Schulz, Jan; Milaković, Mishael

Article — Published Version

How Wealthy are the Rich?

Review of Income and Wealth

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Schulz, Jan; Milaković, Mishael (2021) : How Wealthy are the Rich?, Review of Income and Wealth, ISSN 1475-4991, Wiley, Hoboken, NJ, Vol. 69, Iss. 1, pp. 100-123, https://doi.org/10.1111/roiw.12550

the review of income and wealth

Review of Income and Wealth
Series 69, Number 1, March 2023
DOI: 10.1111/roiw.12550

HOW WEALTHY ARE THE RICH?

BY JAN SCHULZ* and MISHAEL MILAKOVIĆ

University of Bamberg

Underreporting and undersampling biases in top tail wealth, although widely acknowledged, have not been statistically quantified so far, essentially because they are not readily observable. Here we exploit the functional form of power law-like regimes in top tail wealth to derive analytical expressions for these biases, and use German microdata from a popular survey and rich list to illustrate that tiny differences in non-response rates lead to tail wealth estimates that differ by an order of magnitude, in our case ranging from 1 to 9 trillion euros. Underreporting seriously compounds the problem, and we find that the estimation of totals in scale-free systems oftentimes tends to be spurious. Our findings also suggest that recent debates on the existence of scale- or type-dependence in returns to wealth are ill-posed because the available data cannot discriminate between scale- or type-dependence, on one hand, and statistical biases, on the other hand. Yet both economic theory and mathematical formalism indicate that sampling and reporting biases are more plausible explanations for the observed data than scale- or type-dependence.

JEL Codes: C46, C81, D31

Keywords: wealth inequality, stochastic growth, differential non-response, Hill estimator, tail index bias

1. INTRODUCTION

The starting point for this analysis was a conscientious effort to quantify the total wealth of the richest Germans from survey microdata. This seemingly innocuous exercise pointed us to a problem that, to the best of our knowledge, has not yet been adequately addressed in the pertinent literature. The problem arises in the top tail of wealth, generally following power law-like distributions, where survey data apparently suggest total wealth to be orders of magnitude smaller than implied by named rankings of the super-rich, often referred to as rich lists. Extrapolating the power law *backward* from observed top wealth levels to some unobserved minimum is asymptotically unbiased. Severe biases can arise, however, when extrapolating *forward* from relatively low levels to unreliable or missing maximum wealth levels (Cristelli *et al.*, 2012). In survey data, the latter typically leads to strongly downward biased estimates of wealth and

Note: We thank Simone Alfarano, Anna Gebhard, Philipp Mundt, our colleagues at the Bamberg Research Group on Behavioral Macroeconomics, as well as the participants of the 1st Behavioral Macroeconomics Workshop in Bamberg, the 12th International Conference on Computational and Financial Econometrics in Pisa, the 23rd Forum for Macroeconomics and Macroeconomic Policies, the 9th Young Economists Conference, and the 77th Annual Congress of the International Institute of Public Finance for helpful comments and suggestions at crucial junctions of this investigation. Financial support by the Hans-Böckler-Foundation through grant PK045 is gratefully acknowledged.

*Correspondence to: Jan Schulz, Department of Economics, University of Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany (jan.schulz@uni-bamberg.de).

inequality (Eckerstorfer *et al*., 2016; Vermeulen, 2018). As we cannot quantify this effect without data that go beyond the available, we propose two limit interpretations to gauge the potential impact of this bias. In what we term the *data first* limit, we assume both upper and lower truncated samples to deliver unbiased estimates. Put differently, we attribute all observed differences between upper and lower truncated samples to truly existing differences in the data-generating process. In the complementary *theory first* limit, we assume the data-generating process to be homogeneous across samples on the different scales, attributing the entire observed difference to statistical bias. We show that tail wealth estimates differ by an order of magnitude, depending on which of the two pre-analytical visions we use.

The literature so far has implicitly taken a *data first* stance on this issue (Eckerstorfer *et al*., 2016; Vermeulen, 2016, 2018; Bach *et al*., 2019). Our primary goal with this paper is to argue that a *theory first* perspective is at the very least equally plausible. To show this, we introduce different categories of biases that affect measured inequality and provide closed-form expressions that are readily estimated. First, we show that *underreporting* incentives by themselves are insufficient to generate biased estimates, as the estimate is asymptotically unbiased if the entire population unanimously underreports its wealth. Inequality is underestimated only if underreporting is more pronounced for the richest, which seems intuitively plausible as the super-rich have mightier means at their disposal to avoid taxes than the average person or household. Second, we show that differential underreporting by the super-rich indeed leads to downward biased estimates of inequality for the entire population. Finally, and most importantly, the impact of underreporting rates is highly nonlinear. Even if only a fraction of actual wealth is reported, this will greatly reduce the resulting bias compared to when information on a fraction of the super-rich is missing altogether. We call the latter case *undersampling*, which is typical of survey data that essentially use equiprobable sampling and therefore do not adequately capture the richest individuals in power law-like regimes. We also show that *logarithmic sampling* would greatly improve the statistical quality of wealth surveys. The named rich lists, on the contrary, will be subject to reporting biases as they explicitly try to account for the super-rich but typically suffer from data availability and salience issues, as well as adverse (tax) reporting incentives. Without additional information, both the estimated underreporting and undersampling rates remain within plausible bounds; therefore, the polar *data first* and *theory first* perspectives would appear equally plausible at first.

While it is hardly surprising that the two perspectives imply different estimates for top tail wealth, the difference turns out to be enormous. The lowest estimate arising from *data first* is around 1 trillion euros for Germany's top tail wealth, whereas the *theory first* estimates reach about 9 trillion euros. These vast differences, spanning almost one order of magnitude in top tail wealth, are caused by tiny non-response rates on the order of a tenth of a percent. This disconcerting result suggests that aggregate findings within the *data first* framework can become heavily distorted by tiny degrees of undersampling. The severity of the problem extends far beyond the German data set because our results are functions of the power law tail of wealth distributions that applies across many countries and time

periods.[1] Consequently, estimates of total wealth will crucially depend on the pre-analytical perspective and should thus be treated with extreme caution. If total wealth estimates are to be stated, we believe that scientific integrity at least demands to report the range from the smallest estimates of a *data first* perspective to the largest estimates of a *theory first* perspective, especially if these estimates are intended to inform economic policy or public debate.

The ubiquity of power laws has led to numerous suggestions for potential generating mechanisms, reviewed for instance by Gabaix (2009) or Luttmer (2010). In the case of top tail wealth, any candidate mechanism should be based on a property that is common across the various time periods, countries, or proxies of wealth. One common property, at least across the different varieties of capitalism, concerns the primary types of assets in super-rich portfolios, namely entrepreneurial stakes, financial assets, and speculative (that is non-owner occupied) real-estate, which are perpetually reinvested into or reallocated among these asset classes (Davies and Shorrocks, 2000; Wachter and Yogo, 2010).[2] Thus a random growth model featuring a multiplicative component seems to be the most adequate candidate for a sensible generating mechanism. The idea to explain the emergence of power law tails with stochastic multiplicative processes has a long history but has fallen out of fashion in economics for many decades, essentially for its lack of microfoundations. Yet random multiplicative growth has recently regained traction within economically motivated partial and general equilibrium models that endogenously generate power law tails in wealth from stochastic capital or asset accumulation (Levy, 2003; Levy and Levy, 2003; Nirei and Souma, 2007; Nirei, 2009; Benhabib *et al*., 2011; Toda, 2014; Piketty and Zucman, 2015; Hubmer *et al*., 2016; Aoki and Nirei, 2016; Benhabib and Bisin, 2018).

The literature on random multiplicative growth has typically placed weak restrictions on the particular form of return distributions governing the stochastic process. One notable exception, however, is the assumption of an equilibrating tendency for the expected (risk-adjusted) rate of return or, in more technical terms, of a homogeneous return distribution across wealth portfolios. This is consistent with the classical notion of competition, the implications of (semi-strong)

[1]Table 1 in Online Appendix G summarizes the empirical consensus on this distributional structure for several countries and time periods. Judging from the evidence collected there, the power law property of empirical wealth distributions appears to be robust across time, for instance showing up in medieval Hungary (Hegyi *et al*., 2007) or ancient Egypt (Abul-Magd, 2002), and for different proxies of wealth. Power law distributions in tail wealth also appear to be spatially ubiquitous, showing up across the Western world, for instance in Austria, Canada, Germany, Sweden, the UK, and the US (Levy and Solomon, 1997; Levy, 1998, 2003; Castaldi and Milaković, 2007; Drăgulescu and Yakovenko, 2001; Cowell, 2011; Bach *et al*., 2011; Brzezinski, 2014; Coelho *et al*., 2005; Eckerstorfer *et al*., 2016), yet also hold for less developed countries like China, Russia, or India (Sinha, 2006; Ning and You-Gui, 2007; Brzezinski, 2014), again across varying time horizons.

