-)regions are smaller than one. It can also be observed that $\overline{w}$ exceeds one in some cases. This implies that $E[\underline{W}]$ does not exist for the Pareto distribution in these cases as mentioned in Section 3.

In comparison to the summary results of the *Forbes List of Billionaires*, the *Forbes Billionaires Evolution List* has 69 billionaires less per year listed in the world on average. While most results for the summary statistics are almost the same, there are some differences. The biggest difference is for both $\hat{\mathcal{R}}_2$ & $\hat{\mathcal{R}}_3$ of South Korea. The $\hat{\mathcal{R}}_3$ of the *Forbes Billionaires Evolution List* appears to be 0.18 higher and the $\hat{\mathcal{R}}_2$ is almost 0.1 higher. Compared to all the other statistics, these are the biggest differences. The reason for this could simply be the missing values compared to the *Forbes List of Billionaires*. For all other statistics in this table, the results do not deviate much from the summary statistics of the *Forbes List of Billionaires*.

[1] https://www.kaggle.com/datasets/guillemservera/forbes-billionaires-1997-2023

Table 1

(Sub-)Region	Countries
North America	US, Canada
Europe	Greece, Belgium, Czech Republic, Czechia, Guernsey, Hungary, Iceland, Lithuania, Netherlands, Poland, Portugal, Romania, Slovakia , Spain, British Islands, Scandinavia, France, Alps
British Islands	U.K., Ireland
Scandinavia	Sweden, Denmark, Norway, Finland
France	France, Monaco
Alps	Switzerland, Austria, Liechtenstein
China	China, Hong Kong
East Asia	Vietnam, Macao, New Zealand, Southeast Asia, Asian Islands, South Korea, Japan, Australia
Southeast Asia	Thailand, Malaysia, Singapore
Asian Islands	Taiwan, Philippines, Indonesia
Central Eurasia	Armenia, Georgia, Kazakhstan, Ukraine, Russia
South America	Argentina, Belize, Chile, Colombia, Guatemala, Mexico, Peru, Venezuela, Brazil
Middle East	Cyprus, Egypt, United Arab Emirates, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Israel + Turkey
Rest of World	Zimbabwe, Uganda, Algeria, Angola, Tanzania, South Africa, Morocco, Pakistan, Nigeria, Eswatini (Swaziland), Nepal, St. Kitts and Nevis

This table shows the division of (sub-)regions. If a country is present in Table 18 but not on the left hand side of this table, that country/region consists of just that country.

U.S. City Size is also used, this dataset is from the 2000 U.S. Census, the 2001 vintage. This dataset contains data about the 280 most populous metropolitan statistical areas (MSAs) in the U.S.. The data is not divided here, 4 different subsets are used. These are the following sets: the full set, the top 135, the top 100 and the top 50. In Table 2, $\hat{\mathcal{R}}_2$ & $\hat{\mathcal{R}}_3$ are all again

Table 2

Sample	n	ω	$\overline{w}$	$\overline{w^2}$	$\overline{w^3}$	$\hat{\mathcal{R}}_2$	$\hat{\mathcal{R}}_3$
Full	280	10.8	1.93	4.97	15.8	0.668	0.368
Top 135	135	12.6	1.1	2.07	5.06	0.858	0.637
Top 100	100	12.9	1.06	1.89	4.35	0.848	0.615
Top 50	50	13.8	0.851	1.26	2.42	0.871	0.654

This table shows the summary statistics of the U.S. City Size. n is the amount of MSAs in the sample, ω is the lower bound of the city size. w is defined as the log city size - the log lower bound, where the line above the variables implies that it is the mean. $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ are the proposed statistics as defined in Section 3.

lower than one. Hence, the Pareto assumption already does not seem to hold. Even though the statistics are increasing for the upper tail, the values don't exceed 0.871.

7 Testing Pareto and Weibull via the methodology of Teulings & Toussaint (2023)

In this section, the methodology of Teulings & Toussaint (2023) is repeated for the *Forbes Billionaires Evolution List.*

7.1 Testing Pareto

For Top Wealth, there are 378 observations (21 years x 18 regions). Teulings & Toussaint (2023) show that the variance of the test statistics is predicted to be proportional to $1/N$. This is corrected by using WLS where the weights are defined as $\sqrt{N}$, which represents the square root of the number of billionaires. The results of these regressions are presented in Table 3.

It is directly observed that the null hypotheses ($\hat{\mathcal{R}}_2 = \hat{\mathcal{R}}_3 = 1$) is rejected in all cases. All intercepts differ from unity. This result remains consistent after performing the following robustness checks: For regressions (2) and (4), all observations which contain less than 64 billionaires are dropped. This is due to the small-sample bias in $\hat{\mathcal{R}}_2$ & $\hat{\mathcal{R}}_3$. For regression (3) and (6), the top and bottom 5% of the 75 observations where $N \geq 64$ are also dropped, this leaves 67 observations.

Table 3

Dependent variables:	$\hat{\mathcal{R}}_2$			$\hat{\mathcal{R}}_3$		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Constant	0.82***	0.84***	0.85***	0.60***	0.65***	0.64***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics						
Observations	378	75	67	378	75	67
RMSE	0.128	0.079	0.068	0.200	0.134	0.115
Theoretical RMSE	1	1		$\sqrt{10}$	$\sqrt{10}$	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*'. This table shows the WLS regressions on $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$. Regression (1) and (4) contain all 378 observations. For regression (2) and (5), only the regions with a minimum of 64 billionaires is used and for regression (3) and (6), the top and bottom 5% are removed, leaving 67 observations. RMSE stands for root mean squared error.

When comparing these results to the corresponding table in the paper of Teulings & Toussaint (2023), it is obvious that the results do not align perfectly. The most clear differences are the bigger RMSEs and standard errors which they observe. The RMSE values which are obtained for the *Forbes Billionaires Evolution List* are at least two times as small as those of Teulings & Toussaint (2023). This is most likely the consequence of using another dataset.

Figure 6 in the Appendix presents the distribution of these statistics for all six different cases. It is clear that most of the statistics are smaller than unity.

7.2 Testing Weibull

For this part, only the observations where $N \geq 64$ holds are used. Figure 5 presents the distribution of $\hat{\gamma}$ where the vertical red line represents the $\hat{\gamma}_{median} \approx 0.28$.

Equation (10) shows that $\mathcal{R}_k(\alpha\gamma)$ is a declining function for ln α + ln γ. Because of this, Teulings & Toussaint (2023) do a rough test of this prediction. They regress $\hat{\mathcal{R}}_2$ & $\hat{\mathcal{R}}_3$ on both ln $\hat{\alpha}$ and ln $\hat{\gamma}$. They state that equation (10) predicts that the coefficients on both of the variables should be non-positive and equal for both test statistics.

Again, they observe significantly bigger standard errors and RMSEs. Even though all predictions are confirmed in Table 20, one could argue that these regressions alone are not sufficient to assume that Top Wealth is distributed Weibull. This is why this research will conduct more tests regarding the distribution of Top Wealth.

As a final test, Teulings & Toussaint (2023) run a horse race between Pareto and Weibull for the prediction of mean wealth in 2018. The median gamma is used as a common value for all observations. As a result, both distributions only have a single parameter to estimate. For Pareto, following equation (3), $\hat{\alpha}$ is simply equal to $\overline{w}$. For Weibull, $\hat{\alpha}$ is estimated via equation (7). The results for using the $\hat{\gamma}_{median}$ of the *Forbes Billionaires Evolution List* can be found in Table 4.

