Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operations Research

Who wants to be a Billionaire? - Modelling Extreme Wealth with the Generalized Pareto Distribution

Lars Jelmer van Lingen (613030)

Supervisor:	P. Wan
Second Assessor:	J. Klooster
Date final version:	1st July 2024

Abstract

This research examines the quality of the Generalized Pareto Distribution (GPD) compared to traditional extreme-value models in predicting mean billionaire wealth. We use data from the *Forbes List of Billionaires* from 2000 to 2021. A previous work from Teulings and Toussaint (2023) implements the Pareto distribution and shows that the Weibull distribution is a better model. In our research, we additionally explore the GPD. Utilising maximum likelihood estimation to estimate parameters, and the Kolmogorov-Smirnov test to assess the goodness-of-fit of the three models. We find the Weibull distribution to give the best predictions, with the GPD giving a comparably strong predictive accuracy, but the fit of the GPD to be better than that of the other two models. The results strongly support the adoption of the GPD for analyzing extreme wealth. This analysis shows that by integrating the GPD, we can enhance the accuracy of mean wealth predictions in economic studies. Further research could explore the integration of economic factors within these distributions to understand their impact on wealth prediction accuracy.

1 Introduction

Heavy-tailed distributions are of great interest in many economic aspects, such as CEO salaries (Gabaix and Landier, 2008), income and wealth (Atkinson et al., 2011; Vermeulen, 2018). For years, the typical instrument to analyze these right-skewed distributions has been the Pareto distribution. Blanchet et al. (2022) shows us that predictions made under the Pareto assumption for for example mean wealth and income in the upper tail are not accurate.

Under the null hypothesis of a Pareto distribution, the maximum likelihood estimator of the Pareto coefficient is the most efficient. Furthermore, it is easily calculated, by simply setting it equal to the mean of log wealth (Hill, 1975). For this reason, in this paper, we opt for the maximum likelihood estimator as opposed to applying the log-rank regression, of which Rosen and Resnick (1980) shows an application in the distribution of city sizes.

We apply tests similar to those of the work of Teulings and Toussaint (2023), on the second and third moments of our dataset, which we divide into 18 subregions for each year from 2001 up until 2021. For each of these observations, the null of a Pareto distribution is rejected, which is in line with the aforementioned paper. The test statistics are consistently statistically different from unity, which is what one would expect for the second and third moments under a Pareto distribution. The data seem to be less skewed to the right, which suggests that there might be a better alternative to the Pareto distribution.

The first alternative distribution that we investigate, is the Left Truncated-Weibull model. The log of a Weibull distribution is the Gompertz distribution. We make use of the Truncated Gompertz model, of which from now on the word "Truncated" is left out. The crucial difference between the log of the Pareto variable and the log of a Weibull variable lies in their hazard rates (Chandraa and Abdullaha, 2022). Where the log of a Pareto distributed variable is Exponentially distributed, and hence has a constant hazard rate, the Gompertz distribution, which is the log of the Weibull distribution, has an increasing hazard rate above some lower bound. This fact allows for a potential better fit to the data.

The parameters of the Weibull distribution are calculated by making use of maximum likelihood estimation, and we find the estimate of the parameter γ, the increasing hazard above the lower bound, to be roughly 0.282 Mean wealth is then predicted using this estimate of γ as common for each sub-region. We observe that the Weibull model provides much better predictions, thus proving to be a better fit on the dataset than the Pareto model.

This research is furthermore extended by the addition of the Generalized Pareto Distribution (GPD), as well as testing by making use of the Kolmogorov-Smirnov test. Hanafy et al. (2020) have found the GPD model to perform better in tail modelling, which can also be backed by the Pickands-Balkema-De Haan Theorem (Balkema and De Haan, 1974; Pickands III, 1975). In this research, we find the GPD model to have results similar to those of the Weibull model with minor differences in some instances, thus outperforming the standard Pareto model.

The Kolmogorov-Smirnov test is implemented, to give a more formal comparison between the models, as opposed to the methods of testing used in Teulings and Toussaint (2023), which rely more heavily on intuition instead of sound statistics. The GPD is the model with the best fit of the three having a p-value of 0.345, thus beating the Pareto and Weibull distribution in terms of goodness-of-fit on the dataset.

The contributions of this paper are twofold. First, we compare the models from the original research to an additional distribution, the GPD. Second, we assess the performance of these models in a more formal manner than that of the original work by Teulings and Toussaint (2023), through the Kolmogorov-Smirnov test.

The remainder of this paper is structured as follows. Section 2 covers literature on extreme value distributions, especially focused on the GPD, as well as literature on the Kolmogorov-Smirnov test. Subsequently, Section 3 describes the statistical models used as well as the techniques used to estimate parameters and test these models. This is followed by Section 4, which describes the dataset that is used in this paper. Thereafter, Section 5 presents the results of the replication of Teulings and Toussaint (2023), combined with the results of our contributions. Finally, Section 6 provides a summary of this paper.

2 Literature Review

Extreme Value Theory (EVT) has been a cornerstone in the statistical analysis of rare events and tail risks for quite a while. EVT plays a big role in risk management, especially in the financial industry (Embrechts et al., 1999; McNeil, 1999; Diebold et al., 2000). Studies by Balkema and De Haan (1974), Gumbel (1958), and Pickands III (1975) are foundational in this field. There are numerous methods of estimating the tail index, many of which are explained in Fedotenkov (2020). In our research, we focus on log-likelihood estimation in particular, as also discussed in the works of Hill (1975) and Dekkers et al. (1989).

We relate these heavy-tailed distributions to the real world, by investigating their applications in billionaire wealth. Various studies, including those of Atkinson et al. (2011), Vermeulen (2016), and Vermeulen (2018) have implemented the Pareto distribution in the field of income and wealth. Additionally, studies like that of Huisman et al. (2001) apply tail-distributions in the cross-section of stock returns, providing insight into the tail behaviour of financial assets, while studies like those of Autor et al. (2020) and Luttmer (2011) utilise these distributions to examine the size of firms, linking economic theory with empirical data effectively.

Works like the one of Kopczuk (2015) discuss how billionaire wealth has evolved over the years, and Bagchi and Svejnar (2013) discusses the impact of income inequality on economic growth, with a particular focus on billionaire wealth. In the past two decades, we have seen a major increase in the number of billionaires all around the world. This can be explained by technological advancements as well as new emerging markets, especially in regions like Asia.

Works of Capehart (2014), Blanchet et al. (2022) and Teulings and Toussaint (2023) question the Pareto assumption in income and wealth analysis, proposing alternative methods like the Weibull distribution. This underscores a broader scepticism within the field regarding traditional assumptions and encourages the exploration of other distributions. Teulings and Toussaint (2023) implements an approach similar to that of Capehart (2014), in estimating measurement error of top wealth.

The Weibull distribution is utilised in various applications, such as in wind energy (Aljeddani and Mohammed, 2023) and in the modelling of failure data (Zhang and Xie, 2007). Teulings and Toussaint (2023) proposes the Weibull distribution as an excellent alternative to the Pareto distribution in modelling top wealth, whereas Campolieti (2018) challenges the statement that top wealth would be distributed Weibull, arguing that the power law distribution provides a better fit. We make use of Maximum Likelihood Estimation in estimating the parameters of the Weibull distribution, following Wingo (1989).