[2]From an accounting standpoint, this perpetual reallocation and investment is closely related to saving, and there is a consensus in the literature that propensities to save are strongly positively correlated with (lifetime) income or wealth (Dynan *et al*., 2004; Jappelli and Pistaferri, 2014). This also holds for entrepreneurial households (Quadrini, 1999). As a major reason for this relationship, Deaton (2003) identifies credit constraints that are only binding for low wealth households and individuals. Alan *et al*. (2015), on the contrary, provide a critical discussion of the identification strategy and find no differential savings behavior with respect to long-term income. The major limitation of Alan *et al*. (2015) is the exclusion of the wealthiest 1 percent that we are primarily concerned with here.

informationally efficient capital markets, and the idea that investors' superior talent in either fundamental or technical analysis cannot lead to excess returns over extended periods of time (Fama, 1965, 1970, 1991). Indeed, as Levy (2003) and Levy and Levy (2003) show both experimentally and via Monte Carlo simulations, the scope for differential talent is very limited in light of power law distributed top wealth. If one group of investors were to consistently outperform another group of less talented investors in terms of their expected returns by only a tiny margin, the functional form of the emergent stationary distribution would differ significantly from a power law and exhibit concavity on double-logarithmic scale.[3] Therefore, the defining characteristic of the *theory first* perspective is to assume a homogeneous return distribution, thereby implying equivalent data-generating processes across samples.

A more recent strand of literature has started to challenge the homogeneity hypothesis on both theoretical and empirical grounds. Bach *et al*. (2017) and Fagereng *et al*. (2020) find excess risk-adjusted returns for the wealthiest portfolios, the latter even claiming persistence in abnormal returns, indicating persistent heterogeneity in financial information and talent *if we take the data at face value*. From a more theoretical perspective, Luttmer (2011) and Gabaix *et al*. (2016) build on the well-known limitation of random growth models to typically generate very slow transitions. The former puts this in terms of the stationary distribution of assets, with a half-life of assets that would be way too high from an empirical point of view, whereas the latter argue (formally equivalently) that the rate of convergence to the new stationary distribution after a shock to the variance in the permanent component of earnings is too slow to account for the observed rapid rise in top-level income inequality. Gabaix *et al*. (2016) and Jones and Kim (2018) thus put forward the hypothesis of heterogeneous returns to explain the observed rise in income and wealth inequality, whereby excess returns are either correlated with wealth levels ("scale-dependence") or result from differential talent ("type-dependence"). In informationally efficient capital markets, scale-dependence can occur only when the set of investment opportunities increases in wealth. Hedge funds and some private banks perhaps provide anecdotal evidence, as hedge funds typically require high minimum investment inlays (King and Maier, 2010), whereas some private banks like JP Morgan Chase require their private clients to hold at least 10 million dollars (Glazer, 2016). Concerning type-dependence, Gabaix *et al*. (2016) circumvent the formal problem that differential talent is inconsistent with a Pareto distribution by essentially assuming that "high growth types" only stay in the high growth regime for a limited amount of time and cannot return there. This idea not only lacks theoretical appeal but also introduces another degree of freedom into any empirical investigation that now has to justify after how many periods of abnormally high returns one can safely claim type-dependence.

Moreover, given that our data lack information on investors' sophistication, this notion of type-dependence is phenomenologically equivalent to scale-dependence as we cannot control for investors' ability. Put differently, we cannot

[3]They consider two Gaussian return distributions that merely differ in expected value, showing that already a difference by 1 percentage point in expected returns leads to a stationary distribution that significantly differs from the Pareto type.

distinguish between the hypothesis that individuals are rich because of their excess returns, and the alternative hypothesis that they have excess returns because they are rich. We will thus only focus on testing for scale-dependence. This hypothesis corresponds to the *data first* interpretation, as observed differences between high and low scale samples are then assumed to reflect true differences in the data-generating process, that is to say scale-dependent random growth. We will argue, however, that the idea of scale-dependent growth is not only problematic from a formal point of view, but also violates economic intuitions like informational efficiency or the classical concept of competition that predicts a tendency for the equalization of returns. *Theory first* leaves these economic intuitions intact by attributing observed deviations in the data to statistical biases arising from under-sampling and underreporting, and also casts a different light on the apparently reversed risk-return trade-off that we observe in the data.

The remainder of this paper is organized as follows: Section 2 derives the biases in estimates of the tail exponent that arise from underreporting and under-sampling, respectively. Throughout the paper we have relegated all derivations to the appendix to emphasize important conceptual differences over technical detail. Section 3 introduces the data and discusses our estimation procedure. Our results are presented in Section 4, where we put forward two mutually exclusive yet on their own reasonably plausible explanations for the observed behavior in the data. Section 5 discusses the implications of our results for existing work on top tail wealth and concludes with the suggestion to improve future surveys through logarithmic sampling.

2. Model

The *data first* and *theory first* interpretations are purposefully designed to be antithetical, although we will show that the formal explication of both interpretations can be reduced to conceptually closely related mechanisms that affect measured tail inequality at different stages of empirical estimation. For both interpretations, Zipf (1949) law with a tail exponent of unity is an attractor for a parsimonious stochastic multiplicative process that does not exhibit scale-dependence in accumulation or reporting. Consequently, an observed tail exponent that differs from unity implies scale-dependent behavior in both frameworks. *Data first* attributes this to scale-dependent stochastic growth at the level of accumulation, whereas *theory first* assumes that it is fully caused by scale-dependent reporting behavior at the level of measurement. Although the mechanisms are formally quite similar, the pre-analytical vision obviously differs substantially between narratives.

Within the *data first* framework, the measured tail inequality is a sufficient statistic for both the true (snapshot) inequality among the richest and scale-dependence within the wealth accumulation process over time, because tail inequality is intricately linked to the nature of the underlying stochastic process of multiplicative growth. In Appendix A, we consider a standard drift-diffusion process to show that the tail index of the stationary power law distribution is uniquely determined by the expected return and variance of the stochastic growth process. As Gabaix (1999) shows, and we rederive in greater detail in Appendix A, the

stationary distribution of the right tail for this type of general process is a power law, with tail index α given by

$$\alpha(w;\gamma(\cdot),\sigma(\cdot)) = 1 - 2\frac{\gamma(w)-\overline{\gamma}}{\sigma^2(w)} + \frac{w}{\sigma^2(w)}\frac{\partial\sigma^2(w)}{\partial w} \text{ with } w \in \mathbb{R}^+, \tag{1}$$

where $\overline{\gamma}$ is the average wealth growth rate and $\gamma(w)$ the (normalized) mean growth rate for a given wealth level w. Expression (1) has intuitive comparative statics with respect to the degree of scale-dependence in both mean growth rates γ and variance σ^2. Whenever the expected (excess) mean growth rate $\gamma(w) - \overline{\gamma}$ increases in wealth, the tail exponent decreases and stationary inequality rises. Thus positive scale-dependence in expected returns increases measured inequality. When variance exhibits positive scale-dependence, $\partial\sigma^2(w)/\partial w > 0$, tail exponents increase and system-wide inequality therefore decreases. Zipf's law with $\alpha = 1$ is an interesting limit case for a situation without any scale-dependence (positive or negative), that is, $\gamma(w) = \overline{\gamma}$, $\forall w \in \mathbb{R}^+$, and $\partial\sigma^2(w)/\partial w = 0$. These two conditions are typically called Gibrat's law after the seminal study by Gibrat (1931). Therefore, Gibrat's law in growth rates is a sufficient condition for Zipf's law to hold in wealth levels. Córdoba (2008a, 2008b) proves that it is also a necessary condition. The *data first* interpretation takes the Zipf benchmark in the stationary distribution as an indication for scale-independence, while statistically significant deviations are evidence to the contrary.

The *data first* approach thus implicitly assumes that the measured tail inequality $\widehat{\alpha}$ is equivalent to the true stationary α or, at least, that the estimate is not systematically biased in any direction. The polar *theory first* interpretation assumes no systematic scale-dependence of either type, that is, $\alpha = 1$, and attributes significant deviations from this Zipf benchmark to underreporting and undersampling biases. While the relevance of distorted or missing observations has already been argued on empirical grounds and in Monte Carlo simulations (Vermeulen, 2016, 2018), we derive closed-form expressions here that quantify the resulting bias in the tail exponent when the number of observations, denoted N, becomes large.[4] In addition, we will differentiate between unanimous and differential reporting behavior, on one hand, and underreporting versus undersampling, on the other hand. As undersampling or underreporting rates are impossible to estimate by the very nature of the problem, we consider three stylized scenarios that are analytically tractable: (i) unanimous (proportional) underreporting, (ii) differential (proportional) underreporting, and (iii) undersampling.

