DISCUSSION PAPER SERIES

DP18634
(v. 2)

TOP WEALTH IS DISTRIBUTED WEIBULL, NOT PARETO

Coen Teulings and Simon Toussaint

INTERNATIONAL TRADE AND REGIONAL ECONOMICS, LABOUR ECONOMICS AND MACROECONOMICS AND GROWTH

CEPR

ISSN 0265-8003

TOP WEALTH IS DISTRIBUTED WEIBULL, NOT PARETO

Coen Teulings and Simon Toussaint

Discussion Paper DP18634
First Published 23 November 2023
This Revision 11 June 2024

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- International Trade and Regional Economics
- Labour Economics
- Macroeconomics and Growth

TOP WEALTH IS DISTRIBUTED WEIBULL, NOT PARETO

Abstract

We study the shape of the global wealth distribution, using the Forbes List of Billionaires. We develop simple statistics based on ratios of log moments to test the default assumption of a Pareto distribution, which is strongly rejected. Hazard rates show that the log-transformed data instead follow a Gompertz distribution, which means that the data in levels follow a truncated-Weibull distribution. We further apply our model to the U.S. city size distribution and the U.S. firm size distribution. These distributions also show a rejection of Pareto in favor of (truncated-)Weibull. We discuss some theoretical and practical implications of our results.

JEL Classification: D3, E2, G5

Keywords: N/A

Coen Teulings - c.n.teulings@outlook.com
Utrecht University and CEPR

Simon Toussaint - s.j.toussaint@uu.nl
Utrecht University

Acknowledgements
We are grateful to Tommaso Tulkens for research assistance. We thank Bas van Bavel, Richard Blundell, Xavier Gabaix, Tomer Ifergane, Christian Kleiber, Rutger-Jan Lange, Wouter Leenders, Ben Moll, Thomas Piketty, Maarten de Ridder, Yasmine van der Straten, and Chen Zhou, as well as seminar participants for helpful comments. We also thank Gustav Munch and River Chen for their coding advice.

Top Wealth Is Distributed Weibull, Not Pareto*

Coen N. Teulings[†] Simon J. Toussaint[‡]

June 4, 2024

Abstract

We study the shape of the global wealth distribution, using the *Forbes List of Billionaires*. We develop simple statistics based on ratios of log moments to test the default assumption of a Pareto distribution, which is strongly rejected. Hazard rates show that the log-transformed data instead follow a Gompertz distribution, which means that the data in levels follow a truncated-Weibull distribution. We further apply our model to the U.S. city size distribution and the U.S. firm size distribution. These distributions also show a rejection of Pareto in favor of (truncated-)Weibull. We discuss some theoretical and practical implications of our results.

JEL Classification: D3, E2, G5
Keywords: Wealth distribution; city and firm size distribution; tail distributions.

*We are grateful to Tommaso Tulkens for research assistance. We thank Bas van Bavel, Richard Blundell, Riccardo Cioffi, Xavier Gabaix, Tomer Ifergane, Christian Kleiber, Rutger-Jan Lange, Wouter Leenders, Robin Lumsdaine, Ben Moll, Thomas Piketty, Maarten de Ridder, Yasmine van der Straten, David Seim, Phyllis Wan, Max Welz, and Chen Zhou, as well as seminar participants for helpful comments. We also thank Gustav Munch and River Chen for their coding advice.

[†]Utrecht School of Economics and CEPR. E-mail: c.n.teulings@outlook.com
[‡]Utrecht School of Economics. E-mail: s.j.toussaint@uu.nl

1 Introduction

Right-skewed distributions pervade many aspects of economic life (Gabaix 2009, 2016).[1] The standard instrument for analyzing these phenomena is the Pareto distribution, of which the characteristics are well known for a long time (Pareto 1896; Van der Wijk 1939). This distribution has one parameter governing the fatness of its right tail, the *tail index* α^{-1}. The higher α, the more mass is concentrated at the extreme end of the distribution. However, there is increasing theoretical and empirical evidence that Pareto provides a poor fit to the data in many applications. Blanchet, Fournier, and Piketty (2022) show that the predictions of Pareto for mean wealth and income among the upper tail are heavily at odds with reality. In the city size literature, Eeckhout (2004, 2009) and Rossi-Hansberg and Wright (2007) have challenged the Pareto assumption, arguing for respectively log Normal and a non-specified log-concave distribution function. For firm size, Jones (2023) argues that even thin-tailed firm productivity distributions can give rise to exponential growth, given the combinatorial nature of endogenous growth. So far, however, Pareto remains the default assumption in theoretical and empirical analysis.

Most empirical research on the right tail uses the well known tool of the log rank regression: the relation between the log rank in a sample of the rich/cities/firms and log wealth (for the rich) or log size (for cities and firms) should be linear for a Pareto distribution, see Rosen and Resnick (1980) for an early application. The use of this tool, however, is unfortunate. The log rank is an order statistic. Its construction requires the researcher to order the data by the magnitude of the variable of interest. This introduces correlation between observations, which invalidates many standard techniques and causes the OLS coefficients to be biased. Gabaix and Ibragimov (2011) set out to correct the log rank regression for this bias. This complication, however, can easily be avoided by applying maximum likelihood estimation. If the null of a Pareto distribution is correct, the maximum likelihood estimator of the tail index is most efficient. Moreover, it is extremely simple: the mean of log wealth, see Hill (1975). However, since the predictions of the Pareto model for mean wealth are very bad, see our discussion of Blanchet, Fournier, and Piketty (2022) above, the maximum likelihood estimator has been discredited. Hence, Blanchet, Fournier, and Piketty (2022) give up the idea of characterizing the actual distribution by some simple functional form. They argue that top income and wealth shares in the United States and France exhibit a tail index *curve*, with one parameter for every bracket in the distribution rather than a single tail index parameter.

This paper provides empirical support for a more positive view. We have two main contributions. First, we develop a simple set of tests for the null hypothesis of Paretianity. If a variable is Pareto, its log is Exponential. Since all moments of the Exponential distribution are well-defined – unlike for Pareto, where only moments up to α^{-1} are defined – working with logs has distinct advantages. Our test statistics work as follows. For Exponential distributions, the mean is a maximum likelihood estimator of α. Furthermore, the kth higher moment equals k factorial times α to the kth power. Hence, we can scale higher moments by appropriate powers of the sample mean to obtain a ratio statistic $\mathcal{R}_k$ that is independent from α and is asymptotically equal to unity.

We apply two versions of this test (based on the second and third moment respectively) to the data from the *Forbes List of Billionaires* from 2001 to 2021. Our data unequivocally reject the Pareto null for all years and for all 18 sub-regions of the world that we are able to distinguish in our data. We show that this rejection cannot be driven by measurement error, since measurement error scales up the density proportionately. In addition, we derive the

1. For example, heavy-tailed distributions feature prominently in the distributions of city-size (e.g. Gabaix 1999, Eeckhout 2004,Rossi-Hansberg and Wright 2007), firm-size (Luttmer 2011; Autor et al. 2020; Jones 2023), CEO salaries (Gabaix and Landier 2008), income and wealth (e.g. Atkinson, Piketty, and Saez 2011; Vermeulen 2018), and on financial markets (e.g. Huisman, Koedijk, Kool, and Palm 2001).

small-sample properties of $\mathcal{R}_k$ and show that small-sample bias cannot account for these results. Remarkably, both test statistics have very similar values for all years and all sub-regions, being consistently less than unity, the more so for the test statistic based on the third rather than the second moment. This suggests some systematic pattern in the data not covered by the Pareto prior. In particular, the data seem to be less skewed to the right. It fosters the hope that there might be a simple alternative for Pareto that describes the data well.

Our second contribution is to develop a suitable alternative. We hypothesize that the actual distribution is not Pareto but *truncated-Weibull*. While the Pareto distribution is invariant to the choice of the lower bound, the Weibull distribution is not. The research therefore has to estimate the truncation point of the part of the distribution that is fitted to the data. Where the log of a Pareto variable is distributed exponentially, the log of a Weibull variable follows a *Gompertz* or counter-Gumbel distribution.[2] Like for Weibull, we use the truncated rather than the full Gompertz distribution; for the sake of brevity, we drop the word "truncated" from now on. The Gompertz distribution is not often used in economics, but it is a standard tool for modelling life expectancy in demography. The practical difference between Pareto and Weibull can be understood most easily by analysing the distributions of their logs, that is: Exponential and Gompertz respectively. The Exponential distribution is characterized by a constant hazard rate, equal to α^{-1}, implying that for every unit increase in a particular lower bound of log wealth, the number of people that is even richer than that lower bound decreases by a constant percentage. The Gompertz distribution, instead, is characterized by not a constant but an increasing hazard above some lower bound, equal to $\alpha^{-1}e^{\gamma\omega}$, where γ is a positive parameter and ω is this lower bound: for every .01 increase in the lower bound, the hazard rises by γ%. As $\gamma \to 0$, we recover the Exponential distribution with a constant hazard rate.

The hazard rate of the log-transform can be directly computed from the data and compared to the parametric hazards of some candidate distributions. Figure 1 demonstrates this for the global distribution of billionaires in 2018. The black line is the empirical hazard, estimated with a simple Kaplan-Meier estimator. The colored lines represent parametric cumulative hazards for the log-transform for the Exponential (red) and Gompertz (blue) distribution. The Exponential distribution has a constant hazard α^{-1} and hence a linearly increasing cumulative hazard $\alpha^{-1}w$. The prediction of a constantly increasing cumulative hazard of the Exponential distribution is strongly at odds with the data. Instead, the exponentially increasing (cumulative) hazard of Gompertz fits the data well.

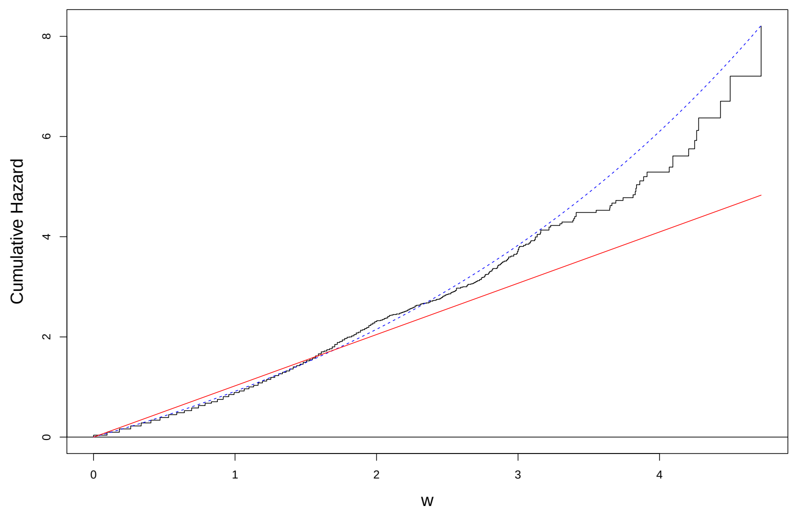
Figure 1 also reveals why Weibull has gone hitherto unnoticed: the divergence between Exponential and Gompertz only occurs far in the upper tail. Visual inference based on log-log plots is therefore prone to miss these departures from Paretianity; in contrast, our test statistics $\mathcal{R}_k$ and the hazard-based plots provide sharp evidence.

Whenever Pareto has previously been challenged, the response has been that one has to go further into the right to arrive at the exponential tail (the log of Pareto), where the hazard rate would converge to a constant. Figure 1 shows that this response is mistaken. The hazard rate never converges but continues to diverge to infinity for higher levels of log wealth. In fact, the hazard rate of the Gompertz distribution (the log of Weibull) diverges even faster than that of the Normal distribution (the log of log Normal): the Gompertz hazard diverges exponentially, while the hazard rate of the Normal distribution diverges only linearly asymptotically.

We estimate the parameters of the Weibull distribution by maximum likelihood and find a rather stable value of $\gamma \approx 0.25$. Weibull has another advantage over Pareto: all its moments exist for all values of the parameters. We use our model to predict mean wealth for all sub-regions, using a common value of γ for all sub-regions. Where Pareto fails miserably, the Weibull distribution provides an almost perfect fit.

2. When $\underline{\boldsymbol{W}}$ is Weibull, $\ln \underline{\boldsymbol{W}}$ is distributed Gompertz and $-\ln \underline{\boldsymbol{W}}$ is distributed Gumbel. There is some confusion in the literature about the definition of the Gompertz distribution, where Gompertz is sometimes defined as Gumbel. See our discussion in footnote 9.

Figure 1: Empirical and Parametric Hazard Rates, Billionaire Log Wealth Distribution



Notes: Figure plots the Kaplan-Meier estimator of the empirical cumulative hazard of the billionaire log wealth distribution, using the 2018 *Forbes List of Billionaires* and pooling all observations. The blue dotted line is a fitted Gompertz hazard, and solid red is an Exponential hazard.

Our results raise the question whether they are typical for the Forbes data on billionaires' wealth or for the wealth distribution, or that they apply more generally to phenomena that have been presumed to be Paretian. We show evidence for two other variables: city and firm size. The shape of the U.S. city size distribution is subject of a long-–standing debate. The default of Pareto (specifically Zipf) is assumed by, among others, Krugman (1996) and Gabaix (1999), who use the 135 largest Metropolitan Statistical Areas. Eeckhout (2004, 2009) challenges this result. We find that even for the upper tail, Pareto is strongly rejected by our test-statistics and that there is solid support for Weibull. The same applies to the U.S. firm size distribution. Again, this distribution is typically assumed to be Pareto (e.g., Axtell 2001; Luttmer 2011). We use data gathered by Kwon, Ma, and Zimmermann (2023). They use generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022) to go from tabulations to a full distribution. This tool can also be used to create synthetic samples of the top 1% of firms. Again, we find overwhelming evidence against Pareto and in favor of Weibull.

Weibull can easily replace Pareto in many economic models. We discuss some avenues for further research. For example, the Gompertz distribution figures in stochastic networks. The length of a *self-avoiding walk* along a graph (that is, once a node is visited it cannot be revisited), is distributed Gompertz (Tishby, Biham, and Katzav 2016). This model structure has interesting parallels with the variables we study; for instance, city size is determined by area already used by existing cities.

Related Literature: Our paper is related to three strands of literature. First, our paper adds to the literature on tail index estimation. This literature, which arose as a consequence of the development of Extreme Value Theory (e.g., Gumbel 1958; Balkema and De Haan 1974; Pickands 1975), is highly multidisciplinary and extensive in scope, and we will not be able to do it justice here; see Bingham, Goldie, and Teugels (1989) and Beirlant, Goegebeur, Segers, and Teugels (2006). Within the tail index estimation literature – reviewed in Fedotenkov (2020) – our work relates most closely to estimation on the basis of moments of the log-density function (Hill 1975; Dekkers, Einmahl, and De Haan 1989).

Second, we contribute to the study of heavy-tailed distributions within the field of economics. Examples of heavy-tailed distributions abound (Gabaix 2009, 2016), ranging from the distributions of income and wealth (Atkinson, Piketty, and Saez 2011; Vermeulen 2018), city size (Gabaix 1999), firm size (Luttmer 2011; Autor et al. 2020), and the cross-section of stock returns (Huisman, Koedijk, Kool, and Palm 2001). A lot of theoretical literature, however, simply assumes the Pareto distribution without properly testing this assumption, or looks at a log rank – log size plot to conclude that a particular distribution looks Pareto-ish. Within this literature, our work relates to studies which question the Pareto assumption. Examples include Eeckhout (2004) and Rossi-Hansberg and Wright (2007) for city size, Jones (2023) and Kondo, Lewis, and Stella (2023) for firm size, and Blanchet, Fournier, and Piketty (2022) for income and wealth. Our work also relates to papers which use the heavy-tailedness of a distribution as a basis to construct causal estimators of macroeconomic aggregates, such as Gabaix and Koijen (2023).

Finally, our paper relates to economic studies on the dynamics of size distributions. A theoretical basis for studying the dynamics of top income and wealth is given by Gabaix, Lasry, Lions, and Moll (2016), who show that the Bewley-Aiyagari-Huggett random-growth models of wealth accumulation (Benhabib and Bisin 2018) fail to deliver Pareto tails with the same speed as observed in the data. Their framework has spurred much theoretical and empirical work. Similarly, Luttmer (2011) shows that only deviations from Gibrat's Law (i.e., firm growth is not independent of firm size) can generate a Pareto-shaped distribution of firms within a reasonable timeframe. Perhaps closest to our focus on billionaires is Gomez (2023), who uses tools from stochastic calculus to decompose the growth of the American Forbes 400 wealth share into growth by incumbents, growth by new entrants, and entry/exit effects; and Blanchet (2022), who also uses empirical data on income and wealth to study stochastic properties of economic models. In a companion paper, we use our Weibull framework to study the dynamics of billionaire numbers since 2000, finding that a simple model can account for most of the time-series and cross-sectional variation (Teulings and Toussaint 2023). Since the number of billionaires is more sensitive to variation in the lower bound under Weibull, an increase in wealth in a given region (and hence a decline in the effective lower bound) will lead to a larger increase in billionaire numbers than predicted by Pareto.

Paper Outline: The rest of this paper is structured as follows. Section 2 presents our Pareto framework. In Section 3, we discuss our main data, the *Forbes List of Billionaires*. Section 4 present the empirical results for the test of the Pareto assumption. Section 5 develops the Weibull distribution. In Section 6, we test this distribution empirically. In Section 7, we further apply our framework to cities and firms. In Section 8, we discuss some theoretical implications of our results. Section 9 concludes.

