

TOP WEALTH IS DISTRIBUTED WEIBULL, NOT PARETO^{*}

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

September 18, 2025

Abstract

We study the shape of the global top wealth distribution, using the *Forbes List of Billionaires*. By constructing test statistics from ratios of log wealth moments, we assess whether the upper tail follows a Pareto distribution. We reject the Pareto null in all region-year combinations with sufficient observations. We then compare log-normal and Weibull alternatives; our moment ratio test results are compatible with both. Introducing a hazard rate-based test for log wealth distributions reveals pronounced right-tail differences: the log-normal distribution is rejected in favor of Weibull. The Weibull distribution accurately predicts regional averages of billionaire wealth, where Pareto performs poorly. Extending the analysis, we find that U.S. city size and firm size distributions are also Weibull-distributed.

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 Arun Advani, 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, and numerous 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

[‡]Leiden University. E-mail: s.j.toussaint@law.leidenuniv.nl

1 Introduction

Right-skewed distributions pervade many aspects of economic life (Gabaix 2009, 2016).¹ The Pareto distribution is the standard model for analyzing these phenomena. This distribution has one parameter governing the fatness of its right tail, the inverse *tail index* σ . The higher σ , the more mass is concentrated at the extreme end of the distribution.

Recent theory and evidence, however, show that Pareto provides a poor fit to the data in many applications. Blanchet, Fournier, and Piketty (2022) demonstrate 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. Despite these challenges, Pareto remains the default assumption in most analyses.

At face value, estimating the inverse tail index σ is straightforward. Let W be the variable of interest (say: wealth) and let us assume that we analyze all observations above some lower bound, normalized to 1 without loss of generality.² Then, the maximum likelihood estimator of σ is simply the mean of $w := \ln W$, see Hill (1975). However, there has been widespread reluctance to use this estimator, maybe largely because this estimator "fits the data badly" when used to predict expected wealth among the richest people, see Blanchet, Fournier, and Piketty (2022). Often, mean log wealth exceeds unity. In that case, expected top wealth is infinite.

As an alternative, most empirical research uses log-rank regressions: 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. Scatter plots and the high R^2 -values of these regressions, usually way above 95%, suggest that this relation is indeed linear, see Figure 1 for some examples. 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. One wonders why going the way of trying to correct an estimator which is known to suffer from biases due to ordering of the data, while a consistent estimator for the tail index is available. Might not be the problem that the starting point of a Pareto distribution for top wealth is at odds with the data? If so, trying to correct a biased estimator of the Pareto coefficient might just ignore the source of the problem.

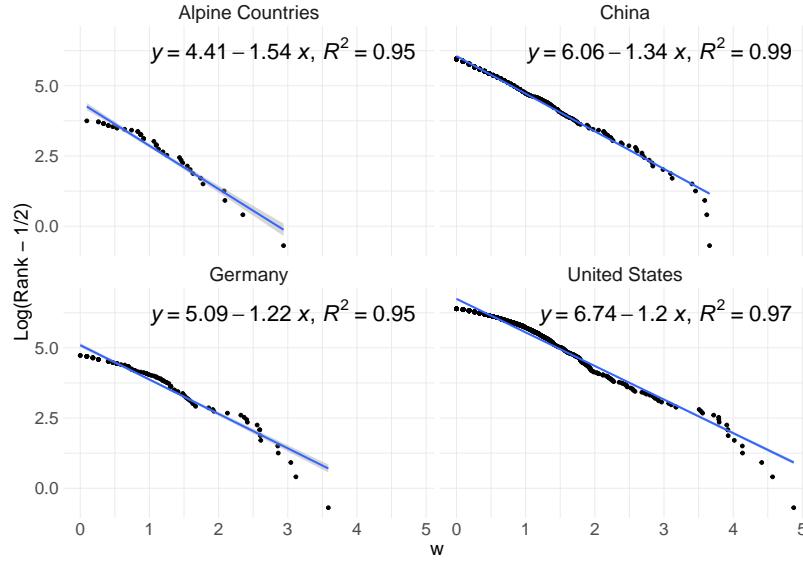
Our first contribution is that we provide conclusive evidence that top wealth is not Pareto distributed. We construct test statistics based on ratios of log-wealth moments; under Pareto, these ratios converge to one regardless of the inverse tail index σ . Because Pareto implies an exponential distribution for $\ln W$, all exponential moments exist, whereas Pareto moments larger than σ^{-1} do not. Applying the second- and third-order moment-ratio tests to the *Forbes List of Billionaires* for 18 regions (2001–2021), we reject Pareto for every region-year combination with enough observations. A joint test across all combinations – allowing each combination to have its own tail index – rejects Pareto even more decisively. We show that this rejection cannot be attributed to classical measurement error.

One might argue that this just implies that one has to dive deeper in the right tail before arriving at the Paretian

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).

2. Any other positive number for the lower bound can be accounted for by dividing the data by this number.

Figure 1: Spurious Log-Rank – Log-Size Regressions



Notes: Figure shows Gabaix and Ibragimov (2011) regressions for the year 2019, for four regions; see Table 3 for further details. All four regressions have very high R^2 's despite all distributions being formally incompatible with a Pareto distribution, as measured by our test statistics R_k being below 1 in all cases.

tail, using the estimation method proposed by Clauset, Shalizi, and Newman (2009) and applied e.g. by Gaillard, Hellwig, Wangner, and Werquin (2023), which estimates simultaneously the inverse tail index σ and threshold above which the upper tail of the distribution is reasonably close to Pareto. Two points invalidate this view. First, if the true distribution were Pareto, log wealth would be exponential with a constant hazard rate; we observe no convergence of the hazard rate to a constant at any threshold. Second, Gaillard, Hellwig, Wangner, and Werquin (2023) find that approximately 12% of respondents lie above their estimated cutoff (based on consumption data rather than wealth). Billionaires constitute roughly one in a million of the population, so the fraction above any plausible Pareto threshold is far smaller; even this tiny fraction lies below the cutoff, implying Pareto would apply to essentially nobody.

Blanchet, Fournier, and Piketty (2022) acknowledge these empirical problems and respond by abandon parametric modeling, proposing a “tail-index curve” with a separate parameter for each income bracket. Our second contribution is to show that this step is unnecessary. We test two alternative distributions for the Pareto/exponential distribution: (i) log-normal/normal and (ii) Weibull/Gompertz for wealth and log wealth respectively, both truncated at the lower support. Where the Pareto/exponential distribution has a single parameter, both alternatives have two. We therefore need a second moment ratio statistic to test whether these alternatives fit the data. Again, we use the moment ratio based on the second and third moment of log wealth. Both the normal and the Gompertz distribution for log wealth predict the log moment ratios to be in a particular ratio to each other; this relation happens to be almost perfectly linear, for both the normal and Gompertz distribution. This is highly convenient for the construction of a test statistic. The positive news is that the empirical moments do fit this linear relation for the alternative distributions almost perfectly for all region-year combinations. The negative news is that both alternative distribution do. The linear relations between the log moment ratios almost coincide. The moment ratios do

therefore not allow us to discriminate between the normal and Gompertz distributions.

Our third contribution is to offer a solution for this problem, using the distinct shapes of the candidates' hazard rates. The exponential distribution has a constant hazard rate, equal to the tail index σ^{-1} . The normal distribution is characterized by an approximately linearly increasing hazard. Its two parameters determine the hazard's level at the lower bound and its slope. The Gompertz distribution, instead, is characterized an exponentially increasing hazard. We nonparametrically estimate the empirical hazard rates and test by means of OLS which of the three candidate distributions fits the data, again allowing for separate parameters for each region-year combination capturing the level and the linear slope of the hazard rate. Unsurprisingly, this test strongly rejects the exponential distribution, confirming the previous conclusion based upon the moment-ratio tests. However, contrary to the moment-ratio test, the hazard rate test also strongly rejects the normal distribution: the second order term is clearly positive, indicating that the hazard rate is convex. Apparently, where the test based on moment-ratios has insufficient power in the extreme right tail, the hazard rate test is able to distinguish between both distributions. This test supports our previous claim that the evidence is inconsistent with the convergence of the hazard rate to some constant, as is required for convergence to the Pareto distribution.

Our results raise the question whether they are typical for the Forbes data on billionaires' wealth or for the wealth distribution in general, or that they apply more generally to other phenomena that have been presumed to be Paretoian. We present evidence for two other variables: city and firm size. The shape of the U.S. city size distribution is subject to a long-standing debate. The default of Pareto (specifically Zipf) is assumed by e.g. Krugman (1996) and Gabaix (1999), who use the 135 largest Metropolitan Statistical Areas. Eeckhout (2004, 2009) challenges this result, and argues for log-normal. 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). Again, we find overwhelming evidence against Pareto and in favor of Weibull.

We conclude that Weibull must replace Pareto in many economic models. We explore some possible theoretical rationalisations of this conclusion. For example, the Gompertz distribution figures in stochastic networks (Tishby, Biham, and Katzav 2016). This model structure has interesting parallels with the variables we study; for instance, city size is constrained by the land area or population 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 Kojen (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 2024). 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 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, and our tests to discriminate between log-normal and Weibull. In Section 5, we further apply our framework to cities and firms. Section 6 concludes.

2 Framework

2.1 Candidate distributions

Consider a random variable \underline{W} (stochastic variables will be underlined) with a support $[1, \infty)$; the normalization of the lower support to one is without loss of generality. Let $\underline{w} := \ln \underline{W}$ be the corresponding log, hence with a support $[0, \infty)$. To economize on notation, we suppress the support in what follows, but all moments and functions of moments we present are *conditional* on $\underline{W} \geq 1$ or $\underline{w} \geq 0$.

Our first candidate is the Pareto distribution with a single parameter, the inverse tail index $\sigma > 0$. When \underline{W} follows a Pareto distribution, its log \underline{w} follows an exponential distribution. The literature often uses its inverse σ^{-1} , which is commonly referred to as the tail index or Pareto coefficient (Jones 2015, Gabaix 2009, 2016). Both conventions have advantages and disadvantages. Our choice to work with σ rather than its inverse is motivated by the fact that σ has a common interpretation as a scale parameter of w that applies to all three candidate distributions that we consider.

If $\sigma = 1$, the Pareto distribution specializes to the Zipf distribution. The first row of Table 1 shows the counter-distribution for Pareto; for the Zipf distribution, this expression simplifies to $\Pr[\underline{W} \geq W] = W^{-1}$. This expression is easy to interpret, since it predicts the probability of a “large” observation to scale with size like a power law.

Table 1: Three right tail distributions and the corresponding distribution for its log

Level (W)			Log (w)		
Distribution	Function		Distribution	Function	
	Counter-distribution	Density		Counter-distribution	Density
Pareto	$W^{-1/\sigma}$	$\frac{1}{\sigma} W^{-\frac{1+\sigma}{\sigma}}$	exponential	$e^{-w/\sigma}$	$\frac{1}{\sigma} e^{-w/\sigma}$
log-normal	$\frac{\Phi(-\psi - \ln W/\sigma)}{\Phi(-\psi)}$	$\frac{\phi(\psi + \ln W/\sigma)}{\sigma W \Phi(-\psi)}$	normal	$\frac{\Phi(-\psi - \frac{w}{\sigma})}{\Phi(-\psi)}$	$\frac{\phi(\psi + \frac{w}{\sigma})}{\sigma \Phi(-\psi)}$
Weibull	$e^{-\psi(W^{1/\sigma}-1)}$	$\frac{\psi}{\sigma} W^{\frac{1-\sigma}{\sigma}} e^{-\psi(W^{1/\sigma}-1)}$	Gompertz	$e^{-\psi(e^{\frac{w}{\sigma}}-1)}$	$\frac{\psi}{\sigma} e^{\frac{w}{\sigma}} - \psi(e^{\frac{w}{\sigma}}-1)$

Notes: $\sigma > 0$ and $\psi > 0$. $\Phi(\cdot)$ and $\phi(\cdot)$ are the distribution and density function, respectively, of the standard normal distribution.