First, we consider the case of *unanimous underreporting*, that is, all respondents only report a fraction ρ of their wealth. Call this fraction the *reporting rate*. We show in Appendix B that this leads to an unbiased estimator of the tail index; therefore, unanimous underreporting does not pose problems for the estimation of inequality. This holds symmetrically for unanimous overreporting, $\rho > 1$, also

[4]Like the assumption of $t \to \infty$ for the drift-diffusion process, the assumption of $N \to \infty$ is necessary to make the problem analytically tractable. Our qualitative results are not materially sensitive to this assumption, as we verified by Monte Carlo simulations that the bias did not significantly differ from the large N limit in finite samples.

showing that the estimator is invariant with respect to inflation. Whenever there are differential reporting rates, however, the bias is unambiguously positive and thus underestimates inequality. We call this case *differential underreporting*. For this, consider the case where the upper q-quantile of the wealth distribution only reports a fraction ρ of their wealth, from whence we show that for large N the estimated tail index, now denoted $\widehat{\alpha}^{du}$, will differ from the Zipf benchmark such that

$$\widehat{\alpha}^{du}(q,\rho) = \frac{1}{1+q\cdot ln(\rho)}, \tag{2}$$

with q and $\rho \in (0,1)$, and the additional restriction that $\rho > q$. The latter restriction is needed to preserve the minimum of the true power law distribution on which the maximum likelihood estimator (MLE) is anchored. It is easily verified that $\widehat{\alpha}^{du}$ is always upward biased compared to Zipf's law for these parameter restrictions, implying that true inequality is underestimated. The effect of varying the parameters is also quite intuitive: an increase in q for a given ρ and a decrease in ρ for a given q increase the bias, as in both cases relatively less wealth is reported for the richest.[5] Furthermore, α is only unity when either $q = 0$ or $\rho = 1$, so there is no differential behavior to begin with. Thus, when it comes to underreporting, the differential behavior of the very richest compared to the relatively less wealthy is necessary to cause upward biases from the *theory first* perspective.

While we cannot derive analytical expressions for $\rho < q$ in general, this is possible for the limit case of $\rho = 0$. In our stylized scenario, this would correspond to a case where the upper q-quantile is non-respondent and the wealth distribution is therefore q-truncated. In this case of *differential undersampling* or *non-response*, the richest quantile is not included at all in the sample, corresponding to a reporting rate of zero. This scenario actually appears to be empirically relevant, and connected to sampling and social desirability biases (Kennickell and Woodburn, 1999; Eckerstorfer *et al*., 2016; Vermeulen, 2016, 2018). As we show in Appendix B, non-response leads asymptotically to a (strong) upward bias in the MLE of the tail exponent, now denoted by

$$\widehat{\alpha}^{nr}(q) = \frac{1-q}{1-q+q\cdot ln\ (q)}, \tag{3}$$

for large N and $q \in (0,1)$. For this parameter range of q, $\widehat{\alpha}^{nr}$ is always upward biased compared to the Zipf benchmark, and monotonically increasing in the quantile q of non-respondents. The quantile q of upper non- or underreporting individuals is thus the only formal difference between the competing narratives of *data first* and *theory first*.[6] If $\widehat{\alpha} \neq 1$, *data first* implicitly assumes $q = 0$ and therefore attributes all the observed deviation from the Zipf benchmark to scale-dependence in either mean or expected returns. In contrast, *theory first* takes $\widehat{\alpha} \neq 1$ to imply $q \neq 0$ and therefore differential reporting behavior according to sample inclusion rates and the level of wealth.

[5]By the same token, for $\rho > 1$, an increase in q increases the downward bias of the estimate and thus overestimates inequality relative to the true (Zipfian) distribution.

[6]We would like to believe that it is not entirely trivial to reduce the impact of the two pre-analytical visions to a single parameter.

3. Data and Estimation

To test the hypothesis of scale-dependence, we examine two samples covering distinct scales in the upper tail of the German wealth distribution. We need two non-overlapping samples that both exhibit power law-like top tails, as is often the case for surveys and rich lists. The German data described below comfortably meet this requirement as the minimum wealth level in the rich lists is about thrice as large as the maximum wealth level in the surveys. Non-overlapping samples are necessary to isolate potential scale effects in the accumulation of wealth, and to ensure that we consider two distinct sets of wealth portfolios. The latter condition minimizes potential Type II error in hypothesis testing, because failure to reject the null hypothesis of insignificant scale differences could otherwise arise from the simultaneous presence of identical wealth portfolios, thereby affecting the estimated parameters in both sample types.

3.1. *Data*

The *Socio-Economic Panel* (SOEP), compiled by *Deutsches Institut für Wirtschaftsforschung*, is probably the most prominent source for microdata on German households and individuals. The 2002, 2007, and 2012 waves of the panel include items on personal wealth that we use in our analysis. Assuming different weighing and imputation techniques for the market value and disaggregation to individual values, the *SOEP* sample claims to be representative of the entire German population, implying that each person or household in Germany is chosen with equal probability (Frick *et al.*, 2007). With a total population of 82.5 million in Germany and about 25,000 individuals in the sample, the sampling ratio thus corresponds to about 0.035 percent (Statistisches Bundesamt, 2017).

While the *SOEP* sample probably provides a reasonable approximation to the distribution of wealth for the majority of Germans, it is well known that wealth data from household surveys become increasingly inaccurate for the tails of the distribution (see, e.g., Davies and Shorrocks, 2000). Casual empiricism indeed suggests that the reported maximum wealth level in the *SOEP* of around 70 million euros is far from being "representative" of the richest Germans, whose fortunes are about three orders of magnitude larger according to the rich lists compiled by *manager magazin*. These named lists rank the 500 richest Germans according to their net wealth in the years 2010–2016. Because the rich lists are not curated for statistical inference, the data likely suffer from numerous issues regarding their consistency both in the time-series and cross-sectional domain. We discuss both data sets and their respective limitations in Online Appendix H.

3.2. *Estimation*

Our empirical analysis starts with the parameter estimates of the power law distributions in the upper tail of the *SOEP* and *manager magazin* samples. We interpret these as the stationary distributions resulting from a general random growth process, as described in Appendix A. The assumption that the empirically observed state coincides with the stationary state of the distribution for time $t \to \infty$ is frequently challenged though. Especially Gabaix *et al.* (2016) and Luttmer

(2011, 2018) show that the convergence to a new stationary distribution from a shock resulting in deviations from the steady-state is extremely slow. A back-of-the-envelope calculation in Luttmer (2018) suggests that for a firm size distribution close to Zipf, but with slightly thinner tails, a shock to the aggregate capital stock would be extremely persistent with a half-life of around 70 years, implying unrealistically low rates of recovery.[7] Given slow convergence, it is questionable whether the empirical distribution truly reflects the dynamics of an underlying random growth process or whether it is merely in a transient state to stationarity. On the contrary, as Levy and Levy (2003) show, the convergence to the *approximate* power law is much faster even though convergence to the *asymptotic* distribution is indeed very slow for these types of random growth processes. Levy and Levy (2003) understand approximate convergence as convergence to a distribution that cannot be statistically distinguished from the stationary state using Kolmogorov–Smirnov (KS) test. If the parameter estimates are at least approximating the true stationary state of the random growth process, the pronounced differences we find between samples will not be mere artifacts of one distribution being in a transient state but not the other, thus reflecting genuine differences in reporting, sampling, or the underlying growth process.

Given the diffusion in Appendix A, we reject the null hypothesis of scale-independence for $\hat{\alpha}$ significantly different from unity. This procedure is advantageous in the sense that it relies on observables to test for scale-dependence and thus allows inferences about the (at least partially) unobservable random growth process. In addition, we also consider the distribution of growth rates in wealth to judge whether scale-(in)dependence characterizes the wealth accumulation process. Notice that the diffusion in Appendix A requires us to consider scale-dependence in both expected value and risk, which are readily measured by the MLEs of the location and dispersion parameters of the growth rate distributions.

We estimate the tail exponent of the power law using maximum likelihood.[8] For the estimation of the minimum wealth level $w_{\min}$ in the *SOEP* sample, we use the standard suggested by Clauset *et al.* (2009), yielding $\hat{w}_{\min} = 280,000$ euros for the 2002 sample, $\hat{w}_{\min} = 200,000$ euro for the 2007 sample, and $\hat{w}_{\min} = 180,000$ euro for the 2012 sample.[9] It seems reasonable to assume that a net worth of around 200,000 euros already gives rise to primarily multiplicative returns, especially considering that most households hold their wealth in the form of owner-occupied housing. As the rich lists should be characterized by power laws, we do not estimate $w_{\min}$ but rather take it directly from the data; therefore, $w_{\min}$ simply corresponds to

[7]For a distribution that is exactly Zipf, there would be no recovery at all.

[8]Clauset *et al.* (2009) show that an MLE fit is the least biased method to estimate the tail index (or *characteristic exponent*) of a power law, compared to OLS methods or a linear fit on double-logarithmic scale. See Goldstein *et al.* (2004) for a more rigorous analysis of different graphical methods and their respective shortcomings. Even though we estimate from a discrete data set, we estimate the power law for its continuous analogue, as the analytical results for different biases in the results section are based on the continuous version. The continuous MLE is introduced and discussed in Appendix B in more detail. Our estimates are not materially sensitive to the choice between the discrete and continuous estimator. Standard errors are determined by exploiting the asymptotic normality of the MLE (De Haan and Resnick, 1997).