Gomez (2023) and Blanchet (2022) both focus on billionaire wealth, coming close to the research of Teulings and Toussaint (2023). Although the findings of the latter are interesting, we could argue that the research is mainly based on intuitive tests and standard distributions. The works of Balkema and De Haan (1974) and Pickands III (1975) motivate us to explore the Generalized Pareto Distribution (GPD) as an alternative. The GPD, particularly effective in modelling exceedances over a threshold (Giles et al., 2016; Chu et al., 2019), offers a robust tool for analyzing extremes in economic data, which is of great interest in our research.

Applications of the GPD can be found in modelling extreme rainfall (Iyamuremye et al., 2019), extreme wind speeds (Holmes and Moriarty, 1999) and extreme temperatures (Alonso et al., 2014). These applications demonstrate the GPD's robustness in capturing the tail behaviour of extreme weather events. Furthermore, the GPD is used in the modelling of non-life insurance claims (Dzidzornu and Minkah, 2021) and in applications in insurances and finance (Villa, 2017). In these applications, the usage of a GPD model increases precision and reliability. Lastly, the GPD can also be found used in actuarial sciences, as shown in (Brazauskas and Kleefeld, 2009).

Parameters of the GPD are estimated using maximum likelihood estimation. However, this could raise some issues with this method, since there is no closed-form solution for the first derivative of the GPD log-likelihood function (Purwani and Ibrahim, 2024). Other papers therefore propose different methods of estimation, including simple fixed-point iterations (Purwani and Ibrahim, 2024) or shrinkage methods (Pels et al., 2023). These methods make for direct and efficient estimates of the parameters of the GPD model.

Čabla and Habarta (2019) is similar to our paper, because, in their research, the GPD framework is also applied to the wealth of the richest people in the world according to estimates from the CEOWORLD magazine. The finding that the wealth distribution of the world's richest is heavy-tailed and that it closely follows a GPD, confirms its applicability in modeling extreme wealth.

Works such as (Massey Jr, 1951) and (Berger and Zhou, 2014) discuss the Kolmogorov-Smirnov (KS) test, and why it is so powerful in assessing the goodness-of-fit of distributions. Massey Jr (1951) highlights the robustness of the KS-test in measuring goodness-of-fit, emphasizing the non-parametric nature of the test. Berger and Zhou (2014) discusses the asymptotic properties of the KS test, demonstrating its effectiveness in large samples. These discussions underline the critical role of the KS test in validating the fit of statistical models.

3 Methodology

This section presents the theory behind the right-skewed distributions used in this research, as well as descriptions of the methods used to obtain parameter estimates and how to compare the different models. All programming is done in R (2023.03.1+446), on a 2020 MacBook Air with a M1 chip.

Since part of this research involves the replication of Teulings and Toussaint (2023), the structure of this section is similar to that of the original research when it comes to the description of the models used and how the testing is conducted. Of course, there are still differences, in the addition of the Generalized Pareto Distribution (GPD) and the usage of a Kolmogorov-Smirnov test to assess the quality of the models used.

3.1 The Pareto Framework

Consider $\underline{X} \geq \Omega$ a random variable, with $\Omega > 0$ a known parameter. In this research, we take $\underline{X}$ to be an individual's net worth (in billions) and Ω as a billion USD. If $\underline{X}$ is Pareto distributed, we obtain the following complementary distribution function and moments:

$$\mathrm{P}[\underline{X} \geq X | \underline{X} \geq \Omega] = (X/\Omega)^{-1/\alpha}, \tag{1}$$

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

with α^{-1} the Pareto coefficient, which indicates the heaviness of the tail. In this paper, we work with α since its maximum likelihood estimator is unbiased whereas that of α^{-1} is not.

When we look at Equation (2), we see that the moments of a Pareto distribution $\mathrm{E}[\underline{X}^k]$ do not exist for $k \geq \alpha^{-1}$. For this reason, we choose to continue working with the log transformation of $\underline{X}$. Define $\underline{W} := \underline{X}/\Omega$, and $\underline{w} := \ln \underline{W} = \underline{x} - \omega$, where $\underline{x} := \ln \underline{X}$ and $\omega := \ln \Omega$.

The log-transformation of a Pareto-distributed random variable is exponentially distributed. This gives us the following complementary distribution functions and moments:

$$\mathrm{P}[\underline{w} \geq w] = e^{-w/\alpha}, \tag{3}$$

$$\mathrm{E}[\underline{w}^k] = \alpha^k \Gamma(k+1) = \alpha^k k!, \quad k > 0, \tag{4}$$

where $\Gamma(.)$ is the Gamma function, and where we assume k to be integer. This distribution is suitable for testing since now a moment exists for all k.

Pareto predictions of mean wealth are calculated by using Equation (2), setting k equal to 1. This prediction of actual mean wealth is then compared to those of the other distributions, which are explained in the upcoming Sections. For the Pareto distribution, it is assumed that observations are independent and identically distributed, which might be a wrong assumption considering consequent years often contain the same billionaires.

3.2 The Truncated Weibull Framework

An alternative to the Pareto distribution is the (Truncated) Weibull distribution. Where the log distribution of a Pareto-distributed variable is exponentially distributed, the log distribution of a Weibull-distributed variable is Gompertz distributed. For convenience, we choose to work with the Gompertz distribution for $\underline{w}$ rather than the Truncated Weibull distribution for $\underline{W}$.

The complementary distribution function of the Gompertz distribution is:

$$P[\underline{w} \geq w] = \exp\left(-\frac{e^{\gamma w} - 1}{\alpha\gamma}\right), \tag{5}$$

with γ a parameter, and moments:

$$E[\underline{w}^k] = \gamma^{-k} h(\alpha\gamma, k), \tag{6}$$

where

$$h(\alpha\gamma, k) := k e^{(\alpha\gamma)^{-1}} \int_0^{\inf} q^{k-1} exp(-(\alpha\gamma)^{-1} e^{-q}) \mathrm{d}q,$$

and where $q := \gamma w$.

The main difference between the exponential and Gompertz distribution is that the hazard function of an exponential distribution remains constant whereas that of a Gompertz distribution is increasing (Chandraa and Abdullaha, 2022). Therefore, the Gompertz distribution has a thinner tail than the exponential distribution.

To compare the Truncated Weibull model against the other models used, we again calculate the expected mean wealth. We do this in the following manner:

$$E[\underline{W}] = (\alpha\gamma)^{1/\gamma} e^{(\alpha\gamma)^{-1}} \Gamma(1 + 1/\gamma, (\alpha\gamma)^{-1}), \tag{7}$$

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

When we compare this to the calculation of the mean of the Pareto distribution in Equation (2), we observe a crucial difference. Where the expected value of the Pareto distribution is only finite for certain parameter estimates ($\alpha < 1$), the expected value of the Truncated Weibull moment can be calculated for all parameter estimates. Once again, data are assumed to be independent of this distribution, which might be a wrong assumption.