Table 4

	Mean Wealth			$\hat{\alpha}$	
Sub-Region	Data	Weibull	Pareto	Weibull	Pareto
United States	5.29	**5.48**	∞	**1.52**	1.16
Canada	3.23	**3.22**	5.96	**1.04**	0.832
Germany	4.7	**5.57**	∞	**1.53**	1.18
British Isles	3.83	**4.26**	∞	**1.29**	1.01
Scandinavia	3.51	**4.06**	227	**1.24**	0.996
France	7.44	**8.12**	∞	**1.89**	1.37
Alpine countries	3.8	**4.66**	∞	**1.37**	1.1
Italy	3.96	**4.29**	∞	**1.29**	1.02
China	3.31	**3.10**	5	**1.01**	0.8
South-East Asia	3.33	**3.52**	9.43	**1.12**	0.894
Asian Island	3.17	**3.22**	6.18	**1.04**	0.838
South Korea	2.88	**2.78**	3.8	**0.912**	0.737
Japan	3.95	**3.82**	12.4	**1.19**	0.919
Australia	2.74	**2.75**	3.86	**0.903**	0.741
India	3.7	**3.90**	24.1	**1.21**	0.959
Russia	4.05	**4.06**	21.3	**1.25**	0.953
Brazil	4.2	**4.41**	∞	**1.32**	1.03
Israel + Turkey	2.17	**2.24**	2.64	**0.727**	0.622

This table consists out of the predicted and realised mean wealth in 2018. The last two columns represent the estimated $\hat{\alpha}$ which is used for the prediction. For the estimation of $\hat{\alpha}$ for Weibull, $\hat{\gamma}_{median}$ ≈ 0.282 (the $\hat{\gamma}_{median}$ from the dataset which is utilized in this research).

The 2 columns corresponding to Weibull predictions are highlighted. If these are compared to those of Teulings & Toussaint (2023), the results do not perfectly align. Again, this is most likely the result of using another dataset. This results in another $\hat{\gamma}_{median}$ which leads to different estimations of $\hat{\alpha}$ which on that part leads to different estimations for mean wealth in 2018. The results for using the $\hat{\gamma}_{median}$ of the *Forbes List of Billionaires* can be found in Table 19. Here, the results align for the most part. For all but 2 regions, the exact same $\hat{\alpha}$ is observed. The biggest differences are for the Asian Islands and for South Korea.

8 Classes of methods

In this section, new classes of methods are proposed for assessing the distribution of Top Wealth and U.S. City Size.

8.1 Simulation of test distributions of $\hat{\mathcal{R}}_2$ & $\hat{\mathcal{R}}_3$

Recall the test statistics proposed by Teulings & Toussaint (2023), $\mathcal{R}_k$ for $k = 2$ and 3. Where Teulings & Toussaint (2023) simply reject for statistics which are significantly different from unity, we simulate these statistics with random generated numbers from several distributions where the parameters are estimated from the data. This results in the distribution of the test statistics. With these distributions, the critical values corresponding to their significance levels can be calculated. With this information, one can perform a statistical test to determine whether a sample of data follows a certain distribution.

To generate random numbers, the inverse transform sampling method is applied. First, $F(U)^{-1}$ (the inverse CDF) is derived. After specifying the parameters, a random number X is calculated by $X = F(U)^{-1}$ where $U \sim$ Uniform(0,1). One random number is generated by first generating a random uniform number between 0 and 1 and using this value in the inverse CDF.

The CDF of the Pareto distribution is:

$$F(x) = 1 - \left(\frac{\Omega}{x}\right)^{\alpha}, \tag{15}$$

where Ω is the shape parameter and α is the scale parameter.

This results in:

$$X = \Omega * (1 - U)^{-(1/\alpha)}, \tag{16}$$

where $U \sim$ Uniform(0,1).

To simulate the test distributions of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$, all parameters have to be specified. These estimates vary between Top Wealth and U.S. City Size. As already mentioned, Ω is the minimum value of the data. The parameter $\hat{\alpha}$ is estimated by equation (3) over the whole sample.

The CDF of the Weibull distribution is:

$$F(x; \alpha, \gamma) = 1 - e^{-(\frac{x}{\alpha})^{\gamma}}, \tag{17}$$

where $\alpha > 0$ is the scale parameter and $\gamma > 0$ is the shape parameter and $x \geq 0$. This results in:

$$X = \alpha(-ln(U))^{1/\gamma}, \tag{18}$$

where $U \sim$ Uniform(0,1). For the Weibull distribution, the parameters are estimated using the **fitdist** function within the **fitdistrplus** package in Rstudio. The parameters are estimated by using maximum likelihood.

One can now simulate the test statistics various amount of times where the amount of random numbers used for one statistic is set equal to the length of the full dataset. This results in the test distributions of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$. With these test distributions, the critical values according to their significance levels can be calculated. With this information, one can statistically reject/not reject the null hypothesis of a certain distribution. All critical values are rounded to three decimals in stead of two. This is because of the small differences between some critical values.

This could be displayed as:

$$\text{Reject distribution if } \hat{\mathcal{R}}_2 < F_{S,\hat{\mathcal{R}}_2} \text{ \& Reject distribution if } \hat{\mathcal{R}}_3 < F_{S,\hat{\mathcal{R}}_3},$$

where S is the significance level and Y is the distribution for which is tested.

For these tests:

Null Hypothesis: The data follows a Pareto/Weibull distribution.

In this context, high p-values correspond to significant results, because significant results are obtained when the null is not rejected. The null is not rejected if the p-value is higher than the given significance level.

8.2 Statistical tests

Two different statistical tests are discussed. These tests are used to determine whether a certain sample of data comes from a specified distribution. The tests can also be used to determine if two or more samples arise from the same distribution. The tests are the Anderson-Darling test (AD test) and the Cramér-von Mises test (CvM test).

The first form is used to determine if the data follows a Weibull distribution. First, the data is fitted to a Weibull distribution using the **fitdist** function within the **fitdistrplus** package in Rstudio. After this, the tests are conducted. In this context (when the parameters are estimated from the data), one must notice that the critical values have to be adjusted. The functions **ad.test** and **cvm.test** within the **goftest** package are used to conduct the tests. The CDFs of the null distribution of the test statistic are computed respectively by the **pAD** package with the algorithm of CSöRgő & Faraway (1996) for the Anderson-Darling test and by the **pCvM** package using the algorithm of Marsaglia & Marsaglia (2004) for the Cramér-von Mises test. These functions also exist within the **goftest** package. The quantiles are computed by **qAD** and **qCvM** using one-dimensional root finding with the **uniroot** function. The function then uses Braun's method (Braun, 1980) because parameter estimation is used. Because his method makes use of randomly splitting the data in two sets, the p-value can slightly differ each time these tests are used. This is why the p-value is calculated 1000 times and the mean p-value is used. This also applies to the test statistics.

Null Hypothesis: The data follows a Weibull distribution.