[9]Clauset *et al.* (2009) show that their method, based on iteratively increasing the reverse order statistic until a goodness-of-fit test (like the KS test) rejects a power law distribution, outperforms other possible procedures, such as minimizing the Bayesian information criterion (BIC).

the minimal wealth level in each rich list, ranging from 200 to 250 million euros in the different years.[10] The minimum in the rich lists is thus three orders of magnitude larger than in the surveys.

Finally, we would expect the distribution of wealth growth rates to be Laplacian (or double-exponential) because we measure wealth growth by the logarithmic difference in wealth levels, that is, $r_{i,t} = log(w_{i,t}) - log(w_{i,t-1})$ for agent i during the period t to $t-1$. It can be shown that $log(w)$ follows an exponential distribution if w follows a power law, and that the difference between two exponentially distributed variables is Laplacian (Kotz *et al.*, 2001). The symmetric Laplace distribution for returns r then has a probability density function (PDF) that is given by

$$f(r;m,\sigma) = \frac{1}{2\sigma} e^{-\left|\frac{r-m}{\sigma}\right|}, \tag{4}$$

where $m \in \mathbb{R}$ and $\sigma > 0$ are location and dispersion parameters, respectively.[11] From a conventional point of view, m measures the expected return in a set of wealth portfolios, whereas σ measures the associated risk in these portfolios.

Our estimation strategy considers the cross-sectional distributions of wealth in both samples, each interpreted as the outcome of a parsimonious random growth process like the one described in Appendix A, whose realizations are at least partially unobservable. The estimated tail index then allows us to infer scale-(in)dependence within this unobservable process. Moreover, using the Laplace estimates from the actual growth rate distributions, we can test parametrically for scale-(in)dependence in expected returns or risk, and we also use several nonparametric tests.

4. Results

Our parametric estimation strategy is based on the two distributional regularities in the upper tail of cross-sectional wealth portfolios, namely the power law distribution in wealth levels and the Laplace distribution of portfolio returns, because the respective empirical densities are reasonably in line with the theoretically expected functional forms. The observed complementary cumulative distribution functions (complementary CDFs) above the minimum thresholds $w_{\min}$ are approximately linear on a double-logarithmic scale, indicating power law-like patterns for the richest individuals in both samples (see Online Appendices I and J). The empirical densities of returns to wealth portfolios are also reasonably well approximated by the expected Laplace distribution. This is readily indicated by their (symmetric) tent shape on semi-logarithmic scale that is characteristic of the Laplace and shown in Appendix C. Apart from mere visual inspection, the

[10]This procedure is also advantageous in so far as the respective $w_{\min}$ levels ensure that the estimated power laws always span at least two orders of magnitude, usually considered to be a minimum requirement for significantly claiming a power law distribution in the first place (Stumpf and Porter, 2012).

[11]In a strict sense, symmetry of the Laplace distribution is not guaranteed because the parameter values of the power law tail might be time-varying. The correct distribution would then be an asymmetric Laplace distribution of wealth returns, yet we find the empirical distributions to have skewness that is not statistically different from zero, indicating that the symmetric version is empirically useful.

TABLE 1
TAIL INDEX ESTIMATES $\hat{\alpha}$ FOR THE POWER LAW REGION OF THE *SOEP* SURVEY WITH STANDARD ERRORS IN PARENTHESES

SOEP	2002	2007	2012
Tail exponent estimate $\hat{\alpha}$	1.3144	1.0978	1.2982
	(0.0423)	(0.0324)	(0.0354)
Minimum wealth level estimate $\hat{w}_{\min}$	0.28	0.20	0.18
Maximum wealth level w_{max}	70.55	30.60	16.00
Sample size N	961	1260	1332

Note: The minimum and maximum wealth levels $\hat{w}_{min}$ and w_{max} are stated in millions of euros, deflated with index year 2010, while N denotes the number of observations in the power law tail.

standard procedure to test for a Laplace distribution is to fit an exponential power (or Subbotin) distribution to the data (Subbotin, 1923). As the Subbotin distribution includes the Laplace as a special case when its shape parameter equals unity, an MLE fit of the Subbotin parameters provides a convenient test. As we show in Appendix C, a shape parameter of unity cannot be rejected in any of the considered cases; therefore, our findings should not be distorted by systematic deviations from the parametric forms we impose for the estimations.

4.1. *Distributional Results*

We estimate the parameters for the power law distribution separately for the survey and the rich list. The tail indices are estimated via maximum likelihood, using the respective empirical minima from the rich list, and using the procedure described in Clauset *et al.* (2009) to estimate the respective minima $\hat{w}_{\min}$ in the survey. Tables 1 and 2 report tail index estimates for the survey tails and rich lists, respectively.

Two peculiarities stand out. First, normality of standard errors for the tail index estimates (De Haan and Resnick, 1997) implies that Zipf's law (with $\alpha = 1$) can be rejected with at least 95 percent confidence in all survey years. Wealth in the survey tails therefore appears more equally distributed than scale-independent growth would imply. Second, the wealth maxima in the survey are not even on the same order of magnitude as the wealth minima reported in the rich lists. These implausibly small maxima indicate severe undersampling (or rather the complete absence) of the super-rich in the survey, and are a major reason for the relatively low degree of measured inequality in the survey tails. Tail index estimates for the rich list, on the contrary, stand in stark contrast to those for the survey. As shown in Table 2, we cannot reject Zipf's law at the usual significance levels in any of the years other than 2013 and 2016, with Zipf's law being only barely rejected in 2013.[12] In the language of the stochastic accumulation process, Zipf's law indicates scale-independent returns among the super-rich. Yet significant deviations from Zipf's law in the survey tails point to scale-dependent wealth returns within the survey

[12]According to *manager magazin* staff, 2016 is the only year in which they tried to account for wealth held in foundations or charitable organizations. Thus the 2016 tail index does not measure the same concept, which is why we mostly discard 2016 in our analysis.

TABLE 2

TAIL INDEX ESTIMATES $\hat{\alpha}$ FOR THE *MANAGER MAGAZIN* RICH LIST WITH STANDARD ERRORS IN PARENTHESES

Manager Magazin	2010	2011	2013	2014	2015	2016
Tail exponent estimate $\hat{\alpha}$	0.9983	0.9863	1.0999	0.9874	0.9358	0.7615
	(0.0450)	(0.0442)	(0.0495)	(0.0443)	(0.0419)	(0.0341)
Minimum wealth level estimate $\hat{w}_{\min}$	200	200	250	250	250	200
Maximum wealth level w_{max}	17,100	19,000	23,950	31,000	26,500	30,000
Sample size N	499	498	494	497	500	500

Note: The minimum and maximum wealth levels w_{min} and w_{max} are stated in millions of euros, deflated with index year 2010, while N denotes the number of observations in the power law tail.

TABLE 3
MAXIMUM LIKELIHOOD PARAMETER ESTIMATES FOR THE LAPLACE DISTRIBUTION OF WEALTH RETURNS, WITH STANDARD ERRORS IN PARENTHESES

Laplace Parameter Estimates	$\hat{m}$	$\hat{\sigma}$
Manager magazin 2010–2015	0.1024	0.1458
	(0.0090)	(0.0010)
Manager magazin 2011–2016	0.0824	0.1319
	(0.0117)	(0.0082)
SOEP 2002–2007	0.0280	0.3710
	(0.0144)	(0.0077)
SOEP 2007–2012	0.0215	0.3745
	(0.0149)	(0.0076)

Note: While a location measure or "average" return of zero cannot be rejected for the survey tails, it is significantly greater than zero in the rich lists. Note that the dispersion of returns, that is, the "average" risk across portfolios, is markedly lower in the rich lists than in the survey tails.

populations, and obviously also to scale-dependent returns between the two sample types.

Therefore, we consider the distribution of wealth returns in the two sample types, and to facilitate comparison we construct wealth returns over 5-year intervals. Several nonparametric tests reject the null hypothesis of distributional equivalence between the two sample types in both periods, but fail to reject it within the samples between periods (cf. Online Appendix K). Apparently the data suggest that wealth dynamics are time-invariant but scale-dependent between sample types. The parameter estimates for the Laplace distribution of wealth returns, summarized in Table 3, strengthen the impression from the nonparametric tests. The estimates for the location parameter m (the "average" return) and the dispersion σ (the "average" risk) do not vary much within the respective samples, yet vastly differ between the two sample types. While average returns do not significantly differ from zero in the survey tails, zero can safely be rejected at the 5 percent level in the rich lists, where m is significantly positive, implying that Germany's super-rich on average became wealthier during the considered period. Paradoxically, however, σ is significantly lower in the rich lists than in the survey tails, apparently indicating that super-rich portfolios are less risky than the ones in the survey tails.[13] So how can we interpret these findings?