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.

Table 1 compares the Pareto distribution with two alternatives: the log-normal and the Weibull distribution, both left truncated at $\underline{W} = 1$. Both distributions have one additional parameter beyond the scale parameter σ . This parameter, ψ , measures the skewness of the distribution. The log of a log-normally distributed variable is distributed normally, the log of a variable with a Weibull distribution has a Gompertz distribution; both distributions are left truncated at $\underline{w} = 0$. The Gompertz distribution is less well known in economics, but commonly used in demography to model life expectancy as a function of age.³

2.2 Moments and Specification Testing

Let $\underline{x} := \underline{w}/\sigma$ be the standardized version of the log transform \underline{w} . As can be checked easily by substituting w/σ for x in the expressions for the counter-distribution function in Table 1, this function does not depend on the scale parameter σ anymore. Evidently, then the same must apply to the corresponding density function, as this is its derivative with respect to x . Define $\mathcal{M}_k^X(\sigma, \psi)$ as the non-central moment of distribution X , $\mathcal{R}_k^X(\psi)$ as its normalized moment ratio, and $\mathcal{H}^X(w, \psi)$ as its hazard rate

$$\begin{aligned}\mathcal{M}_k^X(\sigma, \psi) &:= E[\underline{w}^k] = \sigma^k E[\underline{x}^k], \\ \mathcal{R}_k^X(\psi) &:= \frac{E[\underline{w}^k]}{k! E[\underline{w}]^k} = \frac{E[\underline{x}^k]}{k! E[\underline{x}]^k}, \\ \mathcal{H}^X(w, \sigma, \psi) &:= -\frac{d \Pr[\underline{w} > w]/dw}{\Pr[\underline{w} > w]} = -\sigma \frac{d \Pr[\underline{x} > x]/dx}{\Pr[\underline{x} > x]},\end{aligned}$$

where $k \geq 1$ is an integer and where $X \in \{\mathcal{E}, \mathcal{N}, \mathcal{G}\}$ for the exponential, normal, and Gompertz distribution respectively. While the moments $\mathcal{M}_k^X(\sigma, \psi)$ depend on both σ and ψ , the moment ratio $\mathcal{R}_k^X(\psi)$ only depends on ψ . This is the first advantage of focusing on the distribution of the log transform \underline{w} rather than the original

3. 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 Fréchet. 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.

variable \underline{W} : one can compute a moment ratio directly from the data of which the value does not depend on the scale parameter σ . This is an important advantage when it comes to testing the shape of empirical distributions.

Table 2: Properties of $\mathcal{M}_k^X(\sigma, \psi)$, $\mathcal{R}_k^X(\psi)$, and $\mathcal{H}^X(w, \sigma, \psi)$

Statistic	Distribution		
	exponential	normal	Gompertz
$\mathcal{M}_k^X(\sigma, \psi)$	$\sigma^k k!$	Appendix	$\sigma^k h_k(\psi)$
$\mathcal{R}_k^X(\psi)$	1	Appendix	$\frac{h_k(\psi)}{k! e^{k\psi} \text{Ei}(\psi)^k}$
$\lim_{\psi \rightarrow \infty} \mathcal{R}_k^X(\psi)$	1	1	1
$\lim_{\psi \rightarrow 0} \mathcal{R}_k^X(\psi)$	1	Appendix	$k!^{-1}$
$\mathcal{H}^X(w, \sigma, \psi)$	$\frac{1}{\sigma}$	$\frac{1}{\sigma} \Lambda(\psi + \frac{w}{\sigma}) \cong \frac{\psi}{\sigma} + \frac{w}{\sigma^2}$	$\frac{\psi}{\sigma} e^{w/\sigma}$

Notes: $\text{Ei}(\cdot)$ stands for the exponential integral; $\Lambda(x) := \phi(x)/\Phi(-x)$ is the inverse Mills' ratio. The derivation of these expressions is in the Appendix.

Table 2 summarizes the moment functions of the three candidate distributions of $\underline{x} := \underline{w}/\sigma$, where we use for the moments of the Gompertz distribution:

$$\begin{aligned} h_k(\psi) &:= k e^\psi \int_0^\infty x^{k-1} \exp(-\psi e^x) dx, \\ h_1(\psi) &= e^\psi \text{Ei}(\psi), \end{aligned}$$

see the Appendix for the derivation. The expressions for $\mathcal{M}_k^N(1, \psi)$ provide little insight and are therefore relegated to the Appendix.

The moments exist for all three candidate distributions for all admissible values of the parameters $\sigma > 0$ and $\psi > 0$. Though all moments of \underline{W} exist for all admissible parameter values for the log-normal and Weibull distribution, this is not the case for the Pareto distribution, for which the moments do not exist for $k \geq \sigma^{-1}$. This is the second advantage of focusing on the log transform \underline{w} rather than \underline{W} . Hence, we will work with the log transform \underline{w} and its distributions – exponential, normal, and Gompertz – from now on.

Table 2 shows that for the exponential distribution, the moment ratio $\mathcal{R}_k^E(\psi)$ is equal to unity for all ψ and k . This is not the case for the normal and Gompertz distribution, where $\mathcal{R}_k^X(\psi)$ converges to unity for $\psi \rightarrow \infty$ only, for all k and for $X \in \{N, G\}$. For ψ finite, $\mathcal{R}_k^X(\psi)$ is less than one and a declining function of k .

The reason that the moment $\mathcal{R}_k^X(\psi)$ converges to unity for the normal and Gompertz distribution for $\psi \rightarrow \infty$ is that their moments $\mathcal{M}_k^X(\sigma, \psi)$ converge to the moments of the exponential distribution in this limit. In this sense, these distributions have an ‘asymptotic Pareto/exponential tail’. However, while the moments converge to the exponential distribution, the hazard rate $\mathcal{H}^X(w, \sigma, \psi)$ does not. While $\mathcal{H}^E(w, \sigma, \psi) = \sigma^{-1}$ for all w for the exponential distribution, $\mathcal{H}^X(w, \sigma, \psi)$ is increasing in the limit for $\psi \rightarrow \infty$ for both the normal and Gompertz distribution. In this sense, these distributions do therefore not have an ‘asymptotic Pareto/exponential tail’. The reason that the moments of both distributions nevertheless converge to those of the exponential distribution can be

seen from the survival probability

$$\Pr[\underline{w} > w] = \exp \left[- \int_0^w \mathcal{H}^X(z, \sigma, \psi) dz \right] \cong \exp[-w\mathcal{H}^X(0, \sigma, \psi)],$$

where the latter approximation holds for small w . For high ψ , the initial hazard $\mathcal{H}^X(0, \sigma, \psi)$ is that high, that even for small w , the survival probability is low such that there are not enough survivors for the subsequent increase in the hazard to have a substantial effect on the moments. Therefore, only the initial hazard matters, just as in the case of the exponential distribution.

Much of the literature that uses Pareto distributions as a benchmark does so because many stochastic processes are argued to have an ‘asymptotic Pareto tail.’ This claim is based on 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

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

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: $\mathcal{H}^E(w, \sigma, \psi) = \sigma^{-1}$. 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). Since its hazard rate is constant, the exponential distribution satisfies this criterion, while the normal and Gompertz distribution do not, even while their moments converge to exponential distribution as explained before. The claim that a distribution converges to Pareto/exponential when far enough down into the right tail is therefore misleading: when it applies to the moments, it might not apply to the hazard rates.

The expressions for hazard rates in Table 2 show that the Gompertz distribution is even thinner tailed than the normal: its hazard rates increases exponentially in w , while that of the normal distribution increases only linearly. This analysis provides a cautionary tale for our ability to discriminate empirically between both distributions. Seen through the lens of the hazard rates, the differences between the exponential distribution on the one hand and the normal and Gompertz on the other hand regards the first derivative of the hazard rate: it is zero for the exponential distribution, while it is positive for normal and Gompertz. The difference between the latter two regards its second derivative: it is asymptotically zero for the normal distribution, while it is positive for Gompertz. Establishing this second derivative empirically is a challenge, as it relies on data points far down in the right tail.

2.3 Testing

Let M_k and R_k be the sample moments and their ratios corresponding to the definitions of $\mathcal{M}^X(\sigma, \psi)$ and $\mathcal{R}^X(\psi)$:

$$\begin{aligned} M_k &:= \overline{w^k} = \sigma^k \overline{x^k}, \\ R_k &:= \frac{\overline{w^k}}{k! \overline{w^k}} = \frac{\overline{x^k}}{k! \overline{x^k}}, \end{aligned}$$

where a bar on top of a variable denotes its sample mean, calculated from a sample of size N . M_k and R_k do not depend on parameters can therefore be calculated directly from data on w .

Since the moment ratios $\mathcal{R}_k^E(\psi)$ for the exponential distribution are equal to one for all k , the values for R_k are a suitable starting point for testing whether the data are drawn from an exponential distribution. We give a general expression for the asymptotic variance of R_k in the Appendix. For $X = E$, $\text{plim}(R_k) = 1$ for all k and the variance specializes into a simple expression:

$$\begin{aligned} \text{plim}(N\text{Var}[R_k]) &= \frac{(2k)!}{k!^2} - k^2 - 1, \\ \text{plim}(N\text{Var}[R_1]) &= 0, \quad \text{plim}(N\text{Var}[R_2]) = 1, \quad \text{plim}(N\text{Var}[R_3]) = 10. \end{aligned} \tag{2}$$

We present the variance of R_k for $k = 1$ as a consistency check; since $R_1 = 1$ by definition, its variance must be zero. Since $\mathcal{R}_k^E(\psi)$ does not depend on any parameter, one moment ratio R_k suffices to perform this test. Though the higher asymptotic variance of R_k for higher k suggests that R_2 has most discriminating power, one can use any k . We use the expressions for $\text{Var}[R_2]$ and $\text{Var}[R_3]$ in Section 4 to test whether the realisations of the moment ratios R_2 and R_3 are consistent with the data being drawn from an exponential distribution.