For GPD, a permutation based K-sample AD test and CvM test are used using the **twosamples** package. A K-sample test is used to test whether different samples are from the same distribution. This is why random deviates of the GPD distribution are generated using the **rgpd** function within the **VGAM** package. However the data is first fitted to a GPD distribution using the **fitgpd** function within the **POT** package. After this, the sample of random GPD distributed variables and the real data are compared with these tests. The p-value is calculated by randomly resampling the combined sample into two equal samples. First, the original test statistic is calculated for the given samples. After this, the samples are combined and redivided into two groups for a numerous amount of times. Every iteration of this procedure, the statistic is calculated. At the end, the proportion of statistics which are greater than or equal to the original statistic provides the p-value. Also in this case, the p-values can slightly differ, this is why the mean of 1000 p-values is used. This also applies to the test statistics .

Null Hypothesis: The data follows a Generalized Pareto distribution.

For these tests, if the p-value is smaller than the chosen significance level, the null hypothesis is rejected. This is equivalent to a bigger statistic than the critical value. This means that again, high p-values mean that the null of a certain distribution cannot be rejected. And here, the null should not be rejected for significant results. For all these tests, the data first has to be fitted to the distribution for which is tested. After this, the estimated parameters are used for generating the random numbers (K-Sample test) or for the specification of parameters of the null distribution.

The Anderson-Darling statistic, as well as the Cramér-von Mises statistic are part of the quadratic EDF statistics. These statistics measure the distance between the two empirical distribution functions by

$$n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \, w(x) \, dF(x), \tag{19}$$

where $w(x)$ is the weighting function and n is the number of observations.

8.2.1 Anderson-Darling test

The Anderson-Darling test is used to assess the distribution of a certain sample of data. This test places extra weight on the observations in the tail. In the most basic form of this test, it is assumed that no parameters are to be estimated in the distribution for which is being tested. However, in this research, the basic form is not used, which is often the case.

The Anderson-Darling test is based on:

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)\,(1 - F(x))} \, dF(x), \tag{20}$$

where $F(x)$ is the hypothesized distribution and $F_n(x)$ is the given empirical distribution. In this case, $w(x) = [F(x)(1 - F(x))]^{-1}$,

8.2.2 Cramér-von Mises test

The Cramér–von Mises test also tests the goodness-of-fit of the data but this particular test does it by taking the overall difference between the two distributions into account. It balances the sensitivity over the whole distribution. Like already discussed, the Cramér-von Mises criterion is based on:

$$\omega^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \, dF(x), \tag{21}$$

where $F(x)$ is the hypothesized distribution and $F_n(x)$ is the given empirical distribution. In this case. $w(x) = 1$.

9 Testing for Pareto

For all tests, the parameter of the Pareto distribution first needs to be estimated. After this, 500 test statistics are simulated using random Pareto samples of the same length as the *Forbes Billionaires Evolution List.*

9.1 Top Wealth

In this context, Ω is equal to 1 billion USD. The estimated $\hat{\alpha}$ over the whole sample is approximately equal to 0.94.

The densities of the simulated test statistics are shown beneath in Figure 1 & Figure 2. Both densities exhibit a similar form. The biggest difference is that the x-axis has a greater reach for the simulated $\hat{\mathcal{R}}_3$. As a result, the values in Figure 1 are more centered around one. This is as expected as Teulings & Toussaint (2023) stated that $\hat{\mathcal{R}}_3$ has the bigger variance.

The summary statistics of these simulated statistics are stated below, in Table 5.

Table 5

	Mean	Standard error ($*10^2$)
$\hat{\mathcal{R}}_2$	1.00	0.61
$\hat{\mathcal{R}}_3$	1.00	1.97

This table presents the means and standard errors of the simulated test distributions (for Pareto) using the estimated $\hat{\alpha}$ (0.94) over the whole sample (*Forbes Billionaires Evolution List*) and as minimum, 1 billion USD.

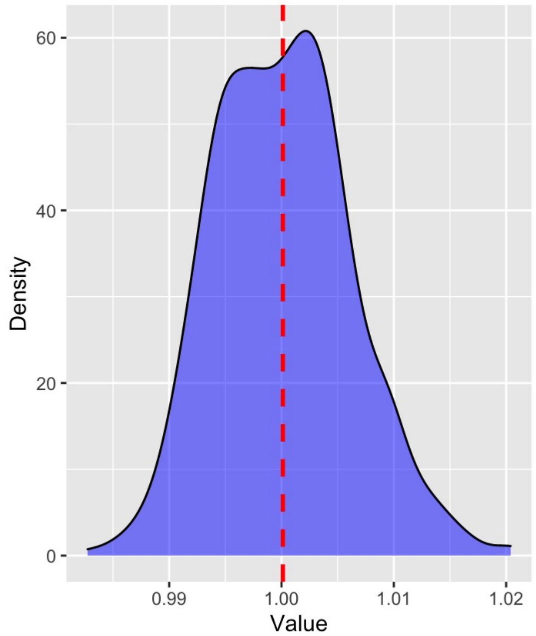
The statistics are centered approximately around 1. It is observed that the density for $\hat{\mathcal{R}}_3$ exhibits a greater standard error which corresponds to its higher variance.

Table 6

Significance levels	$F_{S,\hat{\mathcal{R}}_2}$	$F_{S,\hat{\mathcal{R}}_3}$
1%	0.986	0.956
2.5%	0.988	0.963
5%	0.990	0.969
10%	0.992	0.974

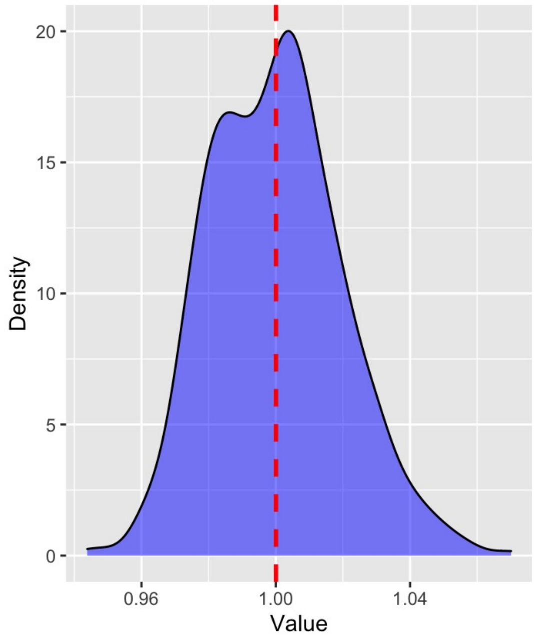
This table displays the critical values corresponding to their significance level from the simulated test distributions (for Pareto) of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ for Top Wealth. S represents the significance level.

Figure 1



This figure shows the density of the simulated $\hat{\mathcal{R}}_2$ (for Pareto) using the estimated $\hat{\alpha}$ (0.94) over the whole sample (*Forbes Billionaires Evolution List*) and as minimum, 1 billion USD.

Figure 2



This figure shows the density of the simulated $\hat{\mathcal{R}}_3$ (for Pareto) using the estimated $\hat{\alpha}$ (0.94) over the whole sample (*Forbes Billionaires Evolution List*) and as minimum, 1 billion USD.

The $\hat{\mathcal{R}}_2$ from the full sample is **0.840**. The $\hat{\mathcal{R}}_3$ from the full sample is **0.622**. These two values are lower than all of their corresponding significance level. Hence, we reject the null hypothesis that Top Wealth follows a Pareto distribution.

9.2 U.S. City Size

For the full sample, the calculated $\overline{w}$ (1.93) is used as $\hat{\alpha}$, as well as the ω (10.8) which is equal to an Ω of approximately 50.000 people. The simulated test distributions can be found as Figure 7 and Figure 8.