4.2. *Data First*

Taken at face value our findings indicate that the accumulation process is scale-dependent in the survey study but scale-independent in the rich list, where we find higher average returns and lower volatility compared to the survey. From a theoretical point of view this is puzzling. How plausible is it that the investment strategies of the super-rich converge to roughly equivalent risk profiles that not only outperform other (still rather) wealthy individuals, but do so at a lower risk? The conventional rationale for the risk-return tradeoff, as for instance in the

[13]This interpretation, although entirely conventional, needs to assume ergodicity in returns to wealth, which is why we use quotation marks for the notions of "average" return or "average" risk.

canonical intertemporal capital asset pricing model of Merton (1973), suggests that the conditional expected excess return should grow linearly with its conditional variance.[14] However, both the nonparametric tests and the parameter estimates for the Laplace distribution of wealth returns indicate that the super-rich enjoy higher expected returns at lower risk. The excess returns of Germany's super-rich cannot be explained by a higher risk tolerance, because this should be reflected in a higher dispersion of returns among the super-rich.

The *data first* interpretation thus suggests not only scale-dependence but also scale-dependence that cannot be explained by heterogeneous risk preferences alone. To explain the estimation results within the framework of random multiplicative growth, we need to assume that financial markets are not fully competitive in the conventional sense. This would suggest that investors' talent or the increased set of possibilities that comes with being very wealthy enables the richest to persistently beat the market and achieve above average risk-adjusted returns at a lower risk. Such an interpretation would also be at odds with empirical findings on risk preferences that observe decreasing risk-aversion in wealth levels, such that higher net worth correlates positively with a higher dispersion in returns to wealth (Guiso *et al.*, 1996; King and Leape, 1998; Calvet and Sodini, 2014). The *data first* interpretation thus poses a challenge to both, the empirically observed risk profiles, and the idea that financial markets with rapid feedbacks and a low degree of informational asymmetries should be close to the benchmark of a fully competitive market.

4.3. *Theory First*

Our central point here is that these "puzzles" can be resolved within the *theory first* interpretation once we agree that estimates of the tail exponent in the two samples suffer from two different sources of bias. Equiprobable sampling in the survey makes it very unlikely to observe the largest wealth levels that are necessary for reliable estimation of the tail index, as we quantify in Appendix E. Note that the probability of including the maximum wealth level for the *SOEP* sampling ratio under equiprobable sampling is 0.035 percent and thus practically equal to zero. Adding to this problem are concerns of social desirability biases, particularly the phenomenon that the super-rich tends not to respond to survey requests. As the probability of non-response is therefore positively correlated with wealth levels, the survey is subject to *differential non-response* (Kennickell and Woodburn, 1999; Eckerstorfer *et al.*, 2016; Vermeulen, 2016, 2018). These two considerations lead to *undersampling*, that is, the largest wealth levels are not included at all in the survey sample. In contrast, the rich list is a carefully selected sample aimed at covering the super-rich, and one can therefore expect that undersampling is not an issue. On the contrary, the *manager magazin* staff relies on public records for their compilation of the

[14]There exists an ongoing debate on whether this relationship can be established empirically (French et al., 1987; Campbell, 1987; Nelson, 1991; Campbell and Hentschel, 1992; Harvey, 2001; Goyal and Santa-Clara, 2003; Brandt and Kang, 2004; Ghysels et al., 2005; Bali and Peng, 2006; Andersen et al., 2006; Guo and Whitelaw, 2006; Lundblad, 2007; Bali, 2008; Gonzales et al., 2012). While most of the studies find at least weak support for the risk-return trade-off for various time frames and markets, the debate seems to have now shifted to the precise functional form of this relationship—as opposed to the linear one implied by the CAPM.

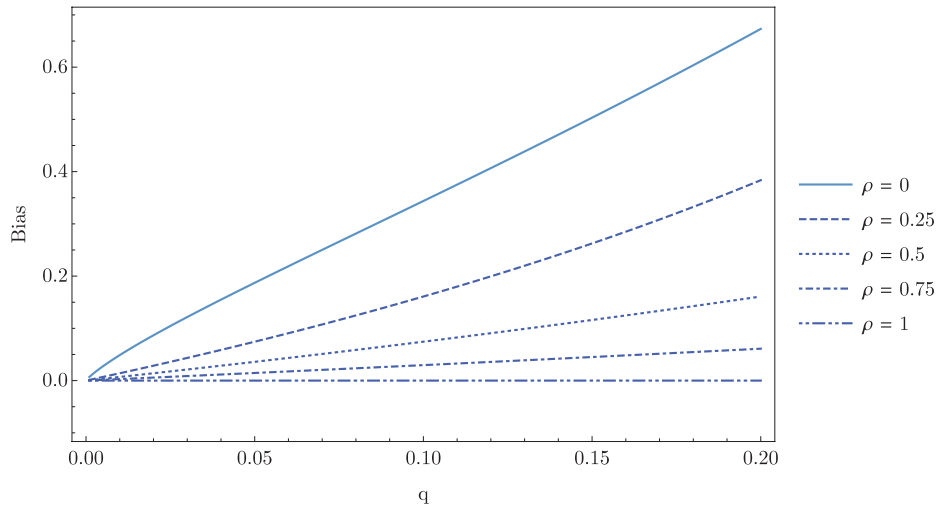

Figure 1. Illustration of the Analytical Bias Resulting from Different Combinations of Reporting Rates ρ and Underreporting Fractions q Compared to the Theoretically Expected α of Unity for Zipf's Law.

Notes: The Strength of the Bias Increases Disproportionately with Decreasing Reporting Rates, Where $\rho = 0$ Leads to Much More Upward Biased Tail Index Estimates, Even Compared to $\rho = 0.25$ [Colour figure can be viewed at wileyonlinelibrary.com]

rich list, likely underestimating the actual wealth levels for Germany's super-rich due to privacy considerations and tax avoidance that is particularly pronounced among the wealthiest (Alstadsæter *et al*., 2019). Consequently, the *manager magazin* sample should be subject to *underreporting*, not undersampling. In more colloquial terms, the upward bias in the survey sample arises because the richest are not included at all in the sample, while the upward bias in the rich list arises because the richest are not included with the full extent of their wealth.

To study the relative biases arising from differential undersampling and underreporting, we plot the upward deviation from the theoretically expected tail exponent of $\alpha = 1$ for different reporting rates ρ and undersampling in the (empirically motivated) quantile $q \in (0, 0.2)$. The case $\rho = 0$ corresponds to undersampling and is also the only case for $\rho < q$ that we can examine along the lines elaborated in Section 2.

Figure 1 supports the intuition that the relative bias is decreasing in the reporting rate ρ, because for smaller ρ a larger fraction of wealth is not reported. When $\rho = 1$ we recover the initial distribution from equation (2), and there is no bias for any q. Compared to the underreporting bias, the undersampling bias is rather unexpected though. If merely 25 percent of wealth were to be reported by the richest q-quantile, this would lead to a disproportionately smaller bias in the estimator, indicating that tail index estimates from the rich list are in all likelihood much less (upward) biased than estimates from the survey. This is reminiscent of the finding by Cristelli *et al*. (2012) that the maximum in a power law is most informative. Our result is more general in the sense that even partial inclusion of these top observations by only a fraction of their true level will greatly reduce the bias in measured

inequality. Given the limited impact of differential underreporting, we conclude that the true inequality of the system is substantially closer to the Zipf benchmark than the survey estimates seem to suggest, as indicated by the less biased estimates for the rich list.

Furthermore, our closed-form expression (2) that quantifies the impact of underreporting on the tail exponent also allows us to back out the reporting rates ρ for the rich lists. We assume that the upper 20 percent quantile exhibits different reporting behavior, in the sense that the richest 100 Germans constitute a rather salient set on the rich list, where the *manager magazin* staff focuses their efforts to compile reliable data (Balz *et al*., 2014), and thus the effect of tax avoidance should not be compounded by rounding errors or selection bias for the considered sources.[15] Therefore, fixing $q = 0.2$ and further assuming that Zipf's law governs the true distribution, we obtain the reporting rates ρ in Table 4.

The implied reporting rates appear to be plausible, except for the 2016 estimate that neatly reflects the qualitative change in the data collection procedure by the *manager magazin* staff.[16] Note the highly nonlinear and perhaps counterintuitive effect of a mere 20 percent decrease in $\hat{\alpha}$ between 2015 and 2016 that requires the implied reporting rate to more than triple, showing that the change in sampling procedures between 2015 and 2016 is a qualitative shift that would easily be missed if we were to exclusively look at the 20 percent increase in measured inequality. Thus the assumption of Zipf's and consequently Gibrat's law along with scale-independence seem entirely plausible in the *theory first* interpretation, especially because we cannot reject Zipf's law in any of the years other than 2013 and 2016.