When this hypothesis is rejected, the moments R_k can then be used to test whether the data are drawn from a normal or Weibull distribution. Since these distributions have an additional parameter ψ , this test requires moment ratios R_k for two values of k . Let $\psi_k^X(R_k)$ for $k \geq 2$ be the inverse function of $\mathcal{R}_k^X(\psi)$, such that

$$\psi_k^X(\mathcal{R}_k^X(\psi)) = \psi.$$

If the distribution of w is indeed X with parameter ψ , then R_k is a consistent estimator of $\mathcal{R}_k^X(\psi)$ and hence $\psi_k^X(R_k)$ is a consistent estimator of ψ . In the Appendix, we show that for the normal distribution, $\psi_2^N(R_2)$ is the maximum likelihood estimator of ψ . Such a simple equivalence does not apply for the Gompertz distribution, but this does not matter for the test we propose. Given that both $\psi_k^X(R_k)$ and $\psi_m^X(R_m)$ for $2 \leq k < m$ are consistent estimates for ψ , $\psi_k^X(R_k)$ and $\psi_m^X(R_m)$ must be asymptotically equal and hence $\psi_k^X(R_k) - \psi_m^X(R_m)$ must be asymptotically equal to zero. In practice, our test statistic takes an even simpler form. We can use the equality $\psi_k^X[\mathcal{R}_k^X(\psi)] = \psi_m^X[\mathcal{R}_m^X(\psi)]$ for establishing a direct functional relation between $\mathcal{R}_k^X(\psi)$ and $\mathcal{R}_m^X(\psi)$:

$$\mathcal{R}_{k,m}^X(R) := \mathcal{R}_k^X[\psi_m^X(R)].$$

Our argument implies that if the distribution of \underline{w} is indeed \mathcal{X} , the following relation holds:

$$\text{plim} \left[R_k - \mathcal{R}_{k,m}^{\mathcal{X}}(R_m) \right] = 0.$$

Empirically, we shall find values for R_2 mostly in the range $R_2 \in (0.75, 0.90)$, which roughly corresponds to $\psi \in (-0.5, 2)$ and $\psi \in (1, 8)$ for the normal and Gompertz distribution respectively. It turns out that the relation between $\ln \mathcal{R}_2^{\mathcal{X}}(\psi)$ and $\ln \mathcal{R}_3^{\mathcal{X}}(\psi)$ implied by $\mathcal{R}_{2,3}^{\mathcal{X}}[\mathcal{R}_3^{\mathcal{X}}(\psi)]$ is almost perfectly linear for this range of ψ . Using this almost perfect log linearity, we obtain a very simple test statistic:

$$\begin{aligned} \text{plim} \left(\ln R_2 - \lambda^{\mathcal{X}} \ln R_3 - \lambda_0^{\mathcal{X}} \right) &\cong 0, \\ \lambda^N &= 0.390, \quad \lambda_0^N = 0.011, \\ \lambda^G &= 0.404, \quad \lambda_0^G = 0.027. \end{aligned} \tag{3}$$

Appendix Figures B.2 and C.2 plot $\ln \mathcal{R}_2^{\mathcal{X}}(\psi) - \lambda^{\mathcal{X}} \ln R_3 - \lambda_0^{\mathcal{X}}$ for both distributions. The deviation is less than 0.001 for the relevant ranges for both distributions. This is far smaller than the standard deviation in the estimated ratios for sample sizes of 1000 observations or more, see the Appendix for an expression for this standard deviation.

Note that these linear functions are very similar for both distributions. This underscores our cautionary tale: distinguishing between both distributions might be a hard task.

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 and their 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 in 18 regions, see Table 3. 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. As a rough threshold for the minimum number of billionaires for a region we use 40 billionaires in 2019. Countries that cannot easily be included in a region are excluded from the region classification. Hence, the region classification excludes some countries; the super-region classification printed in bold covers all countries. Therefore, a super-region can have more billionaires than the sum of its constituent regions.

The super-regions China and India consist of one country and are therefore also classified as a region. The Rest of the World has fewer than 40 billionaires in 2019. Moreover, this super-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.

Table 3: Region Classification & Descriptive Statistics

(Super-)Region	Area	Region Classification		Statistics (Average 2001–2021)		
		R_2	R_3	\bar{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: Super-regions may include billionaires from countries not part of their constituent regions. China and India count both as regions and super-regions. N = total number of billionaires; L is total population in millions; \bar{w} = mean log wealth; R_2 and R_3 are the normalized variance and skewness of mean log wealth.

Table 3 provides summary statistics for all regions and super-regions. We report average numbers for R_2 and R_3 , and \bar{w} , and the average number of billionaires (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 $E[W]$ does not exist for the Pareto distribution. Second, both R_2 and R_3 are consistently smaller than one in all regions, contradicting the prediction of a Pareto distribution.

4 Testing the Candidate Distributions

4.1 Pareto/Exponential

This section tests the Pareto hypothesis using the moment ratio test discussed in Section 2. We have $18 \text{ regions} \times 21 \text{ years} = 378$ region-year combinations available for estimation. We calculate R_2 and R_3 for each observation. The variances of R_2 and R_3 are predicted to be N^{-1} and $10N^{-1}$ respectively, where N is the number of observations in each region-year combination, see equation (2). 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 test whether the intercept of this regression is equal to one and whether the RMSE is in accordance with its predicted value of one for R_2 and $\sqrt{10}$ for R_3 .

This test procedure implies that we allow the scale parameter σ to vary between region-year combinations. This hypothesis nests the stricter hypothesis that all region-year combinations share a common σ . Our procedure offers a simultaneous test for all regions. The simultaneity of this test for all 75 region-year combinations greatly increases its power, compared to testing all 75 combinations separately.

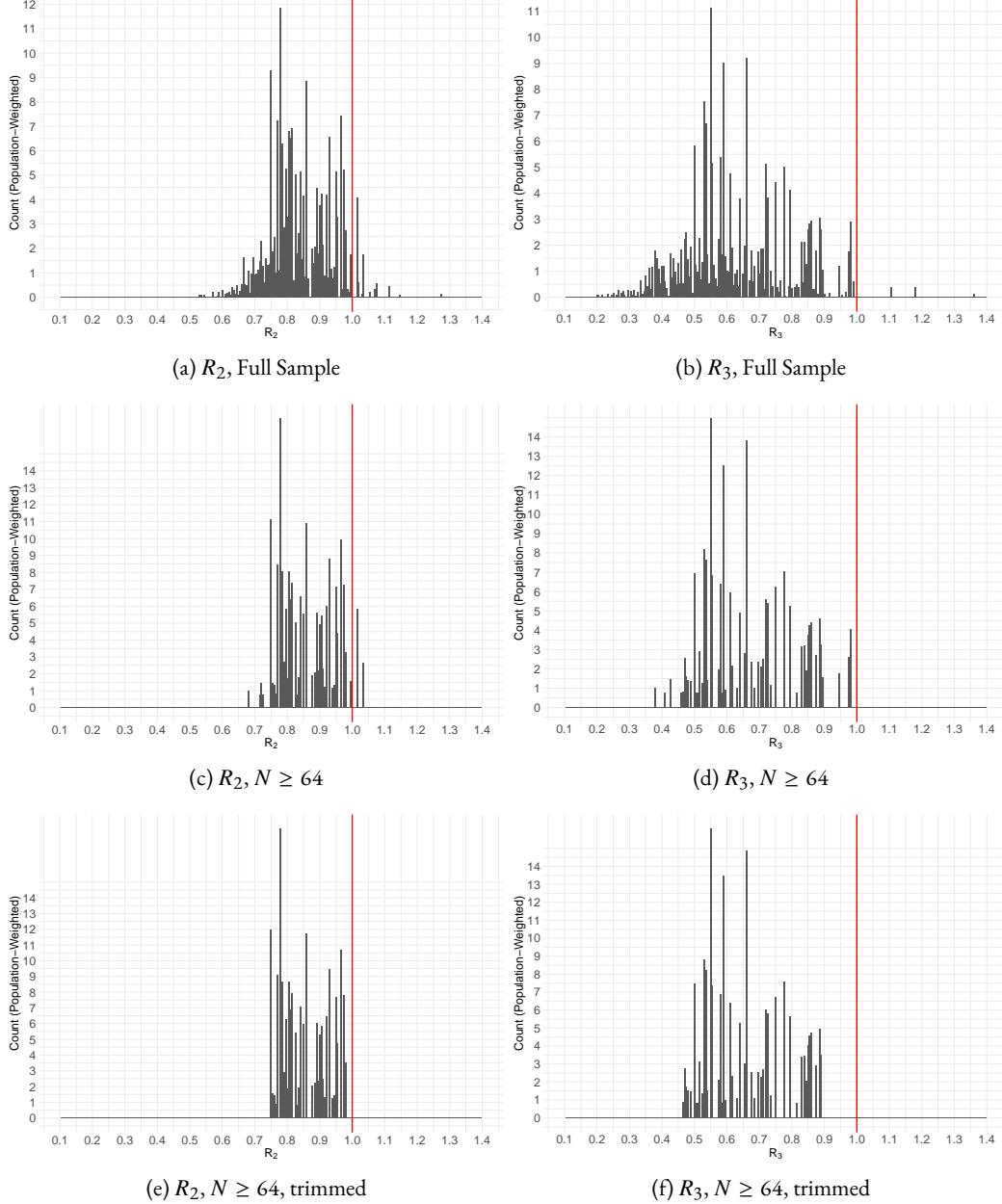
The scale parameters may or may not be correlated between region-year combinations, e.g. within years or regions. Quite likely, they are. For example: the US has a higher Pareto coefficient and hence a lower scale parameter than most European countries during the whole observation period, causing the values of σ to be correlated over time. This correlation is irrelevant for our procedure, since the randomness that is accounted for in our test-procedure comes from the individual observations on billionaires' wealth within each region-year combination with its own scale parameter σ . In panel data wording: the variation in σ is dummied out. In the extreme case where the scale parameters are identical for all combinations (and hence perfectly correlated), our estimations procedure is still valid, although not fully efficient, since we ignore the information that inverse tail index parameters σ are the same for all combinations.

Table 4 reports the results. We clearly reject the null that $R_2 = R_3 = 1$. The intercepts are highly significantly different from one. This remains the case when we do the following robustness checks: drop region-year combinations with fewer than 64 billionaires to reduce the small-sample bias in R_2 and R_3 (column (2) and (5)) and further drop the bottom and top 5% of observations for R_2 and R_3 as a robust regression (column (3) and (6)).

We also report the theoretical root mean squared error (RMSE) derived from equation 2, except for column (3) and (6) where the selective dropping of observations invalidates the prediction for the RMSE. If all variation in R_2 and R_3 were sampling variation (with replacement), the observed RMSE should be equal to the theoretical RMSE, which is equal to one for R_2 and $\sqrt{10}$ for R_3 . This is clearly not the case. Using the sample with $N \geq 64$, the observed RMSE is 31% of the theoretical RMSE for R_2 and $0.54/\sqrt{10} \approx 17\%$ for R_3 . The lower RMSE therefore is further evidence that the actual distribution is not exponential.

Figure 2 shows R_2 and R_3 for all region-year combinations: the great majority is smaller than unity. This holds true when we restrict attention to these combinations with more than 64 billionaires and (obviously) when we further drop the top and bottom 5% of observations. The main takeaway from Table 4 and Figure 2 is that Pareto distribution is strongly rejected across all regions and years.

Figure 2: Distribution of Test Statistics R_2 and R_3



Notes: Figures show the distribution of R_2 and R_3 , with all observation-years pooled. The red line indicates the predicted value under the null that wealth is Pareto-distributed, i.e., $R_2 = R_3 = 1$. 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 4: WLS Regressions, Pareto Test

Dependent Variables:	R_2			R_3		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
<i>Variables</i>						
Constant	0.82*** (0.005)	0.86*** (0.02)	0.86*** (0.01)	0.59*** (0.008)	0.66*** (0.03)	0.65*** (0.02)
Weights	\sqrt{N}	\sqrt{N}	\sqrt{N}	\sqrt{N}	\sqrt{N}	\sqrt{N}
<i>Fit statistics</i>						
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: ***: 0.001, **: 0.01, *: 0.05

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

What about measurement error? Although the *Forbes* list is widely used in research (e.g., Gomez (2023)), its deficiencies as a data source are well known. Could our rejection of Pareto be driven by measurement error? We argue that this is highly unlikely. Let v be a normally distributed classical measurement error with zero mean and variance of θ^2 and $\text{Cov}[\underline{w}, v] = 0$. Observed log wealth satisfies $\tilde{w} := w + v$. The density function $f(\tilde{w})$ of observed log wealth reads

$$f(\tilde{w}) = \int_{-\infty}^{\infty} \theta^{-1} e^{-(\tilde{w}-v)/\sigma} \phi\left(\frac{v}{\theta}\right) dv = e^{-\tilde{w}/\sigma} \int_{-\infty}^{\infty} \theta^{-1} e^{v/\sigma} \phi\left(\frac{v}{\theta}\right) dv = e^{(\theta/\sigma)^2/2} e^{-\tilde{w}/\sigma}. \quad (4)$$

The tail index of actual and observed log wealth is therefore the same, since the proportional constant $e^{(\theta/\sigma)^2/2} > 1$ drops out when normalizing the density function as to integrate to one.

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 \tilde{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 $w = \tilde{w} - |v|$ whose wealth gets overreported to be \tilde{w} , than there are people with log wealth $w = \tilde{w} + |v|$ 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 \tilde{w} due to the exponential distribution of actual log wealth w , this constant of proportionality is the same for all \tilde{w} . Hence, the moment ratios R_k remain unaffected. Measurement error does therefore not affect our test procedure.⁴

The common procedure of using rank-size regressions (Gabaix and Ibragimov 2011), however, is biased under

4. Note that this argument applies only to the extreme right tail of the distribution of wealth. As soon as $\tilde{w} - |v|$ falls below the lower support of 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.

measurement error. First, mismeasured w causes the OLS estimate for σ^{-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.

Gaillard, Hellwig, Wangner, and Werquin (2023) applied the method of Clauset, Shalizi, and Newman (2009) to consumption data. This method is based on the assumption that only the data above some threshold fit the Pareto distribution. The method sets out to simultaneously estimate this threshold and the inverse tail index σ . They find that there is indeed a Pareto tail that applies to 12% of the population. Our data on billionaires cover a fraction of roughly one in a million of the richest people on earth. If the Pareto tail does not apply to this highly selective subset of people, one may wonder to whom it may then apply.

4.2 Normal and Gompertz: Moment Test

Having rejected the Pareto distribution, we turn our attention to the normal and Gompertz distribution. We focus on the 75 region-year combinations with more than 64 observations. We run a WLS regression of $\ln R_2 - \lambda^X \ln R_3 - \lambda_0^X$ for $X \in (\mathcal{N}, \mathcal{G})$ on an intercept, using \sqrt{N} as weights. The theoretical prediction is that this intercept is not significantly different from zero. Again, this is a more lenient hypothesis than that all observations are drawn from the same distribution. We allow σ and ψ to vary between region-year combinations. The tested hypothesis is just that the distribution is either normal or Gompertz for all combinations, where its parameters σ and ψ are allowed to vary between combinations. Again, the simultaneity of this test for all region-year combinations greatly increases its power.

Table 5: $\ln R_2 - \lambda \ln^X R_3$, normal vs. Gompertz

Dependent Variables:	$\ln R_2 - \lambda^N \ln R_3$	$\ln R_2 - \lambda^G \ln R_3$		
Model:	(1)	(2)	(3)	
<i>Variables</i>				
Constant	0.004 (0.002)	0.002 (0.003)	0.002 (0.002)	-0.006** (0.002)
Weights	\sqrt{N}	\sqrt{N}	\sqrt{N}	\sqrt{N}
<i>Fit statistics</i>				
Observations	378	75	378	75
RMSE	0.072	0.086	0.059	0.069

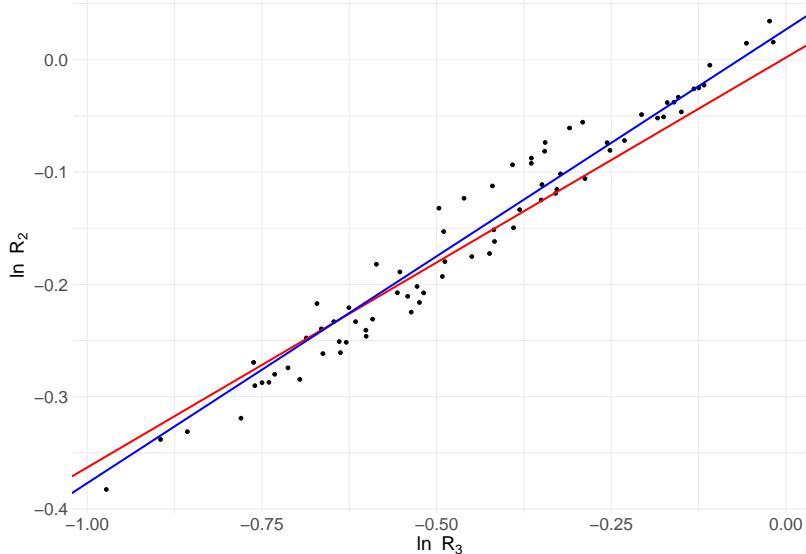
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05

Notes: Driscoll-Kraay (L=2) standard-errors in parentheses

Table 5 reports the results. Column (1) and (3) uses the full set of region-year combinations, column (2) and (4) use only the 75 combinations with more than 64 observations. The intercept is not significantly different from zero for three out of four cases, and in the fourth case the coefficient is very small in magnitude. Hence, we conclude that either normal or Gompertz fits the data well. Based on Table 5, there is little to choose between the two distributions: normal fares best in the sample with at least 64 observations, yet Gompertz fits almost perfectly in the full sample.

At first sight, this result seems to support a negative conclusion for our ambition to establish a common distribution for log wealth. However, we have reasons for a more positive evaluation. Figure 3 plots the functions $\ln \mathcal{R}_2^X(\psi) - \lambda^X \ln \mathcal{R}_3^X(\psi) - \lambda_0^X$ in the $\{\ln R_2, \ln R_3\}$ space for both distributions $X \in \{\mathcal{N}, \mathcal{G}\}$, as well as their realisations for $\{\ln R_2, \ln R_3\}$ for the 75 region-year combinations with more than 64 observations. Figure 3 justifies two conclusion. First, it underscores the conclusion from the discussion of equation (3) that the relations for the normal and Gompertz distribution almost coincide, dwarfing the hope that this test can discriminate between both distributions. However, the individual realisations for $\{\ln R_2, \ln R_3\}$ are closely lined up to the theoretical relation. On an individual basis, many region-year combinations are less than two standard deviations of the RMSE away from zero. The results seem to be somewhat more favorable for Gompertz than normal, but we do not want to make too much of a big deal of this conclusion, in particular since the t -test on the intercept rejects Gompertz when using only the combination with more than 64 observations. Normal and Gompertz clearly fit the data much better than the common presumption that they are drawn from the exponential distribution. In the latter case, all $\{\ln R_2, \ln R_3\}$ combinations should be scattered randomly around the point $\{0, 0\}$, which is clearly at odds with the data.

Figure 3: $\ln R_2$ vs. $\ln R_3$



Notes: Figure plots the empirical values of $\ln R_3$ as a function of $\ln R_2$ for the 75 region-year combinations with at least 64 observations. The blue line gives the Gompertz predicted value, and the red line the Normal predicted value.

Remarkably, the root mean squared error hovers around 0.065 in the four regressions, which is about one third of the theoretical standard deviation of R_2 , see Figure C.3 in the Appendix. This suggests that the wealth of individual billionaires are not independent draws with replacement from the normal or Gompertz distribution, but that there is some interaction between wealth-levels of different billionaires. For one well known example, when Mark Zuckerberg had a dispute with some other students at Harvard about the intellectual ownership of the ideas for Facebook, it was either Mark Zuckerberg or these students who won the day. Their draws from the wealth distribution were therefore interrelated. There was room for one Facebook, not two. This observation might have important

implications for the type of theory used to model the distribution of top wealth.

As a further exploration of the fit of the theoretical distributions to the data and the ability to discriminate between the normal and the Gompertz distribution, Figure 4 plots the empirical distributions of log wealth w against the theoretical prediction for the three candidate distributions. For normal and Gompertz, the distribution of w depend on two parameters, σ and ψ . We use the fact that $w/\text{E}[w]$ depends only on ψ and not on σ . Hence, by dividing the data for each region-year combination by their mean, w/\bar{w} , their theoretical distribution depends on ψ only, which implies a particular value of $\text{plim } R_2 = \mathcal{R}_2^X(\psi)$. We focus on three values for R_2 in the empirically relevant range: 0.75, 0.80, and 0.90. For the empirical distribution, we combine the data from the ten region-year combinations for which R_2 is closest to these three reference values, so that we have a sufficient number of observations on w/\bar{w} for a reliable plot. For each distribution, we set $\sigma = \mathcal{M}_1^X[1, \psi_2^X(R_2)]^{-1}$. This makes the theoretical prediction comparable to the data since $\mathcal{M}_1^X[\sigma, \psi_2^X(R_2)] = \sigma \mathcal{M}_1^X[1, \psi_2^X(R_2)]$. For the exponential distribution, this implies $\sigma = 1$ for all three values of R_2 , for normal and Gompertz, of σ varies between the three reference values for R_2 .

Where the counter-distribution has most visual power just above the lower bound, its log has most power in the extreme right tail, where the number of survivors is low and hence relative changes are more informative than absolute changes.

Figure 4 plots the results. The top three panels ((a)–(c)) show the counter-distributions for $R_2 = \{0.75, 0.80, 0.90\}$. For the 0.75 and 0.80 cases, the empirical data align quite closely to either the normal (blue) or Gompertz (red) theoretical curves. In contrast, the exponential distribution (green) fits much worse. For $R_2 = 0.90$, all three parametric curves fit the data well.

The middle panels ((d)–(f)) show the log-transform of the counter-distribution. While the exponential distribution now quite clearly diverges from the normal and Gompertz, these two candidates remain almost indistinguishable, in particular for $R_2 = 0.90$; only for $R_2 = 0.75$, the log distribution is visibly different. The parametric curves for normal and Gompertz overshoot for $R_2 = 0.75$, fit very well for $R_2 = 0.80$ and undershoot for $R_2 = 0.90$. It should therefore not come as a surprise that the relations between $\ln R_2$ and $\ln R_3$ in Figure 3 are very similar and that it is difficult to discriminate between both distributions using these moment ratios.

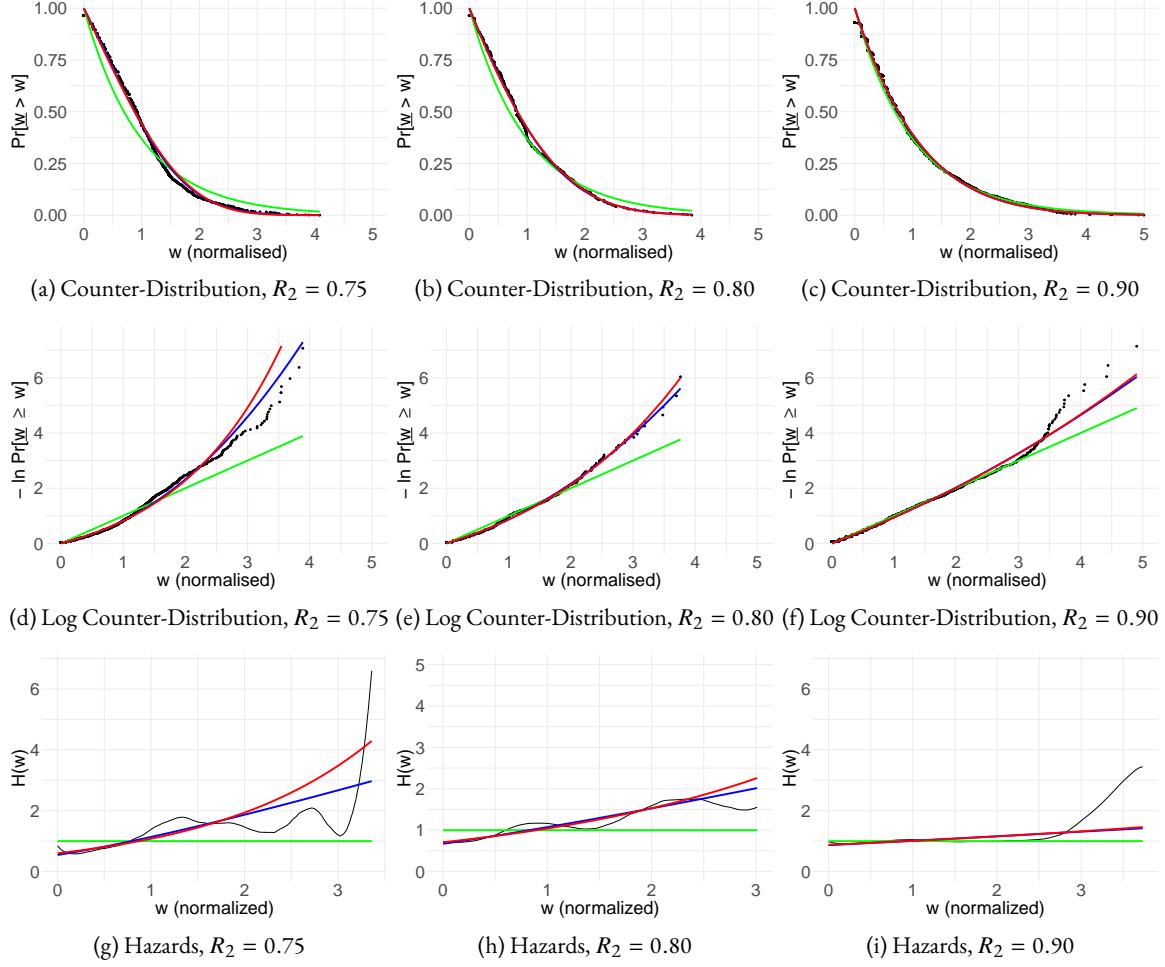
The bottom panels ((g)–(i)) show a potential resolution to this conundrum. There, we plot the parametric and empirical hazard rates $H(w)$. We estimate the empirical hazard rates nonparametrically using an Epanechnikov kernel⁵; this results in smooth hazard curves over the wealth distribution. The three candidate distributions have starkly different predictions regarding $H(w)$: it should be constant (exponential), linearly increasing (normal), or convexly increasing (Gompertz). While the nonparametric estimation means that none of the three hazard curves plotted directly fits with the predicted curves, visual inspection does indicate a convex increase in the hazard rates. This is most evident for $R_2 = 0.75$ and $R_2 = 0.90$; for $R_2 = 0.80$, the multimodal hazard curve could also be consistent with the linearly increasing normal hazard rate.

4.3 Normal and Gompertz: hazard-based tests

The plots of the log distributions in Figure 4 offers a glimpse of hope, focusing on the extreme right tail of the data, that there might be some prospect of distinguishing between normal and Gompertz. This section will do so by

5. Specifically, we use the `muhaz` package in R, with the right bound of w set equal to the maximum value in each sample (i.e., no truncation) and with a global bandwidth (i.e., no smoothing of the curve based on local MSE minimization).

Figure 4: (Log) Counter-Distribution and Hazards, Exponential vs. Normal vs. Gompertz



Notes: Figure plots the counter-cumulative distribution functions (top panels), log counter-cumulative distribution functions (middle panels), and hazard rates (bottom panels) of the exponential (green), normal (blue) and the Gompertz distribution (red), normalized such that their respective values of σ and ψ are consistent with our mean value of $R_2 = 0.75$ (left panels), $R_2 = 0.80$ (middle panels), and $R_2 = 0.90$ (right panels). Hazard rates are estimated with an Epanechnikov kernel.

analyzing the hazard rates $\mathcal{H}^X(w; \sigma, \psi)$. As shown in Table 2, the hazard rate of the normal distribution grows (approximately) linearly with w , while it grows exponentially for the Gompertz distribution.

We exploit this difference in the shape of the hazard rates for a test. We compute the empirical hazard rate $H_{rt}(w_i)$ nonparametrically using an Epanechnikov kernel; since every billionaire is treated as an independent observation, the observation index i implies the region-year combination rt . We use $H_{rt}(w_i)$ as the dependent variable for a regression of the form

$$H_{rt}(w_i) = \beta_{0rt} + \beta_{1rt}w_i + \beta_2 w_i^2 + \varepsilon_{it}. \quad (5)$$

This regression includes fixed effects β_{0rt} for each region-year combination, separate first-order terms $\beta_{1rt}w_i$ for each combination, and a common second-order term in $\beta_2 w_i^2$.

The hazard rate for the exponential distribution is equal to σ^{-1} and independent of w_i . The variation in σ between region-year combinations that we allowed for in Section 4.1 is captured by the region-year combination fixed effects β_{0rt} . The independence of the hazard rate from w_i implies $\beta_{1rt} = \beta_2 = 0$. The hazard rate for the normal distribution is approximately equal to $\psi/\sigma + w/\sigma^2$, see Table 2. Here, the variation in σ and ψ between region-year combinations is captured by combination-specific values β_{0rt} and β_{1rt} . The approximate linearity of the hazard rate in w_i implies that β_2 should be zero if the data were generated by the normal distribution. The hazard rate of the Gompertz distribution reads $\psi e^{w/\sigma}$. The convexity of this function is captured by $\beta_2 > 0$. If β_2 is positive, this is evidence in favor of the Gompertz distribution. The parameter β_2 can be interpreted as the second derivative of the function $\psi e^{w/\sigma}$ evaluated at $w = 0$, which is $\sigma^{-2}\psi$. Strictly speaking, also this parameter varies between region-year combinations, since σ and ψ do so. We choose not to let β_2 vary between region-year combinations in our specification for the sake of parsimony. Finding $\beta_2 > 0$ provides evidence in favor of the Gompertz distribution.

Table 6: Hazard Tests, log-normal vs. Weibull

Dependent Variable:	$H(w)$			
	(1)	(2)	(3)	(4)
<i>Variables</i>				
w	0.27 (0.15)		0.007 (0.37)	
w^2	0.42*** (0.05)	1.6*** (0.04)	0.47*** (0.11)	1.5*** (0.10)
<i>Fixed-effects</i>				
Region \times Year	✓	✓	✓	✓
<i>Varying Slopes</i>				
w (Region \times Year)		✓		✓
<i>Fit statistics</i>				
Observations	38,178	38,178	7,575	7,575
R^2	0.160	0.311	0.151	0.289
Within R^2	0.140	0.119	0.119	0.095
RMSE	3.41	3.09	3.51	3.21

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05

Notes: Driscoll-Kraay (L=2) standard-errors in parentheses.

Table 6 reports the results. Column (1) and (2) report the results for all 378 region-year combinations; column (3) and (4) report the results for the 75 combinations with more than 64 billionaires. Since the unit of observation is the empirical hazard $H_{rt}(w_i)$ for each billionaire, the number of observations is equal to the total number nonparametric interpolation points for each region-year combination. Column (1) and (3) report the results for a regression with a common β_1 for all region-year combinations. Columns (2) and (4) report the results when including separate coefficients for each combination for both β_{0rt} and β_{1rt} . Since the error terms are likely to be correlated across observations, we use Driscoll-Kraay standard errors. Across specifications, the quadratic term w^2 is highly statistically significant and positive, evidence of a convex increasing hazard, and it is orders of magnitude greater than the coefficient on the linear term, indicating that the degree of convexity in the data is strong. We conclude from these regressions that Gompertz is to be preferred above normal.

We have one further theoretical argument against our data on top log wealth being drawn from a truncated normal distribution. A theoretical justification for this hypothesis is that the distribution of top log wealth is a truncated version of the distribution of log wealth for the total population. If the latter distribution is normal, it is natural to assume that the distribution of top log wealth is truncated normal. The value of $\psi = 0.17$ corresponding to the observed empirical value for $R_2 \approx 0.80$ implies that $\Phi(-\psi) = 43\%$ of the population is a billionaire. The actual number of billionaires as a share of the population is one in a billion, corresponding to $\psi = 4.5$ and a value of $\mathcal{R}_2^N(\psi) = 0.96$, way above the observed empirical values. The theoretical justification for using a truncated normal distribution for top log wealth therefore fails. To obtain the lower values of R_2 of around 0.75 observed in the data, the value of ψ has to be even negative, implying that the density of the distribution of top log wealth is initially increasing and that the modus of distribution does not coincide with the lower bound.

4.4 Predicting expected wealth: Pareto versus Weibull

We have provided strong evidence against the common presumption that top wealth follows a Pareto distribution. Instead, we found that the Gompertz distribution provides a fair description of log wealth and therefore that the Weibull distribution is an adequate representation of the level of wealth. Our final test runs a horse race between Weibull and Pareto for the prediction of mean billionaire wealth in 2018. Fitting two parameters per region for the Gompertz distribution (σ and ψ), while fitting only a single parameter for Pareto (σ) would stack the deck against Pareto. We use therefore a common value $\psi_2^X(\bar{R}_2)$ for ψ for the average \bar{R}_2 across all regions for 2018. Next, we use this value for ψ to fit values of σ for Gompertz and Pareto for each region such that the expected log wealth is equal to its mean in the data for 2018.

Table 7 reports actual and estimated mean wealth among billionaires for Weibull and Pareto for all regions, together with the estimated values of σ for both distributions. Actual mean wealth varies widely across regions. Billionaires have a mean wealth of 2.1 billion in the region Israel + Turkey, whereas this is almost 5.3 billion in the US. 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 7 shows σ to be below one for seven

Table 7: Predicted vs. Realised Values of Mean Billionaire Wealth, 2018

Sub-Region	Mean Wealth			$\widehat{\sigma}$	
	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 3. For the Weibull prediction, we take a common $\psi = 1.84$, and we calculate σ using maximum likelihood. The Pareto prediction is made using \bar{w} as the maximum likelihood estimator of σ .

regions, including the United States. This immediately results in infinite values for mean wealth for these seven regions. We note a further five cases where σ is just slightly above one, 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 Weibull. We conclude that Pareto is not a useful model to predict mean wealth, whereas Weibull performs well.

5 Additional Applications

5.1 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 8.

Table 8: Statistics for the U.S. City Size Distribution

Sample	N	Lower bound	\bar{w}	$\overline{w^2}$	$\overline{w^3}$	R_2	R_3	$\ln R_2 - \lambda^N R_3$	$\ln R_2 - \lambda^G R_3$
Full	280	10.8	1.93	4.97	15.8	0.668	0.368	-0.0412	-0.028
Top 135	135	12.6	1.1	2.07	5.06	0.857	0.636	0.009	0.0002
Top 100	100	12.9	1.06	1.89	4.35	0.848	0.615	0.011	0.004
Top 50	50	13.8	0.851	1.26	2.42	0.871	0.654	0.015	0.006

Notes: Data are for the 2000 U.S. Census, 2001 vintage. N is the sample size; the lower bound is in logs.

The R_2 and R_3 statistics are nowhere close to unity, 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 R_k statistics are again far from unity. Remarkably, R_2 and R_3 hover around the same values found for billionaire wealth, namely $R_2 \approx 0.85$ and $R_3 \approx 0.65$.

The final two columns of Table 8 show the fit of the predicted linear relationship between $\ln R_2$ and $\ln R_3$. The gap is small regardless of cutoff point, but is clearly smallest for the top 135, the traditional sample chosen to study the upper tail. The Gompertz prediction is perfect, and the normal prediction is only off by a percentage point. Going further in the upper tail slightly worsens these fits, but the model continues to perform well.

Like with wealth, we distinguish between normal and Gompertz by computing hazard rates for each of the sub-samples reported in Table 8. We regress these hazards on a quadratic polynomial in w (here interpreted as log city size in excess of the chosen threshold). Table 9 reports the results. Across specifications, both w and w^2 are highly significant. This is again evidence against Pareto. Likewise, the quadratic term is highly significant, providing evidence favoring Gompertz over normal.

Table 9: Hazard Tests, City Size

Model:	Dependent Variable: $H(w)$			
	(1)	(2)	(3)	(4)
<i>Variables</i>				
Constant	0.85*	1.6***	1.7***	2.2***
	(0.34)	(0.40)	(0.40)	(0.48)
w	-0.65	-1.6*	-1.9*	-3.3**
	(0.43)	(0.71)	(0.79)	(1.2)
w^2	0.19*	0.55**	0.69**	1.4**
	(0.09)	(0.21)	(0.25)	(0.47)
<i>Fit statistics</i>				
Sample	Full	Top 135	Top 100	Top 50
R ²	0.381	0.351	0.356	0.363
Adjusted R ²	0.369	0.338	0.343	0.350
RMSE	1.31	1.61	1.67	1.95

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05

Notes: Heteroskedasticity-robust standard-errors in parentheses

Hence, using the same data as Eeckhout (2004), we corroborate his conclusion that Pareto provides a poor fit. However, we find no evidence of convergence of the hazard rate to a constant, as is required for the exponential distribution to hold for log city size (equivalently, for Pareto to hold for city size). Moreover, we contradict Eeckhout's preference for log-normal, based on the clearly convexly increasing hazard rates of log city size.

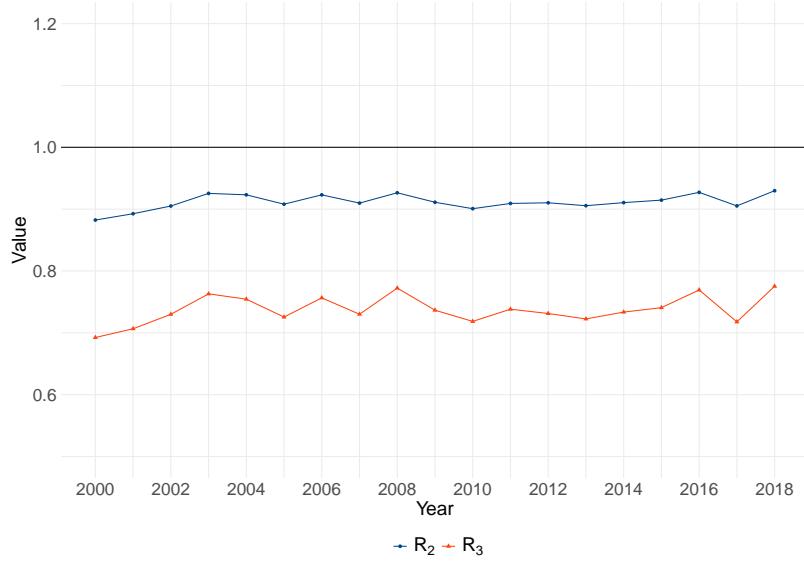
5.2 Firm Size

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 (2024). 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 2000–2018.⁶ We keep the top 1% of the distribution, resulting in 10,000 observations per year. The results are plotted in Figure 5.

Figure 5: Tests for Pareto, U.S. Firm Size Distribution



Notes: 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.

We observe that the value of R_2 and R_3 are constant over time. The values of R_2 , while not far from 1, are clearly different and hover around 0.9. The values for R_3 , moreover, are clearly lower than for R_2 and are about 0.75. We conclude that firm size, too, does not follow a Pareto distribution.

We proceed with our tests to distinguish normal from Gompertz. Table 10 shows the results of regression $\ln R_2 - \lambda^X \ln R_3$ on a constant. We allow for the standard errors to be serially correlated (Newey and West 1987).

The results are as before. The gaps are small in absolute magnitude, but significant. This is of course a consequence of the synthetic nature of our sample, which gives us arbitrary precision. Nevertheless, the quantitative patterns conform to our conclusions with wealth and city size. Like with city size, Gompertz appears to fit slightly better, since the gap is significant only at the 1% level and an order of magnitude smaller than the normal gap.

Our final test computes hazard rates for the firm size distribution in all years. We regress these on a quadratic in w , including year fixed effects to capture only within-year variation. Table 11 shows the results. Column (1) uses a

6. 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.

Table 10: $\ln R_2 - \lambda^X \ln R_3$, Firm Size

Dependent Variables:	$\ln R_2 - \lambda^N \ln R_3$	$\ln R_2 - \lambda^G \ln R_3$
Model:	(1)	(2)
<i>Variables</i>		
Constant	0.02*** (0.001)	0.002** (0.001)
<i>Fit statistics</i>		
Observations	19	19
RMSE	0.033	0.028

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05

Notes: Newey-West (L=0) standard-errors in parentheses

common coefficient on w , column (2) allows for the coefficient to vary across years.

Table 11: Hazard Tests, Firm Size

Dependent Variable:	$H(w)$	
Model:	(1)	(2)
<i>Variables</i>		
w	-0.15*** (0.03)	
w^2	0.03*** (0.004)	0.04*** (0.005)
<i>Fixed-effects</i>		
Year	✓	✓
<i>Varying Slopes</i>		
w (Year)		✓
<i>Fit statistics</i>		
Observations	1,919	1,919
R^2	0.142	0.178
Within R^2	0.139	0.055
RMSE	1.51	1.48

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05

Notes: Driscoll-Kraay (L=2) standard-errors in parentheses

The results are again strongly in favor of Gompertz. The quadratic term is highly significant. Like in Table 10, part of this precision is artificial due to the synthetic nature of the sample; however, the magnitude of the quadratic coefficient, which is on the same order as the linear term, shows its importance. This is a stark conclusion, since we are now examining the full top 1% of the firm size distribution, a much larger part of the upper tail than we examined for wealth. Yet, even for this larger sample, the convexity of the hazard rates are clear, reinforcing our conclusion that Gompertz is a better fit than normal.

6 Conclusion

We test the default assumption of Pareto for right-skewed distributions, using the *Forbes List of Billionaires*. We develop test statistics R_k based on normalizations of k th-order moments of the log-transformed data. We find strong evidence against the Pareto distribution. 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 strong linear fit of $\ln R_2$ on $\ln R_3$ and most importantly from the strongly convex hazard rates of log wealth, which allow us to discriminate between Weibull and log-normal. A comparable analysis of data on city- and firm-size yields the same conclusion: a clear rejection of Pareto and a strong preference for Weibull. We view these results as conclusive evidence against Pareto (or: exponential for its log), and strong evidence in favoring Weibull above log-normal (or: Gompertz above normal for its log). 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 R_k are easily calculated and our Weibull alternative yields easily testable predictions.

This conclusion comes with two caveats. First, *Forbes* rounds the data, resulting in ties between billionaires. This causes imprecisions in the estimated hazard rates. Second, hazard rate estimation requires rank-ordering the data. This is subject to the same criticisms as we levelled against the rank-size regressions. Hence, the preference for Weibull above log-normal is less firmly grounded than the rejection of Pareto.

The truncated-Weibull distribution differs 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 right tail. Pareto has no defined moments for $k \geq \sigma^{-1}$, which is the empirically relevant constraint even for the expectation ($k = 1$). In contrast, the moments of Weibull are always defined. These two differences make the Weibull distribution particularly suited for the study of upper tails.

The clarity of our results raises 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 implicitly reasoning from the stochastic convergence of diffusion processes to Pareto discussed in Section 2. This prior is at odds with the data in all cases we study in this paper. A second reason might be that the primary diagnostic to check Paretianity is the log rank regression. This tool has weak discriminatory power between Pareto and Weibull for the relevant range of ψ , introduces correlations in the error term due to ordering the data, and is sensitive to measurement error. Our test statistics 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. 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 is generated with probability p . 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 until one cannot proceed to a 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”: ulti-

mately, 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 in the number of available people and on the available land; their size is therefore naturally bounded by the total number of people and the available land mass. Likewise, firms 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. We explore these arguments further in a companion paper, Teulings and Toussaint (2024), where we show that the parameter σ depends on the total number of people in region. Wealth inequality in region depends therefore on its population size: the larger the population, the greater wealth inequality.

A final implication of our results concerns optimal taxation of top incomes, assuming that top income also follows a Weibull distribution. 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 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:⁷⁸

$$\tau^* = \frac{\sigma}{\sigma + \eta}, \quad (6)$$

where σ is the inverse tail index, and η the elasticity of earned income. The parameter σ appears in the formula because the optimal tax rate equates the cost of a marginal tax at W (which is $\eta \times W \times \Pr[\underline{W} = W]$) and the additional tax revenue on an even higher income $\underline{W} > W$ enabled by this a higher marginal rate (which is $\Pr[\underline{W} > W]$). For Pareto, this ratio is constant, so that the top tax rate should converge to a constant. For Weibull, the corresponding expression reads:

$$\tau^* = \frac{\frac{\sigma}{\psi} e^{-w/\sigma}}{\frac{\sigma}{\psi} e^{-w/\sigma} + \eta},$$

so that the optimal tax rate converges to zero. These findings mirror the conclusions in Mankiw, Weinzierl, and Yagan (2009), who argue that the log-normal distribution provides an equally good fit of the income data as Pareto. We have shown that in fact, log-normal provides a better fit than Pareto but Weibull an even better one. 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. Fully modelling the optimal taxation implications of our results is an important avenue for further research.

7. 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 are lower.

8. As is often done in the literature, we implicitly assign a marginal social weight of zero to the richest individual, which can be justified using utilitarian arguments.

References

- 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 Applications.” *Review of Income and Wealth* 68 (1): 263–288.
- Clauzel, Aaron, Cosma Rohilla Shalizi, and Mark E.J. Newman. 2009. “Power-Law Distributions in Empirical Data.” *SIAM Review* 51 (4): 661–703.
- 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.
- Deckhout, 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.
- . 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.
- Gaillard, Alexandre, Christian Hellwig, Philipp Wangner, and Nicolas Werquin. 2023. "Consumption, Wealth, and Income Inequality: A Tale of Tails." *Working Paper*.
- 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.
- 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.
- . 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.
- 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. 2024. “100 Years of Rising Corporate Concentration.” *American Economic Review* 114 (7): 2111–2140.
- Luttmer, Erzo G. J. 2011. “On the Mechanics of Firm Growth.” *Review of Economic Studies* 78 (3): 1042–1068.
- Mankiw, N Gregory, Matthew Weinzierl, and Danny Yagan. 2009. “Optimal taxation in theory and practice.” *Journal of Economic Perspectives* 23 (4): 147–174.
- 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.
- Newey, Whitney K., and Kenneth D. West. 1987. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica* 55 (3): 703–708.
- 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.
- 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. 2024. “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.
- 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.

A The asymptotic variances of R_k and $R_k R_m^{-\lambda}$

Since $R_k := M_k / (k! M_1^k)$, its asymptotic variance reads:

$$\begin{aligned}\text{plim}(N \text{Var}[R_k]) &= \text{plim}(NR_k^2 \text{Var}[\ln M_k - k \ln M_1]) \\ &= \text{plim}[NR_k^2 (\text{Var}[\ln M_k] + k^2 \text{Var}[\ln M_1] - 2k \text{Cov}[\ln M_k, \ln M_1])] \\ &= \text{plim}\left(R_k^2 \left[\frac{M_{2k}}{M_k^2} + k^2 \frac{M_2}{M_1^2} - 2k \frac{M_{k+1}}{M_k M_1} - (k-1)^2 \right]\right).\end{aligned}$$

The asymptotic variance of $\ln R_k - \lambda \ln R_m$ reads:

$$\begin{aligned}&= \text{plim}(N \text{Var}[\ln R_k - \lambda \ln R_m]) \\ &= \text{plim}(N \text{Var}[\ln M_k - \lambda \ln M_m - (k - \lambda m) \ln M_1]) \\ &= \text{plim}\left[N \left(\begin{array}{c} \text{Var}[\ln M_k] + \lambda^2 \text{Var}[\ln M_m] + (k - \lambda m)^2 \text{Var}[\ln M_1] - \\ 2\lambda \text{Cov}[\ln M_k, \ln M_m] - 2(k - \lambda m) \text{Cov}[\ln M_k, \ln M_1] + \\ 2\lambda(k - \lambda m) \text{Cov}[\ln M_m, \ln M_1] \end{array} \right)\right] \\ &= \text{plim}\left(\begin{array}{c} \frac{M_{2k}}{M_k^2} + \lambda^2 \frac{M_{2m}}{M_m^2} + (k - \lambda m)^2 \frac{M_2}{M_1^2} - \\ 2\lambda \frac{M_{k+m}}{M_k M_m} - 2(k - \lambda m) \frac{M_{k+1}}{M_k M_1} + 2\lambda(k - \lambda m) \frac{M_{m+1}}{M_m M_1} - \\ (k + \lambda - m\lambda - 1)^2 \end{array}\right).\end{aligned}$$

B The normal distribution

B.1 Derivation of $E[\underline{x}^k] = \mathcal{M}_k^N(1, \psi)$

Define the truncated-normal $\underline{x} := \underline{q} - \psi$, where \underline{q} follows a standard-normal distribution. We use complete induction:

$$\begin{aligned}\int_{-\psi}^{\infty} q^0 \phi(q) dq &= \int_{-\psi}^{\infty} \phi(q) dq = \Phi(-\psi), \\ \int_{-\psi}^{\infty} q \phi(q) dq &= [-\phi(q)]_{q=-\psi}^{\infty} = \phi(\psi), \\ \int_{-\psi}^{\infty} q^k \phi(q) dq &= [-q^{k-1} \phi(q)]_{q=-\psi}^{\infty} + (k-1) \int_{-\psi}^{\infty} q^{k-2} \phi(q) dq \\ &= \psi^{k-1} \phi(\psi) + (k-1) \int_{-\psi}^{\infty} q^{k-2} \phi(q) dq,\end{aligned}$$

using $\phi(-\psi) = \phi(\psi)$. Hence:

$$\begin{aligned}E[\underline{q}^k] &= \frac{\int_{-\psi}^{\infty} q^k \phi(q) dq}{\Phi(-\psi)} = \frac{\psi^{k-1} \phi(\psi) + (k-1) \int_{-\psi}^{\infty} q^{k-2} \phi(q) dq}{\Phi(-\psi)} \\ &= \psi^{k-1} \Lambda(\psi) + (k-1) E[\underline{q}^{k-2}],\end{aligned}$$

where $\Lambda(\psi)$ is the inverse Mills' ratio.

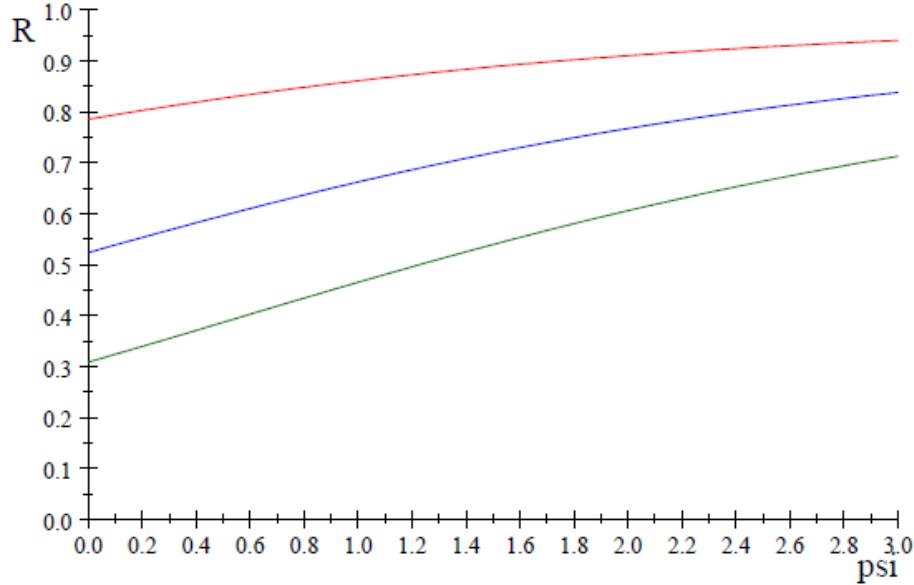
Using $\underline{x} := \underline{q} - \psi$, we obtain:

$$\begin{aligned}
E[\underline{x}^0] &= E[1] = 1, \\
E[\underline{x}] &= \Lambda(\psi) - \psi, \\
E[\underline{x}^k] &= E[(\underline{q} - \psi)^k] = E[(\underline{q} - \psi)(\underline{q} - \psi)^{k-1}] \\
&= E[q(\underline{q} - \psi)^{k-1}] - \psi E[\underline{x}^{k-1}] = \frac{\int_{\psi}^{\infty} q(q - \psi)^{k-1} \phi(q) dq}{\Phi(-\psi)} - \psi E[\underline{x}^{k-1}] \\
&= \frac{[-(q - \psi)^{k-1} \phi(q)]_{q=\psi}^{\infty} + (k-1) \int_{\psi}^{\infty} (q - \psi)^{k-2} \phi(q) dq}{\Phi(-\psi)} - \psi E[\underline{x}^{k-1}] \\
&= (k-1) E[\underline{x}^{k-2}] - \psi E[\underline{x}^{k-1}].
\end{aligned} \tag{7}$$

Hence:

$$\begin{aligned}
E[\underline{x}^2] &= E[\underline{x}^0] - \psi E[\underline{x}] = -\psi \Lambda(\psi) + 1 + \psi^2, \\
E[\underline{x}^3] &= 2E[\underline{x}] - \psi E[\underline{x}^2] = (\psi^2 + 2) \Lambda(\psi) - 3\psi - \psi^3, \\
E[\underline{x}^4] &= 3E[\underline{x}^2] - \psi E[\underline{x}^3] = -(\psi^3 + 5\psi) \Lambda(\psi) + 3 + 6\psi^2 + \psi^4.
\end{aligned}$$

Figure B.1: $\mathcal{R}_k^N(\psi)$ for $k = 2, 3, 4$ (red, blue, green) for the relevant range of ψ



B.2 Maximum likelihood estimator of ψ

The log likelihood for the normal distribution reads

$$\begin{aligned}\frac{\log L}{N} &= \overline{\ln \phi(w/\sigma + \psi)} - \ln \sigma - \ln \Phi(-\psi) \\ &= -\frac{1}{2} \overline{(w/\sigma + \psi)^2} - \ln \sigma + \ln \phi(0) - \ln \Phi(-\psi)\end{aligned}$$

The first order conditions for σ_{ml} and ψ_{ml} read

$$\begin{aligned}\frac{d \log L}{Nd\psi} &= -(\sigma_{\text{ml}}^{-1} \bar{w} + \psi_{\text{ml}}) + \Lambda(\psi_{\text{ml}}) = 0 \\ \frac{d \log L}{Nd\sigma} &= \sigma_{\text{ml}}^{-3} \bar{w^2} + \sigma_{\text{ml}}^{-2} \psi_{\text{ml}} \bar{w} - \sigma_{\text{ml}}^{-1} = 0\end{aligned}$$

These conditions can be written as:

$$\begin{aligned}\bar{w} &= \sigma_{\text{ml}} (\Lambda(\psi_{\text{ml}}) - \psi_{\text{ml}}) \\ \bar{w^2} &= \sigma_{\text{ml}}^2 + \sigma_{\text{ml}} \psi_{\text{ml}} \bar{w} \Rightarrow \\ \frac{\bar{w^2}}{2\bar{w}^2} &= R_2 = \frac{1 + \psi_{\text{ml}} \Lambda(\psi_{\text{ml}}) - \psi_{\text{ml}}^2}{2 [\Lambda(\psi_{\text{ml}}) - \psi_{\text{ml}}]^2}\end{aligned}$$

R_2 is therefore a sufficient statistic for the calculation of ψ_{ml} and $\psi_{\text{ml}} = \psi_2^N(R_2)$.

B.3 Linearization of the relationship between $\mathcal{R}_2^N(\psi)$ and $\mathcal{R}_3^N(\psi)$

Define:

$$\lambda^N := \lim_{\psi \rightarrow 0} \frac{\ln \mathcal{R}_2^N(\psi)}{\ln \mathcal{R}_3^N(\psi)} = \frac{\ln \left(\frac{1}{8} \phi(0)^{-2} \right)}{\ln \left(\frac{1}{12} \phi(0)^{-2} \right)} = \frac{\ln \left(\frac{\pi}{4} \right)}{\ln \left(\frac{\pi}{6} \right)} = 0.3733.$$

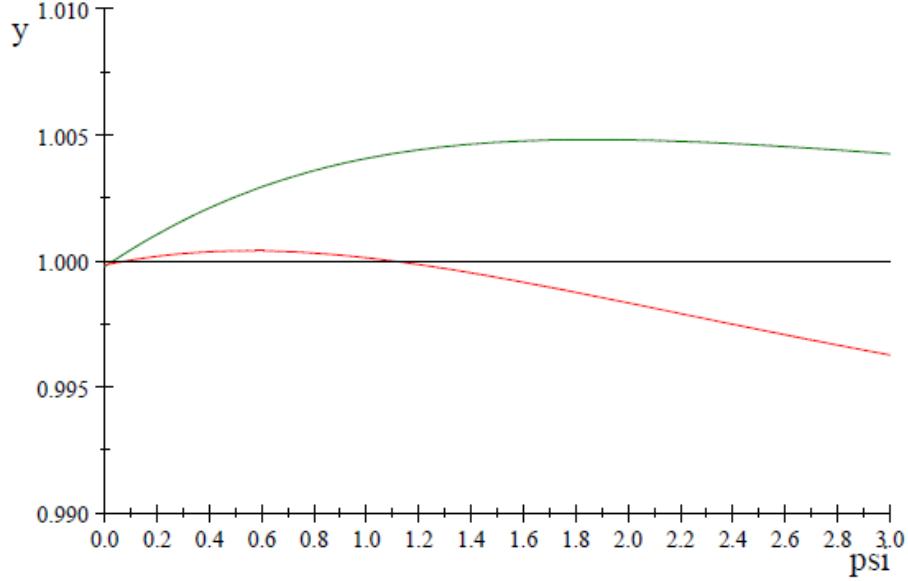
This definition implies that $\ln \mathcal{R}_2^N(\psi) - \lambda^N \ln \mathcal{R}_3^N(\psi)$ satisfies:

$$\lim_{\psi \rightarrow \infty} [\ln \mathcal{R}_2^N(\psi) - \lambda^N \ln \mathcal{R}_3^N(\psi)] = \lim_{\psi \rightarrow 0} [\ln \mathcal{R}_2^N(\psi) - \lambda^N \ln \mathcal{R}_3^N(\psi)] = 0.$$

Presuming that this relation also holds approximately for intermediate values of ψ , we obtain $\ln \mathcal{R}_2^N(\psi) \cong \lambda^N \ln \mathcal{R}_3^N(\psi)$. Figure 2 shows that $\ln \mathcal{R}_2^N(\psi) - \lambda^N \ln \mathcal{R}_3^N(\psi)$ is indeed close to zero for the relevant range of ψ (note the scale of the y-axis). Some finetuning (where we focus on the most relevant range $\psi \in (0, 3)$) yields an even more precise relationship, which is also shown in Figure B.2:

$$\ln \mathcal{R}_2^N(\psi) \cong 0.0105 + 0.390 \ln \mathcal{R}_3^N(\psi).$$

Figure B.2: $\ln \mathcal{R}_2^N(\psi) = 0.373 \ln \mathcal{R}_3^N(\psi)$ (green) and $\ln \mathcal{R}_2^N(\psi) = 0.011 + 0.390 \ln \mathcal{R}_3^N(\psi)$ (red)



C Gompertz distribution

C.1 Derivation of $E[\underline{x}^k] = \mathcal{M}_k^G(1, \psi)$

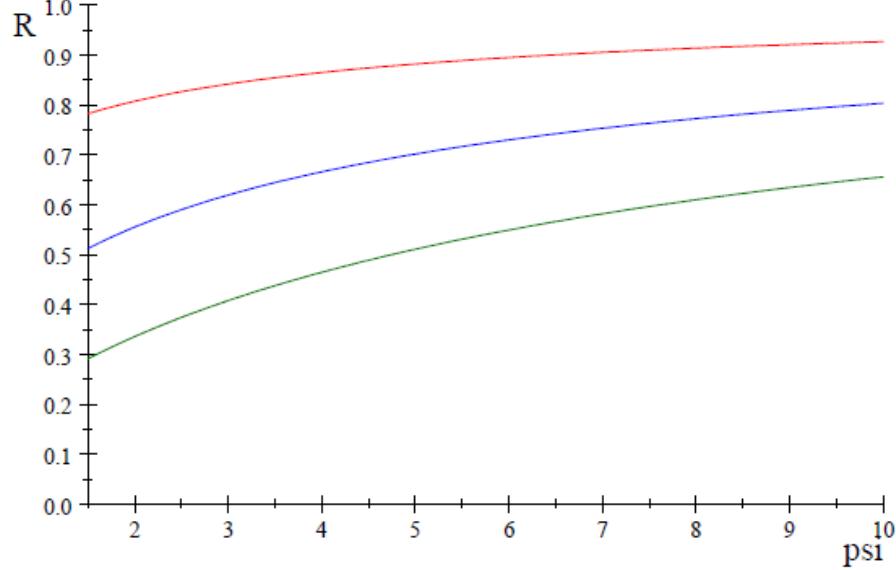
$$\begin{aligned}
E[\underline{x}^k] &= \int_0^\infty \psi x^k \exp(x - \psi(e^x - 1)) dx \\
&= \underbrace{[-e^\psi x^k \exp(-\psi e^x)]_{x=0}^\infty}_{=0} + \underbrace{k e^\psi \int_0^\infty x^{k-1} \exp(-\psi e^x) dx}_{=: h_k(\psi)} \\
h_k(\psi) &:= k e^\psi \int_0^\infty x^{k-1} \exp(-\psi e^x) dx, \\
h_1(\psi) &= e^\psi \int_0^\infty \exp(-\psi e^x) dx = e^\psi \text{Ei}(\psi).
\end{aligned}$$

C.2 Maximum likelihood estimator of ψ

The log likelihood function reads:

$$\frac{\log L}{N} = \bar{w}/\sigma - \psi \overline{e^{w/\sigma}} + \psi + \ln \psi - \ln \sigma.$$

Figure C.1: $\mathcal{R}_k^G(\psi)$ for $k = 2, 3, 4$ (red, blue, green) for the relevant range of ψ



The first order conditions for σ_{ml} and ψ_{ml} read:

$$\begin{aligned} \frac{d \log L}{N d \psi} &= -\overline{e^{w/\sigma_{\text{ml}}}} + 1 + \psi_{\text{ml}}^{-1} = 0, \\ \frac{d \log L}{N d \sigma} &= -\sigma_{\text{ml}}^{-2} \overline{w} + \sigma_{\text{ml}}^{-2} \psi_{\text{ml}} \overline{e^{w/\sigma_{\text{ml}}}} - \sigma_{\text{ml}}^{-1} = 0. \end{aligned}$$

Hence, using $x := w/\sigma$ we obtain:

$$\begin{aligned} \overline{e^x} &= 1 + \psi_{\text{ml}}^{-1}, \\ \psi_{\text{ml}} \overline{x e^x} &= 1 + \overline{x}. \end{aligned}$$

Hence, $\psi_{\text{ml}} \neq \psi_k^G(R_k)$ for any positive integer k .

C.3 Linearization of the relationship between $\mathcal{R}_2^G(\psi)$ and $\mathcal{R}_3^G(\psi)$

Define:

$$\lambda^G := \lim_{\psi \rightarrow 0} \frac{\ln \mathcal{R}_2^G(\psi)}{\ln \mathcal{R}_3^G(\psi)} = \ln(2!) / \ln(3!) = 0.3869.$$

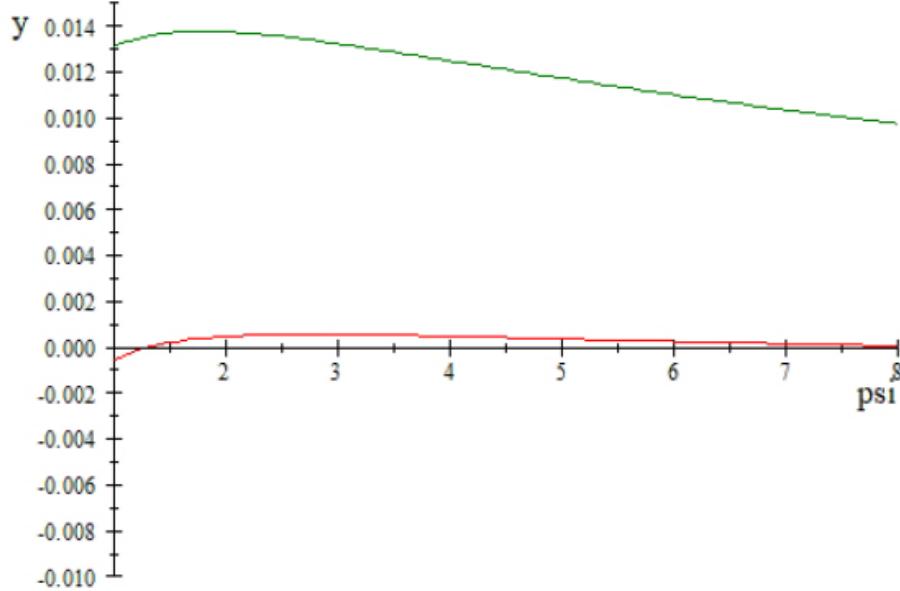
The definition implies that $\ln \mathcal{R}_2^G(\psi) - \lambda^G \ln \mathcal{R}_3^G(\psi)$ satisfies:

$$\lim_{\psi \rightarrow \infty} [\ln \mathcal{R}_2^G(\psi) - \lambda^G \ln \mathcal{R}_3^G(\psi)] = \lim_{\psi \rightarrow 0} [\ln \mathcal{R}_2^G(\psi) - \lambda^G \ln \mathcal{R}_3^G(\psi)] = 0.$$

Presuming that this relation also holds approximately for intermediate values of ψ , we obtain $\ln \mathcal{R}_2^G(\psi) \cong \lambda^G \ln \mathcal{R}_3^G(\psi)$. Figure 4 shows that $\ln \mathcal{R}_2^G(\psi) - \lambda^G \ln \mathcal{R}_3^G(\psi)$ is indeed close to zero for the relevant range of ψ . Some finetuning yields an even more precise relationship, which is also shown in Figure C.2:

$$\ln \mathcal{R}_2^G(\psi) = 0.027 + 0.404 \ln \mathcal{R}_3^G(\psi).$$

Figure C.2: $\ln \mathcal{R}_2^G(\psi) = 0.387 \ln \mathcal{R}_3^G(\psi)$ (dark green), $\ln \mathcal{R}_2^G(\psi) = 0.027 + 0.404 \ln \mathcal{R}_3^G(\psi)$ (red)



C.4 The variance of $\ln R_2 - \lambda \ln R_3$

For the Gompertz distribution, we obtain:

$$\begin{aligned} & \text{plim } (N \text{Var} [\ln R_2 - \lambda \ln R_3]) \\ &= e^{-\psi} \left(\begin{array}{l} \frac{\int_0^\infty x^3 \exp(-\psi e^x) dx}{(\int_0^\infty x \exp(-\psi e^x) dx)^2} + \lambda^2 \frac{2 \int_0^\infty x^5 \exp(-\psi e^x) dx}{3 (\int_0^\infty x^2 \exp(-\psi e^x) dx)^2} + (2 - 3\lambda)^2 \frac{2 \int_0^\infty x \exp(-\psi e^x) dx}{\text{Ei}(\psi)^2} - \\ 2\lambda \frac{5 \int_0^\infty x^4 \exp(-\psi e^x) dx}{6 \int_0^\infty x \exp(-\psi e^x) dx \int_0^\infty x^2 \exp(-\psi e^x) dx} - 2(2 - 3\lambda) \frac{3 \int_0^\infty x^2 \exp(-\psi e^x) dx}{2 \int_0^\infty x \exp(-\psi e^x) dx \text{Ei}(\psi)} + \\ 2\lambda(2 - 3\lambda) \frac{4 \int_0^\infty x^3 \exp(-\psi e^x) dx}{3 \int_0^\infty x^2 \exp(-\psi e^x) dx \text{Ei}(\psi)} - (1 - 2\lambda)^2 \end{array} \right). \end{aligned}$$

For the exponential distribution, we obtain:

$$\begin{aligned} & \text{plim } (N \text{Var} [\ln R_2 - \lambda \ln R_3]) \\ &= \frac{4!}{2!^2} + \lambda^2 \frac{6!}{3!^2} + (2 - 3\lambda)^2 2! - 2\lambda \frac{5!}{2!3!} - 2(2 - 3\lambda) \frac{3!}{2!} + 2\lambda(2 - 3\lambda) \frac{4!}{3!} - (1 - 2\lambda)^2 \\ &= 10\lambda^2 - 6\lambda + 1 \end{aligned}$$

Figure C.3 presents the variance and the standard deviation of $\ln R_2 - \lambda \ln R_3$ for the exponential and Gompertz distribution, for $\lambda = 0.404$. In order to document the slow convergence to its limiting for the exponential distribution we plot these function for a longer range, $\psi \in [1, 50]$.

Figure C.3: Variance (solid) and standard deviation (dashed) of $\ln R_2 - \lambda \ln R_3$ for exponential (green) and Gompertz (red) as a function of ψ