The distribution of $\hat{\mathcal{R}}_3$ exhibits a larger standard error, similar to the previous section. The means of both distributions are again approximately 1. However, in comparison to Section 9.1, the standard errors are almost ten times as big. This can be seen in Table 7.

Table 7

	Mean	Standard Error ($*10$)
$\hat{\mathcal{R}}_2$	0.99	0.59
$\hat{\mathcal{R}}_3$	0.98	1.93

This table shows the means and standard errors of the simulated test distributions (for Pareto) using the estimated $\hat{\alpha}$ (1.93) over the whole sample (*U.S. City Size*) and as minimum, $\omega = 10.8$ (minimum population of 50.000 people).

The relevant critical values are calculated according to the simulated test distributions. These can be found in Table 8.

Table 8

Significance levels	$F_{S,\hat{\mathcal{R}}_2}$	$F_{S,\hat{\mathcal{R}}_3}$
1%	0.884	0.704
2.5%	0.891	0.723
5%	0.909	0.754
10%	0.923	0.789

This table shows the critical values corresponding to their significance level from the simulated test distributions (for Pareto) of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ for U.S. City Size. S represents the significance level.

The $\hat{\mathcal{R}}_2$ from the full sample of 280 MSAs is **0.668**. The $\hat{\mathcal{R}}_3$ from the full sample is **0.368**. These values can be found in Table 2. Both values are smaller than all critical values in Table 8. Hence, the null hypothesis that U.S. City Size follows a Pareto distribution is rejected.

10 Testing for Weibull

In this Section, multiple tests are conducted to assess whether the Weibull distribution is a good fit. For all tests, the parameters have to be specified. These are estimated via the **fitdist** function by maximum likelihood. First, the results regarding the *Forbes Billionaires Evolution List* are discussed. After this, the results considering U.S. City Size are discussed.

10.1 Top Wealth

In this section, Top Wealth is considered.

10.1.1 Simulated test distribution

The following parameters are estimated: $\hat{\alpha} \approx 3.92$ and $\hat{\gamma} \approx 1.01$. 500 test statistics are simulated where 26395 (length of full *Forbes Billionaires Evolution List*) random generated Weibull numbers are used per calculation per test statistic. The distributions can be found below as Figure 3 and Figure 4.

The summary statistics of these test distributions are also stated below, in Table 9.

Table 9

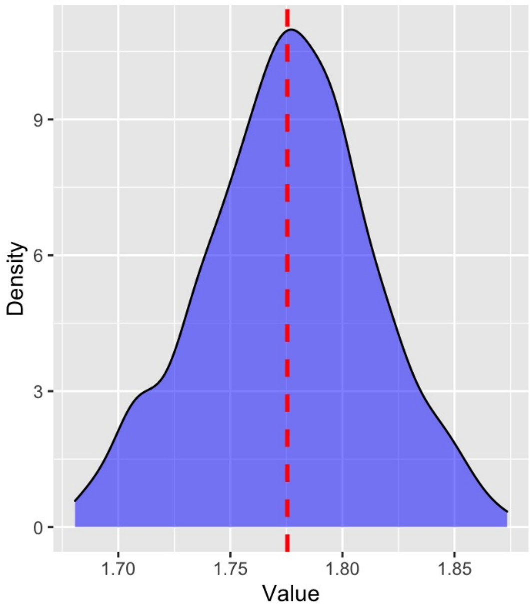
	Mean	Standard Error ($*10^2$)
$\hat{\mathcal{R}}_2$	1.78	3.71
$\hat{\mathcal{R}}_3$	0.67	2.36

This table shows the means and standard errors of the simulated test statistics (for Weibull) using the estimated $\hat{\alpha}$ (3.92) and the estimated $\hat{\gamma}$ (1.01) over the whole sample (*Forbes Billionaires Evolution List*) and as minimum, 1 million USD.

It is clear that the means of these two distributions differ from unity. If this table is compared to Table 5, one can observe that the standard errors in this case are larger.

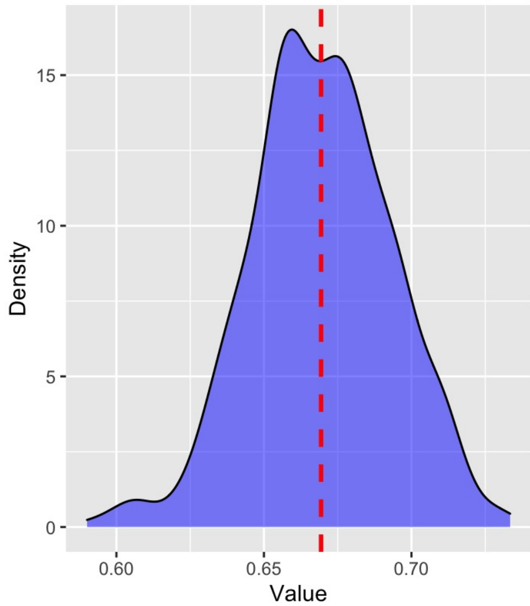
The $\hat{\mathcal{R}}_2$ from the full sample is **0.840**, this is why the null is rejected here. The $\hat{\mathcal{R}}_3$ on the other hand is **0.622**. It must be noted that the null is not rejected here on a significance level of 1%. One could argue that a significance level of 1% is very strict, which might increase the probability of a type 2 error. Also given the fact that the null is rejected for the distribution of

Figure 3



This figure shows the density of the simulated $\hat{\mathcal{R}}_2$ using Weibull random numbers the estimated $\hat{\alpha}$ (3.92) over the whole sample(*Forbes Billionaires Evolution List*) and the estimated $\hat{\gamma}$ (1.01) over the whole sample.

Figure 4



This figure shows the density of the simulated $\hat{\mathcal{R}}_3$ using Weibull random numbers the estimated $\hat{\alpha}$ (3.92) over the whole sample(*Forbes Billionaires Evolution List*) and the estimated $\hat{\gamma}$ (1.01) over the whole sample.

Table 10

Significance levels	$F_{S,\hat{\mathcal{R}}_2}$	$F_{S,\hat{\mathcal{R}}_3}$
1%	1.687	0.607
2.5%	1.704	0.623
5%	1.709	0.632
10%	1.727	0.640

This table shows the critical values corresponding to their significance level from the simulated test distributions (for Weibull) of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ for Top Wealth. S represents the significance level.

$\hat{\mathcal{R}}_2$, one could question this result. In most applications, researchers use a 5% or 10% significance level.

10.1.2 Statistical Tests

The tests are conducted per year for the *Forbes Billionaires Evolution List*. The fitted parameters can be seen in Table 21. The statistical tests provide the following results as stated in Table 11.

For the Anderson-Darling test, the p-values are significant for 2001-2013 on a 5% level, and for the Cramér-von Mises test this holds for 2001-2012. The results show that the Weibull distribution was a good fit but because of the increase of the number of billionaires, the distribution of Top Wealth is shifting to another distribution. In the beginning of this century, the Weibull distribution seemed to approximately model Top Wealth very well. Lately, for example in 2021, the Weibull distribution is rejected on a 5% significance level. This could be the result of the rising amount of billionaires. For example, in 2001 there were 335 billionaires listed, as in 2021 this amount was 2755. This could lead to another shape of the distribution. Table 21 shows that the shape parameter $\hat{\gamma}$ declines over the years. This could suggest that the distribution in 2021 is more spread out. The scale parameter $\hat{\alpha}$ does not differ much except for 2021. It is much larger which suggests that the billionaires have, on average, more money nowadays than back in the day.