3.3 The GPD Framework

A problem with the aforementioned models is that they often lack flexibility and fit, especially in the upper tail of distributions. Therefore, we propose the Generalized Pareto Distribution (GPD) as a suitable alternative. The usage of this distribution is motivated by the Pickands-Balkema-De Haan Theorem, which states that the distribution of exceedances above a certain threshold can be approximated by the usage of a GPD (Balkema and De Haan, 1974; Pickands III, 1975).

Unlike the standard Pareto distribution, which is restricted by its shape and scale parameter, the GPD introduces an additional location parameter, making for a better fit for a broader range of datasets. The GPD is used for modelling excesses over a certain threshold. Furthermore, the GPD can model a variety of tail behaviours because of its shape parameter, making it superior to the Weibull model.

The complementary distribution function of the GPD is as follows:

$$P[\underline{W} \geq W] = (1 + \alpha\frac{W - \mu}{\sigma})^{-1/\alpha}. \tag{8}$$

In this function, α is the shape parameter which influences the tail behaviour, σ is the scale parameter, controlling the spread of the data, and μ is the location parameter. To calculate the expected value of mean wealth, we use the following formula:

$$\mathrm{E}[\underline{W}] = \mu + \frac{\sigma}{1-\alpha}, \quad \alpha < 1. \tag{9}$$

A problem arises when we want to calculate higher moments of the GPD function since it does not exist for all integer values of k. For this reason, combined with the fact that the log of the GPD is not another well-known distribution unlike in the case of the Pareto and Weibull model, we furthermore implement an additional method of testing than in the original paper of Teulings and Toussaint (2023), which is explained in the subsequent. Finally, for this distribution, we once more assume independence of the observations.

3.4 Estimation and Testing

After all models are constructed, the parameters of these models are estimated.

We start by estimating the parameters of the Pareto distribution. Hill (1975) shows us that the maximum likelihood estimator of α is simply the mean log wealth, $\overline{w}$. This estimator is also efficient and unbiased if the model is correct. However, it performs poorly if the data are not Pareto-distributed, especially when α is close to unity.

We use Equation (4) to derive the following simple test statistic:

$$\mathcal{R}_k := \frac{\mathrm{E}[\underline{w}^k]}{k!\mathrm{E}[\underline{w}]^k}, \quad \hat{\mathcal{R}}_k := \frac{\overline{\underline{w}^k}}{k!\overline{\underline{w}}^k}. \tag{10}$$

Under the null of a Pareto distribution for $\underline{W}$, we expect $\underline{w}$ to be distributed exponentially. In that case, it should hold for all integer $k \geq 1$ that:

$$\mathcal{R}_k = 1.$$

For this reason, to test for Paretianity, we investigate if the theoretical quantities are significantly different from one by regressing them on a constant.

The second model for which we estimate its parameters is the Weibull model. When wealth $\underline{W}$ is distributed Weibull, its log $\underline{w}$ is Gompertz distributed. We know that the hazard rate of this function increases with γ for every unit increase in α, as aforementioned in Section 3.2. However, the value of the parameter α does not depend on this γ, but rather only on the lower bound of our dataset, which is in our case equal to one billion USD. For this reason, we opt for a common γ parameter, but with α differing depending on the sub-region/time observation.

We again make use of maximum likelihood estimation in estimating the parameters α and γ by using the density function of w. We then obtain the following expressions:

$$\hat{\alpha} = \gamma^{-1}(\overline{e^{\gamma w}} - 1), \tag{11}$$

$$\hat{\gamma} = \frac{\overline{e^{\hat{\gamma} w}} - 1}{e^{\hat{\gamma} w}} (\hat{\gamma}^{-1} + w). \tag{12}$$

The derivation of these expressions can be found in Appendix B. We start by estimating a common $\hat{\gamma}$ for all sub-region/year observations, and then we use this common $\hat{\gamma}$ to determine the maximum likelihood estimates of α in each sub-region/year observation.

We test if our data are Weibull-distributed by regressing the same test statistic from Equation (10). The expected asymptotic value for the Weibull distribution is:

$$\mathcal{R}_k(\alpha\gamma) = \frac{\int_0^\infty q^{k-1} \exp\left(-(\alpha\gamma)^{-1} \mathrm{e}^{-q}\right) \mathrm{d}q}{(k-1)! \mathrm{e}^{(\alpha\gamma)^{k-1}} \mathrm{Ei}\left((\alpha\gamma)^{-1}\right)^k}, \tag{13}$$

with Ei the exponential integral. From this equation, we see that $\mathcal{R}_k(\alpha\gamma)$ is a declining function of $\alpha\gamma$, or of $\ln\alpha + \ln\gamma$. For this reason, we regress the theoretical quantities from Equation (10) on a constant, $\ln\alpha$ and $\ln\gamma$. Under the null of a Weibull distribution, we would expect the sign of these to both be negative and also roughly equal.

The third model for which we estimate its parameters is the GPD. As we see in Section 3.3, this distribution contains three parameters: shape α, location μ and scale σ. We set the location parameter μ equal to one billion USD since that is the lower bound of our research. We then obtain parameter estimates for α and σ by using maximum likelihood estimation.

Additionally, we apply the Kolmogorov-Smirnov (KS) test to the different models to make for a more statistically sound test for goodness-of-fit. The KS is a non-parametric test to determine if a sample comes from a population with a specific distribution. This can in our case either be from the Pareto, Weibull or GPD model.

We calculate the empirical CDFs for each of the three distributions based on the dataset and then perform the KS test by computing the maximum distance between the empirical CDF and the theoretical CDF. The test statistic is thus calculated as follows:

$$D = \sup|F_n(\underline{W}) - F(\underline{W})|, \tag{14}$$

in which $F_n(\underline{W})$ is the empirical distribution function and $F(\underline{W})$ is the theoretical CDF. Under the null hypothesis that the sample is drawn from the theoretical distribution, the KS test statistic follows the Kolmogorov distribution, of which the CDF is given by:

$$\mathrm{P}[D \leq d] = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1} e^{-2k^2 d^2}. \tag{15}$$

In finite samples, this is only an approximation of the distribution of D. In larger samples, $\sqrt{N}D$ converges to the Kolmogorov distribution in the limit. We make use of Equation (15) to calculate our p-values.

4 Data

This analysis is conducted on the *Forbes List of Billionaires*[1] dataset from 2001 until 2021. This dataset contains variables like the names of the billionaires, their wealth (net worth), age, country of citizenship and country of origin. In this paper, we are interested in the country of citizenship of these billionaires.

[1]The exact dataset used in this report can be downloaded from Kaggle through `https://www.kaggle.com/datasets/guillemservera/forbes-billionaires-1997-2023`

The countries of citizenship are firstly aggregated in regions and then separated into different sub-regions. These sub-regions are constructed if the regions can be divided into coherent geographical units, containing few countries, having at least 40 billionaires in the year 2019.

Table 1 on the next page presents some summary statistics for these regions and sub-regions. These summary statistics are the average numbers of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$, as well as the average log wealth $\overline{w}$ and the average number of billionaires N.