2 Pareto: Theory and Estimation

2.1 Theory

We consider a random variable $\underline{X} \geq \Omega$, where $\Omega > 0$ is a parameter (stochastic variables will be underlined). We assume Ω to be known. For example, when using data from the *Forbes List of Billionaires*, Ω is a billion USD.[3] The complementary distribution function of the Pareto distribution and its moments are given by:

$$\Pr\left[\underline{X} \geq X \,\middle|\, X \geq \Omega\right] = (X/\Omega)^{-1/\alpha}, \tag{1}$$

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

where $\alpha > 0$ is a parameter; α^{-1} is commonly referred to as the tail index or Pareto coefficient (Jones 2015).[4] If $\alpha = 1$, the Pareto distribution specializes to the Zipf distribution. Equation (1) is easy to interpret, since it predicts the probability of a "large" observation to scale with size like a power law. This simple rule, together with its apparent fit of the distribution of the right tail of many empirical phenomena, has contributed to the popularity of Pareto.

Equation (2) reveals a problem that comes with Pareto: its moments $\mathrm{E}\left[\underline{X}^k\right]$ for $k \geq \alpha^{-1}$ do not exist. Since α^{-1} is estimated to be between 1 and 2 in many economic applications[5], this implies that the variance does not exist, let alone higher moments. For many phenomena, even the expectation does not exist. This makes the interpretation of α problematic.

It is therefore more convenient to work with the log-transform of $\underline{X}$, as we shall do from now on. Define $\underline{W} := \underline{X}/\Omega$ as the variable of interest divided by its lower bound, or using lower cases for logs, $\underline{w} := \ln \underline{W} = \underline{x} - \omega$ where $\underline{x} := \ln \underline{X}$ and $\omega := \ln \Omega$. Since $\underline{X} \geq \Omega$, $\underline{W} \geq 1$ and hence $\underline{w} \geq 0$; the realisations w can be directly calculated from the data. The log transform of a Pareto distributed random variable is the exponentially distributed:

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

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

where $\Gamma(\cdot)$ is the Gamma function, and where we imply that k has an integer value whenever we write $k!$. Where only a limited set of moments for $k < \alpha^{-1}$ exists for the Pareto distribution, all moments exist for the distribution of its log transform, making it particularly suitable for testing.

2.2 Estimation and Testing

As a diagnostic to test whether a distribution looks sufficiently Pareto-ish, the log-log (or Zipf) plot is often used. Take the log of equation (1), and equate the complementary distribution function with the empirical distribution's

3. If the lower bound is unknown, one can use the lowest observation on $\underline{X}$ in the data as an estimate for Ω and drop this observation afterwards.

4. It is also common to work with $\zeta = \alpha^{-1}$ (e.g., Gabaix 2009, 2016). Both notations have advantages and disadvantages. Our choice to work with α rather than its inverse is motivated by the fact that the maximum likelihood estimator of α is unbiased and hence that of α^{-1} is not.

5. $\alpha^{-1} \approx 1.5$ for most wealth distributions (Vermeulen 2018), and $\alpha^{-1} = 1.05$ for the U.S. firm size distribution (Luttmer 2011), and is close to 1 for city sizes (Eeckhout 2004).

order statistic ranking. Since

$$\text{plim } [\ln \text{Rank}[w]] = \ln \Pr \left[\underline{w} \geq w \right] = -\alpha^{-1} w,$$

we obtain the regression equation

$$\ln \text{Rank}[w] = \alpha^{-1} - \alpha^{-1} w + \varepsilon, \tag{5}$$

where ε is an error term. When plotted on a double-logarithmic scale, equation (5) predicts the data to form a straight line, with a slope given by $-\alpha^{-1}$. A plot reasonably close to a straight line is typically interpreted as evidence in favour of the Pareto distribution and the regression coefficient is interpreted as an estimate of α^{-1}. Clearly, this does not constitute a formal test of Paretianity; Eeckhout (2009) argues that visual inference based on the log-log plot is dangerous since the double-logarithmic scale distorts observations at the tails, causing many distributions to look almost linear. More importantly, the calculation of the left hand variable $\ln \text{Rank}[w]$ requires ordering the data. Hence, the error terms ε are not i.i.d., invalidating standard diagnostic tests for OLS. Futhermore, this procedure is known to introduce small-sample bias. Hence, the common solution, recommended by Gabaix and Ibragimov (2011), is to use $\ln \text{Rank}[w - 1/2]$ as the dependent variable.

The common alternative to the regression approach is to use the method of moments. Equation (4) for $k = 1$ implies that the mean $\overline{w}$ for the full sample of billionaires can be used as a method-of-moments estimator of α (a bar above a random variable denotes its sample mean).[6] In fact, $\overline{w}$ is the maximum likelihood estimator of α (Hill 1975).[7] Since $\widehat{\alpha}^{\text{Hill}}$ is also the maximum likelihood estimator, it is efficient. Since it is linear in $\overline{w}$, it is unbiased. These features make the moment approach preferable to the regression approach *if* the model is correctly specified. If the data-generating process is not Pareto, in contrast, maximum likelihood gives very bad results, in particular when used to predict expected wealth $\text{E}\left[\underline{W}\right] = (1 - \alpha)^{-1}$ for the case where α is close to unity, or even below unity is biased. However, a rank order regression is likely to yield the same problem.

Test Statistics for Pareto: Equation (4) can be used for the construction of simple statistics to test the null hy–pothesis that the distribution of $\underline{w}$ is indeed Exponential (and by implication: the distribution of $\underline{W}$ is Pareto). Our proposed test statistics read:

$$\mathcal{R}_k := \frac{\text{E}\left[\underline{w}^k\right]}{k!\text{E}\left[\underline{w}\right]^k}, \quad \widehat{\mathcal{R}}_k := \frac{\overline{w^k}}{k!\overline{w}^k}. \tag{6}$$

Our statistics $\widehat{\mathcal{R}}_k$ can be interpreted as normalizations of higher non-central sample moments of $\underline{w}$ by corresponding powers of the sample mean. We refer to $\widehat{\mathcal{R}}_k$ for $k = 2$ and 3 to as the normalized variance and skewness, respectively. Appendix A derives the following general formula for the expectation and the variance of $\widehat{\mathcal{R}}_k$:

6. The sample mean of a stochastic variable is a stochast itself, so we should write $\overline{\underline{w}}$. We save on notation by omitting the underbar.

7. This is a consequence of the well-known fact that for distributions in the exponential family, moment estimators are equivalent to maximum-likelihood estimators (van der Vaart 2000).

Proposition 1. *Assume $\underline{w}$ is exponentially distributed. Then, for any integer $k \geq 1$, $\mathcal{R}_k$ defined in equation (6) satisfies:*

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

while $\mathrm{E}\left[\widehat{\mathcal{R}}_k\right]$ *has expectation and variance*

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

where N is the sample size.

Proof. See Appendix A. □

$\widehat{\mathcal{R}}_k$ converges asymptotically to unity for all k. In this paper, we specifically work with $\mathcal{R}_2$ and $\mathcal{R}_3$. Their expectation and variance read:

$$\begin{aligned} \mathrm{E}\left[\widehat{\mathcal{R}}_2\right] &= 1 + O\left(N^{-2}\right) & \mathrm{Var}\left[\widehat{\mathcal{R}}_2\right] &= N^{-1} + O\left(N^{-2}\right) \\ \mathrm{E}\left[\widehat{\mathcal{R}}_3\right] &= 1 + 7N^{-1} + O\left(N^{-2}\right) & \mathrm{Var}\left[\widehat{\mathcal{R}}_3\right] &= 10N^{-1} + O\left(N^{-2}\right) \end{aligned} \tag{7}$$

Apart from being unbiased up to terms of order N^{-2}, the variance of $\widehat{\mathcal{R}}_2$ is much smaller than that of $\widehat{\mathcal{R}}_3$. Hence, $\widehat{\mathcal{R}}_2$ is a more powerful test statistic than $\widehat{\mathcal{R}}_3$. Nevertheless, both $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ will prove to be useful for purpose of this paper. Our test of Pareto boils down to a two-sided test for the value of $\widehat{\mathcal{R}}_k$ for $k \geq 2$, where we reject Pareto if $\widehat{\mathcal{R}}_k$ significantly differs from unity.

What about measurement error? The data on top wealth are likely to be subject to substantial measurement error. We now show that classical measurement error affects neither the tail index α^{-1} of the distribution of top wealth, nor our test statistics. To show this, let $\underline{v}$ be a normally distributed measurement error with zero mean and variance of σ^2 and $\mathrm{Cov}\left[\underline{w}, \underline{v}\right] = 0$, so that observed log wealth $\widetilde{w}$ satisfies $\widetilde{w} := w + v$. The density function $f(\widetilde{w})$ of observed log wealth satisfies for large w (the upper tail)

$$f(\widetilde{w}) = \int_{-\infty}^{\infty} \sigma^{-1} \mathrm{e}^{-(\widetilde{w}-v)/\alpha} \phi\left(\frac{v}{\sigma}\right) \mathrm{d}v = \mathrm{e}^{-\widetilde{w}/\alpha} \int_{-\infty}^{\infty} \sigma^{-1} \mathrm{e}^{v/\alpha} \phi\left(\frac{v}{\sigma}\right) \mathrm{d}v = \mathrm{e}^{(\sigma/\alpha)^2/2} \mathrm{e}^{-\widetilde{w}/\alpha}, \tag{8}$$

where $\phi(\cdot)$ denotes the standard-Normal density function. The tail index of actual and observed log wealth is there–fore the same, though the density function of the latter is scaled up by a proportional constant $\mathrm{e}^{(\sigma/\alpha)^2/2} > 1$.

Intuitively, measurement error causes some individuals' log wealth to be overreported and other individuals' wealth to be underreported. Focus on one particular level of observed log wealth $\widetilde{w}$ and on one particular level of the absolute value of the measurement error $|v|$. Since the density function of w is declining, there are more people with actual log wealth $w = \widetilde{w} - |v|$ whose wealth gets overreported to be $\widetilde{w}$, than there are people with log wealth $w = \widetilde{w} + |v|$ whose wealth gets underreported. Hence, the density of the right tail of the distribution of observed log wealth is increased by some proportional factor compared to the distribution of actual log wealth. However, since the ratio of underreporters to overreporters is the same for all $\widetilde{w}$ due to the Exponential distribution of actual log wealth w, this constant of proportionality is the same for all $\widetilde{w}$. Hence, the moment ratios $\widehat{\mathcal{R}}_k$ remain unaffected.

Measurement error does therefore not affect our test procedure.[8]

However, the common procedure of using rank-size regressions as in equation (5) *will* be biased under measurement error. First, mismeasured w causes the OLS estimate for α^{-1} to be attenuated toward zero. A subtler problem is that this method depends on order statistics, and in effect assumes these are known perfectly. In empirical applications, however, measurement error introduces uncertainty about precise rankings, making order statistics an object to be estimated rather than something known *a priori* (Mogstad, Romano, Shaikh, and Wilhelm 2023). This introduces additional uncertainty into the estimator.

3 Data

For our main application, we use the *Forbes List of Billionaires* for the years 2001–2021. The dataset provides the names of billionaires, net worth, country of origin, age and citizenship. We classify billionaires according to their citizenship.

Forbes calculates net worth at the individual level, but aggregates family wealth, unless each family member has USD 1 billion or more after the split. On the one hand, as observed by Piketty (2014), this is likely to create an upward bias on individual fortunes around the threshold. On the other hand, given the difficulty for *Forbes* to estimate wealth components that are not publicly observed, some fortunes may well be biased downward. Moreover, they use available documentation and sometimes data provided by billionaires themselves to estimate their net worth. The number of billionaires and their wealth is likely to be underestimated in less developed countries or for wealth derived from nefarious activities. *Ex ante*, it is unclear which direction the measurement error goes; "rounding up" to 1 billion may put too much weight on the bottom of the list, whereas the difficulty of measuring liabilities and other poorly observable wealth components may overstate wealth for fortunes at the top. We follow the existing literature which uses rich lists like Forbes, as other sources are likely to underestimate the number of billionaires (Vermeulen 2016; Novokmet, Piketty, and Zucman 2018; Piketty, Yang, and Zucman 2019; Gomez 2023).

We cluster countries first in regions and then separate out sub-regions from some of these regions. The guiding principle for this clustering is to merge countries that are geographically connected and close in terms of GDP per capita, and have sufficient billionaire numbers to make the estimation of our statistics precise. With this in mind we define nine regions, which together cover all countries. The sub-region classification subdivides regions in coherent geographical units consisting of one or a small number of countries. As a rough threshold for the minimum number of billionaires for a country or a group of countries to be considered as a sub-region we use 40 billionaires in 2019. Countries that cannot easily be included in a sub-region are excluded from the sub-region classification. Hence, where region-classification encompasses the whole world, the sub-region classification excludes some countries. Therefore, a region can have more billionaires than the sum of its constituent sub-regions. Equation (6) implies that, under the assumption of top wealth being Pareto, a minimum of 40 billionaires in 2019 yields a maximal variance for our estimate of $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ of 0.025 and 0.25 respectively; likewise, the upward bias of $\widehat{\mathcal{R}}_3$ is at most 0.175. Since the total number of billionaires has gone up over time, the number of billlionaires and hence the precision of $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ are lower for the early years in our data. The regions China and India consist of one country and are therefore also classified as sub-region. The region Rest of the World has fewer than 40 billionaires in 2019. Moreover,

8. Note that this argument applies only to the extreme right tail of the distribution of wealth. As soon as $\widetilde{w} - |v|$ falls below the lower support of $\underline{w}$, the argument no longer applies. However, the literature typically assumes that the Pareto tail would start at a threshold much below a billion USD, perhaps at several million USD (e.g., Albers, Bartels, and Schularick (2022) assume it starts at the 99th percentile), so measurement error will not affect the distribution of billionaires' wealth.

this region is rather heterogeneous, with some really poor countries in Sub-Saharan Africa, as well as Afghanistan and Bangladesh, but also some middle income countries like South Africa. We therefore exclude it from our analysis from now on. We end up with 18 sub-regions. We use this sub-region classification for our regressions. Table 1 gives an overview of these regions and sub-regions.

Table 1 provides summary statistics for all regions and sub-regions. We report average numbers for $\widehat{\mathcal{R}}_2$, $\widehat{\mathcal{R}}_3$, and $\overline{w}$, and the average number of billionaires N (both raw and normalized by total population in millions). The summary statistics immediately reveal some striking facts. First, mean log wealth in Europe exceeds unity, implying that $\mathrm{E}\left[\underline{W}\right]$ does not exist for the Pareto distribution. Second, both $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ are consistently smaller than one in all regions and sub-regions, contradicting the prediction of a Pareto distribution. These estimates might be affected by a small sample bias, but equation (7) shows the bias for $\widehat{\mathcal{R}}_2$ to be of order $O\left(N^{-2}\right)$ only and for $\widehat{\mathcal{R}}_3$ to be upward rather than downward, making the low reported values for both statistics even more striking. Given the small number of billionaires in many of these sub-region/year observations (in particular for early years), it might be hard to reject the null $\widehat{\mathcal{R}}_k = 1$ for an individual observation. However, since we have multiple observations, we can ask: what is the likelihood that we observe this *distribution* of statistics? This we do in the next subsection.

4 Testing the model

This section formally tests the Pareto assumption discussed in Section 2. The results are reported in Table 2. We have 18 sub-regions $\times$ 21 years = 378 observations available for estimation. From equation (7), the variance of $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ is predicted to be proportional to N^{-1}, the inverse number of billionaires in a sub-region. Hence, the model exhibits heteroskedasticity and OLS is inefficient. We correct for this by using weighted least squares (WLS) with $\sqrt{N}$ as weights.

We clearly reject the null that $\widehat{\mathcal{R}}_2 = \widehat{\mathcal{R}}_3 = 1$. The intercepts are highly significantly different from unity. This remains the case if we do the following robustness checks: drop observations with fewer than 64 billionaires to reduce the small-sample bias in $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ (column (2) and (5)) and further drop the bottom and top 5% of observations for $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ as a robust regression (column (3) and (6)). We also report the theoretical root mean squared error (RMSE) derived from equation (7), except for column (3) and (6) where the selective dropping of observations invalidates the prediction for the RMSE. If all variation in $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ were sampling variation, the observed RMSE should be equal to the theoretical RMSE. This is clearly not the case. Using the sample with observations with $N \geq 64$, the observed RMSE is 31% of the theoretical RMSE for $\widehat{\mathcal{R}}_2$ and $0.54/\sqrt{10} \approx 17\%$ for $\widehat{\mathcal{R}}_3$. This is a further indication that the Pareto distribution does not fit the data.

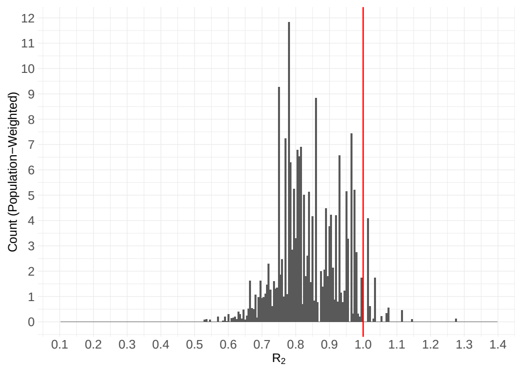
Figure 2 shows $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ for all observations: the great majority is smaller than unity. This holds true when we restrict attention to observations with more than 64 billionaires and (obviously) when we further drop the top and bottom 5% of observations. The main takeaway from Table 2 and Figure 2 is that Pareto distribution is strongly rejected across all sub-regions and years.