Table 11

Year	AD Statistic		CvM Statistic	
2001	3.66	(0.26)	0.71	(0.24)
2002	3.84	(0.22)	0.76	(0.19)
2003	3.83	(0.22)	0.76	(0.20)
2004	4.05	(0.20)	0.81	(0.17)
2005	4.33	(0.18)	0.86	(0.15)
2006	4.55	(0.15)	0.91	(0.13)
2007	4.82	(0.13)	0.98	(0.10)
2008	5.11	(0.11)	1.04	(0.08)
2009	4.94	(0.11)	0.99	(0.09)
2010	5.46	(0.08)	1.10	(0.07)
2011	5.86	(0.06)	1.17	(0.05)
2012	5.83	(0.06)	1.15	(0.05)
2013	6.01	(0.05)	1.22	(0.04)
2014	6.34	(0.04)	1.26	(0.04)
2015	6.63	(0.03)	1.31	(0.03)
2016	6.56	(0.03)	1.30	(0.03)
2017	6.98	(0.02)	1.35	(0.02)
2018	6.84	(0.03)	1.34	(0.03)
2019	6.98	(0.02)	1.38	(0.02)
2020	7.09	(0.02)	1.38	(0.02)
2021	7.83	(0.01)	1.52	(0.01)

The two tests represent the Anderson-Darling test and the Cramer-von Mises test respectively where in all tests, the null hypothesis is: the data follows a Weibull distribution. The numbers between parentheses represent the p-values.

10.2 U.S. City Size

In this Section, the results for U.S. city size are discussed.

10.2.1 Simulated test distributions

For U.S. City Size, the full sample consists out of 280 MSAs. Because of the low amount of observations in the sub samples, only the full sample is used. When this data is fitted to a Weibull distribution, NaN values are produced. This could be a consequence of the large reach of the data. The smallest observation is 57.831 and the biggest is 2.119.9865. This is why the data is re-scaled. Essentially, because of scaling, the value of the scale parameter is reduced. The U.S. City Size data is re-scaled by dividing every number by the mean to reduce the reach of the data. Because the scaling only affects the scale parameter, it only affects the spread of the data but not its shape. Hence, if the scaled data fits a Weibull distribution, the original data does also fit a Weibull distribution.

The following parameters are estimated: $\hat{\alpha} \approx 0.77$ and $\hat{\gamma} \approx 0.74$. 500 test statistics are simulated where 280 (length of full U.S. City Size) random Weibull numbers are used per test statistic. The distributions can be found as Figure 9 and Figure 10.

The summary statistics of these simulated statistics are stated below, in Table 12. These

Table 12

	Mean	Standard Error ($*10^2$)
$\hat{\mathcal{R}}_2$	0.51	0.12
$\hat{\mathcal{R}}_3$	0.18	0.14

This table shows the means and standard errors of the simulated test statistics (for Weibull) using the estimated $\hat{\alpha}$ (0.77) and the estimated $\hat{\gamma}$ (0.74) over the whole sample(*U.S. City Size*) and as minimum, $\omega = 10.8$ (minimum population of 50.000 people).

two distributions exhibit extremely small standard errors in comparison to the other test distributions. This shows that these tests are very strict.

Table 13

Significance levels	$F_{S,\hat{\mathcal{R}}_2}$	$F_{S,\hat{\mathcal{R}}_3}$
1%	0.507	0.174
2.5%	0.508	0.175
5%	0.508	0.175
10%	0.508	0.176

This table shows the critical values corresponding to their significance level from the simulated test distributions (for Weibull) of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ for U.S. City Size. S represents the significance level.

The $\hat{\mathcal{R}}_2$ from the full sample of 280 MSAs is **0.505**. The $\hat{\mathcal{R}}_3$ from the full sample is **0.171**. These values are calculated with the re-scaled data. Both values are smaller than all critical values displayed in Table 13. Hence, the null hypothesis that U.S. City Size follows a Weibull distribution is rejected here. One must note that the standard errors are extremely low within this distribution. This makes the test very strict and maybe not fully trustworthy. This is why statistical tests are conducted in Section 10.2.2.

10.2.2 Statistical Tests

As already discussed in the previous section, the U.S. City Size data is re-scaled by dividing every observation by the mean. As a result, all observations are closer together and the reach of the data reduces significantly. Subsequently, if the data is fitted to a Weibull distribution, it provides the estimates which are displayed in Table 14. It can be observed that the parameters

Table 14

	Full sample	Top 135	Top 100	Top 50
$\hat{\alpha}$	0.77(0.07)	0.93(0.10)	0.95(0.06)	1.15(0.11)
$\hat{\gamma}$	0.75(0.03)	0.89(0.05)	0.97(0.11)	1.06(0.14)

This table shows the estimated parameters per sub sample of U.S. City Size for the Weibull distribution. $\hat{\alpha}$ is the scale parameter and $\hat{\gamma}$ is the shape parameter. The standard errors are between parentheses.

are increasing.

The statistical tests are conducted for the scaled sample. The results are stated in Table 15. Both the Anderson Darling and the Cramér-von Mises test provide significant results on a

Table 15

	Full Sample	Top 135	Top 100	Top 50
AD test statistic	3.52	3.02	2.67	2.18
AD p-value	(0.28)	(0.35)	(0.45)	(0.48)
CvM test statistic	0.70	0.58	0.47	0.43
CvM p-value	(0.25)	(0.31)	(0.42)	(0.40)

This concerns the U.S. City Size. The two tests represent the Anderson-Darling test and the Cramér-von Mises test respectively where in all tests, the null hypothesis is: the data follows a Weibull distribution.

10% significance level for all samples. This shows that the test in the previous section is not reliable for U.S. City Size. This could be the consequence of the low amount of observations in the sample. To conclude, for all samples, the null hypothesis that U.S. City Size follows a Weibull distribution is not rejected.

11 Testing for Generalized Pareto

Because of the definition of the Pickands-Balkema-De Haan theorem, U.S. City Size is not considered here. The full sample of the *Forbes Billionaires Evolution List* is used first. These values are already above the threshold 1. The exceedances are tested here. Because first the whole sample is used, the threshold is 1. Hence the exceedances are simply the *Forbes Evolution Billionaires List* minus the value of the threshold, which is 1.

Table 16

	Top Wealth full sample
AD test statistic ($*10^2$)	0.27
AD p-value	(<0.001)
CvM test statistic	7.91
CvM p-value	(<0.001)

The two tests represent the Anderson-Darling test and the Cramér-von Mises test respectively where in all tests, the null hypothesis is: the data follows a Generalized Pareto distribution.

All p-values are smaller than 0.001. Hence, the null hypothesis is rejected here. Because the theorem applies to extreme values above a high enough threshold T, the top decile of Top Wealth is also tested to follow a Generalized Pareto distribution. This corresponds to the top 10% of the *Forbes Billionaires Evolution List.* The top 10% corresponds to a minimum net worth of 7.5 billion USD. This means that the exceedances $Y_i = X_i - 7.5$ are tested here. The results are stated in the table below: The null is not rejected here for both tests on a 5% significance level. Hence, there is not enough evidence to suggest that the data significantly deviates from a Generalized Pareto distribution.