Table 1: Region Classification & Descriptive Statistics

Region Classification		Statistics (Average 2001–2021)			
(Sub-)Region	Area	$\hat{\mathcal{R}}_2$	$\hat{\mathcal{R}}_3$	$\overline{w}$	N
North America	U.S. and Canada	**0.826**	**0.614**	**0.988**	**440.5**
- U.S.		0.827	0.616	0.995	413.0
- Canada		0.788	0.541	0.891	27.5
Europe	excl. former USSR but incl. Baltics	**0.775**	**0.510**	**1.042**	**247.7**
- Germany		0.723	0.430	1.125	67.9
- British Islands	U.K. and Ireland	0.761	0.475	0.922	38.1
- Scandinavia	Sweden, Denmark, Norway and Finland	0.771	0.481	1.087	30.0
- France	France and Monaco	0.797	0.530	1.286	25.1
- Alps	Switzerland, Austria and Liechtenstein	0.663	0.347	1.064	23.2
- Italy		0.791	0.532	1.022	22.9
China	incl. Hong Kong, but excl. Taiwan	**0.933**	**0.792**	**0.793**	**185.9**
East Asia	Asia East of India and South-East of China; incl. Australia	**0.841**	**0.606**	**0.816**	**129.9**
- Southeast Asia	Thailand, Malaysia and Singapore	0.790	0.515	0.938	29.6
- Asian Islands	Taiwan, Philippines and Indonesia	0.806	0.560	0.761	37.8
- South Korea		0.952	0.810	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	former USSR exc. Baltics	**0.868**	**0.621**	**0.980**	**75.6**
- Russia		0.867	0.614	1.009	66.8
South America	incl. middle America and Mexico	**0.832**	**0.618**	**1.002**	**56.8**
- Brazil		0.817	0.594	0.903	29.2
Middle East	Middle East incl. Turkey and Egypt excl. Iran	**0.868**	**0.659**	**0.740**	**56.1**
- Israel and Turkey		0.882	0.644	0.599	34.7
Rest of World	mainly Africa excl. Egypt, incl. Iran, Afghanistan, Pakistan, Bangladesh	**0.829**	**0.583**	**0.920**	**5.7**
World		**0.840**	**0.622**	**0.939**	**1257.0**

Notes: This Table presents summary statistics for different regions as well as for sub-regions belonging to these regions. Some billionaires do not fall into a certain sub-region, while they do fall into a region, therefore the sum of N, the average count of billionaires per year, of particular sub-regions do not need to sum up to the N of their respective region. The regions *China* and *India* are also considered to be sub-regions. Furthermore, $\overline{w}$ is the average mean log wealth per year, with $\hat{\mathcal{R}}_2$ its normalized variance, and $\hat{\mathcal{R}}_3$ its normalized skewness, which are calculated using Equation (10).

5 Results

In this section, we discuss the empirical findings of the models. We mainly follow Teulings and Toussaint (2023) in the order and presentation of our empirical results, together with some additional findings. The different distributions are discussed in the same order as they are presented in Section 3. That is, the first model to be discussed is the Pareto distribution, followed by the Weibull model and the General Pareto Distribution (GPD) respectively. In between each model, their accuracy is assessed on the data set.

5.1 Testing the Pareto Assumption

We start by investigating the standard Pareto model. There are a total of 18 subregions over a timespan of 21 years, providing us with a total of 18x21 = 378 observations. We construct density plots of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$, which are depicted in Figure 1. The plots of the full sample are presented in Figures 1a and 1b. In addition to the full sample, we also investigate the observations that only contain at least 64 billionaires, to reduce the small sample bias. The density plots of these observations are depicted in Figures 1c and 1d. Lastly, the top and bottom 5% of the theoretical quantities of these observations are dropped to ensure a robust regression. These plots can be found in Figures 1e and 1f. As a final remark, all counts have been weighted with the number of billionaires in each observation.

When investigating the plots, one can observe that most observations are below unity, which is the value expected under the null of a Pareto distribution. This observation continues to hold when we drop all counts that contain less than 64 billionaires, and also when we drop the top and bottom 5% of these observations. This finding already raises our suspicion of no Paretianity.

To confirm our suspicion, we test more formally. The theoretical quantities $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ are regressed on a constant in a total of six regressions, three for each test statistic. Results of which can be found in Table 2. Regressions (1) and (4) are based on the full sample, (2) and (5) only consider observations with at least 64 billionaires, and lastly (3) and (6) drop the top and bottom 5%.

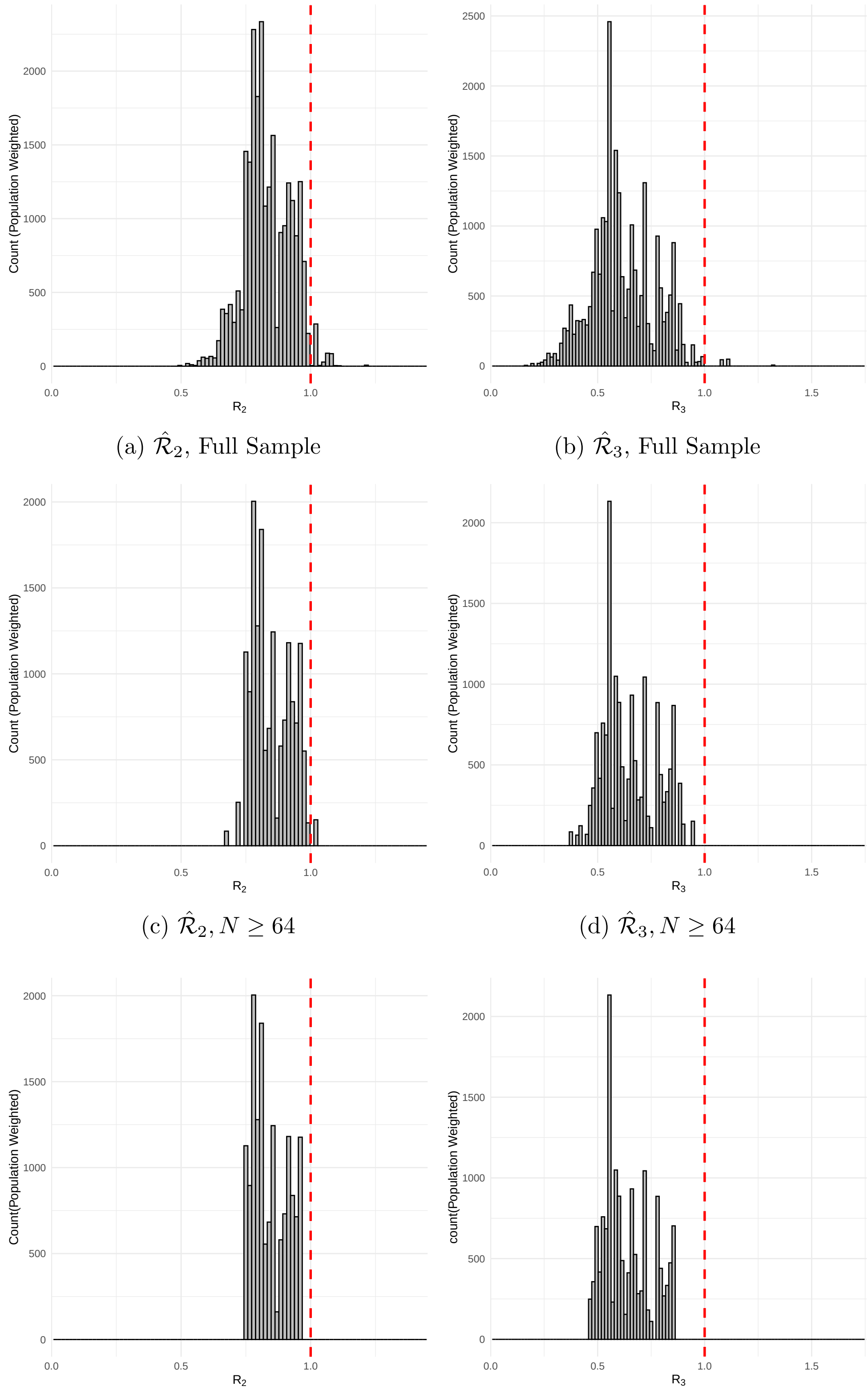
Table 2: WLS Regressions, Pareto Test

	$\hat{\mathcal{R}}_2$			$\hat{\mathcal{R}}_3$		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Constant	0.82***	0.85***	0.85***	0.58***	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	0.199	0.135	0.115
Theoretical RMSE	1	1	1	$\sqrt{10}$	$\sqrt{10}$	$\sqrt{10}$

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: This Table shows the results of the Weighted Least Squares (WLS) regression of the theoretical quantities in Equation (10) on a constant, using the square root of N, the number of billionaires in a sub-region-year observation, as weights. The numbers in parentheses denote standard errors. Specifications (1) and (4) contain the whole sample, whereas (2) and (5) drop observations with fewer than 64 billionaires. Lastly, specifications (3) and (6) drop the top and bottom 5% of observations of the dependent variable in the samples of (2) and (4). The Theoretical Root Mean Squared Errors (RMSEs) of each dependent variable are calculated using Equations (18), and (19) from Appendix A.

Figure 1: Distribution of Test Statistic $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$



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

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

Notes: The Figures above depict the distribution of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$, for each sub-region-year. The red line in each figure is the value that one would expect of these quantities under the null of a Pareto distribution. We weigh each count with the number of billionaires. Panels (a) and (b) contain the full sample, whereas panels (c) and (d) only contain those observations that contain at least 64 billionaires. Lastly, panels (e) and (f) drop the top and bottom 5% of observations from (c) and (d) respectively.

Regression is done through Weighted Least Squares (WLS), where the weights are $\sqrt{N}$. The reason for the usage of WLS instead of Ordinary Least Squares (OLS), is that the model is heteroskedastic because the variance of the test statistics is proportional to N^{-1}. Proof of this can be found in Appendix A.

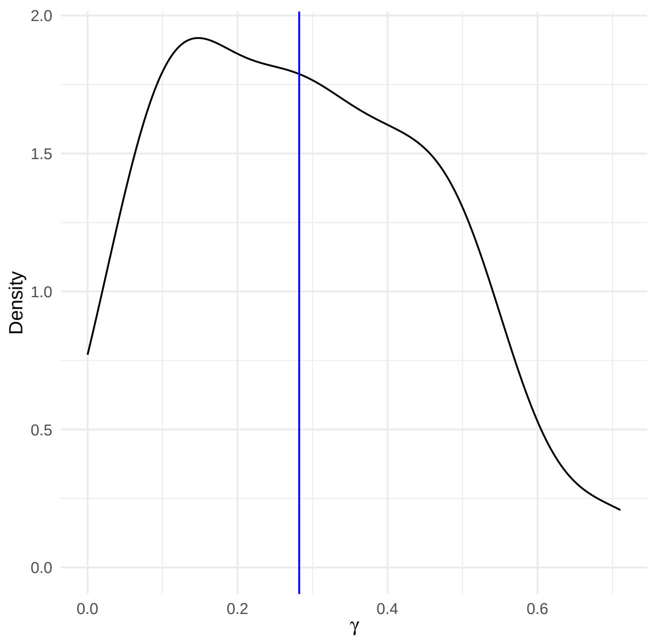
Following from the regression results in Table 2, we reject the null of a Pareto distribution. This is because the coefficients are always statistically significantly different from 1. This is backed up by the Root Mean Squared Errors (RMSEs) that are also provided in the Table. These are way lower than what they should be theoretically, being at the lowest only 6.8% of the theoretical RMSE for $\hat{\mathcal{R}}_2$, and 3.6% for $\hat{\mathcal{R}}_3$. All these facts confirm our suspicion that the data are likely not Pareto distributed.

5.2 Investigating the Alternative: Weibull

In Section 5.1, we observed that our data is extremely likely not Pareto distributed. Therefore, we now shift our focus to an alternative distribution: we investigate if wealth follows the Weibull distribution. We estimate the parameters α and γ of this distribution by applying maximum likelihood estimation on the density function of log wealth, the Gompertz distribution. We use equations 11 and 12 to obtain these parameter estimates. First, we lay our focus on observations with at least 64 billionaires. We plot the distribution of $\hat{\gamma}$ in Figure 2.

From this Figure, we observe the median value of γ to be roughly 0.282, which is depicted with the blue line in the plot.

Figure 2: Distribution of γ Across Sub-Region-Years



Notes: This figure depicts the distribution of γ across all sub-region-years observations of the *Forbes List of Billionaires* from 2001, 2021, focusing only on those observations with at least 64 billionaires included. The blue line represents the median value of γ, which is 0.282.

Next, we again regress the test statistics $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ using WLS. Since the theoretical $\mathcal{R}_k$ values are declining in the product $(\alpha\gamma)$, and thus its log to be declining in $\ln(\alpha) + \ln(\gamma)$, we expect to see a negative sign for these parameters. Therefore, we choose to regress on a constant, $\ln(\alpha)$ and $\ln(\gamma)$. We focus on those observations that contain at least 64 billionaires, and where the top and bottom 5% are dropped. These are the observations that correspond to panels 1e and 1f of Figure 1.

The results of these regressions can be found in Table 3 below:

Table 3: Weibull Predictions

Dependent Variables:	$\hat{\mathcal{R}}_2$	$\hat{\mathcal{R}}_3$
Model:	(1)	(2)
Variables		
Constant	0.778***	0.521***
	(0.011)	(0.010)
$\ln\hat{\alpha}$	-0.119**	-0.211***
	(0.016)	(0.014)
$\ln\hat{\gamma}$	-0.061***	-0.104***
	(0.006)	(0.005)
$\hat{R}_2$		
Weights	$\sqrt{N}$	$\sqrt{N}$
Fit statistics		
Observations	67	67
R^2	0.947	0.986
Adjusted R^2	0.945	0.986
RMSE	0.018	0.015

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: This Table shows the results of the Weighted Least Squares (WLS) regression of the theoretical quantities in Equation (10) on a constant, and the natural logarithm of the parameters of the Weibull distribution, using the square root of N, the number of billionaires in a sub-region-year observation, as weights. The numbers in parentheses denote standard errors. The regressions are based on observations with at least 64 billionaires, and in which the top and bottom 5% of observations are dropped.