Regarding the survey, *theory first* suggests that the deviation from Zipf's law in the *SOEP* data originates from undersampling ($\rho = 0$) such that the super-rich are entirely absent in the sample. Using the closed-form expression (3), we can infer the q-quantiles of non-respondents from the estimated tail exponents both in the survey tail, denoted q_{pl}, and also for the survey as a whole, $q_{tot} = q_{pl}(n_{pl}/n_{tot})$, where n_{tot} denotes the size of the *SOEP* sample and n_{pl} denotes the size of the survey tail (reported in the last row of Table 1). The results are summarized in Table 5.

The implied non-response rates relative to the size of the survey sample are remarkably low. The effects of equiprobable sampling combined with differential non-response quite plausibly lead to non-response rates q_{tot} of 0.1–0.4 percent. Consequently, the survey data are not inconsistent with the interpretation of scale-independent multiplicative growth and therefore Zipf's law in wealth levels. As the mixture of non-overlapping Zipf samples is also distributed as a Zipf law, the *theory first* interpretation supports scale-independence across the entire tail of the German wealth distribution. After all, our results underline the importance of maximum wealth levels for the estimation of tail indices because failing to account

[15]This is supported by apparent "digit-preferences" or "heaping effects" that we can observe below rank 100, where the data seem abnormally clustered in increments of 50 million euro (Heitjan and Rubin, 1991; Schneeweiß et al., 2010).

[16]Reporting rates $\rho > 1$ could, at least in principle, arise from the salience of the richest quantile through intensified compilation efforts that lead to a relative overestimation of top wealth. So whenever $\rho > 1$, the salience bias would outweigh the tax avoidance and social desirability biases. More importantly, however, a reporting rate of $\rho = 1$ can be safely rejected only in 2016, while in 2013 we are just very slightly above the conventional 5 percent level. In all other years, we cannot reject $\rho = 1$ at this significance level.

TABLE 4

IMPLIED REPORTING RATES ρ FOR DIFFERENTIAL UNDERREPORTING IN THE RICH LISTS, WITH STANDARD ERRORS IN PARENTHESES

Manager Magazin	2010	2011	2013	2014	2015	2016
Implied reporting rate ρ	1.0086	1.0719	0.6350	1.0659	1.4092	4.7878
	(0.1959)	(0.2095)	(0.1517)	(0.2083)	(0.2884)	(1.1743)

Note: For illustrative purposes, we assume that $q = 0.2$ and that the true distribution follows Zipf's law exactly. Except for 2016, the estimates suggest little differential reporting behavior, adding to the plausibility of the *theory first* interpretation.

TABLE 5
Implied Non-response Rates in the Survey Tail, $\hat{q}_{pl}$, and in the Entire Survey, $\hat{q}_{tot}$, Calculated from Equation (3) Under the Assumption of Zipf's Law, with Standard Errors in Parentheses

SOEP	2002	2007	2012
Implied non-response rate PL tail $\hat{q}_{pl}$	0.0906	0.02310	0.08538
	(0.0135)	(0.0094)	(0.0114)
Implied non-response rate full sample $\hat{q}_{tot}$	0.0030	0.0010	0.0039
	(0.0005)	(0.0004)	(0.0005)

Note: Non-response rates are tiny and imply that missing merely 25 to 100 of the super-rich in the survey can already explain the observed deviations from Zipf's law.

for merely 0.1–0.4 percent of the richest individuals already leads to substantial biases—and the descriptive statistics for the two samples clearly indicate that the actual response rate of the super-rich in the survey is zero. As we show in the upcoming subsection, differences in tail index estimates translate into substantial differences in estimated top tail wealth, and therefore also lead to enormous differences in measures of wealth inequality.[17]

4.4. *Total Wealth Estimates*

How much wealth is concentrated in the power law tail? The most recent literature on this matter extrapolates the estimates from survey studies to a maximum determined from rich lists (Vermeulen, 2018; Bach *et al*., 2019). Even within this established methodology, three very different kinds of answers emerge depending on the pre-analytical vision one uses. In line with the literature, we use the continuous analogue of the power law distribution and integrate to derive a measure for the total power law wealth W. The minima correspond to the estimates for the survey study, while we take the maxima from the rich lists. Within the *data first* interpretation, we need to choose between the estimated tail exponents from the survey study and the rich list corresponding to the respective belief that either the inequality within the *SOEP* or the *manager magazin* sample is more representative of the power law tail as a whole. The *theory first* perspective suggests Zipf's law and thus leaves no such degree of freedom. The estimation strategy is elaborated in more detail in Appendix D, where we also detail how to estimate the population n inhabiting the power law tail.

In the *data first* estimations, we essentially extrapolate the power law population in-sample to the entire German population of $N = 82.5$ million (Statistisches Bundesamt, 2017). This simple extrapolation is justified because *data first* assumes no systematic non-response rates for the richest. The estimates for the population from the survey study reveal a relatively large power law population with a relatively homogeneous wealth distribution, while the estimates for the rich list imply a very small population characterized by an extremely heterogeneous wealth distribution. The *theory first* perspective implies Zipf's law for the entire top tail and attributes observed differences from this benchmark in the survey to differential

[17]This will of course also be true for measured inequality with respect to the entire population, and not just within the power law tail that we focus on here.

TABLE 6
Estimated Wealth in the Power Law Tail for Combinations of Minima and Maxima from the Respective Survey and Rich List Samples in Billions of Euros (Inflation-Adjusted with Base Year 2010)

mm $\widehat{W}$						
SOEP years/mm years	2010	2011	2013	2014	2015	2016
2002	1096	1079	1819	1409	1280	1171
2007	1129	1109	1910	1447	1305	1177
2012	1140	1118	1939	1458	1313	1179
SOEP $\widehat{W}$						
SOEP years/mm years	2010	2011	2013	2014	2015	2016
2002	3114	3117	3124	3131	3126	3130
2007	5420	5447	5506	5571	5532	5563
2012	2884	2887	2894	2900	2896	2900
Zipf $\widehat{W}$						
SOEP years/mm years	2010	2011	2013	2014	2015	2016
2002	8498	8579	8757	8956	8835	8931
2007	8185	8261	8428	8614	8500	8590
2012	7882	7955	8114	8292	8184	8269

Note: Details regarding the underlying estimation strategies and parameter constellations are described in Appendix D. The estimates exhibit tremendous variation, almost spanning one order of magnitude.

non-response. We thus correct our population estimates for the survey by the estimated non-response rates. Unanimously, we find the largest estimated power law populations for this *theory first* perspective (see Appendix D). Both the corrected and the uncorrected estimates for the survey study differ up to one order of magnitude with respect to the estimates for the rich list. The correction within the *theory first* approach has a very limited effect on the estimated total population, resulting from the fact that the estimated non-response rates are tiny. This leaves us with three estimation strategies, each with 18 possible combinations of $\widehat{w}_{\min}^{SOEP}$ and w_{max}^{mm} for all sampling years. Table 6 shows how the differences in the estimated power law populations and tail indices translate into differences in total wealth.

We note first that especially the 2007 estimates for the *SOEP* are in remarkably close agreement with the latest estimates in Bach *et al*. (2019) based on the Household Finance and Consumption Survey (HFCS), even though our samples differ substantially from theirs.[18] This is also the case where the estimation procedure for the total power law population is closest to theirs. We take this as evidence that our results are not driven by idiosyncrasies in our data and instead testify to the external validity of our approach.

Second, and more importantly, the results differ enormously between the two pre-analytical visions. The estimates for the pure Zipf case are higher than the rich list estimates by at least a factor of six, in some cases even by one order of magnitude. This is primarily caused by the huge differences in estimated population, with both estimation strategies appearing to be plausible. Even when populations are not differing too much, the pre-analytical vision has a large effect on estimated total wealth, as the uncorrected and corrected estimates for the pure survey and

[18]The main difference is that the estimation by Bach *et al*. (2019) can exploit survey weights of the HFCS for oversampling which are not available for the *SOEP*.

Zipf case show, differing by up to a factor of three. Therefore, even state-of-the-art methods for this kind of estimation will likely severely underestimate the degree of inequality both within the richest group and also between the top tail and the rest of the population. A case can be made (more or less convincingly) for all three estimation strategies, and it seems fair to say that total wealth estimates are influenced at least as much by pre-analytical belief as they are by the data used for estimation.

5. Discussion

We have shown that the pre-analytical vision decisively informs the research agenda as well as the conclusions drawn from it. So how wealthy are the rich, and are returns to wealth scale-dependent for them? As we have argued here, the proposed mutually contradicting interpretations of *data first* versus *theory first* are observationally equivalent to each other. *Data first* interprets the observed deviations from Gibrat's law in wealth returns, and consequently Zipf's law in wealth levels, as evidence for scale-dependence. *Theory first*, on the contrary, explains these deviations through sampling and reporting biases that affect the two sample types differently. Ultimately, we cannot discriminate between the two narratives based on the data alone and seem to face a classic instance of the underdetermination of scientific theory by evidence, featuring prominently in the philosophy of science at least since the turn of the 20th century (Quine, 1975). On the contrary, dearly held convictions in economic theory, such as the risk-return trade-off, informationally efficient markets, and the classical notion of competition that requires an equalization of rates of return, patently suggest that *theory first* is a more plausible explanation for the data.