Table 17

	Top Wealth exceedances above $T = 7.5$
AD test statistic ($*10^3$)	0.55
AD p-value	(0.07)
CvM test statistic	1.64
CvM p-value	(0.21)

The two tests represent the Anderson-Darling test and the Cramér-von Mises test respectively where in all tests, the null hypothesis is: the data follows a Generalized Pareto distribution.

12 Conclusion

Many features of economic life are assumed to follow a Pareto distribution. This assumption is tested for various data. However, the Pareto distribution is actually a poor fit in many cases. This is why, in this research, such right-skewed distributions are tested against two alternative distributions. The Weibull distribution and the Generalized Pareto distribution (GPD) are considered. Next to this, the Pareto assumption is also tested using proposed statistics by Teulings & Toussaint (2023).

We test the assumption of Pareto by simulating test distributions using their proposed test statistics. All tests provide strong evidence against the Pareto distribution for both Top Wealth and U.S. City Size. However, multiple statistical tests like the Anderson-Darling test and the Cramér-von Mises test show that there is not enough evidence to reject the Weibull distribution for both samples of data. Specifically, Weibull is not rejected every year for Top Wealth from 2001 to 2012 and the p-values are still declining. This demonstrates that, at the beginning of this century, Weibull described Top Wealth well but the distribution of Top Wealth appears to be shifting toward another, unknown distribution. This could be the result of the rising amount of billionaires compared to twenty years ago. For U.S. City Size, the Weibull distribution is not rejected for the every sub-sample. However, for the test with the simulated distributions, U.S. City Size is rejected for the Weibull distribution. Although it must be stated that these tests are not the most reliable for Weibull because of the small standard errors of the distribution. This makes the test very strict. Finally, the Pickands-Balkema-De Haan theorem is used as motivation to test if Top Wealth follows a Generalized Pareto Distribution using the Anderson-Darling test and the Cramér-von Mises test. The full *Forbes Billionaires Evolution List* does not follow a Generalized Pareto Distribution, but the exceedances above the top decile of this list does.

In conclusion, both right-skewed distributions are rejected for the Pareto distribution. On the other hand, it is shown that the Weibull distribution can replace the Pareto assumptions in certain economic models. However, the distribution of Top Wealth appears to be shifting towards another unknown distribution in the last 10/15 years. This could be the result of the rising amount of billionaires. For example, in 2001 there were 335 listed billionaires while there were 2755 billionaires listed in 2021. Also, currently the average billionaire is wealthier. In 2021, the average net worth from billionaires was approximately 3.6 billion USD while in 2021 this number already was approximately 4.8. This could potentially be the reason for this shift.

Further research could try to find the reason for this apparent shift of distribution. One could also try to find a distribution which describes Top Wealth currently.

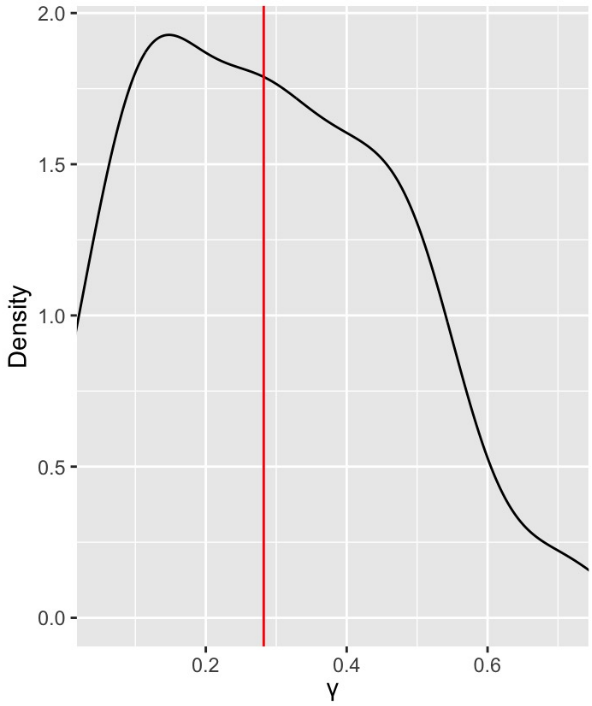
A Appendix

Table 18

(Sub-)Region	$\hat{\mathcal{R}}_2$	$\hat{\mathcal{R}}_3$	$\overline{w}$	N
North America	0.826	0.614	0.988	440.5
US	0.827	0.616	0.995	413
Canada	0.788	0.541	0.891	27.5
Europe	0.773	0.507	1.044	245
Germany	0.723	0.43	1.125	67.9
British Islands	0.761	0.475	0.922	38.1
Scandinavia	0.771	0.481	1.087	30
France	0.797	0.53	1.286	25.1
Alps	0.663	0.347	1.064	23.2
Italy	0.791	0.532	1.023	22.9
China	0.933	0.791	0.794	185.5
East Asia	0.84	0.605	0.817	129.8
Southeast Asia	0.79	0.515	0.938	29.6
Asian Islands	0.806	0.56	0.761	37.8
South Korea	0.952	0.81	0.662	18.2
Japan	0.847	0.618	0.881	22.7
Australia	0.868	0.649	0.793	18.2
India	0.852	0.636	0.941	54.7
Central Eurasia	0.868	0.621	0.98	75.6
Russia	0.867	0.614	1.009	66.8
South America	0.832	0.618	1.002	56.8
Brazil	0.817	0.594	0.903	29.2
Middle East	0.879	0.674	0.745	58.8
Israel + Turkey	0.882	0.644	0.599	34.7
Rest of World	0.777	0.486	0.962	10.2
World	0.84	0.622	0.939	1256

This table presents the summary statistics of all (sub-)regions. The numbers are all averages over 2001-2021. N is the average amount of billionaires per (sub-)region, $\overline{w}$ is the mean log wealth and $\hat{\mathcal{R}}_2$ & $\hat{\mathcal{R}}_3$ are the test statistics for Pareto.

Figure 5



This figure shows the distribution of $\hat{\gamma}$ for the 75 observations where the minimum amount of billionaires is 64. The red vertical line is the $\gamma_{median} \approx 0.28$.

Table 19

	Mean Wealth			$\hat{\alpha}$	
Sub-Region	Data	Weibull	Pareto	Weibull	Pareto
United States	5.29	5.64	∞	1.48	1.16
Canada	3.23	3.25	5.96	1.02	0.832
Germany	4.7	5.76	∞	1.5	1.18
British Isles	3.83	4.36	∞	1.26	1.01
Scandinavia	3.51	4.15	227	1.22	0.996
France	7.44	8.46	∞	1.83	1.37
Alpine countries	3.8	4.8	∞	1.34	1.1
Italy	3.96	4.39	∞	1.26	1.02
China	3.31	3.13	5	0.985	0.8
South-East Asia	3.33	3.57	9.43	1.09	0.894
Asia Island	3.17	3.26	6.18	1.02	0.838
South Korea	2.88	2.8	3.8	0.893	0.737
Japan	3.95	3.88	12.4	1.16	0.919
Australia	2.74	2.77	3.86	0.886	0.741
India	3.7	3.98	24.1	1.18	0.959
Russia	4.05	4.13	21.3	1.21	0.953
Brazil	4.2	4.51	∞	1.29	1.03
Israel + Turkey	2.17	2.25	2.64	0.716	0.622

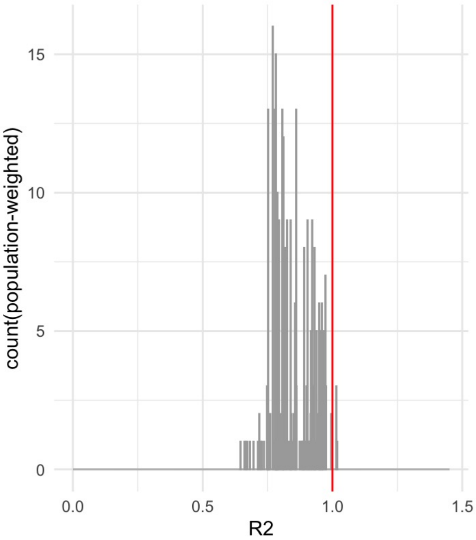
For the estimation of $\hat{\alpha}$ for Weibull, I use $\hat{\gamma}_{median} = 0.257$(the $\hat{\gamma}_{median}$ which the main paper (Teulings & Toussaint, 2023) also uses so I can get their values). This table displays the predictions of mean wealth via Pareto and Weibull.