Table 1: Region Classification & Descriptive Statistics

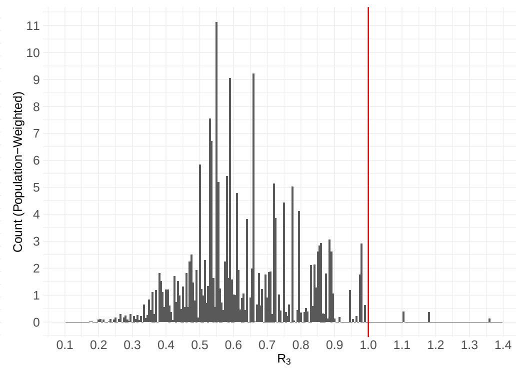
Region Classification		Statistics (Average 2001–2021)				
(Sub–)Region	Area	$\widehat{\mathcal{R}}_2$	$\widehat{\mathcal{R}}_3$	$\overline{w}$	N/L	N
North America	US and Canada	**0.866**	**0.688**	**0.92**	**1.35**	**472**
– U.S.		0.87	0.697	0.923	1.41	443
– Canada		0.797	0.564	0.894	0.832	29.2
Europe	excl. former USSR but incl. Baltics	**0.781**	**0.516**	**1.02**	**0.595**	**264**
– Germany		0.722	0.43	1.12	0.901	74.1
– British Islands	U.K. + Ireland	0.793	0.525	0.828	0.589	40.6
– Scandinavia	Sweden + Denmark + Norway + Finland	0.759	0.461	1.17	1.17	30.6
– France	incl. Monaco	0.779	0.502	1.27	0.402	26.5
– Alps	Switzerland + Austria + Liechtenstein	0.658	0.342	1.05	1.49	25
– Italy		0.792	0.535	1.01	0.405	24.1
China	excl. Taiwan, incl. Hong Kong	**0.929**	**0.771**	**0.794**	**0.135**	**188**
East Asia	Asia East of India and South-East of China; incl. Australia	**0.799**	**0.533**	**0.798**	**0.175**	**135**
– Southeast Asia	Thailand + Malaysia + Singapore	0.741	0.447	0.923	0.3	31.1
– Asian Islands	Taiwan + Philippines + Indonesia	0.766	0.49	0.727	0.101	38.5
– South Korea		0.843	0.63	0.673	0.368	18.7
– Japan		0.819	0.562	0.875	0.209	26.6
– Australia		0.81	0.537	0.736	0.802	19
India		**0.824**	**0.582**	**0.964**	**0.043**	**56.1**
Central Eurasia	former USSR except Baltics	**0.851**	**0.61**	**0.919**	**0.371**	**78**
– Russia		0.843	0.597	0.956	0.478	68.8
South America	incl. middle America and Mexico	**0.821**	**0.597**	**0.964**	**0.0978**	**59.8**
– Brazil		0.809	0.56	0.861	0.148	30
Middle East	Middle East incl. Turkey and Egypt excl. Iran	**0.858**	**0.635**	**0.765**	**0.568**	**60.4**
– Israel + Turkey		0.891	0.652	0.585	0.43	36.2
Rest of World	mainly Africa excl. Egypt, incl. Iran, Afghanistan, Pakistan, Bangladesh	**0.743**	**0.436**	**0.961**	**0.00556**	**11.8**
World		**0.856**	**0.647**	**0.896**	**0.183**	**1325**

Notes: Regions may include billionaires from countries not part of their constituent sub-regions. China and India count both as regions and sub-regions. N = total number of billionaires; L is total population in millions; $\overline{w}$ = mean log wealth; $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ are the normalized variance and skewness of mean log wealth from equation (6).

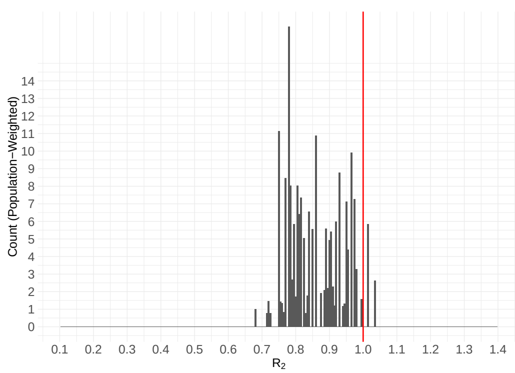
Figure 2: Distribution of Test Statistics $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$



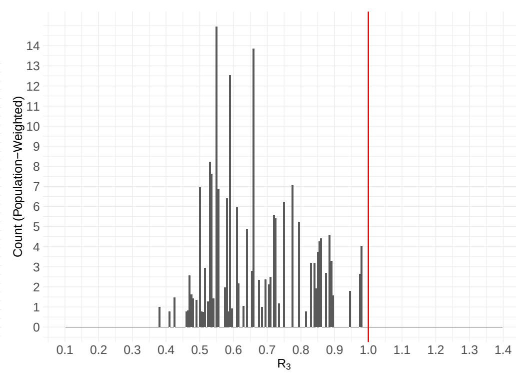
(a) $\widehat{\mathcal{R}}_2$, Full Sample



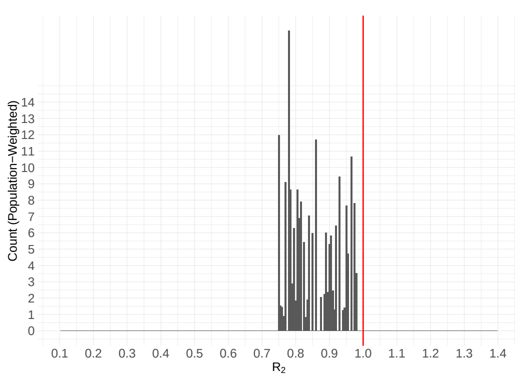
(b) $\widehat{\mathcal{R}}_3$, Full Sample



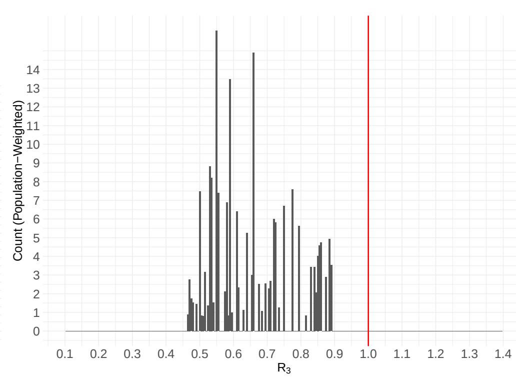
(c) $\widehat{\mathcal{R}}_2$, $N \geq 64$



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



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



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

Notes: Figures show the distribution of $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$, with all observation-years pooled. The red line indicates the predicted value under the null that wealth is Pareto-distributed, i.e., $\widehat{\mathcal{R}}_2 = \widehat{\mathcal{R}}_3 = 1$. All counts are weighted by the number of billionaires. Panels (a) and (b) use the full sample, panels (c) and (d) drop observations with fewer than 64 billionaires, and (e) and (f) further drop the top and bottom 5% of observations from (c) and (d).

Table 2: WLS Regressions, Pareto Test

Dependent Variables:	$\widehat{\mathcal{R}}_2$			$\widehat{\mathcal{R}}_3$		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Constant	0.82***	0.86***	0.86***	0.59***	0.66***	0.65***
	(0.005)	(0.02)	(0.01)	(0.008)	(0.03)	(0.02)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics						
Observations	378	75	67	378	75	67
RMSE	0.274	0.309	0.270	0.451	0.540	0.467
Theoretical RMSE	1	1		$\sqrt{10}$	$\sqrt{10}$	

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

Notes: Table reports regressions using the square root number of billionaires in a sub-region–year observation as weights. Parentheses underneath point estimates denote Driscoll and Kraay (1998) standard errors. Specifications (2) and (5) drop sub-region–year observations with fewer than 64 billionaires; specifications (3) and (6) further drop observations in the top and bottom 5% of the dependent variable. The theoretical root mean squared error is derived from Equation (7).

5 The Alternative: Weibull

5.1 Theory

We have documented that our test statistics $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ strongly and reject the null of a Pareto distribution. Moreover, the value of these statistics are very similar across observations, $\widehat{\mathcal{R}}_2$ being strongly clustered around 0.85 and and $\widehat{\mathcal{R}}_3$ around 0.65, see Table 1 and Figure 2. This suggests that there is a common pattern in the deviation from Paretianety and fosters the hope that our results not merely reject the hypothesis of a Pareto distribution but that there is an alternative distribution that describes the data more accurately.

We claim that the Weibull distribution gives a better description of top wealth than Pareto. Where the log of a Pareto distributed variable follows an Exponential distribution, the log of a Weibull distributed variable follows a Gompertz distribution.[9] Since our interest regards the distribution of top-wealth, we focus on the right tail of both distributions, the part above a lower bound $\underline{W} \geq 1$ in levels or $\underline{w} \geq 0$ in logs. Hence, it would be more precise to refer to the lower truncated Weibull and Gompertz distributions. In the interest of parsimony, we omit this further qualification in the subsequent discussion. Like it was more convenient to work with the Exponential distribution for log wealth $\underline{w} \geq 0$ rather than the Pareto distribution for wealth $\underline{W} \geq 1$, it is more convenient to work with the Gompertz distribution for $\underline{w}$ rather than the Weibull distribution for $\underline{W}$.

9. The Gompertz distribution is identical to the counter-Gumbel distribution: if $\underline{w}$ is Gompertz, $-\underline{w}$ is Gumbel. There is some confusion in the literature about the definition of the Gompertz distribution, where Gompertz is sometimes defined as Gumbel; see Kleiber and Kotz (2003), following Ahuja and Nash (1967). If $\underline{w}$ is Gumbel, $\underline{W}$ is inverse-Weibull or Frèchet. The confusion is permeated on Wikipedia, which gives our definition for Gompertz but then claims that the exponent of Gompertz is inverse-Weibull, citing Kleiber and Kotz (2003). We follow the definition used in demography, which implies $\underline{W}$ to be (truncated-)Weibull. We thank Christian Kleiber for helpful conversations on this point.

The Gompertz distribution for $\underline{w} \geq 0$ reads:

$$\Pr\left[\underline{w} \geq w\right] = \exp\left(-\frac{e^{\gamma w} - 1}{\alpha\gamma}\right), \tag{9}$$

where $\gamma > 0$ is a parameter. Since $\lim_{\gamma\to 0} \frac{e^{\gamma w}-1}{\alpha\gamma} = w/\alpha$, we recover the Exponential distribution for $\gamma \to 0$, compare equation (3).

The difference between the Exponential and the Gompertz distribution is understood most easily by comparing their hazard rates $H(w) := \Pr\left[\underline{w} = w\right] / \Pr\left[\underline{w} \geq w\right]$. Table 3 lists the hazard rates of four distributions: next to the Exponential and Gompertz distribution, we also list the Weibull and Normal distribution. All four distributions are parameterized such that the hazard rate for $w = 0$ is equal to α^{-1}. While the hazard is constant for the Exponential distribution, it is increasing for Gompertz. The latter hazard diverges to infinity for $w \to \infty$. For every unit increase $(\mathrm{d}w)$ in wealth, the relative increase of the hazard is $\gamma \mathrm{d}w$. Conditional on being part of the group billionaires with a log wealth larger than or equal to w, the probability of leaving the group of billionaires with an even higher wealth than w is increasing in w, reducing the probability of finding a billionaire with even higher wealth. The Gompertz distribution therefore has a thinner tail than the Exponential distribution.

Table 3: Hazard Rates for Four Distributions

Distribution	Complement Distribution $\Pr\left[\underline{w} \geq w\right]$	Density $\Pr\left[\underline{w} = w\right]$	Hazard Rate $H(w)$
Exponential	$e^{-w/\alpha}$	$\alpha^{-1}e^{-w/\alpha}$	α^{-1}
Gompertz	$\exp\left(-\frac{e^{\gamma w}-1}{\alpha\gamma}\right)$	$\alpha^{-1}\exp\left(\gamma w - \frac{e^{\gamma w}-1}{\alpha\gamma}\right)$	$\alpha^{-1}e^{\gamma w}$
Weibull	$\exp\left(-\frac{(1+w)^{\gamma+1}-1}{\alpha(\gamma+1)}\right)$	$\alpha^{-1}(1+w)^{\gamma} \Pr\left[\underline{w} \geq w\right]$	$\alpha^{-1}(1+w)^{\gamma}$
Normal	$\Phi\left(-\frac{1+w}{\alpha\psi_\alpha}\right)$	$(\alpha\psi_\alpha)^{-1}\phi\left(\frac{1+w}{\alpha\psi_\alpha}\right)$	$(\alpha\psi_\alpha)^{-1}\frac{\phi\left(\frac{1+w}{\alpha\psi_\alpha}\right)}{\Phi\left(-\frac{1+w}{\alpha\psi_\alpha}\right)}$

Notes: $\Phi(\cdot)$ and $\phi(\cdot)$ are the distribution and density functions respectively of the standard normal distribution; ψ_α solves $\psi_\alpha\Phi\left(-(\alpha\psi_\alpha)^{-1}\right) = \phi\left((\alpha\psi_\alpha)^{-1}\right)$; $\psi_1 = 1.330$.

As discussed, the distributions are parameterized such that the hazard rate is equal to α^{-1} for the lower bound $\underline{w} = 0$, or equivalently, $\underline{x} = \omega$. Since the hazard is constant for the Exponential distribution, the choice of the value of ω does not matter for the value of α. For Gompertz, as for the other two distributions, the hazard rate varies with log wealth relative to its lower bound $\underline{w} := \underline{x} - \omega$. Hence, the choice of the lower bound ω matters for the value of α. To see this, presume the distribution of top log wealth $\underline{x}$ to be Gompertz. Let α_0 be the value of α when we set $\omega = 0$, so $\underline{w} = \underline{x}$. Now consider what happens when we decide to study the distribution of the wealth of bi-billionaires instead of billionaires, so $\omega = \ln 2$ instead of $\omega = \ln 1 = 0$. Then, the hazard rate at the lower bound would equal $\alpha_0^{-1}e^{\gamma \ln 2}$ rather than α_0^{-1}. Hence, a general expression for the relation between α and ω reads

$$\ln \alpha_\omega = \ln \alpha_0 - \gamma\omega. \tag{10}$$

While γ is a structural parameter, the value of the parameter α depends on the lower bound of our data.

Generalizing this argument, one might hypothesize that the sub-region/time observations of the distribution of billionaires follow a Gompertz distribution with a common γ parameter. But then α must differ between these observations, since one has to dive much deeper into the right tail to encounter a billionaire, in 2001 rather than 2021 (just due to growth and inflation), or for that matter, in India rather than the U.S. (with 0.043 and 1.41 billionaires per million respectively, see Table 1). The parameter α can therefore be expected to be lower in 2001 rather than 2021 and lower in India rather than the United States.

For the sake of comparison, we also show the hazard rates of the Weibull and Normal distribution in Table 3.[10] Similar to Gompertz, the hazards of both distributions are increasing and diverge to infinity for $w \to \infty$, however, at a slower rate. For the Weibull distribution, this can be seen directly from the hazard rate, since while $e^{\gamma w}$ and $(1+w)^{\gamma}$ are equal for $w = 0$, $e^{\gamma w}$ exceeds $(1+w)^{\gamma}$ for any $w > 0$ and $e^{\gamma w}/(1+w)^{\gamma}$ diverges to infinity for $w \to 0$ for any $\gamma > 0$.

For the Normal distribution, this is more difficult to see since there is no closed form solution for the hazard rate. Figure 3 plots the hazard rates for $\alpha = 1$ and for $\gamma = 1/2, 1/3$, and $1/4$ (the latter for Gompertz only).[11] For $\gamma > 1/2$, the hazard increases more quickly for the Gompertz rather than the Normal distribution. For $\gamma < 1/3$, the hazard increase more rapidly for Normal initially, but eventually the hazard for Gompertz will dominate for any $\gamma > 0$.[12] The Gompertz distribution therefore has an even thinner extreme right tail than the Normal distribution. The fact that this feature shows up only in the extreme right tail for $\gamma < 1/2$ might explain why our empirical claim that top log wealth is Gompertz rather than Exponential has gone unnoticed in the empirical literature so far.

Moments of Weibull and Gompertz: We gather our results on the moments of Weibull and Gompertz in the following proposition, compare Wingo (1989).

Proposition 2. *Assume $\underline{W}$ is Weibull distributed and hence $\underline{w}$ is Gompertz with the complement distribution function given in equation (9). Then, the moments for $k > 0$ of $\underline{W}$ and $\underline{w}$ read*

$$\mathrm{E}\left[\underline{W}^k\right] = (\alpha\gamma)^{k/\gamma}\, \mathrm{e}^{(\alpha\gamma)^{-1}}\Gamma\left(1 + k/\gamma, (\alpha\gamma)^{-1}\right), \tag{11}$$

$$\mathrm{E}\left[\underline{w}^k\right] = \gamma^{-k} h\left(\alpha\gamma, k\right), \tag{12}$$

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

$$h\left(\alpha\gamma, k\right) \coloneqq k\mathrm{e}^{(\alpha\gamma)^{-1}} \int_0^\infty q^{k-1} \exp\left(-(\alpha\gamma)^{-1}\mathrm{e}^{q}\right) \mathrm{d}q,$$
$$h\left(\alpha\gamma, 1\right) = \mathrm{e}^{(\alpha\gamma)^{-1}} \mathrm{Ei}\left((\alpha\gamma)^{-1}\right),$$

where $q \coloneqq \gamma w$ *and* $\mathrm{Ei}(\cdot)$ *is the exponential integral.*

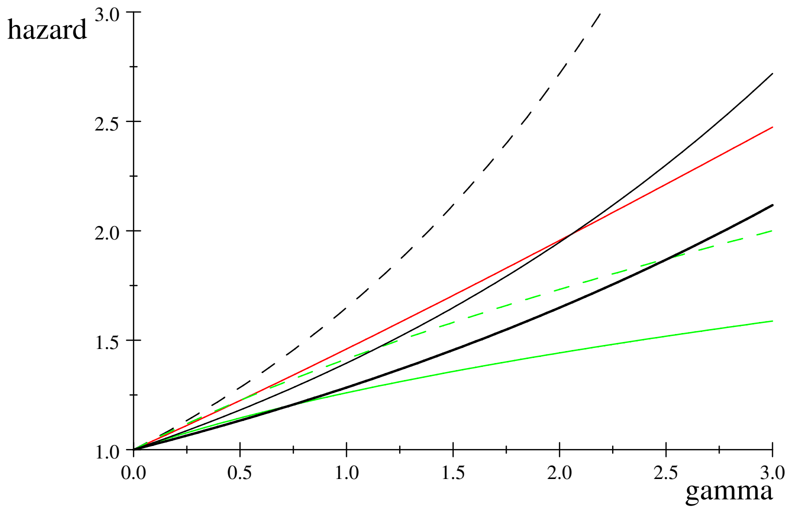
Proof. See Appendix B. □

10. When wealth follows a Weibull distribution, log wealth follows a Gompertz distribution. In Table 3, we analyze the case where *log wealth* rather than *wealth* follows a Weibull distribution. Similarly, when log wealth follows a Normal distribution, the distribution of wealth is log Normal .

11. For the Normal distribution, we must introduce an additional parameter ψ_α that is a function of α that achieves that $H(0) = \alpha^{-1}$, see the footnote to Table 3 for details.

12. This follows from the fact that the hazard of the Normal distribution converges to an increasing linear function since $\lim_{x\to\infty} \frac{\phi(x)}{x\Phi(-x)} = 1$. An increasing exponential function will eventually always dominate an increasing linear function.

Figure 3: Variation of Hazard Rates with γ



Notes: Black lines = Gompertz, green = Weibull, red = Normal; dashed lines: $\gamma = 1/2$, continuous lines: $\gamma = 1/3$, fat line: $\gamma = 1/4$ (Gompertz only).

While the moments of Gompertz are not available in closed form, the mean can be easily computed with routines available in all common statistical software packages and higher moments can be calculated by simple numerical integration. Recall that moments of $\underline{W}$ higher than α^{-1} do not exist when $\underline{w}$ is Exponential and hence $\underline{W}$ is Pareto. In contrast, when $\underline{w}$ is Gompertz and hence $\underline{W}$ is Weibull, all moments of $\underline{W}$ exist.

We can use equations (11) and (12) for the calculation of the asymptotic expection of the $\widehat{\mathcal{R}}_k$-statistics as defined in equation (3), which are a function $\mathcal{R}_k(\alpha\gamma)$ of the parameter product $\alpha\gamma$ for the Gompertz distribution:

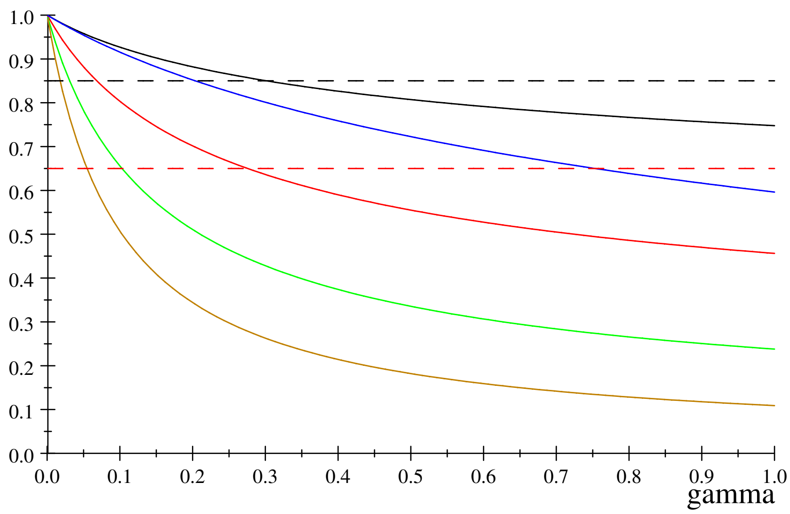
$$\mathcal{R}_k(\alpha\gamma) = \frac{h(\alpha\gamma, k)}{k! h(\alpha\gamma, 1)^k} = \frac{\int_0^\infty q^{k-1} \exp\left(-(\alpha\gamma)^{-1} e^q\right) dq}{(k-1)! e^{(\alpha\gamma)^{k-1}} \mathrm{Ei}\left((\alpha\gamma)^{-1}\right)^k}. \tag{13}$$

Figure 4 plots the values of $\mathrm{E}\left[\underline{w}\right]$ (blue) and $\mathcal{R}_k(\alpha\gamma)$ for k from 2 until 5 (black, red, green, and beige re–spectively). For $\gamma \to 0$ and hence $\alpha\gamma \to 0$, the Gompertz distribution converges to the Exponential distribution. Indeed, $\mathrm{E}\left[\underline{w}\right]$ and $\mathcal{R}_k(\gamma)$ are equal to unity in this case, the same as for the Exponential distribution. For $\alpha\gamma > 0$, $\mathcal{R}_k(\alpha\gamma) < 1$ and $\mathcal{R}_k(\alpha\gamma) > \mathcal{R}_{k+1}(\alpha\gamma)$. We also plotted the mean value of $\widehat{\mathcal{R}}_2 \approx 0.85$ and $\widehat{\mathcal{R}}_3 \approx 0.65$ from Table 1. Figure 4 documents that these values are both consistent with a value of $\alpha\gamma$ between 0.25 and 0.30. Note that these values depend on the *product* of α and γ. Hence, the availability of data on both $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ provides over-identifying restrictions that we can use to test the null hypothesis of top log wealth beign Gompertz. The consistency of both estimates for $\alpha\gamma$ suggest that the null hypothesis fits the data well. We return to this issue in Section 6.

While the hazard rate will continue to diverge to infinity irrespective how far one dives into the right tail of the distribution (that is: for an arbitrarily high lower bound ω) and therefore never converges to a constant, as predicted by the Pareto distribution, $\mathcal{R}_k(\alpha\gamma)$ does converge to its unit value obtained under Pareto. An ever higher lower bound ω implies an ever lower value of α, see equation (10), and $\lim_{\alpha\to 0} \mathcal{R}_k(\alpha\gamma) = 1$, the value of $\mathcal{R}_k$ under

Pareto. The reason for this apparent contradiction is that the increasing values of the hazard rate become increasingly irrelevant as the initial value of the hazard rate is already high, so that by time the hazard rate has gone up noticeably, most of the remaining rich have already left the pool of even richer people. While there is no convergence in the extreme right tail of Weibull to Pareto for hazard rates, there is therefore convergence in the truncated moments.

Figure 4: Variation of Mean Log Wealth and $\mathcal{R}_k(\gamma)$ with γ



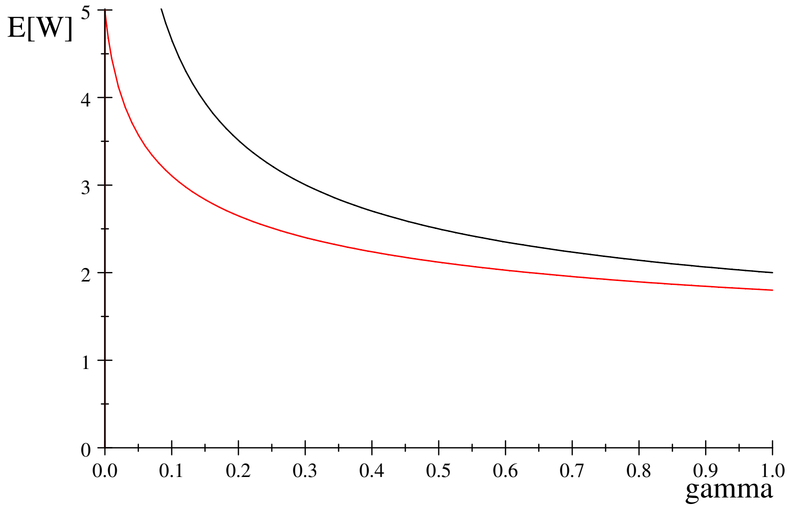
Notes: Values for $\alpha = 1$. Blue = mean log wealth, black = $\mathcal{R}_2$, red = $\mathcal{R}_3$, green = $\mathcal{R}_4$, and beige = $\mathcal{R}_5$. Horizontal dashed lines represent the empirical mean values of $\mathcal{R}_2$ (black, 0.85) and $\mathcal{R}_3$ (red, 0.65).

Figure 5 plots the expected level of wealth $\mathrm{E}\left[\underline{W}\right]$ as a function of γ, for $\alpha = 0.80$ (red) and $\alpha = 1$ (black). Since log wealth is Gompertz, the level of wealth is Weibull. Since Gompertz converges to Exponential for $\gamma \to 0$, Weibull converges to Pareto. For Pareto, $\mathrm{E}\left[\underline{W}\right] = (1-\alpha)^{-1}$, which is 5 for $\alpha = 0.80$ and infinite for $\alpha = 1$. Figure 5 confirms that $\mathrm{E}\left[\underline{W}\right]$ converges to these values when γ converges to zero. However, Figure 5 also shows that $\mathrm{E}\left[\underline{W}\right]$ is highly sensitive to even slight variations in γ for any $\gamma < 0.2$, in particular for values of α close to or above unity.

5.2 Estimation

The parameters α and γ of the Gompertz distribution can be estimated with the method of moments, like in Wingo (1989). We opt to estimate these parameters by maximum likelihood using the density function of w. The derivation of the likelihood function is in Appendix C. The procedure yields an explicit expression for the estimator $\widehat{\alpha}$ as a

Figure 5: Variation of Mean Wealth with γ



Notes: Values for $\alpha = 0.8$ (red) and $\alpha = 1$ (black).

function of the parameter γ and the data on w and it yields an implicit expression for $\widehat{\gamma}$:

$$\widehat{\alpha} = \gamma^{-1}\left(\overline{e^{\gamma w}} - 1\right), \tag{14}$$

$$\mathrm{SE}(\widehat{\alpha}) = \widehat{\alpha}\sqrt{N}^{-1},$$

$$N^{-1}\log\mathcal{L}(\gamma) = \ln\gamma - \ln\left(\overline{e^{\gamma w}} - 1\right) + \gamma\overline{w}. \tag{15}$$

$$\widehat{\gamma}^{-1} = \frac{\overline{w e^{\widehat{\gamma} w}}}{\overline{e^{\widehat{\gamma} w}} - 1} - \overline{w},$$

$$\mathrm{SE}(\widehat{\gamma}) = \sqrt{N\left(\frac{\overline{(w^2+1)\,e^{\widehat{\gamma} w}} - 2\overline{w}\,\overline{w e^{\widehat{\gamma} w}}}{\overline{e^{\widehat{\gamma} w}} - 1} + \overline{w}^2\right)}^{-1}.$$

The first two lines are the expression for $\widehat{\alpha}$ and its standard error, the third line is the concentrated log likelihood, while the fourth and fifth line give the implicit condition for $\widehat{\gamma}$ and its standard error. The latter is an explicit function of $\widehat{\gamma}$ and $\overline{w}$. For $\widehat{\gamma} = 0.25$ and $\overline{w} = 1$, $\mathrm{SE}(\widehat{\gamma}) \approx 0.166\sqrt{N}^{-1}$, so that $\widehat{\gamma}$ can be estimated reliably for $N \geq 64$. The fact that we have a simple expression for $\widehat{\alpha}$ as a function of γ and that we can therefore eliminate α from the likelihood is convenient. As discussed before, the value of α depends on the choice of the lower bound ω. One might presume that the data for various years or subregions are characterized by a common γ but by different values of α, depending how far one has to go into the right tail of in the distribution to arrive at a billionaire. One can first estimate a common $\widehat{\gamma}$ for all sub-region/year observations with $N \geq 64$ and then use the first line to calculate a separate $\widehat{\alpha}$ for each observation setting γ to this common $\widehat{\gamma}$. The restriction of a common γ can be tested using the concentrated likelihood for a likelihood ratio test.

An alternative, discussed in Appendix D, is to use maximum likelihood estimation based on the hazard rates

$H(w) := \Pr\left[\underline{w} = w\right] / \Pr\left[\underline{w} \geq w\right]$, where $\Pr\left[\underline{w} \geq w\right]$ is calculated from the data. Using the likelihood of hazard rates rather than w itself comes with a disadvantage: the calculation of an empirical estimate for $\Pr\left[\underline{w} \geq w\right]$ requires ordering the data, similar to the log rank regression. In fact, $\Pr\left[\underline{w} \geq w\right]$ is the inverse of an order statistic. The observations on the hazard are therefore not mutually independent. Maximum likelihood estimation based on w rather than $H(w)$ is therefore more efficient.

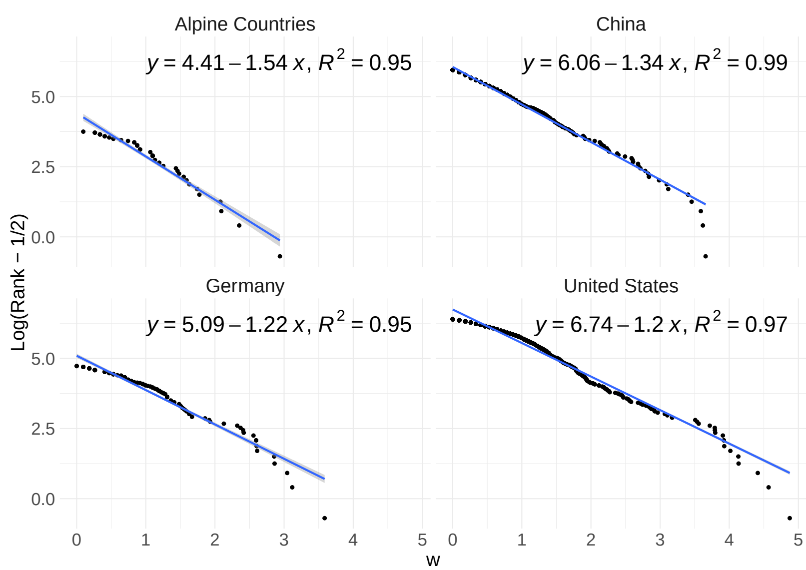
When we use a log rank regression instead we obtain:

$$\ln \text{Rank}\left[w\right] = \ln \Pr\left[\underline{w} \geq w\right] = (\alpha\gamma)^{-1}\left(1 - e^{\gamma w}\right) = -w/\alpha - \frac{1}{2}\frac{\gamma}{\alpha}w^2 + O\left(w^3\right). \tag{16}$$

The quadratic term is close to zero for small γ and given the limited range of w in most empirical settings (w rarely exceeds 4 in the data on top wealth) the second order term is likely to be insignificant. This explains the ubiquitous linearities found in log-log plots in the literature.

We can directly show this spurious linearity in our setting. Figure 6 shows the results of four rank-size regres–sions, for the year 2019. The four regions studied are the Alpine Countries (Austria, Switzerland and Liechtenstein), China, Germany, and the United States. We chose these regions to illustrate that the illusory nature of rank-size re–gressions is an issue across economic and institutional environments and regardless of sample size.

Figure 6: Spurious Log-Rank – Log-Size Regressions



Notes: Gabaix and Ibragimov (2011) regressions for the year 2019. $\widehat{\gamma} = 0.885$, (Alpine Countries), 0.129 (China), 0.503 (Germany), and 0.304 (United States), all statistically significant.

Figure 6 reveals that all four cases have rank-size plots which are almost perfectly linear. This is despite the fact that all cases also have estimated values for γ which are significantly different from 0. The more populous regions of the United States and China have lower values of γ[13] and are therefore closer to the Pareto benchmark. Correspond–

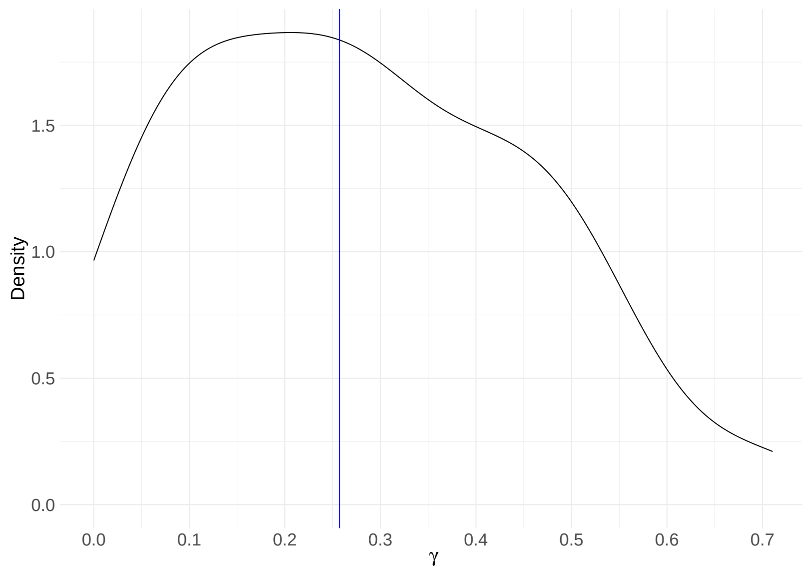
13. This is a systematic finding: more populous regions have higher tail inequality. In Teulings and Toussaint (2023), we explore this issue further.

ingly, their rank-size plots are closer to the perfect straight line than those of the Alpine Countries and Germany. Nevertheless, we significantly reject Pareto and accept Weibull in all cases. This cautions against the widespread use of rank-size regressions as “evidence” for Pareto.

6 Testing Weibull

This section implements the estimators derived in the previous section. We use equation (14) to estimate $\widehat{\gamma}$ and $\widehat{\alpha}$ for the 75 observations with at least 64 billionaires. Figure 7 shows the distribution of $\widehat{\gamma}$. Recall that $\gamma = 0$ recovers the Pareto distribution. $\widehat{\gamma}$ is positive[14] for all but three observations. The vertical blue line shows the median value $\widehat{\gamma}^{\text{Med}} \approx 0.257$.

Figure 7: Distribution of γ Across Sub-Region–Years



Notes: Figure plots the density of all values of γ estimated on the sub-region–year observations of the *Forbes List of Billionaires* with at least 64 billionaires. The vertical blue line indicates the median value $\gamma^{\text{Med}} \approx 0.257$.

Equation (13) shows that $\mathcal{R}_k(\alpha\gamma)$ is a declining function of $\alpha\gamma$, or in logs of $\ln\alpha + \ln\gamma$. As a rough test of this prediction, we regress $\widehat{\mathcal{R}}_k$ for $k = 2, 3$ on $\ln\widehat{\alpha}$ and $\ln\widehat{\gamma}$. The test is only a rough test since the estimates on $\ln\widehat{\alpha}$ and $\ln\widehat{\gamma}$ are contaminated by an estimation error. Hence, their coefficients are biased toward zero by attenuation bias. Equation (13) predicts that the coefficients on both variables should be negative and equal for $\widehat{\mathcal{R}}_2$ and for $\widehat{\mathcal{R}}_3$. Figure 4 shows that $\mathcal{R}_3(\alpha\gamma)$ declines more strongly with $\alpha\gamma$ than $\mathcal{R}_2(\alpha\gamma)$, so that the coefficients are predicted to be larger in absolute value for $\widehat{\mathcal{R}}_3$ rather than $\widehat{\mathcal{R}}_2$. The results from this test are reported in Table 4 for observations with at least 64 billionaires. We drop the top and bottom 5% of observations for robustness.[15]

All predictions are confirmed by the data. The coefficients on $\ln\widehat{\alpha}$ and $\ln\widehat{\gamma}$ are all negative and significant and the

14. Numerically, all estimates are positive, but we treat any observation for which $\widehat{\gamma} < 10^{-6}$ as $\widehat{\gamma} = 0$.

15. These observations are ($\widehat{\gamma}$ in parentheses): The U.S. in 2003 ($< 10^{-6}$) and 2005 ($< 10^{-6}$), China in 2011 ($< 10^{-6}$) and 2012 (0.009), Germany in 2014 (0.709) and 2018 (0.582), the British Islands in 2021 (0.711) and Brazil in 2021 (0.611).

Table 4: Weibull Predictions

Dependent Variables:	$\widehat{\mathcal{R}}_2$	$\widehat{\mathcal{R}}_3$	
Model:	(1)	(2)	(3)
Variables			
Constant	0.781***	0.524***	-0.788***
	(0.024)	(0.027)	(0.064)
$\ln \widehat{\alpha}$	-0.115**	-0.207***	
	(0.038)	(0.032)	
$\ln \widehat{\gamma}$	-0.058***	-0.101***	
	(0.015)	(0.013)	
$\widehat{\mathcal{R}}_2$			1.68***
			(0.085)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics			
Observations	67	67	67
R^2	0.941	0.986	0.950
Adjusted R^2	0.939	0.986	0.949
RMSE	0.066	0.054	0.105
F-Stat	1.212	5.780	

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

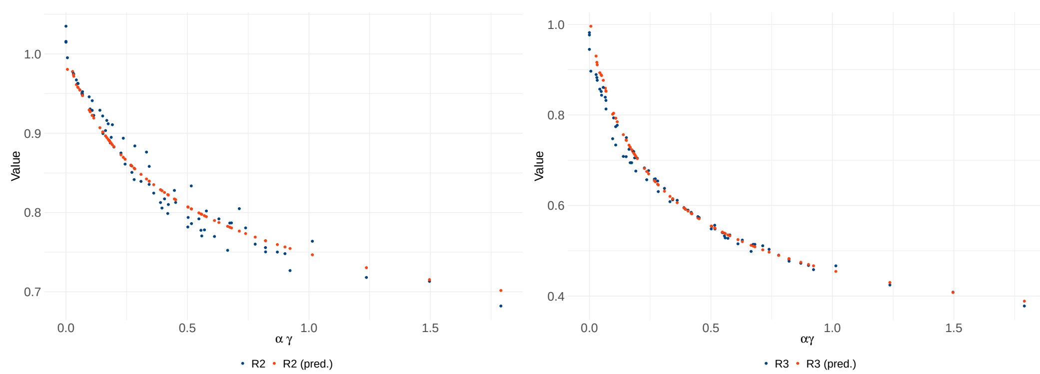
Notes: Driscoll and Kraay (1998) standard errors in parentheses. The regressions are estimated on observations with at least 64 billionaires, with the top and bottom 5% dropped. The bottom line reports F-tests on the restriction that $\ln \widehat{\alpha}$ and $\ln \widehat{\gamma}$ have equal coefficients.

coefficients are indeed more negative for $\widehat{\mathcal{R}}_3$. The restriction that the coefficient on $\ln\widehat{\alpha}$ and $\ln\widehat{\gamma}$ are equal is accepted for both $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$, as evidenced by the insignificant F-statistics. Moreover, the fact that both the coefficient and standard error for $\ln\alpha$ exceeds those for $\ln\gamma$ is predicted by our model.[16]

Similarly, since both $\mathcal{R}_2(\alpha\gamma)$ and $\mathcal{R}_3(\alpha\gamma)$ are increasing functions of the same parameter $\alpha\gamma$, $\mathcal{R}_3$ can be expressed as a function of $\mathcal{R}_2$. Ignoring measurement error and sampling variation, $\widehat{\mathcal{R}}_2$ should therefore be a perfect predictor of $\widehat{\mathcal{R}}_3$. The correlation coefficient between both statistics is indeed close to one (0.97 to be precise). Moreover, we run the regression $\widehat{\mathcal{R}}_3 = \beta_0 + \beta_1\widehat{\mathcal{R}}_2$, reported in column (3). Figure 4 suggests that β_1 should be about 1.5 and Table 1 suggests that $\overline{\widehat{\mathcal{R}}}_3 = \beta_0 + \beta_1\overline{\widehat{\mathcal{R}}}_2 = \beta_0 + 0.85\beta_1 = 0.65$. The estimation results closely fit this observation. Conditional on the empirical value $\widehat{\mathcal{R}}_2$, the prediction for $\widehat{\mathcal{R}}_3$ matches its empirical value almost perfectly.

We now directly test the over-identifying restrictions implied by equation (13). For each observation with at least 64 billionaires, we calculate the theoretical values of $\mathcal{R}_2$ and $\mathcal{R}_3$, based on the estimated values of $\widehat{\alpha}$ and $\widehat{\gamma}$. We then contrast these theoretical values with the values in the data.

Figure 8: Predicted versus Realized Values of $\mathcal{R}_2$ and $\mathcal{R}_3$



(a) $\mathcal{R}_2$ (b) $\mathcal{R}_3$

Notes: Red dots are the theoretical values for an observation based on equation (13), given its estimated $\widehat{\alpha}$ and $\widehat{\gamma}$. Blue dots show the empirical values.

Figure 8 shows the results. Since both panels depend on the same value $\alpha\gamma$, a close fit in both panels provides an over-identifying test. The Figure shows that this test is easily accepted. The data for both $\mathcal{R}_2$ and $\mathcal{R}_3$ are extremely closely in line with their theoretical values. This provides direct evidence in favor of Weibull as a well-fitting distribution.

Our final test runs a horse race between Weibull and Pareto for the prediction of mean wealth in 2018. For the Weibull distribution, we assume that all observations have a common value for $\widehat{\gamma}$, so that both Weibull and Pareto have a single parameter $\widehat{\alpha}$ to fit the data for each observation. We use the median value $\widehat{\gamma}^{\text{Med}} = 0.257$. This horse race provides a fair comparison between both distributions, since we have a single parameter for each observation for both distributions. We test the restriction of the common γ equal to $\widehat{\gamma}^{\text{Med}}$ by means of a likelihood ratio test. We sum the concentrated likelihood (15) across all 75 observations, and compare this to the sum of restricted likelihoods where we impose $\gamma = \widehat{\gamma}^{\text{Med}}$ on all observations. The likelihood ratio statistic (75 degrees of freedom) is 202, rejecting

16. For small $\widehat{\gamma}$, equation (14) implies $\text{SE}(\widehat{\gamma}) \approx \widehat{\gamma}\sqrt{2N}^{-1}$ and hence $\text{SE}(\ln\widehat{\alpha}) = \sqrt{N}^{-1} > \sqrt{2N}^{-1} \approx \text{SE}(\ln\widehat{\gamma})$. In our regression sample, $\text{Var}(\widehat{\alpha}) \approx 0.098$ and $\text{Var}(\widehat{\gamma}) \approx 0.023$, conforming to our predictions.

the restriction of a common γ. Using a common γ therefore works against the Weibull distribution, as we impose a restriction which does not fit the data perfectly. As we shall see, this is of minor concern for the quality of predictions of the Weibull model.

We then use equation (14) to estimate a value of α specific for each observation. For the Pareto estimate of α, we use $\widehat{\alpha}^{\text{Hill}} = \overline{w}$, see Equation (4). Table 5 reports actual and estimated mean wealth among billionaires for Weibull and Pareto for all sub-regions, together with the estimated values of α for both distributions.

Table 5: Predicted vs. Realised Values of Mean Billionaire Wealth, 2018

	Mean Wealth			$\widehat{\alpha}$	
Sub-Region	Data	Weibull	Pareto	Weibull	Pareto
United States	5.29	5.65	∞	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	4.35	∞	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	3.3	3.12	4.97	0.983	0.799
South-East Asia	3.52	3.85	16.7	1.16	0.94
Asia Islands	2.85	2.86	4.17	0.912	0.76
South Korea	2.88	2.8	3.8	0.894	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	3.97	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

Notes: Sub-regions are defined in Table 1. The Weibull prediction is made using Equation (14), where we take the median maximum likelihood estimate of $\widehat{\gamma}$ ($\widehat{\gamma}^{\text{Med}} \approx 0.257$), and where we calculate α using equation (14). The Pareto prediction is made using Equation (4), with $\overline{w}$ as the maximum likelihood estimator of α.

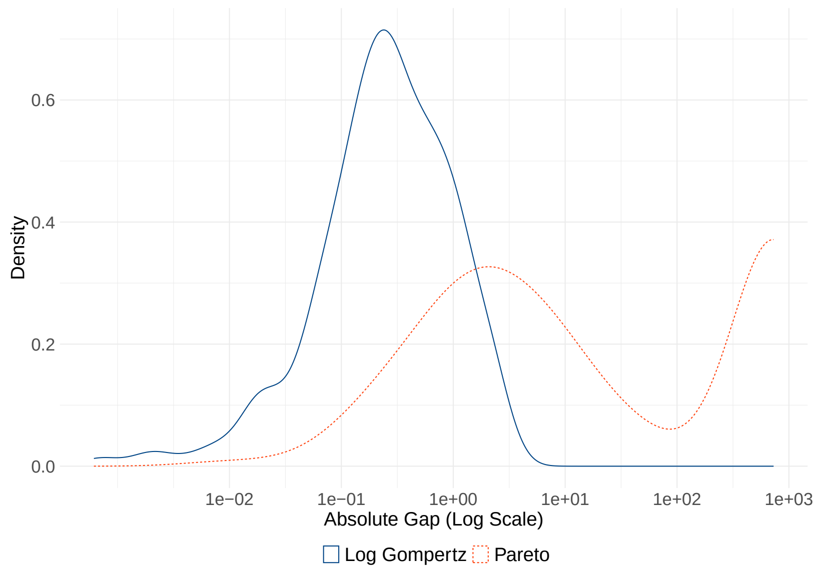
Actual mean wealth varies widely across sub-regions. In the sub-region Israel + Turkey, billionaires have a mean wealth of 2.1 billion, whereas in the U.S., this is almost 5.3 billion. One super-rich billionaire can have an enormous influence on these statistics; for instance, France's mean wealth of 7.4 is heavily affected by the wealth of LVMH owner Bernard Arnault (estimated to be about 72 billion in 2018).

The mean wealth predicted by the Weibull model traces actual mean wealth remarkably well. Four predictions are almost identical to the real value (Canada, Asia Islands, Australia, Israel+Turkey), and a further nine estimates are within half a point of the real value. The remaining five estimates, moreover, are also reasonably close, with the largest gap being for Germany (1.06 points).

Now compare the prediction of the Weibull and the Pareto model. Table 5 shows $\widehat{\alpha}$ to be above 1 for seven sub-regions, including the United States. This immediately results in infinite values for these seven regions. We note a further five cases where $\widehat{\alpha}$ is close to 1, such as Scandinavia. Here, we obtain finite values, but these are obviously so large as to be meaningless. This leaves us with six reasonable estimates. None of these, however, are closer to the real value than truncated-Weibull. We conclude that Pareto is not a useful model to predict mean wealth, whereas Weibull performs well.

This conclusion is not driven by our choice of year. We calculate the absolute value of the gap between real and predicted values for all observation-years, for both Weibull and Pareto. Figure 9 plots the density of these gaps on a log scale. For visual purposes, infinite values for $\mathrm{E}[W]$ have been set to the maximal finite value of 733.

Figure 9: Absolute Prediction Gaps for Mean Wealth, Truncated-Weibull vs. Pareto



Notes: Figure shows the absolute gap between real mean wealth and the predictions made by truncated-Weibull and Pareto. Infinite values for predicted mean wealth have been set to the maximal predicted finite value, 733.

It is immediately obvious that truncated-Weibull performs very well, with the vast majority of gaps being less than 1 in absolute value and a significant number even less than 0.1. This stands in stark contrast to the Pareto predictions. Although Pareto can result in reasonable estimates, the heavy right tail shows a nontrivial fraction of observations where the predictions gaps are between 10 and 100. The jump in the right tail further illustrates the fragility of Pareto; 129 out of 378 observations are predicted to have infinite mean wealth.

7 Additional Applications

7.1 Cities

Might our conclusion that top wealth is distributed Weibull rather than Pareto apply to other phenomena? We check this by applying our testing procedure to two other distributions, U.S. city and U.S. firm size. There has been considerable debate about whether city size is distributed Pareto or something thinner-tailed. Proponents for Pareto, such as Krugman (1996) and Gabaix (1999), typically find evidence for this relationship by observing a linear slope in a log-log plot of the largest 135 or so metropolitan statistical areas (MSAs). Eeckhout (2004, 2009) criticizes this approach on both statistical and substantial grounds. His statistical criticism, much like ours in Section 2, consists of pointing out that a log-log plot distorts observations at the tails of the distribution, and that formal statistical tests such as a Kolmogorov-Smirnov test cannot distinguish between Pareto and log Normal at this range. Substantively, he argues that the MSAs – which must have at least 50,000 inhabitants and typically do not consist of integrated economic units – are an imperfect measure of urban density, and argues for log-normality on the basis of the entire distribution of places, including small towns.

Here we take a different approach. We return to the distribution of MSAs, but show that even for this subset of the data, the Pareto hypothesis fails. For the sake of comparison with Eeckhout (2004), we use the 2000 U.S. Census in the 2001 vintage. Our results are reported in Table 6.

Table 6: Statistics for the U.S. City Size Distribution

Sample	n	ω	$\overline{w}$	$\overline{w^2}$	$\overline{w^3}$	$\mathcal{R}_2$	$\mathcal{R}_3$
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	0.857	0.636
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

Notes: Data are for the 2000 U.S. Census, 2001 vintage. n is sample size, ω is the log lower bound, $w := x - \omega$ is log city size minus log lower bound, overlines refer to sample averages, and the statistics $\mathcal{R}_2$ and $\mathcal{R}_3$ are as defined in the main text.

The $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ statistics are nowhere close to equalling 1, regardless of our sample selection choice. It is often argued that the Pareto hypothesis applies only to the extreme upper tail (e.g., Gabaix (2009) and Jones and Kim (2018)). However, we see that even at narrower slices of the data Pareto is rejected. In particular, for the top 135 MSAs – the traditional cutoff in city-size analyses (Krugman 1996; Eeckhout 2004) – the $\widehat{\mathcal{R}}_k$ statistics are again far from unity. Remarkably, $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ hover around the same values found for billionaire wealth, namely $\widehat{\mathcal{R}}_2 \approx 0.85$ and $\widehat{\mathcal{R}}_3 \approx 0.65$.

We apply our maximum likelihood procedures to estimate the parameters of Weibull distribution at various lower bounds. The results are reported in Table 7. All estimates for $\widehat{\alpha}$ and $\widehat{\gamma}$ are positive and significant. The likeli–hood clearly indicates that the model fits better to the top 135 MSAs than to the full distribution, which is under–scored by the stability of $\widehat{\gamma}$ across the bottom three rows. The estimates for $\widehat{\alpha}$ decrease with the threshold. We test our prediction that this decline is a function of the lower bound and γ, using equation (10). The predictions work well, in particular when applied to the top 135 MSAs; the predictions made for the values of α for the top 100 and top 50 conform almost identically to their estimated values. In sum, γ remains constant while α varies; this further underscores our discussion in Section 5.

We also report the visual fit of the Gompertz model, using hazard rates. This is done in Figure 10, which plots the empirical cumulative hazard (estimated with a nonparametric Kaplan-Meier regression) against the predicted fit from a Gompertz distribution and of an Exponential distribution.

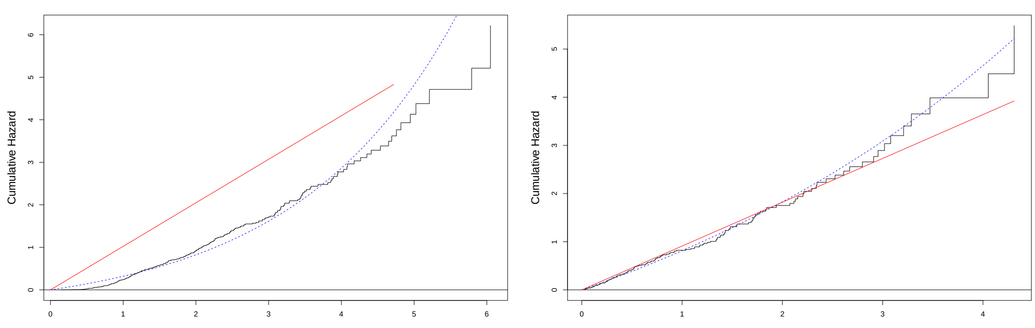
The first thing to note is that the Gompertz model approximates the data remarkably well. This is especially true for the top 135 MSAs (panel (10b)), but the fit for the full sample is also very close until the final observations (panel (10a)). Our results in Table 7 and Figure 10 clearly indicate that Gompertz fits the hazard of the log city size

Table 7: Gompertz Estimation, City Size

Sample	γ	S.E.	α	S.E.	Predicted α	
					$\alpha_0 = 3.94$	$\alpha_0 = 1.82$
Full	0.452^{***}	0.016	3.94^{***}	0.236	–	–
Top 135	0.219^{***}	0.013	1.82^{***}	0.118	2.656	–
Top 100	0.252^{***}	0.017	1.69^{***}	0.135	2.321	1.687
Top 50	0.252^{***}	0.024	1.24^{***}	0.147	1.845	1.345

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
Notes: Parameters γ and α estimated with maximum likelihood; see Appendix C for details. The predictions for α are made with equation (10), where the respective lower bounds ω are taken from Table 6. The first prediction assumes the prediction to apply to the full sample, the second prediction takes the 136th MSA as lower bound.

Figure 10: Cumulative Hazard Rates for the U.S. City Size Distribution



(a) Full Sample (b) Top 135 MSAs

Notes: Figure plots the empirical cumulative hazard rate (black) against the fitted Gompertz hazard (blue, dotted) and an Exponential hazard with $\alpha = 1$ (red). Panel (a) fits the full sample of Metropolitan Statistical Areas, panel (b) only the 135 largest MSAs. Note that w is defined relative to a different lower bound in both panels.

distribution very well, with only mild parameter instability when moving to the extreme upper tail.

In contrast, Figure 10 also shows the hazard rate for an Exponential distribution, in red. It is clear that the Exponential hazard fails to match the empirical hazards. The Weibull hazard fits the data well, in particular for the top 135 MSAs.

7.2 Firm Size

Contrary to our previous two cases, which suffer from relatively small sample sizes, we draw inference from a large (synthetic) sample of U.S. firms. We use data on the tabulated firm size distribution in the U.S. since 1930, compiled by Kwon, Ma, and Zimmermann (2023). These tabulations report the number of firms and average firm size in each bracket. The authors use generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022, `gpinter`) to study the increase in top firm size shares. This interpolation method takes as inputs the bracket lower bound, the percentile corresponding to that lower bound, and the bracket average, and interpolates the entire distribution based on a semiparametric approximation to the Lorenz curve.

We use Kwon et al.'s tabulated datasets, and use the `gpinter` interface on `www.wid.world` to generate the interpolations. We use asset value as our measure of firm size. Once we have these interpolations, we can use the interface to sample from this interpolation. Effectively, this procedure assumes the interpolation to be the correct data-generating process, and proceeds to draw synthetic samples of a given sample size which are representative of the interpolated data. As a result, assuming that the interpolated data are a good approximation to the true distribution, we can effectively study the upper tail of the firm size distribution with arbitrary precision. Hence, we do not report standard errors.

We draw a million observations from the full distribution for all years between 2000–2018.[17] We keep the top 1% of the distribution, resulting in 10,000 observations per year. The results are plotted in Figure 11.

We observe that the value of $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ are constant over time. The values of $\widehat{\mathcal{R}}_2$, while not far from 1, are significantly different and hover around 0.9. The values for $\widehat{\mathcal{R}}_3$, moreover, are clearly lower than for $\widehat{\mathcal{R}}_2$ and are about 0.75. These values are consistent with the plots of $\mathcal{R}_2$ and $\mathcal{R}_3$ in Figure 4 for $\alpha\gamma \approx 0.1$. We conclude that firm size too does not follow Pareto and fits Weibull.

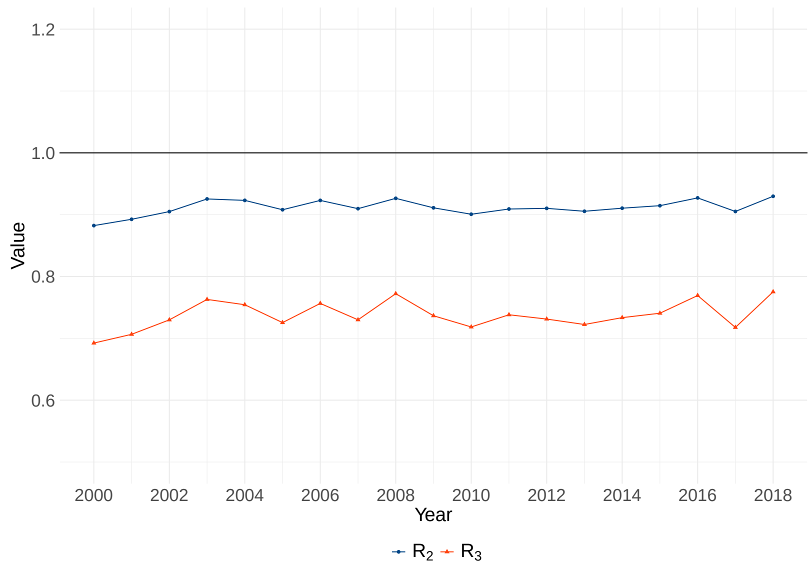
The Gompertz coefficients are reported in Table 8, where we again use maximum likelihood. We observe stable coefficients for both parameters. The values for γ are much lower than in our other two applications. Firm size is therefore closer to the Pareto benchmark than wealth or city size. $\alpha\gamma \approx 0.1$ in most years, conforming to the predictions made based on Figure 4.

Finally, we show the visual fit of the Gompertz model to the log firm size data using the hazard-based approach. We report the data for 2018 in Figure 12.

We see that Gompertz fits the log hazard rates very well. Despite the estimated low value of γ, the Gompertz hazard still fits much better than the Exponential hazard. The divergence happens beyond $w = 4$; the advantage of this dataset is that the sample size is sufficiently large such that w continues until 12 or so. The empirical failure of the epxonential distribution therefore becomes especially pronounced in Figure 12.

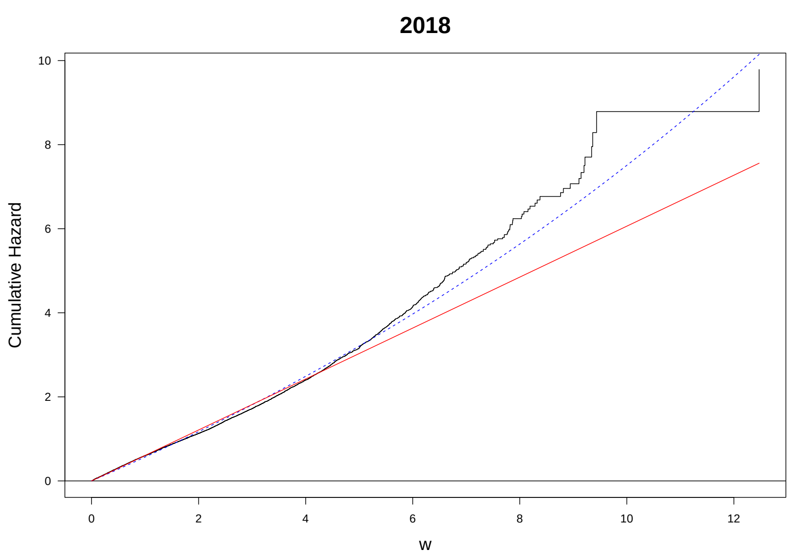
17. Drawing this many observations for all years since 1930 is computationally burdensome; we have drawn samples of 10,000 from all years, with very similar results which are available on request.

Figure 11: Tests for Pareto, U.S. Firm Size Distribution



Notes: $\underline{\mathcal{R}}_2$ and $\underline{\mathcal{R}}_3$ are defined in equation (6). Each point represents a sample of 10,000 drawn from the top 1% of the U.S. firm size distribution using generalized Pareto interpolation; see the main text for details.

Figure 12: Cumulative Hazard Rates for the U.S. Firm Size Distribution



Notes: Figure plots the empirical cumulative hazard rate (black) against the fitted Gompertz hazard (blue, dotted) and an Exponential hazard with $\alpha = 1$ (red). Data are for the 2018 top 1% of the U.S. firm size distribution.

Table 8: Gompertz Coefficients, U.S. Firm Size

Year	γ	α
2000	0.11	1.8
2001	0.099	1.9
2002	0.084	1.9
2003	0.064	1.8
2004	0.067	1.8
2005	0.08	1.9
2006	0.066	1.8
2007	0.079	1.9
2008	0.062	1.8
2009	0.077	1.9
2010	0.086	2
2011	0.077	1.9
2012	0.079	1.9
2013	0.083	1.9
2014	0.077	1.9
2015	0.075	1.9
2016	0.062	1.8
2017	0.085	1.9
2018	0.059	1.8

Notes: Parameters γ and α estimated with maximum likelihood; see Appendix C for details. We use generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022) to draw a sample of size 10,000 from the top 1% of the firm size distribution for each year, with data from the firm size distribution from Kwon, Ma, and Zimmermann (2023)

.

8 Implications

Wealth, city size, and firm size all follow a Weibull rather than a Pareto distribution. This section explores some of the implications of this finding. We discuss these implications in a wider setting, since our results might carry over to other contexts where Pareto has commonly been assumed, such as income or trade. We first discuss some stochastic properties of Weibull and Pareto, and notions of convergence. Then, we discuss the implications for optimal taxation. Afterwards, we briefly discuss some other theoretical settings where Pareto is often used. Finally, we discuss the implications for the "granular hypothesis" (Gabaix 2011).

Convergence to Pareto and Weibull: The Pareto distribution is the quintessential example of a fat-tailed distribution. These distributions are defined by a concept from extreme-value theory called *regular variation*, which ensures that a function behaves like a Pareto distribution in the limit (cf. Benhabib and Bisin 2018; Bingham, Goldie, and Teugels 1989). Formally, a function $f(x)$ is called regularly varying if

$$\lim_{x\to\infty} \frac{f(xy)}{f(y)} = x^{-1/\alpha}, \quad \forall x > 0. \tag{17}$$

The interest in Pareto has therefore not only arisen because of its simple theoretical properties, but also because it seems to emerge as the stationary distribution for many stochastic processes. Gabaix (2009) gives many examples of diffusion processes which lead to Pareto in the limit. In general, these processes are variations on random growth models (e.g., geometric Brownian motions), with some frictions added to stabilize the distribution.

There is a tight link between convergence to Pareto and the decay of the distribution, which is governed by the hazard rate of its log-transform. Contrary to the Gompertz distribution (the log transform of Weibull) the hazard rate of the Exponential distribution (the log transform of Pareto) is constant and *memoryless*: it has a constant hazard rate regardless of the choice of lower bound: $H(w) = H(x) = \alpha^{-1}, \forall \omega$. It can be shown that a distribution is fat-tailed if and only if the hazard rate of its log-transform converges to a constant (Beirlant, Goegebeur, Segers, and Teugels 2006). Our results, in contrast, show an exponentially increasing hazard.

Stochastic processes can converge to other distributions than Pareto. The exact process for convergence to truncated-Weibull can be modeled using a diffusion of the form $\mathrm{d}W = \boldsymbol{\mu}(W)\mathrm{d}t + \boldsymbol{\sigma}(W)\mathrm{d}\mathcal{B}(t)$, where $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$ are functions governing the drift and diffusion, respectively, and where $\mathcal{B}$ is a standard Brownian motion. We can reverse-engineer this process by inserting the truncated-Weibull distribution into the Kolmogorov Forward Equation, and solving for $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$. Preliminary results indicate that this is possible both for wealth in levels and in logs, but that these functional forms are highly non-linear.

It has recently been proved that Gompertz emerges as the distribution of the length of *self-avoiding walks* (SAW) on stochastic Erdős–Rényi–Gilbert networks (Tishby, Biham, and Katzav 2016). These networks are formed by a graph $ER(n, p)$ with n nodes, where each edge connecting two nodes has probability p of being generated. The connected nodes form a network. A SAW on such a network is defined as follows. Start at any node i that is part of the network. A SAW is a random walk from one node to another node that is connected to that node by the network until one cannot proceed to an new node not visited previously. Tishby, Biham, and Katzav (2016) prove that the distribution of SAW path lengths is Gompertz.

Stochastic networks are increasingly explored by economic theorists, see Goyal (2023). We can interpret the exponentially increasing hazard rate of the next node being the endpoint of a SAW as a "capacity constraint": ultimately, a SAW can never be longer than the number of nodes minus one, $n-1$. This could serve as a microfoundation for the emergence of Gompertz in our settings. For instance, cities can only grow on the available land; hence, their size is naturally bounded by the total land mass. Likewise, firm can only grow by "invading" new parts of the economy; hence, their size is bounded by the size of the economy at large. Similar analogues can be constructed for income and wealth.

Optimal taxation: A well-known result in the Mirrlees (1971) framework is that the marginal tax rate on the richest individual should be zero when the income distribution is bounded (Sadka 1976). A marginal tax rate just yields distortions at its level of incidence. However, it comes with the benefit of a higher tax on on higher income levels. For the highest income, there are no higher income levels, so the benefit of a positive marginal tax rate is zero and hence this marginal tax should be zero. This is not necessarily true when we allow the income distribution to be unbounded, see Saez (2001). When the distribution of top income is Pareto and the uncompensated elasticity of earned income with respect to the marginal tax rate (denoted η) is the same for all income levels, the Diamond (1998)–Saez (2001) formula for the optimal top income tax rate τ^* satisfies:[18]

$$\tau^* = \frac{\alpha}{\alpha + \eta}. \tag{18}$$

τ^* can be calculated from just two statistics: the tail index α^{-1} and the elasticity of earned income η. The parameter α appears in the formula is because the optimal tax rate equates the cost of a marginal tax at W (which is $\eta \times W \times \Pr\left[\underline{W} = W\right]$) and the additional tax revenue on an even higher income $\underline{W} > W$ enabled by this a higher marginal rate (which is $\Pr\left[\underline{W} > W\right]$). For Pareto, this ratio is constant, so that the top tax rate should converge to a constant. For Weibull, the corresponding expression reads:

$$\tau^* = \frac{\alpha e^{-\gamma w}}{\alpha e^{-\gamma w} + \eta},$$

so that the optimal tax rate converges to zero. However, this analysis maintains the assumption that the earned income elasticity for the top earners η converges to a constant. It might be the case that a formal model that generates a Weibull right tail of the income distribution also generates a declining elasticity of earned income with respect to the marginal tax rate since it becomes increasingly difficult to increase one's income or wealth due to "capacity constraints", as in our discussion of networks with SAWs above. In that case, η would also converge to zero for the top earner and τ^* might again be constant.

Firm Productivity and Endogenous Growth: Kortum (1997) models endogenous growth arising from researchers that are heterogeneous in their productivity, where productivity is the highest draw ('best idea') from a distribution. He argues that exponential productivity growth can only be consistent with Pareto-distributed productivity. This result has been challenged by Jones (2023), who shows that if ideas are combinations of existing ideas, productivity need not be fat-tailed: combinatorial draws from a thin-tailed distribution will also result in exponential productivity growth. He arrives at a convenient closed-form expression for the growth rate of the best idea (g_Z) if productivity is Weibull-distributed (in our notation): $g_Z^{\text{Weibull}} = g_N/\gamma$, where g_N is the population growth rate. Jones calls γ the

18. The Diamond-Saez formula ignores externalities of taxing top earners. Jones (2022) argues that top earners generate new ideas that increase productivity growth. In that case, optimal tax rates could be lower.

rate at which ideas are getting harder to find. This mirrors our interpretation of γ.

International trade: Chaney (2018) derives an explanation for the gravity equation based on Pareto-distributed firm productivity. Core to his argument is a linear relationship between log rank and log sizes. In general, Pareto provides closed-form expressions for trade models based on heterogeneous firms in the spirit of Melitz (2003); see also Helpman, Melitz, and Yeaple (2004) and Chaney (2008). Our results show that the actual firm size distribution is described better by Weibull than Pareto.

Granularity: Pareto is also popular because of the *granular hypothesis*, first formulated by Gabaix (2011). Gabaix argues that if firm size is Pareto, idiosyncratic shocks to firms do not wash out in the aggregate, and hence can partly account for aggregate fluctuations. Core to his argument is the fact that if α is sufficiently large, the firm size distribution does not have finite variance, and hence the law of large numbers does not apply as normal (which would imply that volatility decays on the order of $\sqrt{N}^{-1}$). Gabaix shows that if the firm size distribution has infinite variance, the decay is much slower, on the order of $1/\ln N$. Weibull, in contrast, has finite variance, and hence idiosyncratic shocks will decay faster than under Pareto. A Pareto distribution of firm size is no prerequisite for idiosyncratic shocks to generate aggregate fluctuations; as long as the distribution of firm connections is sufficiently right-skewed, shocks can propagate along the production network (Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi 2012).

9 Conclusion

We test the default assumption of Pareto for right-skewed distributions, using the *Forbes List of Billionaires*. We develop test statistics $\mathcal{R}_k$ based on normalizations of kth-order moments of the log-transformed data. We find strong evidence against a Pareto tail based on our statistics $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ in all cases. Our rejection of Pareto conforms to the arguments of Blanchet, Fournier, and Piketty (2022); however, where they argue for a non-parametric distribution, we show that a parametric alternative – (truncated-)Weibull – fits the data remarkably well. Our evidence for Weibull comes from the fit of within region/year maximum likelihood estimates, and from cross-equation restrictions on between region/year observations (the cross-equation restrictions in regressions of $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$ on $\ln\widehat{\alpha}$ and $\ln\widehat{\gamma}$, and of $\widehat{\mathcal{R}}_3$ on $\widehat{\mathcal{R}}_2$). Our model yields sharp testable predictions, which are all accepted. Weibull-based predictions of mean wealth are almost perfect, whereas the predictions based on Pareto are nonsensically large or even infinite.

Our results cannot be driven by measurement error or small-sample bias. Our conclusions also apply to two other settings, city size and firm size. We conjecture our rejection of Pareto and embrace of Weibull to apply to even more settings where Pareto has commonly been used. Further empirical research will easily be able to test this hypothesis, since our statistics $\mathcal{R}_k$ are easily calculated and our Weibull alternative yields easily testable predictions.

The predictions of the truncated-Weibull distribution differ on two accounts from Pareto. First, whereas the Pareto distribution has a constant hazard rate, the truncated-Weibull distribution has an exponentially increasing hazard rate. Hence, the ratio of the density at the threshold and the fraction above the threshold is increasing in the threshold. However, while the Weibull hazard does not converge to Pareto, the moments of Weibull do converge to Pareto far in the upper tail; however, this only applies to the upper-most parts of the distribution. For the majority of the upper tail, Pareto has no defined moments for the empirically relevant values of α^{-1}, while the moments of the truncated-Weibull are always defined. These two differences make the truncated-Weibull distribution particularly suited for the study of upper tails.

The clarity of our results raise the question why they have not been found before. Existing research appears to hold a prior that the hazard rate will converge to a constant, perhaps implictly reasoning from the stochastic convergence of diffusion processes to Pareto discussed in Section 8. This prior is at odds with the data in all cases we study in this paper. A second reason might be because the primary diagnostic to check Paretianity is the log rank regression. This tool has weak discriminatory power between Pareto and Weibull for small γ, introduces correlations in the error term because the data are ordered, and is sensitive to measurement error. Our test statistics $\mathcal{R}_k$ and estimation methods do not suffer from these problems.

Our results have important theoretical and empirical implications. If wealth is Weibull, log wealth is Gompertz. Gompertz emerges as the stationary distribution of self-avoiding walks on stochastic networks. This model structure has intriguing parallels with the environments we study, because we can microfound these SAWs as capacity constraints. Our results further have implications for optimal taxation. Mechanically, the optimal tax rate decreases because the hazard is now an increasing function of the threshold. However, the capacity constraints argument may mean that the behavioral elasticity is also no longer constant; hence, the full implications of our findings are ambiguous and need to be explored in a full model of optimal taxation. Finally, Weibull shows why the number of billionaires has increased since 2000; in effect, the threshold to be a billionaire, Ω, has decreased as countries have become richer and stock market capitalization has increased. Since Weibull is sensitive to the lower bound, these changes translate into a fattened upper tail. We explore these arguments further in a companion paper, Teulings and Toussaint (2023).

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. "The Network Origins of Aggregate Fluctuations." *Econometrica* 80 (5): 1977–2016.

Ahuja, J.C., and Stanley W. Nash. 1967. "The Generalized Gompertz-Verhulst Family of Distributions." *Sankhyā: The Indian Journal of Statistics, Series A,* 141–156.

Albers, Thilo N., Charlotte Bartels, and Moritz Schularick. 2022. "Wealth and its Distribution in Germany, 1895-2018." *CEPR Discussion Paper* DP17269.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49 (1): 3–71.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135 (2): 645–709.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293 (5536): 1818–1820.

Balkema, August A., and Laurens De Haan. 1974. "Residual Life Time at Great Age." *The Annals of Probability,* no. 5 (2): 792–804.

Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef L. Teugels. 2006. *Statistics of Extremes: Theory and Applications.* John Wiley & Sons.

Benhabib, Jess, and Alberto Bisin. 2018. "Skewed Wealth Distributions: Theory and Empirics." *Journal of Economic Literature* 56 (4): 1261–91.

Bingham, Nicholas H., Charles M. Goldie, and Jozef L. Teugels. 1989. *Regular Variation.* Cambridge University Press.

Blanchet, Thomas. 2022. "Uncovering the Dynamics of the Wealth Distribution." *arXiv preprint arXiv:2211.15509.*

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty. 2022. "Generalized Pareto Curves: Theory and Appli–cations." *Review of Income and Wealth* 68 (1): 263–288.

Chaney, Thomas. 2008. "Distorted Gravity: the Intensive and Extensive Margins of International Trade." *American Economic Review* 98 (4): 1707–1721.

———. 2018. "The Gravity Equation in International Trade: An Explanation." *Journal of Political Economy* 126 (1): 150–177.

Dekkers, Arnold L.M., John H.J. Einmahl, and Laurens De Haan. 1989. "A Moment Estimator for the Index of an Extreme-Value Distribution." *The Annals of Statistics* 17 (4): 1833–1855.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88 (1): 83–95.

Driscoll, John C., and Aart C. Kraay. 1998. "Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data." *Review of Economics and Statistics* 80 (4): 549–560.

Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review* 94 (5): 1429–1451.

———. 2009. "Gibrat's Law for (All) Cities: Reply." *American Economic Review* 99 (4): 1676–1683.

Fedotenkov, Igor. 2020. "A Review of More than One Hundred Pareto-Tail Index Estimators." *Statistica* 80 (3): 245–299.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics* 114 (3): 739–767.

———. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics* 1 (1): 255–294.

———. 2011. "The Granular Origins of Aggregate Fluctuations." *Econometrica* 79 (3): 733–772.

———. 2016. "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30 (1): 185–206.

Gabaix, Xavier, and Rustam Ibragimov. 2011. "Rank −1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents." *Journal of Business & Economic Statistics* 29 (1): 24–39.

Gabaix, Xavier, and Ralph S.J. Koijen. 2023. "Granular Instrumental Variables." *Journal of Political Economy,* Forthcoming.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123 (1): 49–100.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica* 84 (6): 2071–2111.

Gomez, Matthieu. 2023. "Decomposing the Growth of Top Wealth Shares." *Econometrica* 91 (3): 979–1024.

Goyal, Sanjeev. 2023. *Networks: An Economics Approach.* MIT Press.

Gumbel, Emil Julius. 1958. *Statistics of Extremes.* Columbia University Press.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300–316.

Hill, Bruce M. 1975. "A Simple General Approach to Inference About the Tail of a Distribution." *The Annals of Statistics* 3 (5): 1163–1174.

Huisman, Ronald, Kees G. Koedijk, Clemens J.M. Kool, and Franz Palm. 2001. "Tail-Index Estimates in Small Samples." *Journal of Business & Economic Statistics* 19 (2): 208–216.

Jones, Charles I. 2015. "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality." *Journal of Economic Perspectives* 29 (1): 29–46.

———. 2022. "Taxing Top Incomes in a World of Ideas." *Journal of Political Economy* 130 (9): 2227–2274.

Jones, Charles I. 2023. "Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail." *Journal of Political Economy* 131 (8): 1994–2031.

Jones, Charles I., and Jihee Kim. 2018. "A Schumpeterian Model of Top Income Inequality." *Journal of Political Economy* 126 (5): 1785–1826.

Kleiber, Christian, and Samuel Kotz. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences.* John Wiley & Sons.

Kondo, Illenin O., Logan T. Lewis, and Andrea Stella. 2023. "Heavy Tailed but not Zipf: Firm and Establishment Size in the United States." *Journal of Applied Econometrics* 38 (5): 767–785.

Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica* 65 (6): 1389–1419.

Krugman, Paul. 1996. "Confronting the Mystery of Urban Hierarchy." *Journal of the Japanese and International Economies* 10 (4): 399–418.

Kwon, Spencer Yongwook, Yueran Ma, and Kaspar Zimmermann. 2023. "100 Years of Rising Corporate Concentration." *University of Chicago, Becker Friedman Institute for Economics Working Paper,* nos. 2023-20.

Luttmer, Erzo G. J. 2011. "On the Mechanics of Firm Growth." *Review of Economic Studies* 78 (3): 1042–1068.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.

Mogstad, Magne, Joseph P. Romano, Azeem M. Shaikh, and Daniel Wilhelm. 2023. "Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries." *Review of Economic Studies* (January): rdad006.

Mood, Alexander McFarlane, Franklin A. Graybill, and Duane C. Boes. 1974. *Introduction to the Theory of Statistics.* McGraw-Hill Kogakusha.

Novokmet, Filip, Thomas Piketty, and Gabriel Zucman. 2018. "From Soviets to Oligarchs: Inequality and Property in Russia 1905–2016." *Journal of Economic Inequality* 16:189–223.

Pareto, Vilfredo. 1896. *Cours d'économie politique: professé á l'Université de Lausanne.* Vol. 1. F. Rouge.

Pickands, James. 1975. "Statistical Inference using Extreme Order Statistics." *The Annals of Statistics* 3 (1): 119–131.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Harvard University Press.

Piketty, Thomas, Li Yang, and Gabriel Zucman. 2019. "Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015." *American Economic Review* 109 (7): 2469–2496.

Rosen, Kenneth T., and Mitchel Resnick. 1980. "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy." *Journal of Urban Economics* 8 (2): 165–186.

Rossi-Hansberg, Esteban, and Mark L.J. Wright. 2007. “Urban Structure and Growth.” *Review of Economic Studies* 74 (2): 597–624.

Sadka, Efraim. 1976. “On Income Distribution, Incentive Effects and Optimal Income Taxation.” *Review of Economic Studies* 43 (2): 261–267.

Saez, Emmanuel. 2001. “Using Elasticities to Derive Optimal Income Tax Rates.” *Review of Economic Studies* 68 (1): 205–229.

Teulings, Coen N., and Simon J. Toussaint. 2023. “Why Has the Number of Billionaires Increased So Much?” *Working Paper.*

Tishby, Ido, Ofer Biham, and Eytan Katzav. 2016. “The distribution of path lengths of self avoiding walks on Erdős–Rényi networks.” *Journal of Physics A: Mathematical and Theoretical* 49 (28): 285002.

Van der Vaart, Aad W. 2000. *Asymptotic Statistics.* Cambridge University Press.

Van der Wijk, J. 1939. *Inkomens- en Vermogensverdeling.* 26. Netherlands Economic Institute.

Vermeulen, Philip. 2016. “Estimating the Top Tail of the Wealth Distribution.” *American Economic Review* 106 (5): 646–50.

———. 2018. “How Fat is the Top Tail of the Wealth Distribution?” *Review of Income and Wealth* 64 (2): 357–387.

Wingo, Dallas R. 1989. “The Left-Truncated Weibull Distribution: Theory and Computation.” *Statistical Papers* 30 (1): 39–48.

A Bias and variance of $\widehat{\mathcal{R}}_2$ and $\widehat{\mathcal{R}}_3$

Let $\underline{g} := w/\alpha$ be distributed exponentially with decay parameter 1 with moments $\mathrm{E}\left[\underline{g}^k\right] = k!$ and let g be a realisation of $\underline{g}$. $\mathcal{R}_k$ is defined in equation (6) as $\mathcal{R}_k = \mathrm{E}\left[\underline{g}^k\right] / \left(k!\mathrm{E}\left[\underline{g}\right]^k\right)$ for any $k \geq 1$. The estimator $\widehat{\mathcal{R}}_k = \overline{g^k}/\left(k!\overline{g}^k\right)$ of $\mathcal{R}_k$ uses the sample means $\overline{g^k}$ as the estimator for $\mathrm{E}\left[\underline{g}^k\right]$. However, since $\widehat{\mathcal{R}}_k$ is a non-linear transform of $\overline{g^k}$ and $\overline{g}$ this estimator is biased: $\mathrm{E}\left[\widehat{\mathcal{R}}_k\right] \neq \mathcal{R}_k$. This section asseses the variance and bias of $\widehat{\mathcal{R}}_k$.

A.1 Variance $\overline{g}^k$

The variance of $\overline{g}^k$ satisfies

$$\mathrm{Var}\left[\overline{g}^k\right] = \mathrm{E}\left[\overline{g}^{2k}\right] - \mathrm{E}\left[\overline{g}^k\right]^2 .$$

The variance of $\mathrm{E}\left[\overline{g}^k\right]$ satisfies:

$$\begin{aligned}\mathrm{E}\left[\overline{g}^k\right] &= \mathrm{E}\left[\left(N^{-1}\sum_i \underline{g}_i\right)^k\right] \\ &= N^{-k}\left(\frac{N!}{(N-k)!}\mathrm{E}\left[\underline{g}\right]^k + \frac{N!}{(N-k+1)!}\binom{k}{2}\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{k-2} + O\left(N^{k-2}\right)\right).\end{aligned}$$

$\left(\sum_i \underline{g}_i\right)^k$ consists of N^k terms $\prod_{i\in\mathcal{K}_n} \underline{g}_i$, where $\mathcal{K}_n$ is one of the $n = \left\{1, ..., N^k\right\}$ feasible combinations of k draws from the set of N draws. We retain only these terms where at most $\underline{g}_i$ are the same.

For $\frac{N!}{(N-k)!} = \prod_{m=0}^{k-1}(N-m)$ terms, all $\underline{g}_i$ are different. The number of these terms is of order $O\left(N^k\right)$. Using $\mathrm{E}\left[\underline{g}_i\underline{g}_j\right] = \mathrm{E}\left[\underline{g}_i\right]\mathrm{E}\left[\underline{g}_j\right]$ the expectation of these terms is $\mathrm{E}\left[\underline{g}\right]^k$.

For $\frac{N!}{(N-k+1)!} = \prod_{m=0}^{k-2}(N-m)$ combinations of $\underline{g}_i$, all $\underline{g}_i$ but one pair are different. $\prod_{m=0}^{k-2}(N-m)$ is of order $O\left(N^{k-1}\right)$. For each combination, the pair of equal $\underline{q}_i'$s can be drawn from the k factors of the product $\left(\sum_i \underline{g}_i\right)^k$ in $\binom{k}{2}$ different ways. The expectation of these terms is $\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{2k-2}$.

The number of terms where more than than two draws are the same is of order N^{k-m} with $m > 1$. These terms are therefore captured in the term $O\left(N^{k-2}\right)$.

Using $\mathrm{E}\left[\underline{g}^k\right] = k!$ and $\left(\frac{N!}{(N-k+1)!}\right)^2 = O\left(N^{2k-2}\right)$ these relations can be simplified

$$\begin{aligned}\mathrm{E}\left[\overline{g}^k\right] &= N^{-k}\left(\frac{N!}{(N-k)!} + \frac{N!}{(N-k+1)!}k(k-1) + O\left(N^{k-2}\right)\right) \\ &= 1 - \frac{k(k-1)}{2N} + \frac{k(k-1)}{N} + O\left(N^{-2}\right) = 1 + \frac{k(k-1)}{2N} + O\left(N^{-2}\right).\end{aligned} \tag{19}$$

using

$$\frac{N!}{(N-k)!} = \prod_{m=0}^{k-1}(N-m) = N^k - N^{k-1}\sum_{m=0}^{k-1} m + O\left(N^{k-2}\right) = N^k - \frac{k(k-1)}{2}N^{k-1} + O\left(N^{k-2}\right),$$

which implies $\frac{N!}{(N-k+1)!} = N^{k-1} + O\left(N^{k-2}\right)$.

The same argument applies to $\mathrm{E}\left[\overline{g}^{2k}\right]$, replacing k by $2k$.

$$\mathrm{E}\left[\overline{g}^{2k}\right] = 1 + \frac{k\,(2k-1)}{N} + O\left(N^{-2}\right).$$

$\mathrm{E}\left[\overline{g}^{k}\right]^2$ satisfies

$$\mathrm{E}\left[\overline{g}^{k}\right]^2 = \left(1 + \frac{k\,(k-1)}{2N} + O\left(N^{-2}\right)\right)^2 = 1 + \frac{k\,(k-1)}{N} + O\left(N^{-2}\right) \tag{20}$$

Combining both terms yields

$$\mathrm{Var}\left[\overline{g}^{k}\right] = 1 + \frac{k\,(2k-1)}{N} - 1 - \frac{k\,(k-1)}{N} - O\left(N^{-2}\right) = \frac{k^2}{N} + O\left(N^{-2}\right). \tag{21}$$

A.2 Variance $\overline{g^k}$

$$\mathrm{Var}\left[\overline{g^k}\right] = N^{-1}\mathrm{Var}\left[\underline{g}^k\right] = N^{-1}\left(\mathrm{E}\left[\underline{g}^{2k}\right] - \mathrm{E}\left[\underline{g}^{k}\right]^2\right) = \frac{(2k)! - k!^2}{N}. \tag{22}$$

A.3 Covariance $\overline{g}^k$ and $\overline{g^k}$

$$\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right] = \mathrm{E}\left[\overline{g^k}\,\overline{g}^k\right] - \mathrm{E}\left[\overline{g^k}\right]\mathrm{E}\left[\overline{g}^k\right]^2.$$

where we use the expression for $\mathrm{E}\left[\overline{g}^k\right]^2$ in equation (19). We have:

$\mathrm{E}\left[\overline{g^k}\right] = \mathrm{E}\left[\underline{g}^k\right]$, while $\mathrm{E}\left[\overline{g^k}\,\overline{g}^k\right]$ satisfies

$$\begin{aligned}\mathrm{E}\left[\overline{g^k}\,\overline{g}^k\right] &= \mathrm{E}\left[\left(N^{-1}\sum_i \underline{g}_i^k\right)\left(N^{-1}\sum_i \underline{g}_i\right)^k\right] \\ &= N^{-k-1}\left(\begin{array}{c} \frac{N!}{(N-k-1)!}\mathrm{E}\left[\underline{g}^k\right]\mathrm{E}\left[\underline{g}\right]^k + \frac{N!}{(N-k)!}\binom{k}{2}\mathrm{E}\left[\underline{g}^k\right]\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{k-2} \\ +\frac{N!}{(N-k)!}k\mathrm{E}\left[\underline{g}^{k+1}\right]\mathrm{E}\left[\underline{g}\right]^{k-1} + O\left(N^{k-2}\right)\end{array}\right).\end{aligned}$$

$\left(\sum_i \underline{g}_i^k\right)\left(\sum_i \underline{g}_i\right)^k$ consists of N^{k+1} terms $\prod_{i\in\mathcal{K}_n} \underline{g}_i \underline{g}_j^k$, where $\mathcal{K}_n$ is one of the $n = \left\{1, \ldots, N^k\right\}$ feasible combinations of k draws from the set of N draws. We retain only these terms where at most on pair of these k factors with $\underline{g}_i$ are the same and where they are different from $\underline{g}_j$ in $\underline{g}_j^k$ or where one these k factors has the same $\underline{g}_i$ as $\underline{g}_j$ in $\underline{g}_j^k$.

For $\frac{N!}{(N-k-1)!} = \prod_{m=0}^{k}(N-m)$ terms, all $\underline{g}_i$ are different from each other and from the $\underline{g}_j$. The number of these terms is of order $O\left(N^{k+1}\right)$. Using $\mathrm{E}\left[\underline{g}_i\underline{g}_j\right] = \mathrm{E}\left[\underline{g}_i\right]\mathrm{E}\left[\underline{g}_j\right]$ the expectation of these terms is $\mathrm{E}\left[\underline{g}\right]^k\mathrm{E}\left[\underline{g}^k\right]$.

For $\frac{N!}{(N-k)!} = \prod_{m=0}^{k-1}(N-m)$ combinations of $\underline{g}_i$, either all $\underline{g}_i$ but one pair are different and they are all different from $\underline{g}_j$ in $\underline{g}_j^k$ or they are all different, but one of them is equal to $\underline{g}_j$ in $\underline{g}_j^k$. For the first group, the pair of equal $\underline{g}_i'$s can be drawn from the k factors of the product $\left(\sum_i \underline{g}_i\right)^k$ in $\binom{k}{2}$ different ways. The expectation of these terms is

$\mathrm{E}\left[\underline{g}^k\right]\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{k-2}$. For the second group, the $\underline{g}_i$ that is equal to $\underline{g}_j$ in $\underline{g}_j^k$ can be drawn from each of the k factors of the product $\left(\sum_i \underline{g}_i\right)^k$. The expectation of these terms is $\mathrm{E}\left[\underline{g}^{k+1}\right]\mathrm{E}\left[\underline{g}\right]^{k-1}$.

The numbers of terms where more than than two observations are the same are of order N^{k-m} with $m > 1$. These terms are therefore captured in the term $O\left(N^{k-1}\right)$.

Using $\mathrm{E}\left[\underline{g}^k\right] = k!$ and previous relations, we obtain

$$\begin{aligned}
\mathrm{E}\left[\overline{g^k}\overline{g}^k\right] &= N^{-k-1}\left(\frac{N!}{(N-k-1)!}k! + \frac{N!}{(N-k)!}k\,(k-1)\,k! + \frac{N!}{(N-k)!}k\,(k+1)! + O\left(N^{k-2}\right)\right) \\
&= k! - \frac{k\,(k+1)}{2N}k! + \frac{k\,(k-1)\,k! + k\,(k+1)!}{N} + O\left(N^{-2}\right) \\
&= k!\left(1 + \frac{k\,(3k-1)}{2N}\right) + O\left(N^{-2}\right).
\end{aligned}$$

Combining these results yields

$$\begin{aligned}
\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right] &= k!\left(1 + \frac{k\,(3k-1)}{2N}\right) - k!\left(1 + \frac{k\,(k-1)}{2N}\right) + O\left(N^{-2}\right) \\
&= \frac{k^2 k!}{N} + O\left(N^{-2}\right).
\end{aligned} \tag{23}$$

A.4 The expectation and variance of $\widehat{\mathcal{R}}_k$

A Taylor expansion for the expectation and variance of a quotient reads (Mood, Graybill, and Boes 1974, pp. 180)

$$\begin{aligned}
\mathrm{E}\left[\frac{\underline{z}}{\underline{y}}\right] &= \frac{\mathrm{E}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]} - \frac{\mathrm{Cov}\left[\underline{y}, \underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]^2} + \frac{\mathrm{E}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]^3}\mathrm{Var}\left[\underline{y}\right] + O\left(N^{-2}\right), \\
\mathrm{Var}\left[\frac{\underline{z}}{\underline{y}}\right] &= \left(\frac{\mathrm{E}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]}\right)^2\left(\frac{\mathrm{Var}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{z}\right]^2} - 2\frac{\mathrm{Cov}\left[\underline{y}, \underline{z}\right]}{\mathrm{E}\left[\underline{z}\right]\mathrm{E}\left[\underline{y}\right]} + \frac{\mathrm{Var}\left[\underline{y}\right]}{\mathrm{E}\left[\underline{y}\right]^2}\right) + O\left(N^{-2}\right).
\end{aligned}$$

Using equation (19), (21), (22) and (23) and

$$\mathrm{E}\left[\overline{g}^k\right]^{-1} = \left(1 + \frac{k\,(k-1)}{2N}\right)^{-1} + O\left(N^{-2}\right) = 1 - \frac{k\,(k-1)}{2N} + O\left(N^{-2}\right),$$

we obtain

$$
\begin{aligned}
\mathrm{E}\left[\widehat{\mathcal{R}}_k\right] &= \frac{\mathrm{E}\left[\overline{g^k}\right]}{k!\mathrm{E}\left[\overline{g}^k\right]} - \frac{\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right]}{k!^2\mathrm{E}\left[\overline{g}^k\right]^2} + \frac{\mathrm{E}\left[\overline{g^k}\right]}{k!^3\mathrm{E}\left[\overline{g}^k\right]^3}\mathrm{Var}\left[\overline{g}^k\right] + \mathcal{O}\left(N^{-2}\right) \\
&= 1 - \frac{k\left(k-1\right)}{2N} - \frac{k^2 k!}{k!N} + \frac{(2k)! - k!^2}{k!^2 N} + \mathcal{O}\left(N^{-2}\right) \\
&= 1 + N^{-1}\left(\frac{(2k)!}{k!^2} - \frac{3k^2 - k + 2}{2}\right) + \mathcal{O}\left(N^{-2}\right).
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}\left[\widehat{\mathcal{R}}_k\right] &= \left(\frac{\mathrm{E}\left[\overline{g^k}\right]}{k!\mathrm{E}\left[\overline{g}^k\right]}\right)^2 \left(\frac{\mathrm{Var}\left[\overline{g^k}\right]}{\mathrm{E}\left[\overline{g^k}\right]^2} - 2\frac{\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right]}{\mathrm{E}\left[\overline{g^k}\right]\mathrm{E}\left[\overline{g}^k\right]} + \frac{\mathrm{Var}\left[\overline{g}^k\right]}{\mathrm{E}\left[\overline{g}^k\right]^2}\right) + \mathcal{O}\left(N^{-2}\right) \\
&= \frac{(2k)! - k!^2}{k!^2 N} - \frac{2k^2 k!}{k!N} + \frac{k^2}{N} + \mathcal{O}\left(N^{-2}\right) \\
&= N^{-1}\left(\frac{(2k)!}{k!^2} - k^2 - 1\right) + \mathcal{O}\left(N^{-2}\right).
\end{aligned}
$$

B Moments of the Gompertz and Weibull distribution

Define the parameter $\theta \coloneqq \alpha\gamma$ and the normalized random variable $\underline{q} \coloneqq \gamma\underline{w}$ for $\underline{q} \geq 0$. Its complement distribution function $S(q)$ reads:

$$\Pr\left[\underline{q} > q\right] =: S(q) = \exp\left(-\frac{e^{q}-1}{\theta}\right).$$

The moments of $\underline{q}$ for $k > 0$ satisfy:

$$\begin{aligned}
\mathrm{E}\left[\underline{q}^{k}\right] &= \int_0^\infty q^k s(q)\,\mathrm{d}q \\
&= -\overbrace{\left[q^k S(q)\right]_{q=0}^{\infty}}^{=0} + \overbrace{k\int_0^\infty q^{k-1} S(q)\,\mathrm{d}q}^{=h(\theta,k)}, \\
h(\theta,k) &\coloneqq k\int_0^\infty q^{k-1}\exp\left(-\frac{e^{q}-1}{\theta}\right)\mathrm{d}q = k e^{1/\theta}\int_0^\infty q^{k-1}\exp\left(-e^{q}/\theta\right)\mathrm{d}q, \\
h(\theta,1) &= e^{1/\theta}\int_0^\infty \exp\left(-e^{q}/\theta\right)\mathrm{d}q = e^{1/\theta}\int_{1/\theta}^\infty y^{-1}e^{-y}\mathrm{d}y = e^{1/\theta}\mathrm{Ei}(1/\theta), \\
y &\coloneqq e^{q-\ln\theta} \Rightarrow q = \ln y + \ln\theta \Rightarrow \frac{\mathrm{d}q}{\mathrm{d}y} = \frac{1}{y},
\end{aligned}$$

where $s(q) \coloneqq -S'(q)$ is the density function and where $\mathrm{Ei}(z)$ is the standard Exponential integral. The second line uses the definition of $S(q)$ and $h(\theta,k)$ and applies integration by parts. The third line gives the definition of $h(\theta,k)$. The fourth line analyses the special case $h(\theta,1)$. The definitions of $\underline{q}$ and θ imply:

$$\mathrm{E}\left[\underline{w}^{k}\right] = \gamma^{-k}\mathrm{E}\left[\underline{q}^{k}\right] = \gamma^{-k}h(\alpha\gamma,k).$$

for the derivation of $\mathrm{E}\left[\underline{w}^{k}\right]$.

The expectation of $\underline{W} = e^{\underline{w}}$ satisfies:

$$\begin{aligned}
\mathrm{E}\left[\underline{W}\right] &= \frac{\gamma}{\theta}e^{1/\theta}\int_0^\infty \exp\left((1+\gamma)w - \theta^{-1}e^{\gamma w}\right)\mathrm{d}w \\
&= \frac{\gamma}{\theta}e^{1/\theta}\int_{\theta^{-1}}^\infty \frac{1}{\gamma y}(\theta y)^{(1+\gamma)/\gamma}e^{-y}\mathrm{d}y \\
&= \theta^{1/\gamma}e^{1/\theta}\int_{\theta^{-1}}^\infty y^{1/\gamma}e^{-y}\mathrm{d}y = \theta^{1/\gamma}e^{1/\theta}\Gamma\left(1+\gamma^{-1},\theta^{-1}\right) \\
y &\coloneqq \theta^{-1}e^{\gamma w} \Rightarrow w = \gamma^{-1}\ln(\theta y) \Rightarrow \frac{\mathrm{d}w}{\mathrm{d}y} = \frac{1}{\gamma y}
\end{aligned}$$

compare the expressions for the moments of $\underline{W}$ for the truncated Weibull distribution in (Wingo 1989).

C Likelihood of the Gompertz Distribution

The density function of the Gompertz distribution reads:

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

Hence, the log likelihood reads:

$$N^{-1}\log\mathcal{L}(\alpha,\gamma) = -\ln\alpha + \gamma\overline{w} - (\alpha\gamma)^{-1}\left(\overline{e^{\gamma w}} - 1\right). \tag{24}$$

The first order condition for α reads:

$$\begin{aligned} N^{-1}\frac{\partial\log\mathcal{L}(\alpha,\gamma)}{\partial\alpha} &= \left(-1 + \frac{\overline{e^{\gamma w}} - 1}{\widehat{\alpha}\gamma}\right)\widehat{\alpha}^{-1} = 0 \Rightarrow \\ \widehat{\alpha} &= \gamma^{-1}\left(\overline{e^{\gamma w}} - 1\right). \end{aligned} \tag{25}$$

The second order condition for α reads:

$$\begin{aligned} N^{-1}\frac{d^2\log\mathcal{L}(\alpha,\gamma)}{(d\alpha)^2} &= \frac{\alpha\gamma - 2\left(\overline{e^{\gamma w}} - 1\right)}{\alpha^3\gamma}, \\ N^{-1}\left.\frac{d^2\log\mathcal{L}(\alpha,\gamma)}{(d\alpha)^2}\right|_{\alpha=\widehat{\alpha}} &= -\widehat{\alpha}^{-2}, \\ \mathrm{SE}(\widehat{\alpha}) &= \widehat{\alpha}\sqrt{N}^{-1}. \end{aligned}$$

In the second line we use equation (25) to simplify the first line.

Substitution of equation (25) in equation (24) yields the concentrated log likelihood (up to a constant):

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

The first (FOC) and second order condition read:

$$\begin{aligned} N^{-1}\left.\frac{d\log\mathcal{L}(\gamma)}{d\gamma}\right|_{\gamma=\widehat{\gamma}} &= \widehat{\gamma}^{-1} + \overline{w} - \frac{\overline{we^{\widehat{\gamma}w}}}{\overline{e^{\widehat{\gamma}w}} - 1} = 0, \\ N^{-1}\frac{d^2\log\mathcal{L}(\gamma)}{(d\gamma)^2} &= -\gamma^{-2} + \left(\frac{\overline{we^{\gamma w}}}{\overline{e^{\gamma w}} - 1}\right)^2 - \frac{\overline{(w^2+1)\,e^{\gamma w}}}{\overline{e^{\gamma w}} - 1}, \\ N^{-1}\left.\frac{d^2\log\mathcal{L}(\gamma)}{(d\gamma)^2}\right|_{\gamma=\widehat{\gamma}} &= \left(2\widehat{\gamma}^{-1} + \overline{w}\right)\overline{w} - \frac{\overline{(w^2+1)\,e^{\widehat{\gamma}w}}}{\overline{e^{\widehat{\gamma}w}} - 1} = -\frac{\overline{(w^2+1)\,e^{\widehat{\gamma}w}} - 2\overline{w}\,\overline{we^{\widehat{\gamma}w}}}{\overline{e^{\widehat{\gamma}w}} - 1} - \overline{w}^2, \\ \mathrm{SE}(\widehat{\gamma}) &= \sqrt{N\left(\frac{\overline{(w^2+1)\,e^{\widehat{\gamma}w}} - 2\overline{w}\,\overline{we^{\widehat{\gamma}w}}}{\overline{e^{\widehat{\gamma}w}} - 1} + \overline{w}^2\right)}^{-1}. \end{aligned}$$

In the third line we use the first line to simplify the second line.

D Gompertz Hazard Rate Estimation

The Gompertz hazard is characterized by

$$H(w) = \frac{1}{\alpha} \exp(\gamma w) \,. \tag{26}$$

The hazard collapses to the constant Exponential hazard as $\gamma \longrightarrow 0$. We use a parametric duration model[19] to estimate γ and α, with maximum likelihood. Specifically, we use the `eha` package in `R`. This package uses two parametrizations for the Gompertz hazard, which the author calls the "rate" and "canonical" parameterizations.[20] The former is closest to our notation, and parameterizes the hazard as

$$H(w) = p \exp(kw) \,, \tag{27}$$

where p is called the `shape` parameter and k the `rate`; $p = \alpha^{-1}$ and $k = \gamma$. To calculate the standard error of α, we use standard Taylor approximation arguments (i.e., $\text{Var}[g(x)] \approx (g'(\text{E}[x]))^2 \cdot \text{Var}[x]$) to calculate $\text{SE}(\alpha) \approx \alpha^{-2} \cdot \text{SE}\left(\alpha^{-1}\right)$.

19. Specifically, we use an accelerated failure time model, since this results in more stable estimations than proportional hazards. Since we do not use covariates, accelerated failure time models are equivalent to proportional hazards models.

20. See: https://cran.r-project.org/web/packages/eha/vignettes/gompertz.html.