The proposed differential biases cast doubt on the validity of conclusions drawn across and within sample types, both in the cross-sectional and in the time series domain. Valid inference in the presence of reporting biases requires stability of parameters over time; otherwise identified trends might become spurious and instead reflect changes in bias. Because the proposed explanation of biases is behavioral and builds on empirically well-established phenomena such as salience, differential tax avoidance, or social desirability rather than being based on sampling method, there is no reason to expect stability. The estimated parameters within the *theory first* framework indeed suggest such variable behavioral responses over time.

Improving data availability and quality, for instance through the use of wealth or capitalized income tax data, might mitigate the severity of undersampling. Data availability then depends on the political willingness to impose such taxes in the first place, while future research will still be confined to the taxed population and a legal definition of wealth that is generally not catered toward the needs of statistical inference (Galbraith, 2019). Our results thus highlight the need to improve on survey and sampling methods, not to abandon them altogether. The recently conducted *SOEP-P* sample, which uses information on stock holdings to target high net worth individuals, is a first step in this direction, but still fails to adequately capture the super-rich in the targeted random sample, and thus fails to include the maximum wealth levels that we show to be crucial for valid inference. This is why data from the *SOEP-P* need to be complemented by the *manager magazin* rich lists

in the hope of adequately capturing tail wealth. Yet our findings strongly suggest that the *SOEP-P* supplement and the resulting composite sample from the rich lists still must be scrutinized along the lines of the fundamental *theory first* versus *data first* distinction. In the end, our results cast serious doubt on simply pooling data from different sample types and comparing trends therein, which has been standard practice so far (see, e.g. Vermeulen, 2018; Bach *et al*., 2019; Schröder *et al*., 2020).

While it is not surprising that the two narratives yield different estimates for total wealth in the top tail, the magnitude of this difference comes probably unexpected for most, because power laws have the counter-intuitive property that supposedly small variations in the tail index lead to enormous variations in totals. This property is substantially compounded by tiny degrees of undersampling, here on the order of a tenth of a percent, that lead to differences in estimated total wealth by a factor of up to three. Such small degrees of undersampling are easily explained by equiprobable sampling from a power law, leading to an inclusion probability of the maximum that is on the order of a hundredth of a percent in our case, and thus practically equal to zero. As we have shown how important the inclusion of an accurately measured maximum is for unbiased tail index estimation, this is disconcerting.

The enormous differences between total wealth estimates suggest that inferences from survey studies regarding the cross-sectional distribution of wealth and its time variation tend to be severely distorted, illustrating that discussions about the notion of "representativeness" in scale-free systems are not discussions about technical subtleties but disagreements in substance. In Appendix F, we conduct a simple analytical thought experiment for an extreme case of unrepresentative oversampling of the rich using *logarithmic sampling* across different orders of magnitude in wealth levels. We show that for Zipf's law, the necessary sampling ratio to surely include the maximum decays extremely fast by a power function. If the maximum is indeed as important as our analytical results on the undersampling bias indicate, even conventional oversampling techniques will be insufficient, and should instead try to implement logarithmic sampling to allow for unbiased estimations. After all, our results show that accurate representations of total wealth require us to be highly "unrepresentative" in the sampling of individuals.

References

Abul-Magd, A. Y., "Wealth Distribution in an Ancient Egyptian Society," *Physical Review E*, 66(5), 1–3, 2002.

Alan, S., K. Atalay, and T. F. Crossley, "Do the Rich Save More? Evidence from Canada," *Review of Income and Wealth*, 61(4), 739–58, 2015.

Alstadsæter, A., N. Johannesen, and G. Zucman, "Tax Evasion and Inequality," *American Economic Review*, 109(6), 2073–103, 2019.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold, "Volatility and Correlation Forecasting," *Handbook of Economic Forecasting*, 1, 777–878, 2006.

Aoki, S., and M. Nirei, "Pareto Distribution of Income in Neoclassical Growth Models," *Review of Economic Dynamics*, 2, 25–42, 2016.

Bach, L., L. E. Calvet, and P. Sodini, "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy," *HEC Paris Research Paper*, 1126, 16-03, 2017.

Bach, S., M. Beznoska, and V. Steiner, "A Wealth Tax on the Rich to Bring Down Public Debt?," *DIW Berlin Discussion Paper*, 113, 1–22, 2011.
Bach, S., A. Thiemann, and A. Zucco, "Looking for the Missing Rich: Tracing the Top Tail of the Wealth Distribution," *International Tax and Public Finance*, 26(6), 1234–58, 2019.
Bali, T. G., "The Intertemporal Relation Between Expected Return and Risk," *Journal of Financial Economics*, 8, 101–31, 2008.
Bali, T. G., and L. Peng, "Is There a Risk-Return Trade-Off? Evidence from High-Frequency Data," *Journal of Applied Econometrics*, 21(8), 1169–98, 2006.
Balz, J., C. Sunstein, and R. Thaler, "Choice Architecture." in E. Shafir, (ed), *The Behavioral Foundations of Public Policy*. Princeton University Press, Princeton, 428–39, 2014.
Benhabib, J., and A. Bisin, "Skewed Wealth Distributions: Theory and Empirics," *Journal of Economic Literature*, 56(4), 1261–91, 2018.
Benhabib, J., A. Bisin, and S. Zhu, "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents," *Econometrica*, 79(1), 123–75, 2011.
Blank, A., and S. Solomon, "Power Laws in Cities Population, Financial Markets and Internet Sites (Scaling in Systems with a Variable Number of Components)," *Physica A: Statistical Mechanics and its Applications*, 287(1–2), 279–88, 2000.
Bottazzi, G., "Subbotools User's Manual," *Laboratory of Economics and Management Working Paper Series, Sant'Anna School of Advanced Studies*, 2004(1), 2004.
Brandt, M. W., and Q. Kang, "On the Relationship Between the Conditional Mean and Volatility of Stock Returns: A latent VAR Approach," *Journal of Financial Economics*, 72(2), 2004.
Brzezinski, M., "Do Wealth Distributions Follow Power Laws? Evidence from 'Rich Lists'," *Physica A: Statistical Mechanics and its Applications*, 40, 155–62, 2014.
Calvet, L. E., and P. Sodini, "Twin Picks: Disentangling the Determinants of Risk-Taking in Household Portfolios," *The Journal of Finance*, 69(2), 867–906, 2014.
Campbell, J. Y., "Stock returns and the term structure," *Journal of Financial Economics*, 18(2), 373–99, 1987.
Campbell, J. Y., and L. Hentschel, "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31(3), 281–318, 1992.
Castaldi, C., and M. Milaković, "Turnover Activity in Wealth Portfolios," *Journal of Economic Behavior & Organization*, 6, 537–52, 2007.
Clauset, A., C. R. Shalizi, and M. E. J. Newman, "Power-Law Distributions in Empirical Data," *SIAM Review*, 5, 661–703, 2009.
Coelho, R., Z. Néda, J. J. Ramasco, and M. M. Augusta Santos, "A Family-Network Model for Wealth Distribution in Societies," *Physica A: Statistical Mechanics and its Applications*, 35, 515–28, 2005.
Córdoba, J. C., "A generalized Gibrat's law," *International Economic Review*, 49(4), 1463–68, 2008a.
Córdoba, J.-C., "On the Distribution of City Sizes," *Journal of Urban Economics*, 63(1), 177–97, 2008b.
Cowell, F. A., "Inequality Among the Wealthy," *Centre for Analysis of Social Exclusion Series*, 15, 1–33, 2011.
Cristelli, M., M. Batty, and L. Pietronero, "There is more than a Power Law in Zipf," *Scientific Reports*, 2, 812, 2012.
Davies, J. B., and A. F. Shorrocks, "The Distribution of Wealth." in A. Atkinson, and F. Bourguignon, (eds), *Handbook of Income Distribution*. Elsevier Science B.V, Amsterdam, 605–75, 2000.
De Haan, L., and S. Resnick, "On Asymptotic Normality of the Hill Estimator, Communications in Statistics," *Stochastic Models*, 14(4), 849–66, 1997.
Deaton, A., "Saving and Liquidity Constraints," *Econometrica*, 59(5), 1221–48, 2003.
Drăgulescu, A. A., and V. M. Yakovenko, "Exponential and Power Law Probability Distributions of Wealth and income in the United Kingdom and the United States," *Physica A: Statistical Mechanics and its Applications*, 29, 213–21, 2001.
Dynan, K. E., J. Skinner, and S. P. Zeldes, "Do the Rich Save More?," *Journal of Political Economy*, 112(2), 397–444, 2004.
Eckerstorfer, P., J. Halak, J. Kapeller, B. Schütz, F. Springholz, and R. Wildauer, "Correcting for the Missing Rich: An Application to Wealth Survey Data," *Review of Income and Wealth*, 62(4), 605–27, 2016.
Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri, "Heterogeneity and Persistence in Returns to Wealth," *Econometrica*, 88(1), 115–70, 2020.
Fama, E. F., "The Behavior of Stock-Market Prices," *The Journal of Business*, 38(1), 34–105, 1965.
_____, "Efficient Capital Markets: A Review of Theory and Empirical Work," *The Journal of Finance*, 25(2), 383–417, 1970.
_____, "Efficient Capital Markets: II," *The Journal of Finance*, 46(5), 1575–617, 1991.
French, K. R., G. W. Schwert, and R. F. Stambaugh, "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19(1), 3–29, 1987.