Table 20

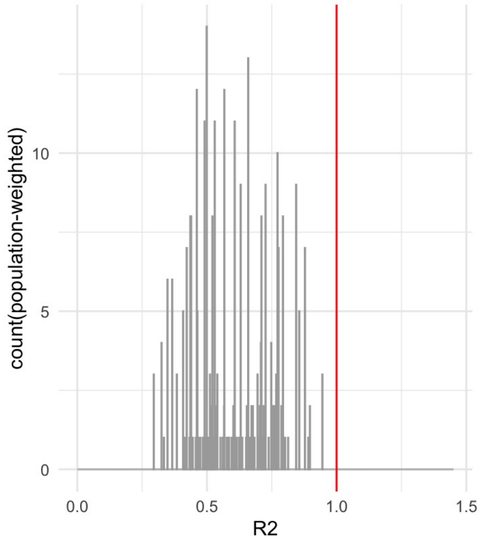
Dependent variables:	$\hat{\mathcal{R}}_2$	$\hat{\mathcal{R}}_3$	
Model:	(1)	(2)	(3)
Variables			
Constant	0.764***	0.503***	-0.754***
	(0.011)	(0.009)	(0.040)
$ln(\hat{\alpha})$	-0.100***	-0.185***	
	(0.014)	(0.012)	
$ln(\hat{\gamma})$	-0.069***	-0.114***	
	(0.006)	(0.005)	
$\hat{\mathcal{R}}_2$			1.642***
			(0.047)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics			
Observations	67	67	67
R^2	0.943	0.987	0.949
Adjusted R^2	0.941	0.986	0.948
RMSE	0.018	0.014	0.03

Signif. codes: 0 '***' 0.001 '**' 0.01 '*'. All regressions are estimated with 67 observations. These are the observations with a minimum of 64 billionaires, where the top and bottom 5% is left out. RMSE stands for root mean squared error.

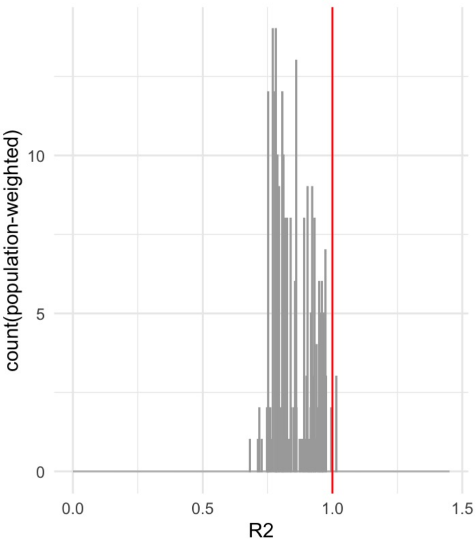
Figure 6



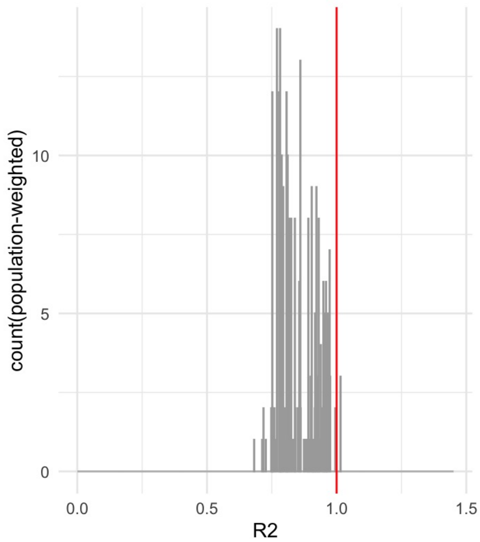
(a) $\hat{\mathcal{R}}_2$, Full Sample



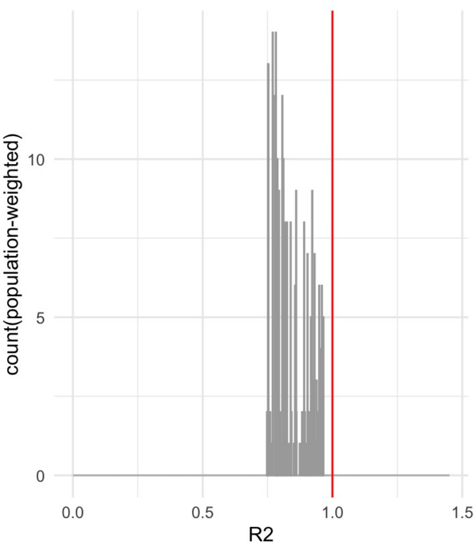
(b) $\hat{\mathcal{R}}_3$, Full Sample



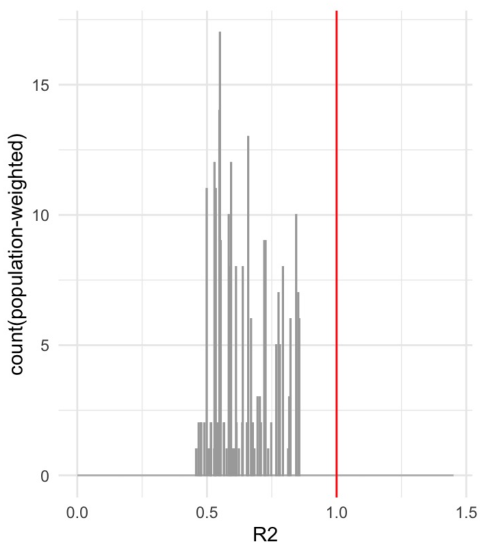
(c) $\hat{\mathcal{R}}_2$, N $\geq$ 64



(d) $\hat{\mathcal{R}}_3$, N $\geq$ 64

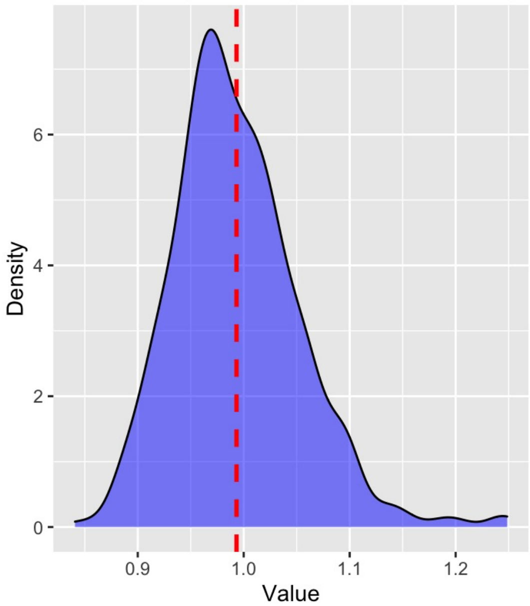


(e) $\hat{\mathcal{R}}_2$, N $\geq$ 64, trimmed



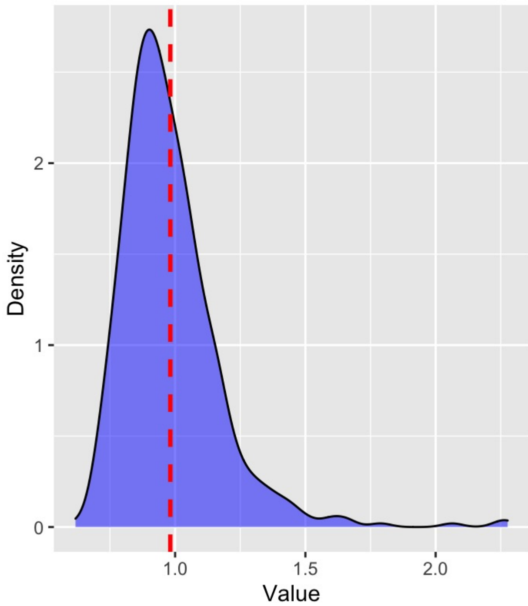
(f) $\hat{\mathcal{R}}_3$, N $\geq$ 64, trimmed

Figure 7



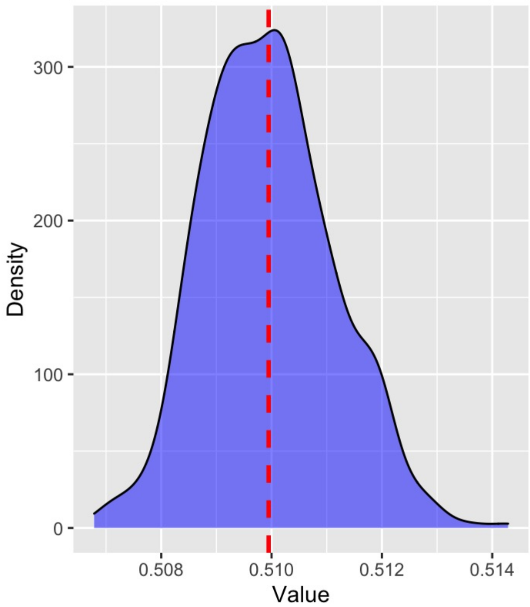
This Figure shows the density of the simulated $\hat{\mathcal{R}}_2$ (for Pareto) using the estimated $\hat{\alpha}$ (1.93) over the whole sample (*U.S. City Size*) and as minimum, 50.000 people, which corresponds to $\omega = 10.8$.

Figure 8



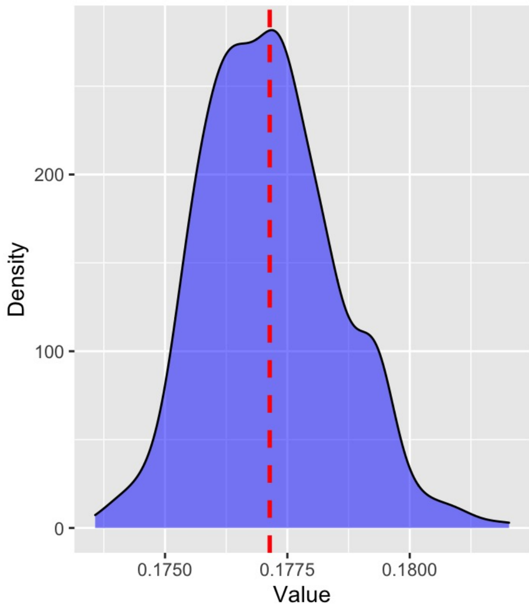
This Figure shows the density of the simulated $\hat{\mathcal{R}}_3$ (for Pareto) using the estimated $\hat{\alpha}$ (1.93) over the whole sample (*U.S. City Size*) and as minimum, 50.000 people, which corresponds to $\omega = 10.8$.

Figure 9



This Figure shows the density of the simulated $\hat{\mathcal{R}}_2$ using Weibull random numbers the estimated $\hat{\alpha}$ (0.77) over the whole sample (*U.S. City Size*) and the estimated $\hat{\gamma}$ (0.74) over the whole sample.

Figure 10



This Figure shows the density of the simulated $\hat{\mathcal{R}}_3$ using Weibull random numbers the estimated $\hat{\alpha}$ (0.77) over the whole sample (*U.S. City Size*) and the estimated $\hat{\gamma}$ (0.74) over the whole sample.

Table 21

Year	$\hat{\alpha}$	$\hat{\gamma}$
2001	3.74	1.06
2002	3.44	1.06
2003	3.27	1.13
2004	3.59	1.10
2005	3.60	1.11
2006	3.74	1.12
2007	4.12	1.08
2008	4.44	1.05
2009	3.28	1.13
2010	3.66	1.08
2011	3.82	1.05
2012	3.84	1.06
2013	3.90	1.04
2014	3.98	1.03
2015	3.90	1.02
2016	3.66	1.04
2017	3.79	1.02
2018	4.08	0.99
2019	3.99	0.99
2020	3.82	0.99
2021	4.49	0.92

This table shows the fitted parameters per year for the Weibull distribution considering Top Wealth. $\hat{\alpha}$ is the scale parameter and $\hat{\gamma}$ represents the shape parameter.

References

Anderson, T. W. & Darling, D. A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 193–212.

Balkema, A. A. & De Haan, L. (1974). *Residual life time at great age, the annals of probability.* &.

Blanchet, T., Fournier, J. & Piketty, T. (2022). Generalized pareto curves: theory and applications. *Review of Income and Wealth*, *68*(1), 263–288.

Braun, H. (1980). A simple method for testing goodness of fit in the presence of nuisance parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *42*(1), 53–63.

Cramér, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, *1928*(1), 13–74.

CSöRgő, S. & Faraway, J. J. (1996). The exact and asymptotic distributions of cramér-von mises statistics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 221–234.

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

Gabaix, X. (1999). Zipf's law and the growth of cities. *American Economic Review*, *89*(2), 129–132.

Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, *1*(1), 255–294.

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, *30*(1), 185–206.

Jones, C. I. (2015). Pareto and piketty: The macroeconomics of top income and wealth inequality. *Journal of Economic Perspectives*, *29*(1), 29–46.

Krugman, P. (1996). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International economies*, *10*(4), 399–418.

Luttmer, E. G. (2011). On the mechanics of firm growth. *The Review of Economic Studies*, *78*(3), 1042–1068.

Marsaglia, G. & Marsaglia, J. (2004). Evaluating the anderson-darling distribution. *Journal of statistical software*, *9*, 1–5.

Mises, R. (1931). *Wahrscheinlichkeitsrechnung und ihre anwendung in der statistik und theoretischen physik.* Mary S. Rosenberg.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 119–131.

Teulings, C. N. & Toussaint, S. (2023). Top wealth is distributed weibull, not pareto.

Vermeulen, P. (2018). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, *64*(2), 357–387.

Wingo, D. R. (1989). The left-truncated weibull distribution: theory and computation. *Statistical papers*, *30*(1), 39–48.