From this regression, we observe the coefficients of the parameters have a negative sign and are statistically significantly different from zero. This test gives us a rough, intuitive idea that the Weibull distribution provides a better fit on the data than the Pareto distribution.

Furthermore, we investigate if the signs are equal using an F-test since that is what one would expect in the case of a Weibull distribution following from Equation (13). We obtain F-statistics of 6.80 (0.011) and32.00 (4.939e-07) for regressions (1) and (2) respectively, with p-values in parentheses. From the p-values, we reject our null hypothesis of equal signs, which is in contradiction with the original work of Teulings and Toussaint (2023). We conclude that the Weibull distribution is a better fit to the dataset than the Pareto distribution, but that it still has its flaws, such that a potential better alternative exists.

5.3 Does the Generalized Pareto Distribution provide better predictions?

We extend the paper of Teulings and Toussaint (2023) by applying a Generalized Pareto Distribution (GPD). This distribution contains three parameters: shape α, location μ and scale σ. We set μ to one billion USD, which is the lower bound of the wealth we focus on. Next, we simultaneously estimate α and γ by maximum likelihood estimation.

We then predict wealth in 2018 for each subregion, with each of the three models. In the case of the Weibull distribution, we apply a common γ, which we set equal to the median value found in Subsection 5.2, 0.282. We then calculate the α for each subregion using Equation (11) with this common γ. For the estimate of the shape parameter α of the Pareto distribution, we make use of the Hill estimator, which is in this case simply equal to the mean log wealth, $\overline{w}$, as derived from Equation 4.

We then calculate the expected wealth for each subregion with the Pareto, Weibull and GPD models, using Equations 2, 7 and 9 respectively. The results of these calculations can be found in Table 4 below, together with the estimates of the α parameters for the different models:

We can immediately observe that Pareto performs poorly in predicting mean wealth for each sub-region. Out of the 18 sub-regions, the Pareto estimate is ∞ for 7 sub-regions, due to the estimate $\hat{\alpha}$ being greater than 1. Furthermore, there are another 4 observations, which are Scandinavia, Japan, India and Russia, for which the estimate $\hat{\alpha}$ lies close to 1 (within 0.1), causing these estimates to be way too high. This leaves us with only 7 reasonable Pareto predictions of mean wealth. However, these estimates are also too high to be considered adequate.

When we look at the estimates of the Weibull model, however, we can see that our estimates are much better. They are quite similar to the actual mean wealth values of the data set. The least accurate prediction is that of Germany, in which we have a difference of 0.87. From this, we would thus conclude that the Weibull model is quite adequate for predicting mean wealth.

Lastly, we take a look at the GPD model. We observe that GPD often makes quite promising predictions, sometimes even better than those of the Weibull model. Take for instance the Alps, where the GPD model is only off by 0.01, whereas the Weibull model is off by 0.86. However, there are also instances where the Weibull model outperforms the GPD model, sometimes even by a large margin. The GPD model performs quite poorly in certain sub-regions. For example, GPD overshoots Russia's mean wealth quite a lot (3.02) whereas the Weibull model is only off by 0.01. In most cases, it is an adequate distribution for predicting mean wealth, but in some cases, it lacks consistency. This could be explained by the fact that countries in which the GPD is not the most accurate at predicting mean billionaire wealth have few observations. We know from the Pickands-Balkema-De Haan Theorem (Balkema and De Haan, 1974; Pickands III, 1975) that data is distributed following the GPD only asymptotically, so it makes sense that in smaller samples this model might be off a bit more.

We therefore observe the Weibull distribution to give the most consistent mean wealth predictions followed by the GPD model, which lacks consistency in certain, smaller, sub-regions. The worst extreme-value distribution in terms of predicting mean wealth is the Pareto distribution, by quite a margin.

Table 4: Predicted vs. Realised Values of Mean Billionaire Wealth, 2018

	Mean Wealth				$\hat{\alpha}$		
Sub-Region	Data	Weibull	Pareto	GPD	Weibull	Pareto	GPD
United States	5.29	5.48	∞	5.12	1.516	1.156	0.456
Canada	3.23	3.22	5.96	3.51	1.039	0.832	0.484
Germany	4.70	5.57	∞	4.85	1.533	1.182	0.334
British Isles	3.83	4.26	∞	4.02	1.287	1.015	0.319
Scandinavia	3.51	4.06	226.69	3.56	1.244	0.996	0.154
France	7.44	8.11	∞	8.67	1.894	1.367	0.644
Alpine Countries	3.80	4.66	∞	3.79	1.367	1.098	0.151
Italy	3.96	4.29	∞	4.08	1.294	1.017	0.452
China	3.31	3.09	4.98	3.71	1.006	0.799	0.615
South-East Asia	3.33	3.51	9.43	3.54	1.117	0.894	0.400
Asian Islands	3.17	3.22	6.18	3.33	1.039	0.838	0.436
South Korea	2.88	2.78	3.80	3.28	0.912	0.737	0.541
Japan	3.95	3.82	12.42	5.17	1.191	0.919	0.694
Australia	2.74	2.75	3.86	2.92	0.903	0.741	0.408
India	3.70	3.90	24.11	2.82	1.209	0.959	0.408
Russia	4.05	4.06	21.35	7.07	1.245	0.953	0.794
Brazil	4.20	4.41	∞	4.39	1.319	1.030	0.451
Israel+Turkey	2.17	2.24	2.64	2.27	0.727	0.622	0.180

Notes: In this Table, we make use of the same sub-regions as in Table 1. We make predictions of mean wealth for each sub-region in the year 2018. For the Pareto distribution, these predictions are made using Equation (2), in which we set Ω equal to 1 (Billion USD) and set the expected value equal to ∞ when $\alpha \geq 1$, where $\hat{\alpha} = \overline{w}$, the maximum likelihood estimator of α. For the Weibull distribution, we use Equation (7) to make mean wealth predictions for each sub-region, setting γ equal to its median value (0.282), and then compute the shape parameter α for each observation individually using MLE. Lastly, we make use of Equation (9) to compute mean wealth predictions under the GPD model, where we set the location parameter μ equal to 1 (Billion USD), and then calculate estimates of the shape parameter α and the scale parameter σ simultaneously for each observation using MLE.

Finally, we assess the model fit of each of the three models. We do so using a Kolmogorov-Smirnov (KS) test, which we apply to the full sample. We use the formulas from Equation (14) and (15). We determine the maximum absolute deviation between the Empirical CDF and the theoretical CDF for each model. For the Pareto distribution, we get a test statistic $D = 0.896$ (<2.2e-16) where the number in parentheses is the p-value. From this result it becomes evident that the dataset is not Pareto distributed, confirming our previous suspicions. For the Weibull distribution, we get more promising results. The test statistic and p-value are $D = 0.093$ (0.003), which is a significant improvement from the Pareto distribution. However, we still reject the null of the dataset following a Weibull distribution at a 5% significance level, but the fit is way better than that of the Pareto distribution. Lastly, we determine the goodness-of-fit of our extension model, the GPD. For the GPD, we obtain a test statistic of 0.828 (0.345), and thus we now do not reject the null of equal distributions.

We find that the GPD has the best fit on the data, proving to be better than the Pareto and Weibull distributions in goodness-of-fit. However, the Weibull distribution gives more accurate mean wealth predictions on average. Overall, the GPD is a strong competitor of the Weibull distribution in the world of mean billionaire wealth, with their predictions coming extremely close to each other, and the fit of the GPD to the dataset being better than the fit of the Weibull distribution.

6 Conclusion

In this paper, we used three different extreme value distributions in predicting mean billionaire wealth, based on data from the *Forbes List of Billionaires.* These distributions are the Pareto Distribution, the Weibull Distribution and the Generalized Pareto Distribution. In particular, we calculate the expected mean wealth of billionaires in different sub-regions in different years by constructing splits based on the geographical location of the countries of citizenship and the number of billionaires in the year 2019.

We then continued by testing the different distributions of the data. We have found the Pareto distribution to perform extremely poorly in predicting mean wealth, as well as having a poor fit to the data. The Weibull distribution performs way better than the Pareto model, showing better mean wealth predictions as well as having a better fit for the data. These findings resonate with the paper of Teulings and Toussaint (2023), who conclude that mean wealth is distributed Weibull rather than Pareto. Weibull predictions of mean wealth were extremely close to actual mean wealth numbers, whereas the Pareto predictions were off extremely, more often than not.

This could be explained due to the difference in hazard rates of both distributions. Where the Pareto distribution has a constant hazard rate, the Weibull distribution has an exponentially increasing hazard rate. Additionally, the Weibull distribution can make predictions for all values of the shape parameter α, whereas the Pareto distribution can only make these predictions for $\alpha < 1$. These two facts explain as to why the Weibull distribution shows to be a better distribution for modelling mean billionaire wealth.

We extend the work of Teulings and Toussaint (2023) by exploring an additional extreme value distribution, the Generalized Pareto Distribution (GPD). Our choice for this distribution as an alternative is motivated by the papers of Balkema and De Haan (1974) and Pickands III (1975), which state that for a large class of underlying distributions, the distribution of exceedances above a certain threshold, which in our case is one billion USD, can be approximated by a GPD.

We fit the GPD to the dataset and observe that its predictions are better than those of the Pareto distribution but somewhat similar to those of the Weibull distribution. Sometimes the GPD gives better predictions, but there are also cases in which the GPD performs worse than the Weibull distribution. This is likely because the observations for which this is the case do not contain a lot of billionaires. In bigger samples, we observe the GPD to perform better at predicting mean wealth. In contrast, in smaller samples it performs slightly worse than the Weibull distribution, resonating with the Pickands-Balkema-De Haan (Balkema and De Haan, 1974; Pickands III, 1975) Theorem.

We also assess the goodness-of-fit of the three models through a Kolmogorov-Smirnov (KS) test. We find the Pareto distribution to have the worst fit on the data, with an extremely small p-value. This is likely because our mean wealth predictions reach ∞ for certain parameter estimates. The Weibull distribution shows to have an improved fit over the Pareto distribution, but still with a poor p-value. Lastly, we find the GDP model to have the most promising results, with a p-value causing us not to reject our null hypothesis, indicating that the data are GP-distributed. Therefore, we conclude that the GPD can certainly outperform traditional extreme-value models in billionaire wealth applications.

Directions for future research

Future research could explore the integration of economic factors, applying regression-based extreme value theory. We could incorporate these economic factors directly into the parameters of the distributions to investigate how shifts in the economy might affect the distribution of billionaire wealth. More specifically, we could model the scale parameter of the distributions using covariates. Potentially relevant economic factors could be the growth rate of the Gross Domestic Product (GDP), interest rates or indices from the stock market.

Furthermore, one could investigate the assumption of independent observations, by simulating values of the theoretical quantities $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$. Then, by inspection of the distribution of p-values, it can be determined whether or not this assumption is correct. If these p-values are distributed roughly uniformly, then the observations are indeed independent of each other, and otherwise, they are likely not.

Finally, exploring Bayesian Time Series methods could offer valuable insights, particularly in capturing the temporal dynamics and uncertainties that are present in economic data influencing billionaire wealth. Bayesian approaches provide a robust framework for incorporating prior knowledge and updating beliefs as new data becomes available. This can be of particular interest when there is not a lot of data or when it is highly variable, allowing for a more statistically sound prediction. By applying Bayesian Time Series analysis, one could enhance their understanding of how economic trends and cycles impact mean billionaire wealth over time.

References

Sadiah M Aljeddani and MA Mohammed. An extensive mathematical approach for wind speed evaluation using inverse weibull distribution. *Alexandria Engineering Journal*, 76:775–786, 2023.

Andrés M Alonso, Patricia de Zea Bermudez, and Manuel G Scotto. Comparing generalized pareto models fitted to extreme observations: an application to the largest temperatures in spain. *Stochastic environmental research and risk assessment*, 28:1221–1233, 2014.

Anthony B Atkinson, Thomas Piketty, and Emmanuel Saez. Top incomes in the long run of history. *Journal of economic literature*, 49(1):3–71, 2011.

David Autor, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709, 2020.

Sutirtha Bagchi and Jan Svejnar. Does wealth inequality matter for growth? the effect of billionaires. *Wealth, Income Distribution, and Poverty*, 2013.

AA Balkema and L De Haan. Limit laws for order statistics. 1974.

Vance W Berger and YanYan Zhou. Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online*, 2014.

Thomas Blanchet. Uncovering the dynamics of the wealth distribution. *arXiv preprint arXiv:2211.15509*, 2022.

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

Vytaras Brazauskas and Andreas Kleefeld. Robust and efficient fitting of the generalized pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics*, 45 (3):424–435, 2009.

Adam Čabla and Filip Habarta. Distribution of the wealth of the richest persons in the world. In *22nd International Scientific Conference on Applications of Mathematics and Statistics in Economics (AMSE 2019)*, pages 158–169. Atlantis Press, 2019.

Michele Campolieti. 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, 2018.

Kevin W Capehart. Is the wealth of the world's billionaires not paretian? *Physica A: Statistical Mechanics and its Applications*, 395:255–260, 2014.

Christian Evan Chandraa and Sarini Abdullaha. Graduating mortality rates by mixture of pareto, loglogistic, and two weibull distributions using bayesian method. *International Journal on Advanced Science, Engineering and Information Technology*, 12(5):1907–1914, 2022.

Jeffrey Chu, O Dickin, and Saralees Nadarajah. A review of goodness of fit tests for pareto distributions. *Journal of Computational and Applied Mathematics*, 361:13–41, 2019.

Arnold LM Dekkers, John HJ Einmahl, and Laurens De Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, pages 1833–1855, 1989.

Francis X Diebold, Til Schuermann, and John D Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *The Journal of Risk Finance*, 1(2):30–35, 2000.

SK-B Dzidzornu and R Minkah. Assessing the performance of the discrete generalised pareto distribution in modelling non-life insurance claims. *Journal of Probability and Statistics*, 2021 (1):5518583, 2021.

Paul Embrechts, Sidney I Resnick, and Gennady Samorodnitsky. Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41, 1999.

Igor Fedotenkov. A review of more than one hundred pareto-tail index estimators. *Statistica*, 80(3):245–299, 2020.

Xavier Gabaix and Augustin Landier. Why has ceo pay increased so much? *The quarterly journal of economics*, 123(1):49–100, 2008.

David E Giles, Hui Feng, and Ryan T Godwin. Bias-corrected maximum likelihood estimation of the parameters of the generalized pareto distribution. *Communications in Statistics-Theory and Methods*, 45(8):2465–2483, 2016.

Matthieu Gomez. Decomposing the growth of top wealth shares. *Econometrica*, 91(3):979–1024, 2023.

Emil Julius Gumbel. *Statistics of extremes*. Columbia university press, 1958.

Mohamed Hanafy et al. Application of generalized pareto in non-life insurance. *Journal of Financial Risk Management*, 9(03):334, 2020.

Bruce M Hill. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, pages 1163–1174, 1975.

JD Holmes and WW Moriarty. Application of the generalized pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 83(1-3):1–10, 1999.

Ronald Huisman, Kees G Koedijk, Clemens J M Kool, and Franz Palm. Tail-index estimates in small samples. *Journal of Business & Economic Statistics*, 19(2):208–216, 2001.

Emmanuel Iyamuremye, Joseph Mung'atu, Peter Mwita, et al. Extreme value modelling of rainfall using poisson-generalized pareto distribution: A case study tanzania. *Int. J. Stat. Distrib. Appl*, 5:67–75, 2019.

Wojciech Kopczuk. What do we know about the evolution of top wealth shares in the united states? *Journal of Economic Perspectives*, 29(1):47–66, 2015.

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

Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

Alexander J McNeil. Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*, 12(5):217–37, 1999.

Wilhemina Adoma Pels, Atinuke O Adebanji, Sampson Twumasi-Ankrah, and Richard Minkah. Shrinkage methods for estimating the shape parameter of the generalized pareto distribution. *Journal of Applied Mathematics*, 2023(1):9750638, 2023.

James Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

Sri Purwani and Riza Andrian Ibrahim. Using simple fixed-point iterations to estimate generalized pareto distribution parameters. *IAENG International Journal of Applied Mathematics*, 54(2), 2024.

Kenneth T Rosen and Mitchel Resnick. The size distribution of cities: an examination of the pareto law and primacy. *Journal of urban economics*, 8(2):165–186, 1980.

Coen N Teulings and Simon Toussaint. Top wealth is distributed weibull, not pareto. 2023.

Philip Vermeulen. Estimating the top tail of the wealth distribution. *American Economic Review*, 106(5):646–650, 2016.

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

Cristiano Villa. Bayesian estimation of the threshold of a generalised pareto distribution for heavy-tailed observations. *Test*, 26(1):95–118, 2017.

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

Tieling Zhang and Min Xie. Failure data analysis with extended weibull distribution. *Communications in Statistics—Simulation and Computation®*, 36(3):579–592, 2007.

Appendix

A Mean and variance of $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$

In this section of the Appendix, we derive the mean and variance of our test statistics, $\hat{\mathcal{R}}_k$. We start with the expected value, which is equal to:

$$\mathrm{E}\left[\hat{\mathcal{R}}_k\right] = 1 + N^{-1}\left(\frac{(2k)!}{k!^2} - \frac{3k^2-k+2}{2}\right) + \mathcal{O}(N^{-2}), \tag{16}$$

where N is the sample size. The variance of $\hat{\mathcal{R}}_k$ is equal to:

$$\mathrm{Var}\left[\hat{\mathcal{R}}_k\right] = N^{-1}\left(\frac{(2k)!}{k!^2} - k^2 - 1\right) + \mathcal{O}(N^{-2}). \tag{17}$$

For $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$, we thus get the following:

$$\mathrm{E}\left[\hat{\mathcal{R}}_2\right] = 1 + \mathcal{O}(N^{-2}) \qquad \mathrm{Var}\left[\hat{\mathcal{R}}_2\right] = N^{-1} + \mathcal{O}(N^{-2}), \tag{18}$$

and:

$$\mathrm{E}\left[\hat{\mathcal{R}}_3\right] = 1 + 7N^{-1}\mathcal{O}(N^{-2}) \qquad \mathrm{Var}\left[\hat{\mathcal{R}}_3\right] = 10N^{-1} + \mathcal{O}(N^{-2}). \tag{19}$$

$\hat{\mathcal{R}}_2$ is a more powerful test statistic than $\hat{\mathcal{R}}_3$, because it is both unbiased up to terms of order N^{-2} and because it has a smaller variance. We will test for Paretianity by regressing the theoretical quantities on a constant, applying a two-sided test in which we investigate if its value differs significantly from unity.

B Derivation of the Likelihood function of the Gompertz Distribution

In this section of the Appendix, we derive the likelihood function of the Gompertz distribution. We start with the density function, which is as follows:

$$f(w) = \alpha^{-1}\exp\left(\gamma w - \frac{e^{\gamma w}-1}{\alpha\gamma}.\right). \tag{20}$$

From this density function, we obtain the following log likelihood:

$$N^{-1}\log\mathcal{L}(\alpha,\gamma) = -\ln\alpha + \gamma\overline{w} - (\alpha\gamma)^{-1}\left(\overline{\mathrm{e}^{\gamma w}} - 1\right), \tag{21}$$

giving us the following first order condition for $\hat{\alpha}$:

$$N^{-1}\frac{\partial\log\mathcal{L}(\alpha,\gamma)}{\partial\alpha} = \left[-1 + (\hat{\alpha}\gamma)^{-1}\left(\overline{\mathrm{e}^{\gamma w}} - 1\right)\right]\hat{\alpha}^{-1} = 0 \Rightarrow$$
$$\hat{\alpha} = \gamma^{-1}(\overline{\mathrm{e}^{\gamma w}} - 1). \tag{22}$$

If we substitute Equation 22 into Equation 21, the log likelihood becomes:

$$N^{-1}\log\mathcal{L}(\gamma) = \ln\gamma - \ln\left(\overline{\mathrm{e}^{\gamma w}} - 1\right) + \gamma\overline{w}.$$

Then, the first order condition for $\hat{\gamma}$ becomes:

$$N^{-1}\left.\frac{\mathrm{d}\log\mathcal{L}(\gamma)}{\mathrm{d}\gamma}\right|_{\gamma=\hat{\gamma}} = \hat{\gamma}^{-1} - \hat{\gamma}\frac{\overline{e^{\hat{\gamma}w}}}{e^{\hat{\gamma}\overline{w}} - 1} + \overline{w} = 0.$$