Frick, J. R., M. M. Grabka, and J. Marcus, *Editing and Multiple Imputation of Item-Non-Response in the 2002 Wealth Module of the German Socio-Economic Panel (SOEP)*. Data Documentation, DIW Berlin, 18, 2007.
Gabaix, X., "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, 114(3), 739–67, 1999.
_____, "Power Laws in Economics and Finance," *Annual Review of Economics*, 255–93, 2009.
Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll, "The Dynamics of Inequality," *Econometrica*, 84(6), 2071–111, 2016.
Galbraith, J. K., "Sparse, Inconsistent and Unreliable: Tax Records and the World Inequality Report 2018," *Development and Change*, 50(2), 329–46, 2019.
Ghysels, E., P. Santa-Clara, and R. Valkanov, "There is a Risk-Return Trade-Off after All," *Journal of Financial Economics*, 76(3), 509–48, 2005.
Gibrat, R., *Les inégalités économiques; applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel*. Librairie du Recueil Sirey, 1931.
Glazer, E., "At J.P. Morgan, $9 Million in Assets isn't Rich Enough," *Wall Street Journal*, 2016. March 18.
Goldstein, M. L., S. A. Morris, and G. G. Yen, "Problems with Fitting to the Power-Law Distribution," *The European Physical Journal B*, 41(2), 255–58, 2004.
Gonzales, M., J. Nave, and G. Rubio, "The Cross Section of Expected Returns with MIDAS Betas," *Journal of Financial and Quantitative Analysis*, 47(1), 115–35, 2012.
Goyal, A., and P. Santa-Clara, "Idiosyncratic Risk Matters!," *The Journal of Finance*, 58(3), 975–1007, 2003.
Guiso, L., T. Jappelli, and D. Terlizzese, "Income Risk, Borrowing Constraints, and Portfolio Choice," *American Economic Review*, 86(1), 158–72, 1996.
Guo, H., and R. F. Whitelaw, "Uncovering the Risk-Return Relation in the Stock Market," *The Journal of Finance*, 61(3), 1433–63, 2006.
Harvey, C. R., "The Specification of Conditional Expectations," *Journal of Empirical Finance*, 8(5), 573–637, 2001.
Hegyi, G., Z. Néda, and M. A. Santos, "Wealth Distribution and Pareto's Law in the Hungarian Medieval Society," *Physica A: Statistical Mechanics and its Applications*, 38, 271–77, 2007.
Heitjan, D. F., and D. B. Rubin, "Ignorability and Coarse Data," *The Annals of Statistics*, 19(4), 2244–53, 1991.
Hubmer, J., P. Krusell, and A. A. Smith Jr, "The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation," *NBER Working Paper Series*, 23011, 2016.
Jappelli, T., and L. Pistaferri, "Fiscal Policy and MPC Heterogeneity," *Macroeconomics*, 6(4), 107–36, 2014.
Jones, C. I., and J. Kim, "A Schumpeterian Model of Top Income Inequality," *Journal of Political Economy*, 126(5), 1785–826, 2018.
Kennickell, A. B., and R. L. Woodburn, "Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth," *Review of Income and Wealth*, 45(2), 193–215, 1999.
King, M. A., and J. I. Leape, "Wealth and Portfolio Composition: Theory and Evidence," *Journal of Public Economics*, 69(2), 155–93, 1998.
King, M. R., and P. Maier, "Would Greater Regulation of Hedge Funds Reduce Systemic Risk?" in R. W. Kolb, (ed), *Lessons from the Financial Crisis: Causes, Consequences, and Our Economic Future*. John Wiley & Sons, Hoboken, New Jersey, 625–32, 2010.
Kotz, S., T. J. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations*. Birkhauser, Boston, 2001.
Levy, M., "Are Rich People Smarter?," *Journal of Economic Theory*, 11, 42–64, 2003.
Levy, M., and H. Levy, "Investment Talent and the Pareto Wealth Distribution: Theoretical and Experimental Analysis," *Review of Economics and Statistics*, 85(3), 709–25, 2003.
Levy, M., and S. Solomon, "New Evidence for the Power Law Distribution of Wealth," *Physica A: Statistical Mechanics and its Applications*, 24, 90–94, 1997.
Levy, S., "Wealthy People and Fat Tails: An Explanation for the Lévy Distribution of Stock Returns," *Finance Working Paper: Anderson Graduate School of Management*, 30(98), 1998.
Lundblad, C., "The Risk Return Tradeoff in the Long Run: 1836–2003," *Journal of Financial Economics*, 85(1), 123–50, 2007.
Luttmer, E. G., "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, 2(1), 547–76, 2010.
_____, "On the Mechanics of Firm Growth," *The Review of Economic Studies*, 78(3), 1042–68, 2011.
Luttmer, E . G., *Slow Convergence in Economies with Organization Capital*. University of Minnesota and Federal Reserve Bank of Minneapolis Working Paper Series, 748, 2018.

Malcai, O., O. Biham, and S. Solomon, "Power-Law Distributions and Levy-Stable Intermittent Fluctuations in Stochastic Systems of Many Autocatalytic Elements," *Physical Review E*, 60(2), 1299, 1999.
Merton, R. C., "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41(5), 867–87, 1973.
Nelson, D. B., "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59(2), 347–70, 1991.
Ning, D., and W. You-Gui, "Power Law Tail in the Chinese Wealth Distribution," *Chinese Physics Letters*, 24(8), 2434–36, 2007.
Nirei, M., "Pareto Distributions in Economic Growth Models," *Institute of Innovation Research Working Paper*, 09(05), 2009.
Nirei, M., and W. Souma, "A Two Factor Model of Income Distribution Dynamics," *Review of Income and Wealth*, 53(3), 440–59, 2007.
Piketty, T., and G. Zucman, "Wealth and Inheritance in the Long Run." in A. Atkinson, and F. Bourguignon, (eds), *Handbook of Income Distribution*. Vol. 2. Elsevier Science B.V, Amsterdam, 1303–68, 2015.
Quadrini, V., "The importance of Entrepreneurship for Wealth Concentration and Mobility," *Review of Income and Wealth*, 45(1), 1–19, 1999.
Quine, W. V., "On Empirically Equivalent Systems of the World," *Erkenntnis*, 9(3), 313–28, 1975.
Ramanujan, S., *The Lost Notebook and Other Unpublished Papers*. Narosa Publishing House, New Delhi, Berlin, New York, 1988.
Schneeweiß, H., J. Komlos, and A. S. Ahmad, "Symmetric and Asymmetric Rounding: A Review and Some New Results," *AStA Advances in Statistical Analysis*, 94(3), 247–71, 2010.
Schröder, C., C. Bartels, K. Göbler, M. M. Grabka, and J. König, "MillionärInnen unter dem Mikroskop: Datenlücke bei sehr hohen Vermögen geschlossen - Konzentration höher als bisher ausgewiesen," *DIW Wochenbericht*, 29, 511–22, 2020.
Sinha, S., "Evidence for Power Law Tail of the Wealth Distribution in India," *Physica A: Statistical Mechanics and its Applications*, 35, 555–62, 2006.
Statistisches Bundesamt, "Bevölkerung in Deutschland voraussichtlich auf 82,8 Millionen gestiegen," *Pressemitteilung*, 33/17, 2017.
Stumpf, M. P. H., and M. Porter, "Critical Truths About Power Laws," *Science*, 335(6069), 665–66, 2012.
Subbotin, M. T., "On the Law of Frequency of Error," *Matematicheskii Sbornik*, 31(2), 296–301, 1923.
Toda, A. A., "Incomplete Market Dynamics and Cross-Sectional Distributions," *Journal of Economic Theory*, 154, 310–48, 2014.
Vermeulen, P., "Estimating the Top Tail of the Wealth Distribution," *American Economic Review*, 106(5), 646–50, 2016.
_____, "How Fat is the Top Tail of the Wealth Distribution?," *Review of Income and Wealth*, 64(2), 357–87, 2018.
Wachter, J. A., and M. Yogo, "Why Do Household Portfolio Shares Rise in Wealth?," *The Review of Financial Studies*, 23(11), 3929–65, 2010.
Zipf, G. K., *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press, Boston, 1949.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site: