Generalized Pareto Curves: Theory and Applications

Thomas Blanchet, Juliette Fournier,
Thomas Piketty

March 2017

WORLD
WEALTH & INCOME
DATABASE
THE SOURCE FOR
GLOBAL INEQUALITY DATA

Generalized Pareto Curves: Theory and Applications*

Thomas BLANCHET (Paris School of Economics)
Juliette FOURNIER (Massachusetts Institute of Technology)
Thomas PIKETTY (Paris School of Economics)

First version: March 24, 2017
This version: September 11, 2017

Abstract

We define *generalized Pareto curves* as the curve of inverted Pareto coefficients $b(p)$, where $b(p)$ is the ratio between average income or wealth above rank p and the p-th quantile $Q(p)$ (i.e. $b(p) = \mathbb{E}[X|X > Q(p)]/Q(p)$). We use them to characterize entire distributions, including places like the top where power laws are a good description, and places further down where they are not. We develop a method to nonparametrically recover the entire distribution based on tabulated income or wealth data as is generally available from tax authorities, which produces smooth and realistic shapes of generalized Pareto curves. Using detailed tabulations from quasi-exhaustive tax data, we demonstrate the precision of our method both empirically and analytically. It gives better results than the most commonly used interpolation techniques. Finally, we use Pareto curves to identify recurring distributional patterns, and connect those findings to the existing literature that explains observed distributions by random growth models.

Keywords: inequality, income, wealth, Pareto, power law, interpolation

JEL codes: D31, C14

1 Introduction

It has long been known that the upper tail of the distribution of income and wealth can be approximated by a Pareto distribution, or *power law* (Pareto, 1896). The economic literature has made a wide use of this fact. On the empirical side, Pareto interpolation methods have been used by Kuznets (1953), Atkinson and Harrison (1978), Piketty (2001, 2003), Piketty and Saez (2003) and the subsequent literature exploiting historical tax tabulations to construct long-run series on income and wealth inequality. On the theoretical side, models with random multiplicative shocks have been suggested to explain the shape of the distribution, starting with the work of Champernowne (1953) and Simon (1955) on income, and Wold and Whittle (1957) on wealth. More recent contributions have shown how such models can account for both the levels and the changes in inequality (Nirei, 2009; Benhabib, Bisin, and Zhu, 2011; Piketty and Zucman, 2015; Jones and Kim, 2017; Jones, 2015; Benhabib and Bisin, 2016; Gabaix, Lasry, et al., 2016).

But while the Pareto approximation is acceptable for some purposes, it is not entirely correct, not even at the top. Some authors have explicitly noted this fact (eg. Atkinson, 2017). But deviations from the Pareto distribution — and what they imply for both empirical and theoretical work — have not yet been studied in a systematic way. If we want to better exploit the data at our disposal, and also to better understand the economic mechanisms giving rise to the observed distributions of income and wealth, we need to move beyond standard Pareto distributions.

In this paper, we develop the flexible notion of *generalized Pareto curve* in order to characterize and estimate income and wealth distributions. A generalized Pareto curve is defined as the curve of inverted Pareto coefficients $b(p)$, where $0 \leq p < 1$ is the rank, and $b(p)$ is the ratio between average income or wealth above rank p and the p-th quantile $Q(p)$ (i.e. $b(p) = \mathbb{E}[X|X > Q(p)]/Q(p)$).

If the tail follows a standard Pareto distribution, the coefficient $b(p)$ is constant, at least above a certain level of income or wealth. For example, if $b(p) = 2$ at the top of the wealth distribution, then the average wealth of individuals above €1 million is €2 million, the average wealth of individuals above €10 million is €20 million, and so on. In practice, we find that $b(p)$ does vary within the upper tail of observed income and wealth distributions (including within the top 10% or the top 1%), but that the curves $b(p)$ are relatively similar (typically U-shaped). In this paper, we use these generalized Pareto curves to study income and wealth distributions.

We make three main contributions. First, we show deep connections between the generalized

Pareto curves and the asymptotic power laws as defined in the theoretical literature, which is based on Karamata's (1930) theory of *regular variations.* This allows us to move away from the tight parametric assumptions that have characterized most applied empirical work on power laws. Previous work would often consider a constant Pareto coefficient above a sufficiently high threshold. But trying to summarize the whole top of the distribution by a single parameter can miss important features. Instead, we use a nonparametric definition of power laws and, using generalized Pareto curves, take it directly to the data, without the need for parametric approximations. Doing so, we confirm that the distributions of income and wealth are power laws in Karamata's (1930) sense, yet there are clear deviations from standard Pareto laws: we find that the distribution of income is more skewed toward the very top than what the standard Pareto model implies, especially in the United States. We further demonstrate that the inverted Pareto coefficients $b(p)$ can be viewed as belonging to a larger family of "local" Pareto coefficients, some of which have already been suggested in the literature (Gabaix, 1999). Those different coefficients tell similar stories, but the one that we use has the advantage of being very easy to compute, especially when access to data is limited (as is often the case with historical tax data).

Second, we show the usefulness of our approach for the estimation of income and wealth distributions using tax data, which is often available solely in the form of tabulations with a finite number of inverted Pareto coefficients $b_1, \ldots, b_K$ and thresholds $q_1, \ldots, q_k$ observed for percentile ranks $p_1, \ldots, p_K$. We show that taking carefully into account how the Pareto coefficient varies in the top of the distribution can dramatically improve the way we produce statistics on income and wealth inequality. Existing methods typically rely on diverse Paretian assumptions (or even less realistic ones) that, by construction, blur or even erase deviations from the standard Pareto distribution. We suggest a new method that can estimate the entire distribution much more faithfully. By using quasi-exhaustive annual micro files of income tax returns available in the United States and France over the 1962–2014 period (a time of rapid and large transformation of the distribution of income, particularly in the United States), we demonstrate the precision of our method. That is, with a small number of percentile ranks (e.g. $p_1 = 10\%$, $p_2 = 50\%$, $p_3 = 90\%$, $p_4 = 99\%$), we can recover the entire distribution with remarkable precision. The new method has other advantages too. First, it leads to smooth estimates of the distribution (meaning at least a differentiable quantile function, and a continuous density), while other methods introduce kinks near the thresholds included in the tabulation. Second, it allows for analytical results on the error, which we can then use to answer questions that are much more

difficult to address with other methods. For example, we can provide approximate error bounds for any tabulation of the data, determine how the percentile ranks in the tabulation should be chosen so as to maximize precision, and how many values we need to achieve a given precision level. Surprisingly, we find that the precision of the method is such that it is often preferable to use tabulations based on exhaustive data rather than individual data from a non-exhaustive subsample of the population, even for subsamples considered very large by statistical standards. For example, a subsample of 100 000 observations can typically lead to a mean relative error of about 3% on the top 5% share, while a tabulation based on exhaustive data that includes the percentile ranks $p = 10\%, 50\%, 90\%$ and 99% gives a mean relative error of less than 0.5%. For the top 0.1% share, the same error can reach 20% with the same subsample, while the same tabulation yields an error below 4%.

Third, and maybe most importantly, we show that studying how the inverted Pareto coefficient $b(p)$ varies within the tail can yield insights on the nature of the processes that generate income and wealth distributions. There is already a sizable literature showing that power laws arise from dynamic models with random multiplicative shocks. Those models, in their current form, do not properly account for the tendency of $b(p)$ to rise within the last percentiles of the distribution. But small adjustments to this framework can explain the U-shaped Pareto curves that we observe in practice. The key feature of models with random multiplicative shocks is their *scale invariance*: the rate of growth remains the same regardless of the initial income level. But if the scale invariance only holds asymptotically, then instead of strict Pareto distributions, we recover power laws in Karamata's (1930) sense that we studied earlier. Generalized Pareto curves then become a natural tool to study those distributions. To account for the shape of $b(p)$ that we observe empirically, we must introduce multiplicative shocks whose mean and/or variance changes with the initial level of income or wealth. Simple calibrations suggest that significant increases in the mean and the variance at the top are required to quantitatively reproduce the data. This implies that people at the top experience higher growth and more risk. This finding is consistent with available individual panel data (Chauvel and Hartung, 2014; Hardy and Ziliak, 2014; Bania and Leete, 2009; Guvenen et al., 2015; Fagereng et al., 2016; Bach, Calvet, and Sodini, 2017), yet we are able to reach using only cross-sectional data, something that can be done for a much larger set of countries and time periods. Gabaix, Lasry, et al. (2016) also showed that scale dependence is necessary to explain how fast inequality is increasing in the United States. Here we show that scale dependence also manifests itself when looking at the distribution in a static way, not only when looking at its evolution. Overall, these

findings are consistent with models where top incomes are driven by "superstar" entrepreneurs and managers (Gabaix and Landier, 2008), and where investors have heterogeneous levels of sophistication (Kacperczyk, Nosal, and Stevens, 2014).

The rest of the paper is organized as follows. In section 2, we provide the formal definition of generalized Pareto curves $b(p)$ and we show how they relate to the quantile function $Q(p)$ and the theory of power laws. In section 3, we present our generalized Pareto interpolation method, which is based on a transformation of $b(p)$. In section 4, we test its precision and compare it to other interpolation methods using individual income data for the United States and France covering the 1962–2014 period. In section 5, we analyze the error associated with our interpolation method, and provide formulas that give approximate bounds on this error in the general case. In section 6, connect our findings to models of income growth and wealth accumulation with multiplicative random shocks.

We believe that the approach and the tools that we develop in this article will be useful to researchers who want to improve the way we construct and analyze data on income and wealth inequality, and we indeed plan to use them to keep expanding the World Wealth and Income Database (`WID.world`). In this spirit, we developed an R package, named `gpinter`, that implements the methods described in this article and make them easily available to researchers. We also provide a web interface built on top of this package, available at `http://wid.world/gpinter`, to estimate and manipulate distributions of income and wealth on the basis of simple tabulated data files (such as those provided by tax administrations and statistical institutes) and generalized Pareto interpolation methods.[1]

2 Generalized Pareto Curves: Theory

2.1 Generalized Pareto Curves

We characterize the distribution of income or wealth by a random variable X with cumulative distribution function (CDF) F. We assume that X is integrable (i.e. $\mathbb{E}[|X|] < +\infty$) and that F is differentiable over a domain $D = [a, +\infty[$ or $D = \mathbb{R}$. We note f the probability density function (PDF) and Q the quantile function. Our definition of the inverted Pareto coefficient follows the one first given by Fournier (2015).

Definition 1 (Inverted Pareto coefficient). *For any income level $x > 0$, the inverted Pareto*

---

[1]R is maintained by the R Core Team (2016). The web interface uses `shiny` (Chang et al., 2017).

coefficient is $b^*(x) = \mathbb{E}[X|X > x]$, or:

$$b^*(x) = \frac{1}{(1 - F(x))x} \int_x^{+\infty} zf(z)\,\mathrm{d}z$$

We can express it as a function of the fractile p with $p = F(x)$ and $b(p) = b^*(x)$:

$$b(p) = \frac{1}{(1 - p)Q(p)} \int_p^1 Q(u)\,\mathrm{d}u$$

If X follows a Pareto distribution with coefficient α and lower bound $\bar{x}$, so that $F(x) = 1 - (\bar{x}/x)^\alpha$, then $b(p) = \alpha/(\alpha - 1)$ is constant (a property also known as van der Wijk's (1939) law), and the top $100 \times (1 - p)\%$ share is an increasing function of b and is equal to $(1 - p)^{1/b}$. Otherwise, $b(p)$ will vary. We can view the inverted Pareto coefficient as an indicator of the tail's fatness, or similarly an indicator inequality at the top. It also naturally appears in some economic contexts, such as optimal taxation formulas (Saez, 2001).

We solely consider inverted Pareto coefficient above a strictly positive threshold $\bar{x} > 0$, because they have a singularity at zero and no clear meaning below that. (They essentially relate to the top tail of a distribution.) We also favor looking at them as a function of the fractile p rather than the income x, because it avoids differences due to scaling, and make them more easily comparable over time and between countries. We call *generalized Pareto curve* the function $b : p \mapsto b(p)$ defined over $[\bar{p}, 1[$ with $\bar{p} = F(\bar{x})$.

Proposition 1. *If X satisfies the properties stated above, then b is differentiable and for all $p \in [\bar{p}, 1[$, $1 - b(p) + (1 - p)b'(p) \leq 0$ and $b(p) \geq 1$.*

The proof of that proposition — as well as all the others in this section — are available in appendix. The definition of $b(p)$ directly imply $b(p) \geq 1$. The fact that the quantile function is increasing implies $1 - b(p) + (1 - p)b'(p) \leq 0$. Conversely, for $0 \leq \bar{p} < 1$ and $\bar{x} > 0$, any function $b : [\bar{p}, 1[\to \mathbb{R}$ that satisfies property 1 uniquely defines the top $(1 - \bar{p})$ fractiles of a distribution with $\bar{p} = F(\bar{x})$.

Proposition 2. *If X is defined for $x > \bar{x}$ by $F(\bar{x}) = \bar{p}$ and the generalized Pareto curve $b : [\bar{p}, 1[\to \mathbb{R}$, then for $p \geq \bar{p}$, the p-th quantile is:*

$$Q(p) = \bar{x}\frac{(1 - \bar{p})b(\bar{p})}{(1 - p)b(p)} \exp\left(-\int_{\bar{p}}^p \frac{1}{(1 - u)b(u)}\,\mathrm{d}u\right)$$

2.2 Pareto Curves and Power Laws

For a strict power law (i.e. a Pareto distribution), the Pareto curve is constant. But strict power laws rarely exist in practice, so that we may want to characterize the Pareto curve when power

law behavior is only approximate. Approximate power laws are traditionally defined based on Karamata's (1930) theory of *slowly varying* functions. In informal terms, we call a function slowly varying if, when multiplied by power law, it behaves asymptotically like a constant under integration.[2]

Definition 2 (Asymptotic power law). *We say that X is an asymptotic power law if for some $\alpha > 0$, $1 - F(x) = L(x)x^{-\alpha}$, where $L :]0, +\infty[\to]0, +\infty[$ is a slowly varying function, which means that for all $\lambda > 0$, $\lim_{x \to +\infty} \frac{L(\lambda x)}{L(x)} = 1$.*

Definition 2 corresponds to the broadest notion of power laws. We call them "asymptotic" power laws to distinguish them from "strict" power laws (i.e. Pareto distributions). Strict power laws are characterized by their *scale invariance*, meaning that for all $\lambda > 0$, $1 - F(\lambda x) = \lambda^{-\alpha}(1 - F(x))$. The requirement that L is slowly varying in definition 2 means that $1 - F$ must be asymptotically scale invariant. That includes in particular situations where $1 - F$ is equivalent to a power law (i.e. $1 - F(x) \sim Cx^{-\alpha}$ for some $C > 0$). But we could also set, for example, $L(x) \propto (\log x)^{\beta}$ with any $\beta \in \mathbb{R}$.

This definition has some economic relevance too. In section 6 (theorem 8), we show that asymptotic power laws arise if the mechanisms traditionally assumed to generate Pareto distribution only hold asymptotically.

We will in general restrict ourselves to situations where $\alpha > 1$ to ensure that the means are finite.[3] With $\alpha > 1$, there is a strong link between generalized Pareto curve and asymptotic power laws.

Proposition 3. *Let $\alpha > 1$. X is an asymptotic power law with Pareto coefficient α, if and only if $\lim_{p \to 1} b(p) = \frac{\alpha}{\alpha-1}$.*

Proposition 3 generalizes van der Wijk's (1939) characterization of Pareto distributions to asymptotic power laws. Because $\alpha > 1 \Leftrightarrow \alpha/(\alpha - 1) > 1$, a distribution is an asymptotic power law if and only if its asymptotic inverted Pareto coefficient is strictly above one. It will tend toward infinity when α approaches one, and to one when α approaches infinity. This behavior is in contrast with distributions with a thinner tail, whose complementary CDF is said to be *rapidly varying*.

---

[2]See Bingham, Goldie, and Teugels (1989) for a full account of this theory.

[3]Hence, we exclude edge cases were the inverted Pareto coefficients are finite, but converge to $+\infty$ as $p \to 1$ (for example $b(p) = 3 - \log(1-p)$). Technically, they correspond to a power law, with $\alpha = 1$, but unlike a strict Pareto distribution with $\alpha = 1$, they have a finite mean. In practice, Pareto coefficients for the distribution of income or wealth are clearly above one, so there is no reason to believe that such cases are empirically relevant.

Proposition 4. $1 - F(x)$ *is rapidly varying (of index $-\infty$), meaning that for all $\lambda > 1$,* $\lim_{x \to +\infty} \frac{1 - F(\lambda x)}{1 - F(x)} = 0$ *if and only if* $\lim_{p \to 1} b(p) = 1$.

Distributions concerned by proposition 4 include the exponential, the normal or the log-normal. More broadly, it includes any distribution that converges to zero faster than any power law (i.e. $1 - F(x) = o(x^{-\alpha})$ for all $\alpha > 0$). For all those distributions, the generalized Pareto curve will eventually converge to one. Looking at the Pareto curve near $p = 1$ can therefore help discriminate fat-tailed distributions from others.

Propositions 3 and 4 imply that probability distributions may be divided into three categories, based on the behavior of their generalized Pareto curve. First, power laws, for which $b(p)$ converges to a constant strictly greater than one. Second, "thin-tailed" distributions, for which $b(p)$ converges to one. The third category includes distributions with an erratic behavior in the tail, for which $b(p)$ may oscillate at an increasingly fast rate without converging toward anything.[4] That last category does not include any standard parametric family of distributions, and its members can essentially be considered pathological. If we exclude it, we are left with a straightforward dichotomy between power laws, and thin tails.

When $\lim_{p \to 1} b(p) > 1$, so that X is an asymptotic power law, the generalized Pareto curve can further be used to observe *how* the distribution converges. If $b(p)$ increases near $p = 1$, the tail is getting fatter at higher income levels. But if $b(p)$ decreases, it is getting thinner.

With a strict power law, so that $b(p)$ is constant, the level of inequality stays the same as we move up through the distribution. The share of the top 10% among the whole population is the same as the share of the top 1% among the top 10% or the share of the top 0.1% among the top 1%. This property is often called the "fractal" nature of inequality. Deviations from a constant $b(p)$ indicate deviations from this rule: if $b(p)$ is increasing for $p > 0.9$, the top 0.1% gets a larger fraction of the income of the top 1% than the top 1% does for the top 10%, so that the top 1% is more unequal than the top 10%. Different profiles of generalized Pareto curves can be linked to a typology suggested by Atkinson (2017) to characterize deviations from strict Pareto behavior. An increasing $b(p)$ near $p = 1$ corresponds to what Atkinson describes as a "regal" distribution, whereas a decreasing $b(p)$ correspond to a "baronial" one. As we see below, the profiles observed for the United States and France over the 1962–2014 period belong to the first category.

[4]For example $b(p) = 3 + \sin(\log(1 - p))$.

2.3 Pareto Curves in Finite Samples

We now consider a sample $(X_1, \ldots, X_n)$ of n iid. copies of X. We write $X_{(r)}$ the r-th order statistic (i.e. the r-th largest value). The natural estimator of the inverted Pareto coefficient may be written:

$$\hat{b}_n(p) = \frac{1}{(n - \lfloor np \rfloor)X_{(\lfloor np \rfloor + 1)}} \sum_{k=\lfloor np \rfloor + 1}^{n} X_{(k)}$$

For $(n-1)/n \leq p < 1$, we have $\hat{b}_n(p) = 1$ regardless of the distribution of X. More generally, when p gets too close to one, it becomes severely biased toward one. This is a pure finite sample effect, due to the fact that here, p depends on n (the sample quantile, for example, raises similar issues). But it means that the asymptotic value of $b(p)$ cannot be directly estimated from the data. That result is not surprising, as the plug-in approach cannot be used to estimate purely asymptotic quantities. In practice, we need to stop looking at inverted Pareto coefficients before p gets too close to one, depending on the sample size.

Yet there is a fundamental difference between cases where the final fall of the Pareto curve toward one is a pure artifact of finite samples, and cases where it reflects a property of the underlying distribution. With the former, the fall will happen for increasingly large p as the sample size increases. (With fiscal data, for which samples are extremely large, we need not be concerned by the problem until extremely narrow top income groups.) With the latter, the fall should always happen at the same point of the distribution.
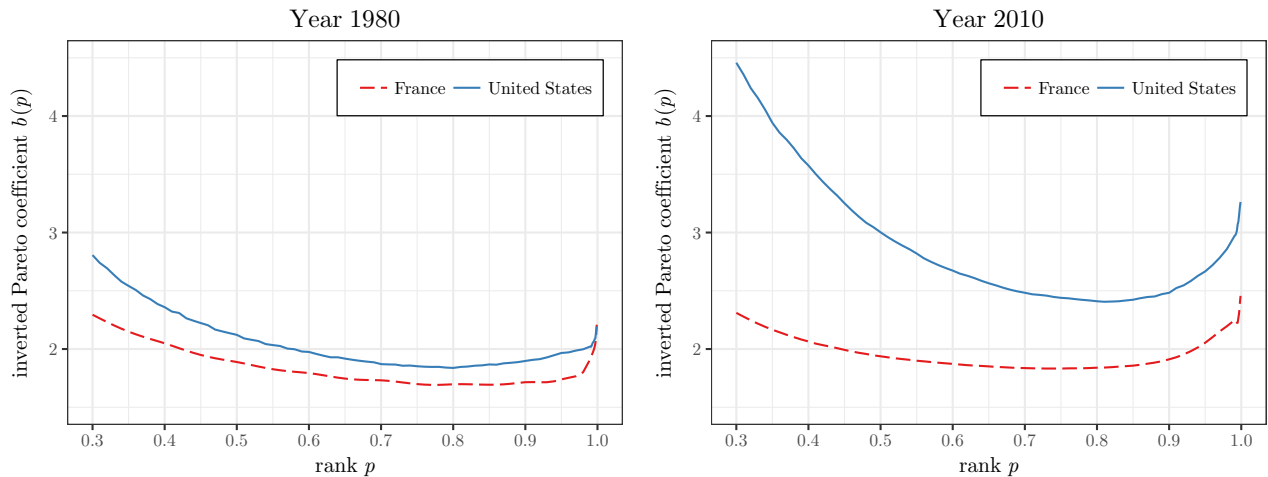
Sources: Piketty, Saez, and Zucman (2016) (United States),
Garbinti, Goupille-Lebret, and Piketty (2016) (France).

FIGURE 1: Generalized Pareto curves of pre-tax national income

Figure 1 shows the empirical generalized Pareto curves for the distribution of pre-tax national

income in France and in the United States in 1980 and 2010, based on quasi-exhaustive income tax data. The curve has changed a lot more in the United States than in France, which reflects the well-known increase in inequality that the United States has experienced over the period. In 2010, the inverted Pareto coefficients are much higher in the United States than in France, which means that the tail is fatter, and the income distribution more unequal.

In both countries, $b(p)$ does appear to converge toward a value strictly above one, which confirms that the distribution of income is an asymptotic power law. However, the coefficients vary significantly, even within the top decile, so that the strict Pareto assumption will miss important patterns in the distribution. Because $b(p)$ rises within the top 10% of the distribution, inequality in both France and the United States is in fact even more skewed toward the very top than what the standard Pareto model suggests. And the amount by which inverted Pareto coefficients vary is not negligible. For the United States, in 2010, at its lowest point (near $p = 80\%$), $b(p)$ is around 2.4. If it were a strict Pareto distribution, it would correspond to the top 1% owning 15% of the income. But the asymptotic value is closer to 3.3, which would mean a top 1% share of 25%.

Though empirical evidence leads us to reject the strict Pareto assumption, we can notice a different empirical regularity: the generalized Pareto curves are U-shaped. We observe that fact for all countries and time periods for which we have sufficient data. This is noteworthy, because even if we know that $b(p)$ is not strictly constant, the fact that it does not monotonically converge toward a certain value is not what the most simple models of the income distribution would predict. In section 6, we show how it can be connected to the properties of the underlying income generating mechanism.

2.4 Other Concepts of Local Pareto Coefficients

The inverted Pareto coefficient $b(p)$ is not the only local concept of Pareto coefficient that can be used to nonparametrically describe power law behavior. Using a simple principle, we can in fact define an infinite number of such coefficients, some of which have already been introduced in the literature (eg. Gabaix, 1999). First, notice that if $G(x) = 1 - F(x) = Cx^{-\alpha}$ is a strict power law, then for $n > 0$:

$$-\frac{xG^{(n)}(x)}{G^{(n-1)}(x)} - n + 1 = \alpha \tag{1}$$

which does not depend on x or n. But when the distribution isn't strictly Paretian, we can always define $\alpha_n(x)$ equal to the left-hand side of (1), which may now depend on x and n.

For example $\alpha_2(x)$ correspond to the "local Zipf coefficient" as defined by Gabaix (1999).[5] For $n = 1$, we get $\alpha_1(x) = xf(x)/(1 - F(x))$. As long as $\alpha > -n+1$, we can also extend formula (1) to zero or negative n, substituting integrals for negative orders of differentiation. More precisely, we set:

$$\forall n < 0 \qquad G^{(n)}(x) = (-1)^n \underbrace{\int_x^{+\infty} \cdots \int_{t_2}^{+\infty}}_{|n| \text{ times}} G(t_1)\, \mathrm{d}t_1 \ \ldots \ \mathrm{d}t_{|n|}$$

The definition above ensures that $G^{(n_1)(n_2)} = G^{(n_1+n_2)}$ for all $n_1, n_2 \in \mathbb{Z}$. We call $\alpha_n(x)$, $n \in \mathbb{Z}$, the *local Pareto coefficient* of order n. We have for $n = 0$:

$$\alpha_0(x) = 1 + \frac{x(1 - F(x))}{\int_x^{+\infty} 1 - F(t)\, \mathrm{d}t}$$

which implies:[6]

$$b(p) = \frac{\alpha_0(x)}{\alpha_0(x) - 1}$$

That formula corresponds to the inverted Pareto coefficient for a strict Pareto distribution $b = \alpha/(\alpha - 1)$. In fact, $b(p)$ is an alternative way of writing $\alpha_0(x)$, with a clearer interpretation in terms of economic inequality. We could similarly define inverted Pareto coefficients $b_n(p) = \alpha_n(x)/(\alpha_n(x) - 1)$ for any order n, and $b(p) = b_0(p)$. But $b_0(p)$ has the advantage of being the most simple to estimate, because it only involves quantiles and averages. Other estimators require estimating the density or one of its successive derivatives, which is much harder, especially when we have limited access to data (see section 3).[7]

The most natural members of the family of local Pareto coefficients are α_0, α_1 and α_2 (other involve many orders of differentiation or integration). Figure 2 shows how these different coefficients compare for the 2010 distribution of pre-tax national income in the United States. There are some differences regarding the range of values taken by the different coefficients. The inverted U-shape is less pronounced for α_1 and α_2 than α_0. But we reach similar conclusions regardless of the one we pick: the coefficient is not constant (including at the very top) and there is an inversion of the slope near the top of the distribution. All these coefficients have fairly regular shapes, and they convey more information about the tail than solely looking at the quantile or the Lorenz curve. This is why it is better to work directly with them rather than with quantiles or shares.

---

[5] Which we can write more simply as $\alpha_2(x) = -xf'(x)/f(x) - 1$.

[6] See lemma 1 in appendix C.1.4.

[7] We can move from one coefficient to the next using the following recurrence relation:

$$\alpha_{n+1}(x) = \alpha_n(x) - \frac{x\alpha_n'(x)}{\alpha_n(x) + n - 1}$$

10
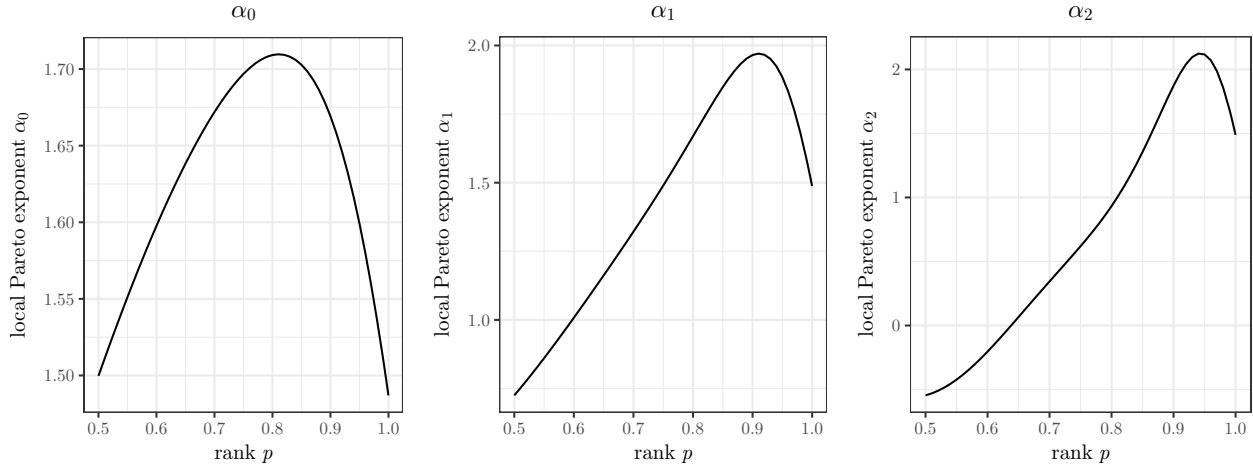
Distribution of pre-tax national income in the United States, 2010. α_0 estimated fitting a polynomial of degree 5 on empirical data. Source: authors' calculations using Piketty, Saez, and Zucman (2016)

FIGURE 2: Different concepts of local Pareto exponent

3 Generalized Pareto Interpolation

The tabulations of income or wealth such as those provided by tax authorities and national statistical institutes typically take the form of K fractiles $0 \leq p_1 < \cdots < p_K < 1$ of the population, alongside their income quantiles $q_1 < \cdots < q_K$ and the income share of each bracket $[p_k, p_{k+1}]$.[8] The interpolation method that we now present estimates a complete generalized Pareto curve $b(p)$ based solely on that information: we call it *generalized Pareto interpolation.*

The tabulations let us compute $b(p_1), \ldots, b(p_K)$ directly. But interpolating the curve $b(p)$ based solely on those points offers no guarantee that the resulting function will be consistent with the input data on quantiles. To that end, the interpolation needs to be constrained. To do so in a computationally efficient and analytically tractable way, we start from the following function:

$$\forall x \geq 0 \qquad \varphi(x) = -\log \int_{1-\mathrm{e}^{-x}}^{1} Q(p) \, \mathrm{d}p$$

which is essentially a transform of the Lorenz curve:

$$\varphi(x) = -\log((1 - L(p))\mathbb{E}[X])$$

with $p = 1 - \mathrm{e}^{-x}$. The value of φ at each point $x_k = -\log(1 - p_k)$ can therefore be estimated

---

[8] That last element may take diverse forms (top income shares, bottom income shares, average income in the brackets, average income above the bracket, etc.), all of which are just different ways of presenting the same information.

directly from the data in the tabulation. Moreover:

$$\forall x \geq 0 \qquad \varphi'(x) = e^{\varphi(x)-x} Q(1 - e^{-x}) = 1/b(1 - e^{-x})$$

which means that the generalized Pareto coefficient $b(p)$ is equal to $1/\varphi'(x)$. Hence, the value of $\varphi'(x_k)$ for $k \in \{1, \ldots, K\}$ is also given by the tabulation.

Because of the bijection between $(p, b(p), Q(p))$ and $(x, \varphi(x), \varphi'(x))$, the problem of interpolating $b(p)$ in a way that is consistent with $Q(p)$ is identical to that of interpolating the function φ, whose value and first derivative are known at each point x_k.

3.1 Interpolation Method

For simplicity, we will set aside sampling related issues in the rest of this section, but we will come back to them in section 5. We assume that we know a set of points $\{(x_k, y_k, s_k), 1 \leq k \leq K\}$ that correspond to the values of $\{(x_k, \varphi(x_k), \varphi'(x_k)), 1 \leq k \leq K\}$, and we seek a sufficiently smooth function $\hat{\varphi}$ such that:

$$\forall k \in \{1, \ldots, K\} \qquad \hat{\varphi}(x_k) = \varphi(x_k) = y_k \qquad \hat{\varphi}'(x_k) = \varphi'(x_k) = s_k \qquad (2)$$

By sufficiently smooth, we mean that φ should be at least twice continuously differentiable. That requirement is necessary for the estimated Pareto curve (and by extension the quantile function) to be once continuously differentiable, or, put differently, not to exhibit any asperity at the fractiles included in the tabulation.

Our interpolation method relies on splines, meaning piecewise polynomials defined on each interval $[x_k, x_{k+1}]$. Although cubic splines (i.e. polynomials of degree 3) are the most common, they do not offer enough degrees of freedom to satisfy both the constraints given by (2) and the requirement that φ is twice continuously differentiable. We use quintic splines (i.e. polynomials of degree 5) to get more flexibility. To construct them, we start from the following set of polynomials for $x \in [0, 1]$:

$$h_{00}(x) = 1 - 10x^3 + 15x^4 - 6x^5 \qquad\qquad h_{01}(x) = 10x^3 - 15x^4 + 6x^5$$
$$h_{10}(x) = x - 6x^3 + 8x^4 - 3x^5 \qquad\qquad h_{11}(x) = -4x^3 + 7x^4 - 3x^5$$
$$h_{20}(x) = \tfrac{1}{2}x^2 - \tfrac{3}{2}x^3 + \tfrac{3}{2}x^4 - \tfrac{1}{2}x^5 \qquad\qquad h_{21}(x) = \tfrac{1}{2}x^3 - x^4 + \tfrac{1}{2}x^5$$

which were designed so that $h_{ij}^{(k)}(\ell) = 1$ if $(i, j) = (k, \ell)$, and 0 otherwise. They are analogous to the basis of cubic Hermite splines (e.g. McLeod and Baart, 1998, p. 328), but for the set of

polynomials of degree up to five. Then, for $k \in \{1, \ldots, K-1\}$ and $x \in [x_k, x_{k+1}]$, we set:

$$\hat{\varphi}_k(x) = y_k h_{00}\left(\frac{x - x_k}{x_{k+1} - x_k}\right) + y_{k+1} h_{01}\left(\frac{x - x_k}{x_{k+1} - x_k}\right)$$
$$+ s_k(x_{k+1} - x_k)h_{10}\left(\frac{x - x_k}{x_{k+1} - x_k}\right) + s_{k+1}(x_{k+1} - x_k)h_{11}\left(\frac{x - x_k}{x_{k+1} - x_k}\right)$$
$$+ a_k(x_{k+1} - x_k)^2 h_{20}\left(\frac{x - x_k}{x_{k+1} - x_k}\right) + a_{k+1}(x_{k+1} - x_k)^2 h_{21}\left(\frac{x - x_k}{x_{k+1} - x_k}\right)$$

for some $a_k, a_{k+1} \in \mathbb{R}$, and $\hat{\varphi}(x) = \hat{\varphi}_k(x)$ for $x \in [x_k, x_{k+1}]$. By construction, we have $\hat{\varphi}(x_k) = y_k$, $\hat{\varphi}(x_k) = y_{k+1}$, $\hat{\varphi}'(x_k) = s_k$, $\hat{\varphi}'(x_{k+1}) = s_{k+1}$, $\hat{\varphi}''(x_k) = a_k$ and $\hat{\varphi}''(x_{k+1}) = a_{k+1}$. Hence, $\hat{\varphi}$ satisfies all the constraints and regularity requirements of the problem.

To pick appropriate values for $a_1, \ldots, a_k$, we follow the usual approach of imposing additional regularity conditions at the jointures. We have a system of $K - 2$ equations, linear in $a_1, \ldots, a_k$, defined by:

$$\forall k \in \{2, \ldots, K-1\} \qquad \hat{\varphi}'''_{k-1}(x_k) = \hat{\varphi}'''_k(x_k)$$

Two additional equations are required for that system to have a unique solution. One solution is to use predetermined values for a_1 and a_K (known as the "clamped spline"). Another, known as the "natural spline", sets:

$$\hat{\varphi}'''_1(x_1) = 0 \qquad \text{and} \qquad \hat{\varphi}'''_{K-1}(x_K) = 0$$

Both approaches are equivalent to the minimization of an irregularity criterion (e.g. Lyche and Mørken, 2002):

$$\min_{a_1, \ldots, a_K} \int_{x_1}^{x_k} \{\hat{\varphi}'''(x)\}^2 \, \mathrm{d}x$$

subject to fixed values for a_1 and a_K (clamped spline) or not (natural spline). Hence, both methods can be understood as a way to minimize the curvature of $\hat{\varphi}'$, which is itself an immediate transform of $b(p)$. That is, by construction, the method aims at finding the most "regular" generalized Pareto curve that satisfies the constraints of the problem.

We adopt a hybrid approach, in which a_1 is determined through $\hat{\varphi}'''_1(x_1) = 0$, but where a_K is estimated separately using the two-points finite difference:

$$a_K = \frac{s_K - s_{K-1}}{x_K - x_{K-1}}$$

Because the function is close to linear near x_K, it yields results that are generally similar to traditional natural splines. But that estimation of $\varphi''(x_K)$ is also more robust, so we get more satisfactory results when the data exhibit potentially troublesome features.

The vector $\boldsymbol{a} = [a_1 \quad \cdots \quad a_K]'$ is the solution of a linear system of equation $\boldsymbol{X}\boldsymbol{a} = \boldsymbol{v}$, where $\boldsymbol{X}$ depends solely on the $x_1, \ldots, x_K$, and $\boldsymbol{b}$ is linear in $y_1, \ldots, y_K$ and $s_1, \ldots, s_K$. Therefore, we find the right parameters for the spline by numerically solving a linear system of equation. We provide the detailed expressions of $\boldsymbol{X}$ and $\boldsymbol{b}$ in appendix.

3.2 Enforcing Admissibility Constraints

The interpolation method presented above does not guarantee that the estimated generalized Pareto curve will satisfy property 1 — or equivalently that the quantile will be an increasing function. In most situations, that constraint need not be enforced, because it is not binding: the estimated function spontaneously satisfy it. But it may occasionally not be the case, so that estimates of quantiles of averages at different points of the distribution may be mutually inconsistent. To solve that problem, we present an *ex post* adjustment procedure which constrains appropriately the interpolated function.

We can express the quantile as a function of φ:

$$\forall x \geq 0 \qquad Q(1 - e^{-x}) = e^{x - \varphi(x)} \varphi'(x)$$

Therefore:

$$\forall x \geq 0 \qquad Q'(1 - e^{-x}) = e^{2x - \varphi(x)}[\varphi''(x) + \varphi'(x)(1 - \varphi'(x))]$$

So the estimated quantile function is increasing if and only if:

$$\forall x \geq 0 \qquad \Phi(x) = \hat{\varphi}''(x) + \hat{\varphi}'(x)(1 - \hat{\varphi}'(x)) \geq 0 \tag{3}$$

The polynomial Φ (of degree 8) needs to be positive. There are no simple necessary and sufficient conditions on the parameters of the spline that can ensure such a constraint. However, it is possible to derive conditions that are only sufficient, but general enough to be used in practice. We use conditions based on the Bernstein representation of polynomials, as derived by Cargo and Shisha (1966):

Theorem 1 (Cargo and Shisha, 1966). *Let $P(x) = c_0 + c_1 x_1 + \cdots + c_n x^n$ be a polynomial of degree $n \geq 0$ with real coefficients. Then:*

$$\forall x \in [0,1] \qquad \min_{0 \leq i \leq n} b_i \leq P(x) \leq \max_{0 \leq i \leq n} b_i$$

where:

$$b_i = \sum_{r=0}^{n} c_r \binom{i}{r} \Big/ \binom{n}{r}$$

14

To ensure that the quantile is increasing over $[x_k, x_{k+1}]$ $(1 \leq k < K)$, it is therefore enough to enforce the constraint that $b_i \geq 0$ for all $0 \leq i \leq 8$, where b_i is defined as in theorem 1 with respect to the polynomial $x \mapsto \Phi(x_k + x(x_{k+1} - x_k))$. Those 9 conditions are all explicit quadratic forms in $(y_k, y_{k+1}, s_k, s_{k+1}, a_k, a_{k+1})$, so we can compute them and their derivative easily.

To proceed, we start from the unconstrained estimate from the previous section. We set $a_k = -s_k(1 - s_k)$ for each $1 \leq k \leq K$ if $a_k + s_k(1 - s_k) < 0$, which ensures that condition (3) is satisfied at least at the interpolation points. Then, over each segment $[x_k, x_{k+1}]$, we check whether the condition $\Phi(x) \geq 0$ is satisfied for $x \in [x_k, x_{k+1}]$ using the theorem 1, or more directly by calculating the values of Φ over a tight enough grid of $[x_k, x_{k+1}]$. If so, we move on to next segment. If not, we consider $L \geq 1$ additional points $(x_1^*, \ldots, x_L^*)$ such that $x_k < x_1^* < \cdots < x_L^* < x_{k+1}$, and we redefine the function $\hat{\varphi}_k$ over $[x_k, x_{k+1}]$ as:

$$
\tilde{\varphi}_k(x) = \begin{cases} \varphi_0^*(x) & \text{if} \quad x_k \leq x < x_1^* \\ \varphi_\ell^*(x) & \text{if} \quad x_\ell^* \leq x < x_{\ell+1}^* \\ \varphi_L^*(x) & \text{if} \quad x_L^* \leq x < x_{k+1} \end{cases}
$$

where the φ_ℓ^* $(0 \leq \ell \leq L)$ are quintic splines such that for all $1 \leq \ell < L$:

$$
\begin{array}{lll}
\varphi_0^*(x_k) = y_k & (\varphi_0^*)'(x_k) = s_k & (\varphi_0^*)''(x_k) = a_k \\
\varphi_L^*(x_{k+1}) = y_{k+1} & (\varphi_L^*)'(x_{k+1}) = s_{k+1} & (\varphi_L^*)''(x_{k+1}) = a_{k+1} \\
\varphi_\ell^*(x_\ell^*) = y_\ell^* & (\varphi_\ell^*)'(x_\ell^*) = s_\ell^* & (\varphi_\ell^*)''(x_\ell^*) = a_\ell^* \\
\varphi_\ell^*(x_{\ell+1}^*) = y_{\ell+1}^* & (\varphi_\ell^*)'(x_{\ell+1}^*) = s_{\ell+1}^* & (\varphi_\ell^*)''(x_{\ell+1}^*) = a_{\ell+1}^*
\end{array}
$$

and $y_\ell^*, s_\ell^*, a_\ell^*$ $(1 \leq \ell \leq L)$ are parameters to be adjusted. In simpler terms, we divided the original spline into several smaller ones, thus creating additional parameters that can be adjusted to enforce the constraint. We set the parameters $y_\ell^*, s_\ell^*, a_\ell^*$ $(1 \leq \ell \leq L)$ by minimizing the L^2 norm between the constrained and the unconstrained estimate, subject to the $9 \times (L+1)$ conditions that $b_i^\ell \geq 0$ for all $0 \leq i \leq 8$ and $0 \leq \ell \leq L$:

$$
\min_{\substack{y_\ell^*, s_\ell^*, a_\ell^* \\ 1 \leq \ell \leq L}} \int_{x_k}^{x_{k+1}} \{\hat{\varphi}_k(x) - \tilde{\varphi}_k(x)\}^2 \, \mathrm{d}x \qquad \text{st.} \qquad b_i^\ell \geq 0 \quad (0 \leq i \leq 8 \text{ and } 0 \leq \ell \leq L)
$$

where the b_i^ℓ are defined as in theorem 1 for each spline ℓ. The objective function and the constraints all have explicit analytical expressions, and so does their gradients. We solve the problem with standard numerical methods for nonlinear constrained optimization.[9,10]

---

[9]For example, standard sequential quadratic programming (Kraft, 1994) or augmented Lagrangian methods (Conn, Gould, and Toint, 1991; Birgin and Martìnez, 2008). See NLopt for details and open source implementations of such algorithms: `http://ab-initio.mit.edu/wiki/index.php/NLopt_Algorithms`.

[10]Adding one point at the middle of the interval is usually enough to enforce the constraint, but more points

3.3 Extrapolation Beyond the Last Bracket

The interpolation procedure only applies to fractiles between p_1 and p_K, but we may want an estimate of the distribution outside of this range, especially for $p > p_K$.[11] Because there is no direct estimate of the asymptotic Pareto coefficient $\lim_{p \to 1} b(p)$, it is not possible to interpolate as we did for the rest of the distribution: we need to extrapolate it.

The extrapolation in the last bracket should satisfy the constraints imposed by the tabulation (on the quantile and the mean). It should also ensure derivability of the quantile function at the juncture. To do so, we can use the information contained in the four values (x_K, y_K, s_K, a_K) of the interpolation function at the last point.

Hence, we need a reasonable functional form for the last bracket with enough degrees of freedom to satisfy all the constraints. To that end, we turn to the generalized Pareto distribution.

Definition 1 (Generalized Pareto distribution). *Let $\mu \in \mathbb{R}$, $\sigma \in]0, +\infty[$ and $\xi \in \mathbb{R}$. X follows a generalized Pareto distribution if for all $x \geq \mu$ ($\xi \geq 0$) or $\mu \leq x \leq \mu - \sigma/\xi$ ($\xi < 0$):*

$$\mathbb{P}\{X \leq x\} = \mathrm{GPD}_{\mu,\sigma,\xi}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0 \\ 1 - e^{-(x-\mu)/\sigma} & \text{for } \xi = 0 \end{cases}$$

μ is called the location parameter, σ the scale parameter and ξ the shape parameter.

The generalized Pareto distribution is a fairly general family which includes as special cases the strict Pareto distribution ($\xi > 0$ and $\mu = \sigma/\xi$), the (shifted) exponential distribution ($\xi = 0$) and the uniform distribution ($\xi = -1$). It was popularized as a model of the tail of other distributions in extreme value theory by Pickands (1975) and Balkema and Haan (1974), who showed that for a large class of distributions (which includes all power laws in the sense of definition 2), the tail converges towards a generalized Pareto distribution.

If $X \sim \mathrm{GPD}(\mu, \sigma, \xi)$, the generalized Pareto curve of X is:

$$b(p) = 1 + \frac{\xi\sigma}{(1-\xi)[\sigma + (1-p)^\xi(\mu\xi - \sigma)]}$$

We will focus on cases where $0 < \xi < 1$, so that the distribution is a power law at the limit ($\xi > 0$), but its mean remains finite ($\xi < 1$). When $\xi\mu = \sigma$, the generalized Pareto curve is constant, and the distribution is a strict power law with Pareto coefficient $b = 1/(1 - \xi)$. That value also corresponds in all cases to the asymptotic coefficient $\lim_{p \to 1} b(p) = 1/(1 - \xi)$. But there are several ways for the distribution to converge toward a power law, depending on the

---

may be added if convergence fails.

[11]It is always possible to set $p_1 = 0$ if the distribution has a finite lower bound.

16

sign of $\mu\xi - \sigma$. When $\mu\xi - \sigma > 0$, $b(p)$ converges from below, increasing as $p \to 1$, so that the distribution gets more unequal in higher brackets. Conversely, when $\mu\xi - \sigma < 0$, $b(p)$ converges from above, and decreases as $p \to 1$, so that the distribution is more equal in higher brackets.

The generalized Pareto distribution can match a wide diversity of profiles for the behavior of $b(p)$. We can use the information at our disposal on the mean and the threshold of the last bracket, plus a regularity condition to identify which of these profiles is the most suitable. The generalized Pareto distribution offers a way to extrapolate the coefficients $b(p)$ in a way that is consistent with all the input data and preserves the regularity of the Pareto curve.

We assume that, for $p > p_K$, the distribution follows a generalized Pareto distribution with parameters (μ, σ, ξ), which means that for $q > q_K$ the CDF is:

$$F(q) = p_K + (1 - p_K)\mathrm{GPD}_{\mu,\sigma,\xi}(q)$$

For the CDF to remain continuous and differentiable, we need $\mu = q_K$ and $\sigma = (1-p_K)/F'(q_K)$, where $F'(q_K)$ comes from the interpolation method of section 3.1. Finally, for the Pareto curve to remain continuous, we need $b(p_K)$ equal to $1 + \sigma/(\mu(1-\xi))$, which gives the value of ξ. That is, if we set the parameters (μ, σ, ξ) equal to:

$$\mu = s_K \mathrm{e}^{x_K - y_K}$$
$$\sigma = (1 - p_K)(a_K + s_K(1 - s_K))\mathrm{e}^{2x_K - y_K}$$
$$\xi = 1 - \frac{(1 - p_K)\sigma}{\mathrm{e}^{-y_K} - (1 - p_K)\mu}$$

then the resulting distribution will have a continuously differentiable quantile function, and will match the quantiles and the means in the tabulation.

4   Tests Using Income Data from the United States and France, 1962–2014

We test the quality of our interpolation method using income tax data for the United States (1962–2014) and France (1994–2012).[12] They correspond to cases for which we have detailed tabulations of the distribution of pre-tex national income based on quasi-exhaustive individual tax data (Piketty, Saez, and Zucman, 2016; Garbinti, Goupille-Lebret, and Piketty, 2016), so that we can know quantiles or shares exactly.[13] We compare the size of the error in generalized Pareto interpolation with alternatives most commonly found in the literature.

---

[12]More precisely, the years 1962, 1964 and 1966–2014 for the United States.

[13]We use pre-tax national income as our income concept of reference. It was defined by Alvaredo et al. (2016) to be consistent with the internationally agreed definition of net national income in the system of national

4.1 Overview of Existing Interpolation Methods

Method 1: constant Pareto coefficient That method was used by Piketty (2001) and Piketty and Saez (2003), and relies on the property that, for a Pareto distribution, the inverted Pareto coefficient $b(p)$ remains constant. We set $b(p) = b = \mathbb{E}[X|X > q_k]/q_k$ for all $p \geq p_k$. The p-th quantile becomes $q = q_k \left(\frac{1-p}{1-p_k}\right)^{-1/\alpha}$ with $\alpha = b/(b-1)$. By construction, $\mathbb{E}[X|X > q] = bq$ which gives the p-th top average and top share.

Method 2: log-linear interpolation The log-linear interpolation method was introduced by Pareto (1896), Kuznets (1953), and Feenberg and Poterba (1992). It uses solely threshold information, and relies on the property of Pareto distributions that $\log(1 - F(x)) = \log(c) - \alpha \log(x)$. We assume that this relation holds exactly within the bracket $[p_k, p_{k+1}]$, and set $\alpha_k = -\frac{\log((1-p_{k+1})/(1-p_k))}{\log(q_{k+1}/q_k)}$. The value of the p-th quantile is again $q = q_k \left(\frac{1-p}{1-p_k}\right)^{-1/\alpha_k}$ and the top averages and top shares can be obtained by integration of the quantile function. For $p > p_K$, we extrapolate using the value α_K of the Pareto coefficient in the last bracket.

Method 3: mean-split histogram The mean-split histogram uses information on both the means and the thresholds, but uses a very simple functional form, so that the solution can be expressed analytically. Inside the bracket $[q_k, q_{k+1}]$, the density takes two values:

$$f(x) = \begin{cases} f_k^- & \text{if} \quad q_k \leq x < \mu_k \\ f_k^+ & \text{if} \quad \mu_k \leq x < q_{k+1} \end{cases}$$

where μ_k is the mean inside the bracket.[14] To meet the requirement on the mean and the thresholds, we set:

$$f_k^- = \frac{(p_{k+1} - p_k)(q_{k+1} - \mu_k)}{(q_{k+1} - q_k)(\mu_k - q_k)} \qquad \text{and} \qquad f_k^+ = \frac{(p_{k+1} - p_k)(\mu_k - q_k)}{(q_{k+1} - q_k)(q_{k+1} - \mu_k)}$$

The means-split histogram does not apply beyond the last threshold of the tabulation.

Comparison Methods 1 and 2 make a fairly inefficient use of the information included in the original tabulation: method 1 discards the data on quantiles and averages at the higher end of

---

accounts. Even though they are mostly based on individual tax data, estimates of pre-tax national income do involves a few corrections and imputations, which may affect the results. That is why we also report similar computations in appendix using fiscal income, which is less comparable and less economically meaningful, but doesn't suffer from such problems.

[14]The breakpoint of the interval $[q_k, q_{k+1}]$ could be different from μ_k, but not all values between q_k and q_{k+1} will work if we want to make sure that $f_k^- > 0$ and $f_k^+ > 0$. The breakpoint q^* must be between q_k and $2\mu_k - q_k$ if $\mu_k < (q_k + q_{k+1})/2$, and between $2\mu_k - q_{k+1}$ and q_{k+1} otherwise. Choosing $q^* = \mu_k$ ensures that the condition is always satisfied.

the bracket, while method 2 discards the information on averages. As a consequence, none of these methods can guarantee that the output will be consistent with the input. The method 3 does offer such a guarantee, but with a very simple — and quite unrealistic — functional form.

Our generalized Pareto interpolation method makes use of all the information in the tabulation, so that its output is guaranteed to be consistent with its input. Moreover, contrary to all other methods, it leads a continuous density, hence a smooth quantile and a smooth Pareto curve. None of the other methods can satisfy this requirement, and their output exhibit stark irregularities that depend on the position of the brackets in the tabulation in input.

Application to France and the United States Using the individual income tax data, we compute our own tabulations in each year. We include four percentiles in the tabulation: $p_1 = 0.1$, $p_2 = 0.5$, $p_3 = 0.9$ and $p_4 = 0.99$.

We interpolate each of those tabulations with the three methods above, labelled "M1", "M2" and "M3" in what follows.[15] We also interpolate them with our new generalized Pareto interpolation approach (labeled "M0"). We compare the values that we get with each method for the top shares and the quantiles at percentiles 30%, 75% and 95% with the value that we get directly from the individual data. (We divide all quantiles by the average to get rid of scaling effects due to inflation and average income growth.) We report the mean relative error in table I:

$$\text{MRE} = \frac{1}{\text{number of years}} \sum_{t=\text{first year}}^{\text{last year}} \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

where y is the quantity of interest (income threshold or top share), and $\hat{y}$ is its estimate using one of the interpolation methods.

The two standard Pareto interpolation methods (M1 and M2) are the ones that perform worst. M1 is better at estimating shares, while M2 is somewhat better at estimating quantiles. That shows the importance not to dismiss any information included in the tabulation, as exhibited by the good performance of the mean-split histogram (M3), particularly at the bottom of the distribution.

Our generalized Pareto interpolation method vastly outperforms the standard Pareto interpolation methods (M1 and M2). It is also much better than the mean-split histogram (M3), except in the bottom of the distribution where both methods work well (but standard Pareto methods M1 and M2 fail badly).

---

[15]We also provide extended tables in appendix with a fourth method, which is much more rarely used.
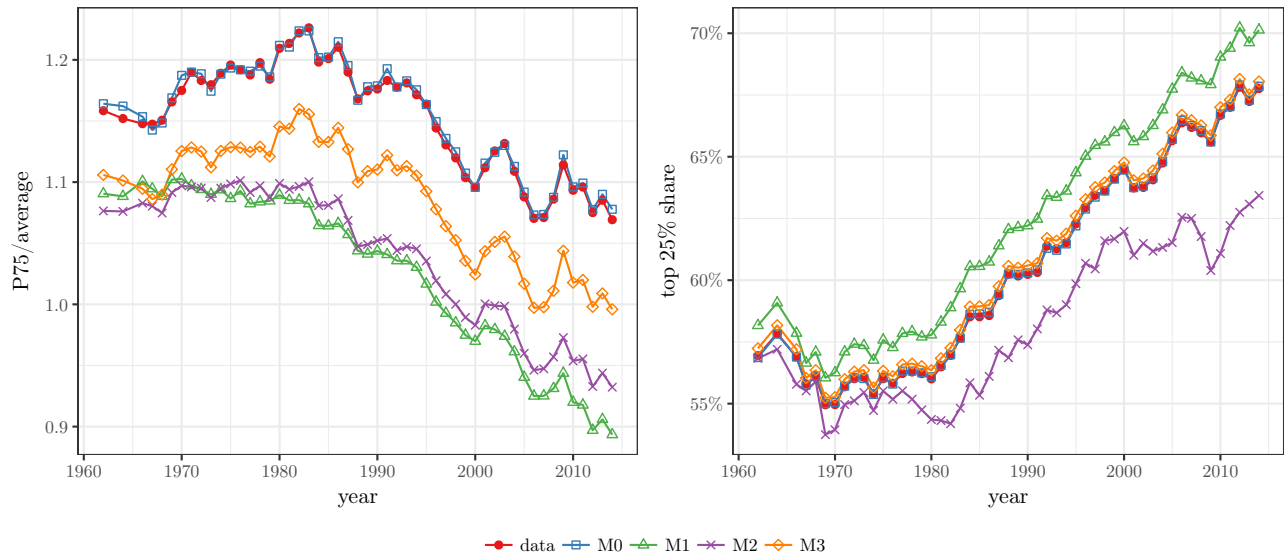
TABLE I: MEAN RELATIVE ERROR FOR DIFFERENT INTERPOLATION METHODS

		mean percentage gap between estimated and observed values			
		M0	M1	M2	M3
United States (1962–2014)	Top 70% share	0.059% (ref.)	2.3% (×38)	6.4% (×109)	0.054% (×0.92)
	Top 25% share	0.093% (ref.)	3% (×32)	3.8% (×41)	0.54% (×5.8)
	Top 5% share	0.058% (ref.)	0.84% (×14)	4.4% (×76)	0.83% (×14)
	P30/average	0.43% (ref.)	55% (×125)	29% (×67)	1.4% (×3.3)
	P75/average	0.32% (ref.)	11% (×35)	9.9% (×31)	5.8% (×18)
	P95/average	0.3% (ref.)	4.4% (×15)	3.6% (×12)	1.3% (×4.5)
France (1994–2012)	Top 70% share	0.55% (ref.)	4.2% (×7.7)	7.3% (×13)	0.14% (×0.25)
	Top 25% share	0.75% (ref.)	1.8% (×2.4)	4.9% (×6.5)	0.37% (×0.49)
	Top 5% share	0.29% (ref.)	1.1% (×3.9)	8.9% (×31)	0.49% (×1.7)
	P30/average	1.5% (ref.)	59% (×40)	38% (×26)	2.6% (×1.8)
	P75/average	1% (ref.)	5.2% (×5.1)	5.4% (×5.3)	4.7% (×4.6)
	P95/average	0.58% (ref.)	5.6% (×9.6)	3.2% (×5.5)	1.8% (×3.2)

Pre-tax national income. Sources: author's calculation from Piketty, Saez, and Zucman (2016) (United States) and Garbinti, Goupille-Lebret, and Piketty (2016) (France). The different interpolation methods are labeled as follows. M0: generalized Pareto interpolation. M1: constant Pareto coefficient. M2: log-linear interpolation. M3: mean-split histogram. We applied them to a tabulation which includes the percentiles $p = 10\%$, $p = 50\%$, $p = 90\%$, and $p = 99\%$. We included the relative increase in the error compared to generalized Pareto interpolation in parentheses. We report the mean relative error, namely:

$$\frac{1}{\text{number of years}} \sum_{t=\text{first year}}^{\text{last year}} \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

where y is the quantity of interest (income threshold or top share), and $\hat{y}$ is its estimate using one of the interpolation methods. We calculated the results over the years 1962, 1964 and 1966–2014 in the United States, and years 1994–2012 in France.

Pre-tax national income. Sources: author's computation from Piketty, Saez, and Zucman (2016). M0: generalized Pareto interpolation. M1: constant Pareto coefficient. M2: log-linear interpolation. M3: mean-split histogram.

FIGURE 3: P75 threshold and top 25% share in the United States (1962–2014), estimated using all interpolation methods and a tabulation with $p = 10\%, 50\%, 90\%, 99\%$
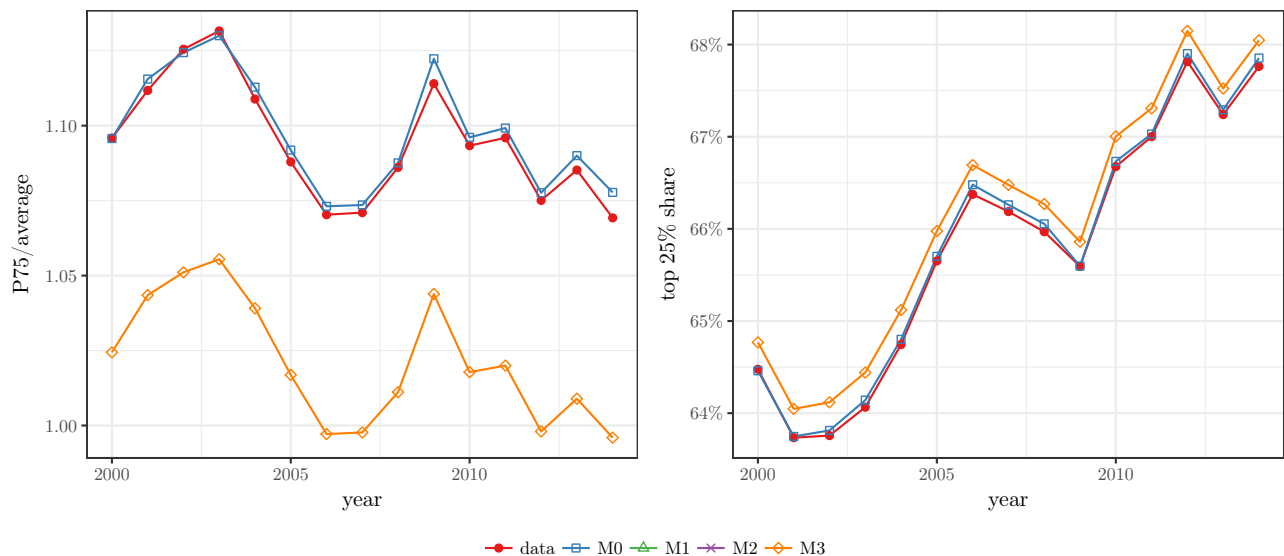
Pre-tax national income. Sources: author's computation from Piketty, Saez, and Zucman (2016). M0: generalized Pareto interpolation. M3: mean-split histogram.

FIGURE 4: P75 threshold and top 25% share in the United States (2000-2014), estimated using interpolation methods M0 and M3, and a tabulation with $p = 10\%, 50\%, 90\%, 99\%$
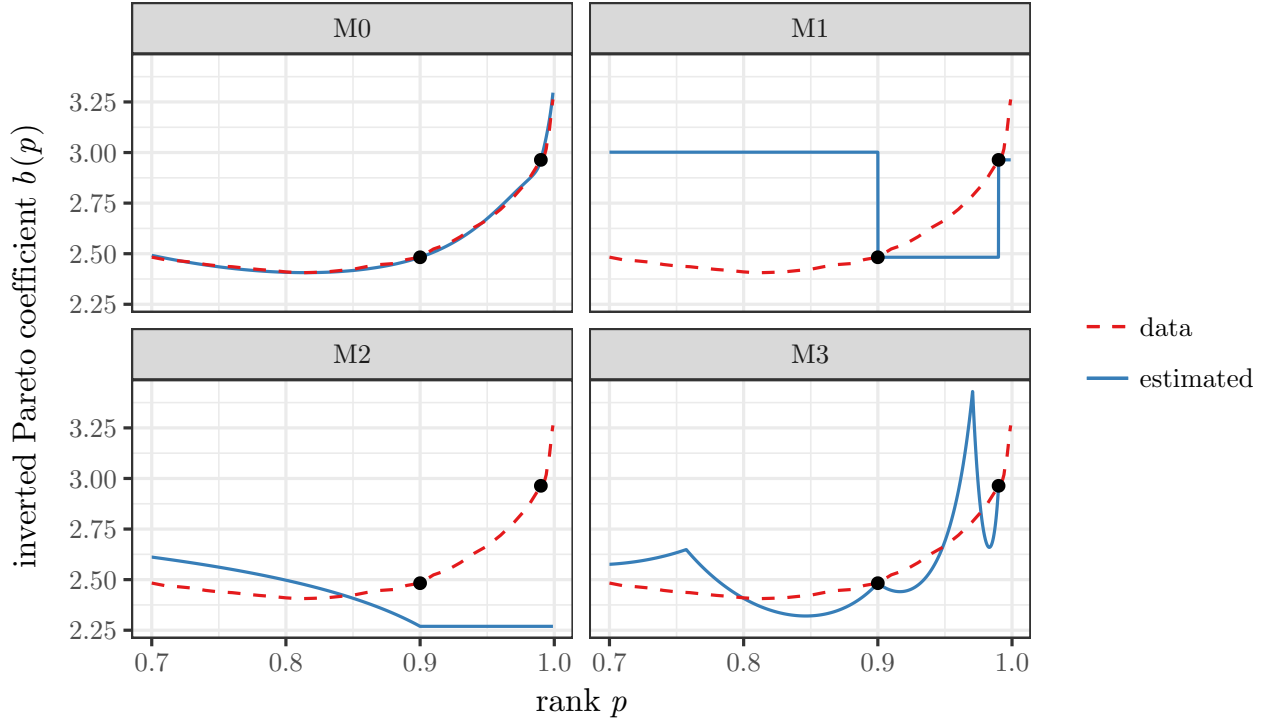
21

Pre-tax national income. Sources: author's computation from Piketty, Saez, and Zucman (2016). M0: generalized Pareto interpolation. M1: constant Pareto coefficient. M2: log-linear interpolation. M3: mean-split histogram.

FIGURE 5: Generalized Pareto curves implied by the different interpolation methods for the United States distribution of income in 2010

Figure 3 shows how the use of different interpolation methods affects the estimation of the top 25% share and associated income threshold. Although all methods roughly respect the overall trend, they can miss the level by a significant margin. The generalized Pareto interpolation estimates the threshold much better than either M1, M2, or M3.

For the estimation of the top 25% share, M3 performs fairly well, unlike M1 and M2. To get a more detailed view, we therefore focus on a more recent period (2000–2014) and display only M0 and M3, as in figure 4. We can see that M3 has, in that case, a tendency to overestimate the top 25% by a small yet persistent amount. In comparison, M4 produces a curve almost identical to the real one.

We can also directly compare the generalized Pareto curves generated by each method, as in figure 5. Our method, M0, reproduces the inverted Pareto coefficients $b(p)$ very faithfully, including above the last threshold (see section 4.2). All the other methods give much worse results. Method M1 leads to discontinuous curve, which in fact may not even define a consistent probability distribution. The M2 method fails to account for the rise of $b(p)$ at the top. Finally,

the M3 leads to an extremely irregular shape due to the use a piecewise uniform distribution to approximate power law behavior.

Overall, the generalized Pareto interpolation method performs well. In most cases, it gives results that are several times better than methods commonly used in the literature. And it does so while ensuring a smoothness of the resulting estimate that no other method can provide. Moreover, it works well for the whole distribution, not just the top (like M1 and M2) or the bottom (like M3).

4.2   Extrapolation methods
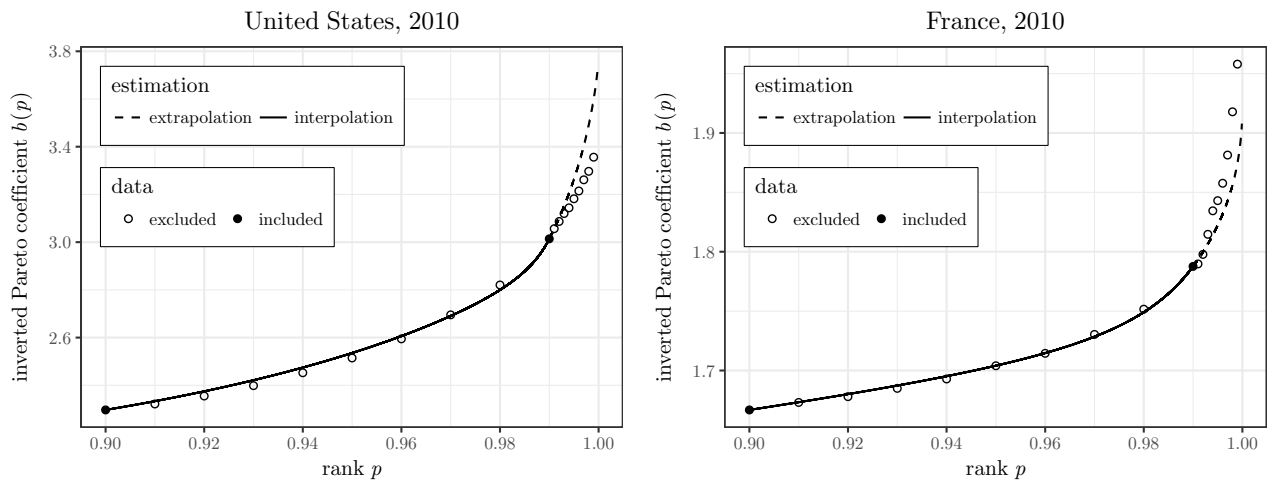
Fiscal income. Sources: author's computation from Piketty, Saez, and Zucman (2016) (for the United States) and Garbinti, Goupille-Lebret, and Piketty (2016) (for France).

FIGURE 6: Extrapolation with generalized Pareto distribution

Of the interpolation methods previously described, only M1 and M2 can be used to extrapolate the tabulation beyond the last threshold. Both assume a standard Pareto distribution. Method M1 estimates $b(p)$ at the last fractile p_K, and assumes a Pareto law with $\alpha = b(p_K)/(b(p_K) - 1)$ after that. Method M2 estimates a Pareto coefficient based on the last two thresholds, so in effect it assumes a standard Pareto distribution immediately after the second to last threshold.

The assumption that $b(p)$ becomes approximately constant for p close to 1, however, is not confirmed by the data. Figure 6 demonstrate this for France and the United States in 2010. The profile of $b(p)$ is not constant for $p \approx 1$. On the contrary, it increases faster than for the rest of the distribution.

In section 3.3 we presented an extrapolation method based on the generalized Pareto distribution that had the advantage of preserving the smoothness of the Pareto curve, use all the information from the tabulation, and allow for a nonconstant profile of generalized Pareto coefficients near the top. As figure 6 shows, this method leads to a more realistic shape of the Pareto curve.

TABLE II: MEAN RELATIVE ERROR ON THE TOP 1% FOR DIFFERENT
EXTRAPOLATION METHODS, KNOWING THE TOP 10% AND THE TOP 5%

		mean percentage gap between estimated and observed values		
		M0	M1	M2
United States (1962–2014)	Top 1% share	0.78% (ref.)	5.2% (×6.7)	40% (×52)
	P99/average	1.8% (ref.)	8.4% (×4.7)	13% (×7.2)
France (1994–2012)	Top 1% share	0.44% (ref.)	2% (×4.6)	11% (×25)
	P99/average	0.98% (ref.)	2.5% (×2.5)	2.4% (×2.4)

Fiscal income. Sources: author's calculation from Piketty, Saez, and Zucman (2016) (United States) and Garbinti, Goupille-Lebret, and Piketty (2016) (France). The different extrapolation methods are labeled as follows. M0: generalized Pareto distribution. M1: constant Pareto coefficient. M2: log-linear interpolation. We applied them to a tabulation which includes the percentiles $p = 90\%$, and $p = 95\%$. We included the relative increase in the error compared to generalized Pareto interpolation in parentheses. We report the mean relative error, namely:

$$\frac{1}{\text{number of years}} \sum_{t=\text{first year}}^{\text{last year}} \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

where y is the quantity of interest (income threshold or top share), and $\hat{y}$ is its estimate using one of the interpolation methods. We calculated the results over the years 1962, 1964 and 1966–2014 in the United States, and years 1994–2012 in France.

Table II compares the performance of the new method with the other ones, as we did in the previous section. Here, the tabulation in input includes $p = 90\%$ but stops at $p = 95\%$, and we seek estimates for $p = 99\%$.[16,17] Method M2 is the most imprecise. Method M1 works quite well in comparison. But our new method M0 gives even more precise results. This because it

---

[16]Here, we use fiscal income instead of pre-tax national income to avoid disturbances created at the top by the imputation of some sources of income in pre-tax national income.

[17]We provide in appendix an alternative tabulation which stops at the top 1% and where we seek the top 0.1%. The performances of M0 and M1 are closer but M0 remains preferable.

can correctly capture the tendency of $b(p)$ to keep on rising at the top of the distribution.

Figure 7 compares the extrapolation methods over time in the United States. We can see M1 overestimates the threshold by about as much as M2 underestimates it, while M0 is much closer to reality and makes no systematic error. For the top share, M1 is much better than M2. But it slightly underestimates the top share because it fails to account for the rising profile of inverted Pareto coefficients at the top, which is why our method M0 works even better.
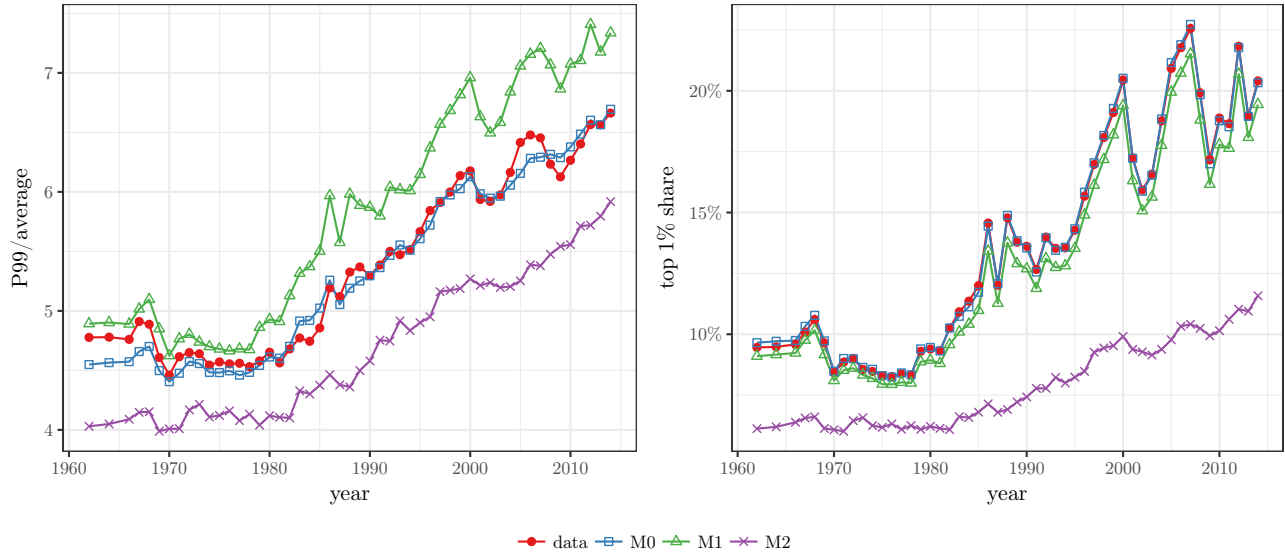
Fiscal income. Sources: author's computation from Piketty, Saez, and Zucman (2016).

FIGURE 7: Comparison of extrapolation methods in the United States for the top 1%, knowing the top 10% and the top 5%

5   Estimation Error

The previous section calculated empirically the precision of our new interpolation method. We did so by systematically comparing estimated values with real ones coming from individual tax data. But whenever we have access to individual data, we do not in fact need to perform any interpolation. So the main concern about the previous section is its external validity. To what extent can its results be extended to different tabulations, with different brackets, corresponding to a different distribution? Is it possible to get estimates of the error in the general case? How many brackets do we need to reach a given precision level, and how should they be distributed?

To the best of our knowledge, none of these issues have been tackled directly in the previous literature. The main difficulty is that most of the error is not due to mere sampling variability (although part of it is), which we can assess using standard methods. It comes mostly from the

discrepancy between the functional forms used in the interpolation, and the true form of the distribution. Put differently, it corresponds to a "model misspecification" error, which is harder to evaluate. But the generalized Pareto interpolation method does offer some solutions to that problem. We can isolate the features of the distribution that determine the error, and based on that provide approximations of it.

In this section, we remain concerned with the same definition of the error as in the previous one. Namely, we consider the difference between the estimate of a quantity by interpolation (e.g. shares or thresholds) and the same quantity defined over the true population of interest. This is in contrast with a different notion of error common in statistics: the difference between an empirical estimate and the value of an underlying statistical model. If sample size were infinite — so that sampling variability would vanish — both errors would be identical. But despite the large samples that characterize tax data, sampling issues cannot be entirely discarded. Indeed, because income and wealth distributions are fat-tailed, the law of large numbers may operate very slowly, so that both types of errors remain different even with millions of observations (Taleb and Douady, 2015).

We consider our definition of the error to be more appropriate in the context of the methods we are studying. Indeed, concerns for the distribution of income and wealth only arise to the extent that it affects actual people, not a probability density function. Moreover, if the distribution of income or wealth in year t depends on the distribution in year $t-1$ (as is the case in models of section 6), then we may need to study the distribution over the finite population to correctly capture all the dynamics at play, not just the underlying random variables from which each population is drawn.

In order to get tractable analytical results, we also focus on the unconstrained interpolation procedure of section 3.1, and thus leave aside the monotonicity constraint of the quantile. That has very little impact on the results in practice since the constraint is rarely binding, and when it is the adjustments are small.[18]

5.1   Theoretical results

Let n be the size of the population (from which the tabulated data come). Recall that $x = -\log(1-p)$. Let $e_n(x)$ be the estimation error on $\varphi_n(x)$, and similarly $e'_n(x)$ the estimation error on $\varphi'_n(x)$. If we know both those errors then we can retrieve the error on any quantity of interest (quantiles, top shares, Pareto coefficients, etc.) by applying the appropriate transforms.

---

[18]For example, the monotonicity constraint is not binding in any of the tabulations interpolated in the previous section.

Our first result decompose the error between two components. Like all the theorems of this section, we give only the main results. Details and proofs are in appendix.

Theorem 5. *We can write $e_n(x) = u(x) + v_n(x)$ and $e'_n(x) = u'(x) + v'_n(x)$ where $u(x), u'(x)$ are deterministic, and $v_n(x), v'_n(x)$ are random variables that converge almost surely to zero when $n \to +\infty$.*

We call the first terms $u(x)$ and $u'(x)$ the "misspecification" error. They correspond to the difference between the functional forms that we use in the interpolation, and the true functional forms of the underlying distribution. Even if the population size was infinite, so that sampling variability was absent, they would still remain nonzero. We can give the following representation for that error.

Theorem 6. *$u(x)$ and $u'(x)$ can be written as a scalar product between two functions ε and φ''':*

$$u(x) = \int_{x_1}^{x_K} \varepsilon(x, t) \varphi'''(t) \, \mathrm{d}t \qquad and \qquad u'(x) = \int_{x_1}^{x_K} \frac{\partial \varepsilon}{\partial x}(x, t) \varphi'''(t) \, \mathrm{d}t$$

where $\varepsilon(x, t)$ is entirely determined by $x_1, \ldots, x_K$.

The function $\varepsilon(x, t)$ is entirely determined by the known values $x_1, \ldots, x_K$, so we can calculate it directly. Its precise definition is given in appendix. The other function, φ''', depends on the quantity we are trying to estimate, so we do not know it exactly. The issue is common in nonparametric statistics, and complicates the application of the formula.[19] But if we look at the value of φ''' in situations where we have enough data to estimate it directly, we can still derive good approximations and rules of thumb that apply more generally.

We call $v_n(x)$ and $v'_n(x)$ the "sampling error". Even if the true underlying distribution matched the functional used for the interpolation, so that there would be no misspecification error, they would remain nonzero. We can give asymptotic approximation of their distribution for large n. We do not only cover the finite variance case ($\mathbb{E}[X^2] < +\infty$), but also the infinite variance case ($\mathbb{E}[X^2] = +\infty$), which leads to results that are less standard. Infinite variance is very common when dealing with distributions of income and wealth.

Theorem 7. *$v_n(x)$ and $v'_n(x)$ converge jointly in distribution at speed $1/r_n$:*

$$r_n \begin{bmatrix} v_n(x) \\ v'_n(x) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{J}$$

---

[19]For example, the asymptotic mean integrated squared error of a kernel estimator depends on the second derivative of the density (Scott, 1992, p. 131).

If $\mathbb{E}[X^2] < +\infty$, then $r_n = \sqrt{n}$ and $\mathcal{J}$ is a bivariate normal distribution. If $\mathbb{E}[X^2] = +\infty$ and $1 - F(x) \sim Cx^{-2}$, then $r_n = (n/\log n)^{1/2}$ and $\mathcal{J}$ is a bivariate normal distribution. If $\mathbb{E}[X^2] = +\infty$ and $1 - F(x) \sim Cx^{-\alpha}$ $(1 < \alpha < 2)$, then $r_n = n^{1-1/\alpha}$ and $\mathcal{J} \overset{\mathcal{D}}{=} (\gamma_1 Y, \gamma_2 Y)$ where Y follows a maximally skewed stable distribution with stability parameter α.

Again, we provide more detailed expressions of the asymptotic distributions in appendix alongside the proof of the result. More importantly, we also show that in practice, we always have $v_n(x) \ll u(x)$ and $v_n'(x) \ll u'(x)$, regardless of the precise characteristics of the underlying distribution. This means that sampling variability is negligible compared to the misspecification error. Therefore, we will apply the result of this section assuming $e_n(x) \approx u(x)$ and $e_n'(x) \approx u'(x)$.

5.2 Applications

5.2.1 Estimation of Error Bounds

TABLE III: OBSERVED MAXIMUM ERROR AND THEORETICAL UPPER BOUND

		maximum absolute error on φ		maximum absolute error on φ'		maximum relative error on top shares	
		actual	bound	actual	bound	actual	bound
United States (1962–2004)	$p = 30\%$	0.0014	0.0030	0.0074	0.0100	0.14%	0.30%
	$p = 75\%$	0.0023	0.0137	0.0048	0.0088	0.23%	1.37%
	$p = 95\%$	0.0020	0.0059	0.0044	0.0077	0.20%	0.59%
France (1994–2002)	$p = 30\%$	0.0054	0.0097	0.0038	0.0231	0.54%	0.97%
	$p = 75\%$	0.0080	0.0208	0.0033	0.0076	0.80%	2.08%
	$p = 95\%$	0.0040	0.0088	0.0060	0.0109	0.40%	0.88%

Given that thee sampling error is negligible, theorem 6 may be used to get bounds on the error in the general case. As an example, imagine that in both France and the United States, we have access to individual data for the most recent ten years, but that we only have access to tabulated data for the years before that. This, in fact, is what happens in the United States (before 1962) and France (before 1970). We can use the envelope of $|\varphi'''|$ over the ten years with individual data as a reasonable upper bound of it for the rest of the period. We write $|\varphi'''(x)| \le M(x)$ for all x. Using the triangular inequality, we get $e_n(x) \le \int_{x_1}^{x_K} |\varepsilon(x, t)| M(t) \, \mathrm{d}t$.

Table III compares the bound on the error calculated as such with reality. The estimates are conservative by construction due to the use of an upper bound for φ''' and the triangular

inequality in the integral. We indeed observe that the theoretical bound is always higher than the actual maximum observed error. Yet in general, the bound that we calculate gives an approximate idea of the error we may expect in each case.

5.2.2 Optimal Choice of Brackets

We now consider the inverse problem: namely, how many brackets do we need to achieve a given precision level, and how should they be placed? Based on theorem 6, we can answer that question for any given φ''' by solving an optimization program. Hence, if we pick a functional form for φ''' which is typical of what we observe, we get the solution of the problem for the typical income distribution.

TABLE IV: OPTIMAL BRACKET CHOICE FOR A TYPICAL DISTRIBUTION OF INCOME

	3 brackets	4 brackets	5 brackets	6 brackets	7 brackets
optimal placement of thresholds	10.0%	10.0%	10.0%	10.0%	10.0%
	68.7%	53.4%	43.0%	36.8%	32.6%
	95.2%	83.4%	70.4%	60.7%	53.3%
	99.9%	97.1%	89.3%	80.2%	71.8%
		99.9%	98.0%	93.1%	86.2%
			99.9%	98.6%	95.4%
				99.9%	98.9%
					99.9%
maximum relative error on top shares	0.91%	0.32%	0.14%	0.08%	0.05%

We assume that we want our tabulation to span from the 10% to the 99.9% percentiles, so we set $p_1 = 0.1$ and $p_K = 0.999$. We pick the median profile of φ''' estimated over all available years for France and the United States. For a given number K of thresholds, we solve the optimization problem:[20]

$$\min_{p_2,\dots,p_{K-1}} \left\{ \max_{t \in [x_1, x_K]} \int_{x_1}^{x_K} \varepsilon(x,t)\varphi'''(t)\, \mathrm{d}t \right\} \quad \text{st.} \quad p_1 < p_2 < \cdots < p_{K-1} < p_K$$

where as usual $x_k = -\log(1 - p_k)$ for $1 \le k \le K$.

Table IV shows that a important concentration of brackets near the top is desirable, but that we also need quite a few to cover the bottom. Half of the brackets should cover the top 20%, most of which should be within just the top 10%. The rest should be used to cover the

---

[20]We solve the problem using the derivative-free Nelder-Mead algorithm.

bottom 80% of the distribution. We can also see that a relatively small number of well-placed brackets can achieve remarkable precision: only six are necessary to achieve a maximal relative error of less than 0.1%.

5.2.3 Comparison with Partial Samples

We have seen that generalized Pareto interpolation can be quite precise, but how does it compare to the use of a subsample of individual data? The question may be of practical interest when researchers have access to both exhaustive data in tabulated form, or a partial sample of individual data. Such a sample could either be a survey, or a subsample of administrative data.

TABLE V: MEAN RELATIVE ERROR USING SUBSAMPLES OF THE FULL POPULATION

	mean percentage gap between estimated and observed values for a survey with simple random sampling and sample size n					
	$n = 10^3$	$n = 10^4$	$n = 10^5$	$n = 10^6$	$n = 10^7$	$n = 10^8$
Top 70% share	0.42%	0.20%	0.10%	0.04%	0.01%	0.00%
Top 50% share	1.26%	0.63%	0.32%	0.13%	0.04%	0.00%
Top 25% share	4.00%	2.04%	1.05%	0.44%	0.15%	0.00%
Top 10% share	9.29%	4.80%	2.50%	1.05%	0.35%	0.00%
Top 5% share	14.32%	7.48%	3.94%	1.65%	0.55%	0.00%
Top 1% share	29.13%	16.01%	8.57%	3.61%	1.21%	0.00%
Top 0.1% share	52.94%	35.23%	19.91%	8.57%	2.89%	0.00%
P30 threshold	4.67%	1.44%	0.45%	0.15%	0.04%	0.00%
P50 threshold	3.29%	1.03%	0.33%	0.10%	0.03%	0.00%
P75 threshold	2.92%	0.91%	0.31%	0.10%	0.03%	0.00%
P90 threshold	3.91%	1.21%	0.39%	0.12%	0.04%	0.00%
P95 threshold	5.86%	1.76%	0.59%	0.18%	0.06%	0.00%
P99 threshold	14.39%	4.79%	1.42%	0.46%	0.14%	0.00%
P99.9 threshold	44.31%	16.29%	5.47%	1.70%	0.49%	0.00%

Original sample of size $N = 10^8$ simulated using the distribution of 2010 pre-tax national income in the United States. Source: author's computations from Piketty, Saez, and Zucman (2016).

We may address that question using an example and Monte-Carlo simulations. Take the 2010 distribution of pre-tax national income in the United States. We can estimate that distribution and use it to simulate a sample of size $N = 10^8$ (the same order of magnitude as the population of the United States).

Then, we create subsamples of size $n \leq N$ by drawing without replacement from the large population previously generated.[21] In the case of surveys, we ignore nonresponse and no mis-

---

[21]This survey design is called simple random sampling.

reporting, a simplification which favors the survey in the comparison. For each of those sub-samples, we estimate the quantiles and top shares at different points of the distribution, and compare it to the same values in the original sample of size N. Table V shows the results for different values of n. We see that even for large samples ($n = 10^5$, $n = 10^6$, $n = 10^7$), the case for using tabulations of exhaustive data rather than subsamples to estimates quantities such as the top 1% or 0.1% share remains strong. Indeed, even with $n = 10^6$ observations, the typical error on the top 1% share is larger than what we get in table IV, even with few thresholds. In practice, the thresholds may not be positioned in an optimal way as in table IV, so may also want to compare the results with table I. The differences in the orders of magnitude are large enough so that the implications of that comparison hold.

6   Pareto Curves and the Dynamics of Income and Wealth

The key mechanism through which power laws emerge is proportional random growth. Its study started with Yule (1925), and it was first applied to the economics of the income and wealth distribution by Champernowne (1953), Simon (1955), and Wold and Whittle (1957).

Models with proportional random shocks can take different forms. They can account for the very high levels of wealth concentration. That is, starting from a given level of labor income inequality, random shocks to rates of return and/or saving rates (possibly coming from shocks to bequest taste or other saving parameters), as well as demographic random shocks (number of children, age at death, etc.), can easily generate large levels of wealth inequality. In such models, the asymptotic inverted Pareto coefficient will typically be a steeply rising function of the variance of the various shocks, as well as of the gap between the average rate of return r and the economy's growth rate g (see e.g. Piketty and Zucman (2015) for simple calibrations). They can also be used to account for the observed level of labor income inequality. In that case, the random shocks to the earning process can be interpreted as promotion or demotion shocks. With the right setting, they can explain the growing labor income inequality (Gabaix, Lasry, et al., 2016).

6.1   Standard models of income growth and wealth accumulation

In discrete time, individual trajectories of income or wealth may be written as:

$$X_t = \omega_t X_{t-1} + \varepsilon_t \tag{4}$$

where X_t is income or wealth at date t, $\{\omega_1, \omega_2, \dots\}$ and $\{\varepsilon_1, \varepsilon_2, \dots\}$ are random iid. shocks. This is called a Kesten (1973) process, and may correspond to different settings depending on how we interpret the parameters. In an intergenerational model of wealth, as in Piketty and Zucman (2015), X_t is the wealth of the current generation, and X_{t-1} the wealth of the previous generation. Then $\omega_t = s_t(1 + r_t)$ captures the rate of return on the wealth of the previous generation and the saving rate of the present generation, while ε_t correspond to the labor income, net of consumption, of the current generation.[22] If X_t is viewed as income, then equation (4) may be viewed as a pseudo-autoregressive model of income, with a random autoregressive coefficient.[23] Kesten's (1973) result, which was further refined by Goldie (1991), is that regardless of the precise distribution of ω_t and ε_t, X_t will converge to a distribution with power-law tail α, where $\mathbb{E}[|\omega_t|^\alpha] = 1$. That is, the closer $|\omega_t|$ is to 1 on average, the fatter the tail.

A few conditions are nevertheless required for that result to hold. Notably, ω_t must be less than 1 on average, otherwise Y_t would generally get larger at each iteration, and there would be no stationary distribution.[24] There must be a positive probability that $\omega_t > 1$. Also, ε_t cannot be identically zero. If it were, then Y_t would be a pure succession of iid. multiplicative shocks: by applying the central limit theorem to its logarithm, we would get that Y_t converges to a log-normal distribution with ever-increasing variance. A nonzero ε_t acts as a stabilizing friction because it prevents Y_t from becoming too small.[25]

6.2 Explaining the Actual Shape of Generalized Pareto Curve

The limitation of Kesten (1973) processes is that they tend to generate flat Pareto curves $b(p)$ at the top of the distribution. But their main idea — random multiplicative shocks with frictions generate power laws — can be generalized to account for the true shape of $b(p)$. Because discrete time models do not in general admit analytical steady-state formula, we move to continuous time. The key mechanisms remain identical, but they make it easier to relax some of the assumptions. In continuous time, random growth takes the form of the *stochastic differential equation*:

$$\frac{\mathrm{d}X_t}{X_t} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{5}$$

---

[22]We assume zero economic growth to simplify exposition. Otherwise wealth would have to be normalized by the economy's growth rate g_t, and $\omega_t = s_t(1 + r_t)/(1 + g_t)$.

[23]In fact, Kesten's (1973) full framework would even allow for autoregressive processes of higher order.

[24]This typically requires normalizing Y_t by the economy's growth rate.

[25]The complete set of assumptions of the theorem is: $\mathbb{E}[|\omega_t|\alpha \max\{\log(\omega_t), 0\}]$, $0 < \mathbb{E}[|\varepsilon_t|\alpha] < +\infty$ and the distribution of $\log|\omega_t|$ given that $\omega_t \neq 0$ is nonlattice. See Gabaix (2009).

where W_t is a Wiener process (i.e. a Brownian motion), which means that the rate of growth of income over a small time period $[t, t + dt]$ is random and independent from X_t with a constant mean $\mu\, dt$ and a constant variance $\sigma^2\, dt$.

If equation (5) holds exactly throughout the entire distribution, then the growth rate is completely independent from the initial level. This corresponds to $\varepsilon_t = 0$ in the discrete time process (4).[26] The consequence is the same in continuous time as in discrete time: X_t is log-normal and its variance grows without bound.[27]

To get a stationary distribution, we need to add some frictions that prevent X_t from becoming too small, the same way a nonzero ε_t did in (4). Such frictions may take different forms: for example, a lower bound on wages (Gabaix, 1999) or a nonzero probability of retiring (Saichev, Malevergne, and Sornette, 2010). But the approach most similar to Kesten (1973) processes is to make the variance of *relative* income growth go to infinity at low income levels, so that the variance of *absolute* income growth remains well above zero: this is the role played by ε_t in (4) (Saichev, Malevergne, and Sornette, 2010, p. 17). To that end, we need to rewrite equation (5) in a more general way as:

$$\frac{dX_t}{X_t} = \mu(X_t)\, dt + \sigma(X_t)\, dW_t \tag{6}$$

We need $\mu(x) < 0$ at the top get a admissible distribution.[28] For labor income, we complement equation (6) with a reflecting barrier that prevents wages from going below zero. We can characterize the stationary distribution of such a process using the following theorem.

Theorem 8. *Let X_t follow the stochastic differential equation (6) with a stationary distribution $\mathcal{D}$. If both $\mu(x)$ and $\sigma^2(x)$ converge toward a constant, $\mathcal{D}$ is a power law in the sense of definition 2.*

Theorem 8 makes the connection between the statistical definition of an asymptotic power law (see definition (2) in section 2), and the economic mechanisms that explain the distributions of income and wealth. The Pareto distribution arises because of scale invariance (above a certain threshold) in the stochastic process that describes the evolution of income or wealth. But if the scale invariance doesn't hold exactly but only asymptotically, then instead of a Pareto distribution we get an asymptotic power law.

---

[26]Indeed, the total growth rate in (4) is equal to $\omega_t + \varepsilon_t / X_{t-1}$, so we need $\varepsilon_t = 0$ for it to be independent from X_{t-1}.

[27]This is known as the *geometric Brownian motion.*

[28]Recall that the distribution is normalized by the overall growth rate. So if the distribution remains the same at the steady state, people at the top must have higher odds of going down than the rest of the population, hence $\mu < 0$.

This shows the economic relevance of power laws in Karamata's (1930) sense. We might have expected that if scale invariance held asymptotically in (6), then we would get a distribution asymptotically equivalent to a power law (i.e. $1 - F(x) \sim Cx^{-\zeta}$ for some $C > 0$). In fact, the set of possible stationary distributions is larger than that. For example, take $\mu(x) = -1 + 1/(2\log x)$ and $\sigma^2(x) = 1$. Then we can show that $1 - F(x) \propto (1 + 3\log x)x^{-3}$, which is not equivalent, in a strict mathematical sense, to any power law. However $x \mapsto 1 + 3\log x$ is a slowly varying function, so we do have a power law in Karamata's (1930) sense.

We showed (see proposition 3) that generalized Pareto curves provide a very direct visual check of asymptotic power law behavior. It makes them particularly suitable for observing such distributions while focusing on the behavior of the tail. With the following theorem, we can further characterize the shape of the stationary distribution.

Theorem 9. *Assume that $\mu(x)$ and $\sigma^2(x)$ converge toward a constant: $\mu(x) \to \mu$, $\sigma^2(x) \to \sigma^2$. Define:*

$$\zeta(x) = 1 - \frac{2\mu(x)}{\sigma^2(x)} \qquad and \qquad \zeta = \lim_{x \to +\infty} \zeta(x) = 1 - \frac{2\mu}{\sigma^2}$$

Let f be the density of the stationary distribution, and F its CDF. We have:

$$f(x) \propto x^{-\zeta-1} \exp\left(-\log(\sigma^2(x)) - \int_1^x \frac{\zeta(t) - \zeta}{t}\,\mathrm{d}t\right)$$

and:

$$1 - F(x) = L(x)x^{-\zeta}$$

where L is a slowly varying function. Therefore, the stationary distribution has an asymptotic inverted Pareto coefficient equal to $1 - \sigma^2/(2\mu)$.

The asymptotic inverted Pareto coefficient $\lim_{p \to 1} b(p)$ is given by the asymptotic value of $\mu(x)$ and $\sigma^2(x)$. For the rest of the distribution, we can express the density in a relatively simple form. A identical result was given by Gabaix (1999) in terms of the "local Pareto exponent". We write that result using the notations of section 2.4:

$$\alpha_2(x) = -\frac{xf'(x}{f(x)} - 1 = 1 - 2\frac{\mu(x)}{\sigma^2(x)} + \frac{x}{\sigma^2(x)}\frac{\mathrm{d}\sigma^2}{\mathrm{d}x}(x)$$

This coefficient $\alpha_2(x)$ is more directly predicted by the theory than $b(p)$, but more difficult to estimate from the data because it involves the density and its derivative. However, as shown in section 2.4, both coefficients have qualitatively similar interpretations. Namely, a higher $b(p)$ at the top of the distribution correspond to a lower $\alpha_2(x)$. So if $\mu(x)$ is constant, we get a higher $b(p)$ by increasing the variance of shocks $\sigma^2(x)$. If $\sigma^2(x)$ is constant, we get a higher $b(p)$ by increasing the mean of shocks $\mu(x)$.

6.3 Calibrations

There are two parameters of the income or wealth generating process that we can adjust in order to match the generalized Pareto curves that we observe in practice: the variance of income shocks, and their mean. It is not possible identify both parameters separately based on the stationary distribution. But by adjusting one while holding the other constant, we can still gain some insights on the parameters of the income-generating mechanism.

Several models can be used as microfoundations for the process (6) (see for example Benhabib and Bisin, 2016; Gabaix, Lasry, et al., 2016). Here, we focus on the reduced-form equation. Our goal is to go from the data to the theory, and to try to identify the key features of the income and wealth generating process that are necessary to produce accurate distributions at the top, in a way that is robust to the precise primitives of the model.

6.3.1 The Distribution of Income

Let X_t be individual incomes normalized by the overall economy growth rate. It follows the stochastic process (6). To ensure a stationarity, we will set for $x \to 0$:

$$\sigma(x) \sim \sqrt{\frac{c_1 + c_2 x^2}{x^2}} \tag{7}$$

as in Saichev, Malevergne, and Sornette (2010). We have $\lim_{x \to 0} x\sigma(x) > 0$, so that absolute income shocks remain non-negligible even at low income levels. We also put a reflective barrier at zero that prevents wages from becoming negative.

Calibration of the variance Here, we assume that $\mu(x)$ does not depend on x, so we normalize it at -1. Then, to account for the rise of $b(p)$ near $p = 1$, we can increase the variance of income shocks at the top. Because of (7), this implies a U-shaped profile of variance, with a high relative volatility of earnings at the bottom and at the top, and a lower one for the middle of the distribution. This profile of variance is in fact strongly suggested by empirical work on panel data using either surveys (Chauvel and Hartung, 2014; Hardy and Ziliak, 2014; Bania and Leete, 2009) or administrative data (Guvenen et al., 2015). Hardy and Ziliak (2014) describe it as the "wild ride" at the top and the bottom of the distribution. We assume the following functional form for $\sigma(x)$, with $c_1, c_2, c_3, c_4 > 0$:

$$\sigma(x) = \sqrt{\frac{c_1 + c_2 x^2}{x^2} + \frac{c_3 x^2}{1 + c_4 x^2}} \tag{8}$$

The second term, $c_3 x^2/(1 + c_4 x^2)$, matches the upward sloping branch of the U-shaped profile of variance. For $x \to +\infty$, we get $\sigma(x) \to \sqrt{1 + c_2 + c_3/c_4}$, so according to theorem 8 the stationary distribution is a power law.
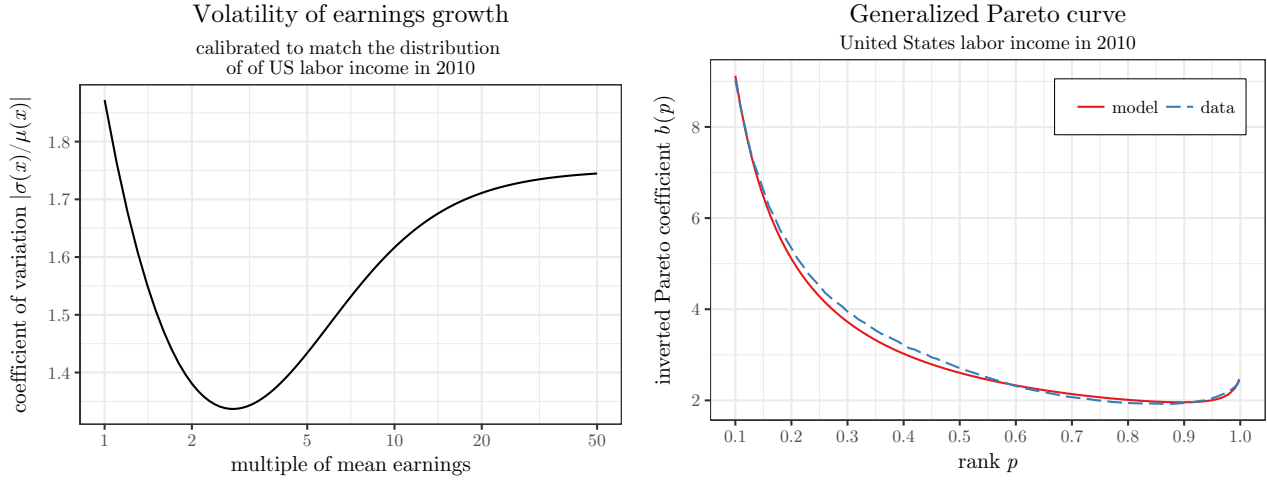
Model calibrated to match the US distribution of labor income in 2010 ($c_1 = 2.341, c_2 = 1.104, c_3 = 0.061, c_4 = 0.031$). The coefficient of variation corresponds to the standard deviation divided by the absolute value of the mean growth of non-reflected units.

FIGURE 8: Calibration of $\sigma(x)$ on the US Distribution of Labor Income

Figure 8 shows the results of the calibration of $\sigma(x)$ to match the actual profile of $b(p)$. The volatility of earnings growth has indeed a U-shaped profile. At the very top of the distribution, the volatility of earnings shocks is about 30% higher than at its lowest point, which occurs around the 90% percentile. Overall, this model is able to match most of the distribution of income, as shown by the two similar Pareto curves in figure 8.

Calibration of the mean   A second solution to match the increase of $b(p)$ at the top is to adjust the mean of income shocks. We set:

$$\mu(x) = -c_1 + \frac{c_2 x^2}{1 + c_3 x^2} \tag{9}$$

with $c_1, c_2, c_3 > 0$. The baseline income growth of non-reflected units is $-c_1$, which is negative because we have normalized income by the overall income growth: since reflected units have positive growth, non-reflected units must have negative growth to compensate (Gabaix, 2009). The other part of the formula, $c_2 x^2/(1 + c_3 x^2)$, is here to make $\mu(x)$ increase at the top of the distribution. It also ensures that μ converges to a constant, so that we get a power law in the

36

end. For the variance, we set:

$$\sigma(x) = \sqrt{\frac{c_4 + x^2}{x^2}}$$

which ensures a stationary process, and normalizes $\sigma(x)$ to 1 at infinity. This normalization is necessary because μ and σ can only be identified up to a scaling constant.
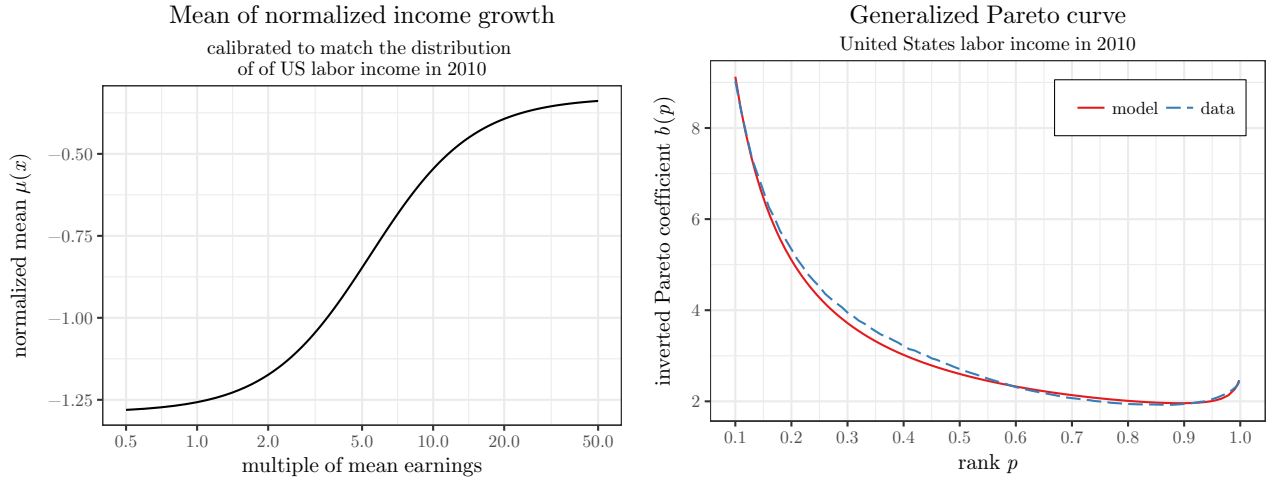
Model calibrated to match the US distribution of labor income in 2010 ($c_1 = 1.289, c_2 = 0.033, c_3 = 0.034, c_4 = 2.574$). $\mu(x)$ corresponds to the difference between the growth of non-reflected units and average income growth, expressed as a multiple of $\sigma(x)$ at infinity.

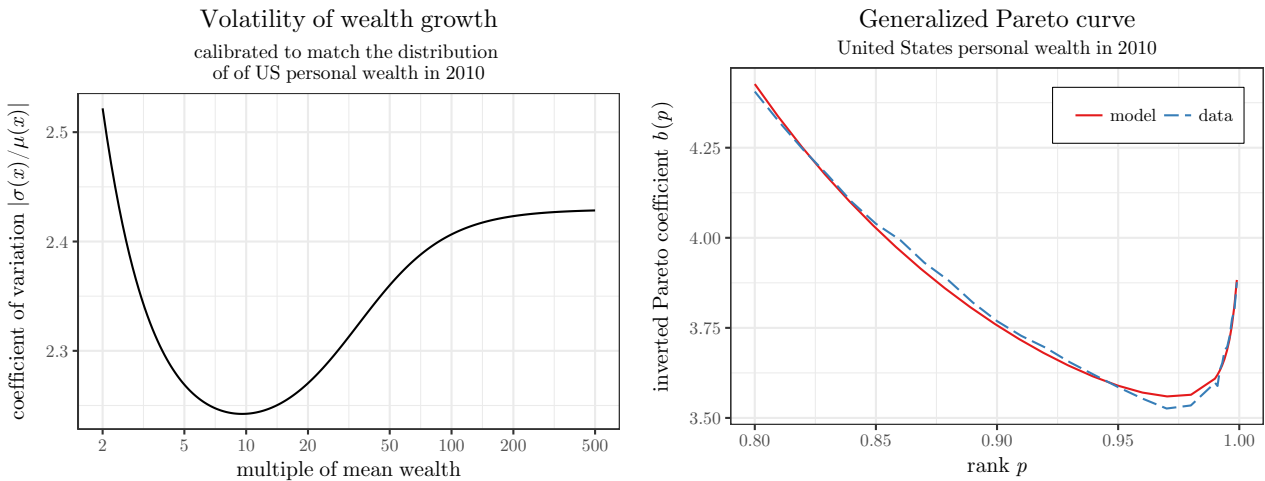FIGURE 9: Calibration of $\mu(x)$ on the US distribution of labor income

Figure 9 shows the increase in the average of income shocks that is necessary to match the increase of $b(p)$ at the top. The value of $\mu(x)$ only concerns the non-reflected units, but units at the top are unlikely to hit the reflecting barrier, so $\mu(x)$ constitute a good indicator of effective average income growth at the top. We can see that the final rise in $b(p)$ is consistent with an increase of $\mu(x)$ of about one standard deviation between the middle and the very top of the distribution. The model with varying mean income growth is also able to precisely match almost the entire income distribution.

Gabaix, Lasry, et al. (2016) suggested that scale dependence in $\mu(x)$ is necessary to account for the speed of the increase in inequality in the United States. Our finding corroborates theirs, showing that scale dependence is can explain the shape of the distribution in a static framework, not just in a dynamic one. Scale dependence of mean income shocks is consistent with models in which income inequality is driven by "superstar" entrepreneurs of managers (Gabaix and Landier, 2008) (see Gabaix, Lasry, et al. (2016, appendix I.3) for a dynamic version).

6.3.2 The distribution of wealth

We can do a similar exercise for the distribution of wealth. The generalized Pareto curves for wealth and income are similar (U-shaped with a smaller increase at the top end). However, for wealth, $b(p)$ is higher overall, and the final increase happens later. What does this mean for the underlying process generating wealth? To answer that question, we consider a wealth generating process like (6). We drop the reflective barrier because wealth can go below zero, but focus on the top 20%.[29]

Calibration of the variance We calibrate the profile of variance using the same formula as for income (8). We see in figure 10 that we also get a U-shaped profile: wealth is more volatile at the very top of the distribution than at the middle. However, the increase starts much later, around ten times average wealth, which correspond roughly to the top 1%. For income, the increase started to happen around the top 10%. The difference between the lowest point and the top of the distribution is also more modest, at about 8% instead of 30%.
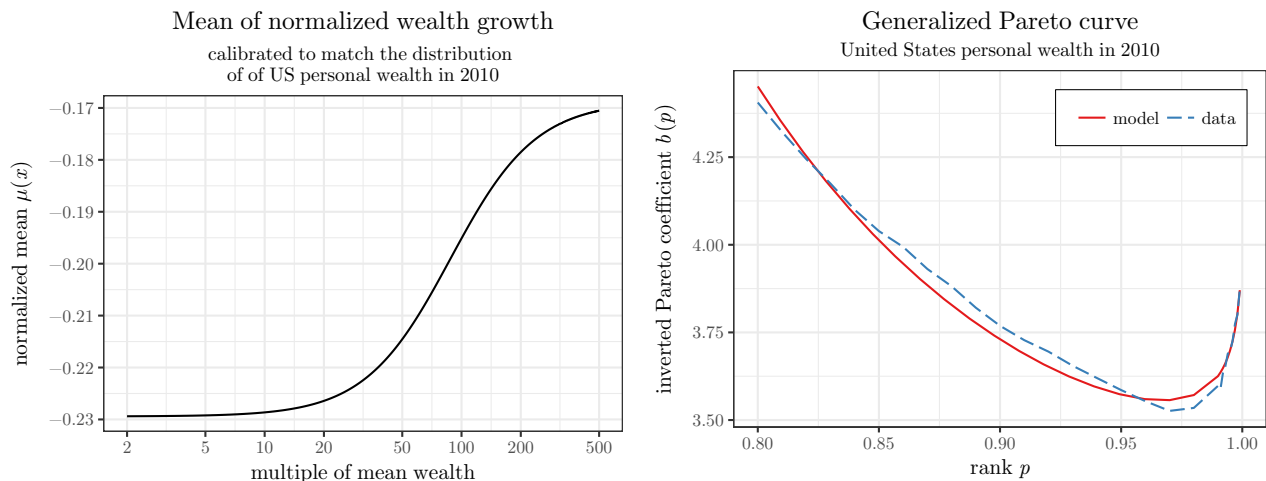

Model calibrated to match the US distribution of personal wealth in 2010 ($c_1 = 5.84, c_2 = 4.90, c_3 = 0.000809, c_4 = 0.000804$). The coefficient of variation correspond to the standard deviation divided by the absolute value of the mean growth.

FIGURE 10: Calibration of $\sigma(x)$ on the US distribution of personal wealth

---

[29]We focus on the top because we view (6) primarily as a model of the top of the distribution, even though it can sometimes fit well the bottom of the distribution too, as we saw for income. Wealth goes to zero fast once we leave top of the distribution, so providing a good fit for the bottom presents more difficulties. We do not explicitly model the negative part of the distribution because it is not necessary for our calibration.

Model calibrated to match the US distribution of labor income in 2010 ($c_1 = 0.229, c_2 = 0.00000788, c_3 = 0.000130, c_4 = 1.43$). The coefficient of variation correspond to the standard deviation divided by the absolute value of the mean growth.

FIGURE 11: Calibration of $\mu(x)$ on the US distribution of personal wealth

Calibration of the mean We use again the formula (9) to model mean wealth growth. Figure 11 shows the result. Again, we do observe an increasing mean of wealth growth (wealthier people experience higher returns, saving rates and/or higher incomes as a fraction of their wealth). But the increase is much more modest than for income (around 6% of a standard deviation). It also happens much later, starting at 10 times the average wealth (which roughly correspond to the top 1%).

This type of scale dependence is consistent with available microdata. Fagereng et al. (2016) document using administrative Norwegian data that returns are positively correlated with wealth. Because these higher returns partly reflect the fact that wealthier people hold riskier assets, it also implies higher variance at the top. Therefore, we have scale dependence for both the variance and the mean. This is also consistent with the work of Bach, Calvet, and Sodini (2017) on Swedish administrative data. The model of Kacperczyk, Nosal, and Stevens (2014), in which investors have different levels of sophistication, can account for these findings. Scale dependence can also arise if the very wealthy have higher saving rates (Saez and Zucman, 2016). Benhabib, Bisin, and Luo (2015) study a model where saving rates increase in wealth because the bequest function is more elastic than the utility of consumption.

Hence, although we do not here have access to individual panel data, it is still possible to identify to some extent the scale dependence of the underlying process that generates income and wealth by looking at the way Pareto coefficients vary in the tail.

7 Concluding comments

In this paper, we introduce the concept of generalized Pareto curve to characterize, visualize and estimate distributions of income or wealth. We show strong connections between those curves and the theory of asymptotic power laws, which makes them a natural tool for analyzing them.

Based on quasi-exhaustive individual tax data, we reveal some stylized facts about the distribution of income that lets us move beyond the standard Pareto assumption. We find that although generalized Pareto curves can vary a lot over time and between countries, they tend to stay U-shaped.

Then we develop a method to interpolate tabulated data on income or wealth — as is typically available from tax authorities and statistical institutes — that can correctly reproduce the subtleties of generalized Pareto curves. In particular, the method guarantees the smoothness of the estimated distribution, and work well over most of the distribution, not just the very top. We show that method to be several times more precise than the alternatives most commonly used in the literature. In fact, it can often be more precise than using non-exhaustive individual data. Moreover, we can derive formulas for the error term that let us approximately bound the error of our estimates, and determine the number of optimally placed brackets that is necessary to achieve a given precision.

Finally, we show how our finding can be connected to the existing literature on the income and wealth distribution that emphasizes the role of random growth in explaining power law behavior. The typical shape of Pareto curves that we observe may be explained by a simple and natural deviation from standard random growth models, which is also backed by theoretical models and empirical studies on panel data. Namely, the very top experience higher growth and/or more risk, meaning the processes that generate income and wealth are not fully scale invariant.

We believe that more empirical work — especially a careful use of administrative data sources — is necessary to study those dynamics in a fully satisfying way. We hope that the interpolation method presented in this paper will allow future researchers make progress in that direction. To that end, we made the methods presented in this paper available as a R package named `gpinter`, and also in the form of an online interface that can be used without any installation or knowledge of any programming language. Both are available at `http://wid.world/gpinter`.

References

Alvaredo, Facundo et al. (2016). "Distributional National Accounts (DINA) Guidelines: Concepts and Methods used in WID.world". In: *WID.world Working Paper Series*.

Atkinson, A. B. (2017). "Pareto and the Upper Tail of the Income Distribution in the UK: 1799 to the Present". In: *Economica* 84.334, pp. 129–156. ISSN: 1468-0335. DOI: 10.1111/ecca.12214. URL: http://dx.doi.org/10.1111/ecca.12214.

Atkinson, A. B. and A. J. Harrison (1978). *Distribution of Personal Wealth in Britan*. Cambridge University Press.

Bach, Laurent, Laurent Calvet, and Paolo Sodini (2017). "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy". URL: https://ssrn.com/abstract=2706207.

Balkema, A. A. and L. de Haan (1974). "Residual Life Time at Great Age". In: *The Annals of Probability* 2.5, pp. 792–804. URL: http://dx.doi.org/10.1214/aop/1176996548.

Bania, Neil and Laura Leete (2009). "Monthly household income volatility in the U.S., 1991/92 vs. 2002/03". In: *Economics Bulletin* 29.3, pp. 2100–2112.

Benhabib, Jess and Alberto Bisin (2016). "Skewed Wealth Distributions: Theory and Empirics". In: *NBER Working Paper Series* 21924, p. 37. DOI: 10.3386/w21924. URL: http://www.nber.org/papers/w21924.

Benhabib, Jess, Alberto Bisin, and Mi Luo (2015). "Wealth distribution and social mobility in the US: A quantitative approach". In: *NBER Working Paper Series* 21721, pp. 1–32. DOI: 10.3386/w21721.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu (2011). "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents". In: *Econometrica* 79.1, pp. 123–157. ISSN: 1468-0262. DOI: 10.3982/ECTA8416. URL: http://dx.doi.org/10.3982/ECTA8416.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press. ISBN: 9780521379434.

Birgin, E. G. and J. M. Martìnez (2008). "Improving Ultimate Convergence of an Augmented Lagrangian Method". In: *Optimization Methods Software* 23.2, pp. 177–195. ISSN: 1055-6788. URL: http://dx.doi.org/10.1080/10556780701577730.

Cargo, G. T. and O. Shisha (1966). "The Bernstein Form of a Polynomial". In: *Journal of Research of the National Bureau of Standards* 60B.1, pp. 79–81.

Champernowne, D. G. (1953). "A Model of Income Distribution". In: *The Economic Journal* 63.250, pp. 318–351. ISSN: 00130133, 14680297. URL: http://www.jstor.org/stable/2227127.

Chang, Winston et al. (2017). *shiny: Web Application Framework for R*. R package version 1.0.3. URL: https://CRAN.R-project.org/package=shiny.

Chauvel, Louis and Anne Hartung (2014). "Dynamics of Income Volatility in the US and in Europe, 1971-2007: The Increasing Lower Middle Class Instability". In: *Paper Prepared for the IARIW 33rd General Conference*.

Conn, Andrew R., Nicholas I. M. Gould, and Philippe L. Toint (1991). "A Globally Convergent Augmented Langrangian Algorithm for Optimization with General Constraints and Simple Bounds". In: *SIAM Journal on Numerical Analysis* 28.2, pp. 545–572. ISSN: 00361429. URL: http://www.jstor.org/stable/2157828.

Fagereng, Andreas et al. (2016). "Heterogeneity and Persistence in Returns to Wealth". URL: http://www.nber.org/papers/w22822.

Feenberg, Daniel and James Poterba (1992). "Income Inequality and the Incomes of Very High Income Taxpayers: Evidence from Tax Returns". In: Working Paper Series 4229. DOI: 10.3386/w4229. URL: http://www.nber.org/papers/w4229.

Fournier, Juliette (2015). "Generalized Pareto curves: Theory and application using income and inheritance tabulations for France 1901-2012". MA thesis. Paris School of Economics.

Gabaix, Xavier (1999). "Zipf's Law for Cities: An Explanation". In: *The Quarterly Journal of Economics* 114.3, p. 739. DOI: 10.1162/003355399556133.

— (2009). "Power Laws in Economics and Finance". In: *Annual Review of Economics* 1.1, pp. 255–294. DOI: 10.1146/annurev.economics.050708.142940.

Gabaix, Xavier and Augustin Landier (2008). "Why Has CEO Pay Increased So Much?" In: *Quarterly Journal of Economics* 123.1, pp. 49–100. ISSN: 0033-5533. DOI: 10.1162/qjec.2008.123.1.49. arXiv: arXiv:1011.1669v3.

Gabaix, Xavier, Jean-Michel Lasry, et al. (2016). "The Dynamics of Inequality". In: *Econometrica* 84.6, pp. 2071–2111. ISSN: 1468-0262. DOI: 10.3982/ECTA13569. URL: http://dx.doi.org/10.3982/ECTA13569.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2016). "Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)". In: WID.world Working Paper. URL: http://wid.world/document/b-garbinti-j-goupille-and-t-piketty-inequality-dynamics-in-france-1900-2014-evidence-from-distributional-national-accounts-2016/.

Goldie, C. M. (1991). "Implicit Renewal Theory and Tails of Solutions of Random Equations". In: *Annals of Applied Probability* 1.1, pp. 126–166.

Guvenen, Fatih et al. (2015). "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" In: *NBER Working Papers*. URL: https://ideas.repec.org/p/nbr/nberwo/20913.html.

Hardy, Bradley and James P Ziliak (2014). "Decomposing Trends in Income Volatility: the "Wild Ride" at the Top and Bottom". In: *Economic Inquiry* 52.1, pp. 459–476.

Jones, Charles I. (2015). "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality". In: *Journal of Economic Perspectives* 29.1, pp. 29–46. ISSN: 0895-3309. DOI: 10.1257/jep.29.1.29. URL: http://dx.doi.org/10.1257/jep.29.L29.

Jones, Charles I. and Jihee Kim (2017). "A Schumpeterian Model of Top Income Inequality". URL: http://search.proquest.com/docview/1629323393?accountid=17248.

Kacperczyk, Marcin, Jaromir B Nosal, and Luminita Stevens (2014). "Investor Sophistication and Capital Income Inequality". URL: http://www.nber.org/papers/w20246.pdf.

Karamata, J. (1930). "Sur un mode de croissance regulière des fonctions". In: *Mathematica* 4.

Kesten, Harry (1973). "Random Difference Equations and Renewal Theory for Products of Random Matrices". In: *Acta Mathematica* 131.1, pp. 207–248. ISSN: 00015962. DOI: 10.1007/BF02392040. arXiv: -.

Kraft, Dieter (1994). "Algorithm 733: TOMP — Fortran Modules for Optimal Control Calculations". In: *ACM Transactions on Mathematical Software* 20.3, pp. 262–281. ISSN: 0098-3500. URL: http://doi.acm.org/10.1145/192115.192124.

Kuznets, Simon (1953). *Shares of Upper Income Groups in Income and Savings*. NBER. URL: http://www.nber.org/books/kuzn53-1.

Lyche, Tom and Knut Mørken (2002). "Spline Methods". URL: http://folk.uio.no/in329/komp.html.

McLeod, R.J.Y. and M.L. Baart (1998). *Geometry and Interpolation of Curves and Surfaces.* Cambridge University Press. ISBN: 9780521321532.

Nirei, Makoto (2009). "Pareto Distributions in Economic Growth Models".

Pareto, Vilfredo (1896). *Cours d'économie politique.*

Pickands, James (1975). "Statistical Inference Using Extreme Order Statistics". In: *The Annals of Statistics* 3.1, pp. 119–131. URL: http://dx.doi.org/10.1214/aos/1176343003.

Piketty, Thomas (2001). *Les hauts revenus en France au XXème siècle.* Grasset.

— (2003). "Income Inequality in France, 1901–1998". In: *Journal of Political Economy* 111.5, pp. 1004–1042. DOI: 10.1086/376955.

Piketty, Thomas and Emmanuel Saez (2003). "Income Inequality in the United States, 1913-1998". In: *The Quarterly Journal of Economics* 118.1, pp. 1–39.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2016). "Distributional National Accounts: Methods and Estimates for the United States". In: Working Paper Series 22945. DOI: 10.3386/w22945. URL: http://www.nber.org/papers/w22945.

Piketty, Thomas and Gabriel Zucman (2015). "Wealth and Inheritance in the Long Run". In: *Handbook of Income Distribution.* Vol. 2. Handbook of Income Distribution. Elsevier, pp. 1303–1368. DOI: http://dx.doi.org/10.1016/B978-0-444-59429-7.00016-9.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Saez, Emmanuel (2001). "Using Elasticities to Derive Optimal Income Tax Rates". In: *The Review of Economic Studies* 68.1, pp. 205–229. ISSN: 0034-6527. DOI: 10.1111/1467-937X.00166. URL: https://academic.oup.com/restud/article-lookup/doi/10.1111/1467-937X.00166.

Saez, Emmanuel and Gabriel Zucman (2016). "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data". In: *Quarterly Journal of Economics* 131.May, pp. 519–578. DOI: 10.1093/qje/qjw004.Advance.

Saichev, Alexander, Yannick Malevergne, and Didier Sornette (2010). *Theory of Zipf's Law and Beyond.* Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-02946-2. DOI: 10.1007/978-3-642-02946-2. URL: http://dx.doi.org/10.1007/978-3-642-02946-2.

Scott, David W. (1992). *Multivariate Density Estimation.* John Wiley & Sons, Inc. ISBN: 9780470316849. URL: http://dx.doi.org/10.1002/9780470316849.

Simon, Herbert (1955). "On a Class of Skew Distribution Functions". In: *Biometrika* 42.3-4, pp. 425–440. ISSN: 00199958. DOI: 10.1093/biomet/42.3-4.425. URL: http://biomet.oxfordjournals.org.libproxy.lib.unc.edu/content/42/3-4/425.full.pdf+html?sid=3d2e52aa-23bb-4338-83be-d71ec5197171.

Taleb, N. N. and R. Douady (2015). "On the super-additivity and estimation biases of quantile contributions". In: *Physica A Statistical Mechanics and its Applications* 429, pp. 252–260. DOI: 10.1016/j.physa.2015.02.038. eprint: 1405.1791.

van der Wijk, J. (1939). *Inkomens- En Vermogensverdeling.*

Wold, H. O. A. and P. Whittle (1957). "A Model Explaining the Pareto Distribution of Wealth". In: *Econometrica* 25.4, pp. 591–595. ISSN: 00129682. DOI: 10.2307/1905385.

Yule, G. Udny (1925). "A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S." In: 213.402-410, pp. 21–87. ISSN: 0264-3960. DOI: 10.1098/rstb.1925.0002.

Generalized Pareto Curves: Theory and Applications[*]

Technical appendix — not for publication

Thomas BLANCHET (Paris School of Economics)
Juliette FOURNIER (Massachusetts Institute of Technology)
Thomas PIKETTY (Paris School of Economics)

First version: March 24, 2017
This version: September 11, 2017

Abstract

This appendix supplements our paper "Generalized Pareto Curves: Theory and Applications". It provides technical details on the interpolation method, extended tables for comparison of interpolation methods, and the proofs of the theorems.

A   Generalized Pareto curves: Theory

A.1   Proof of proposition 1

That $b(p) > 1$ follows directly from the definition. For the rest of the proposition, we have for $p \geq \bar{p}$:

$$(1-p)Q(p)b(p) = \int_p^1 Q(u)\,\mathrm{d}u$$

We differentiate that equality with respect to p:

$$(1-p)Q(p)b'(p) + (1-p)b(p)Q'(p) - b(p)Q(p) = -Q(p)$$

We assumed that $Q(p) > 0$ for $p \geq \bar{p}$, so we can divide both sides by $Q(p)$:

$$(1-p)b'(p) + (1-p)b(p)\frac{Q'(p)}{Q(p)} - b(p) = -1$$

Hence:

$$(1-p)b(p)\frac{Q'(p)}{Q(p)} = b(p) - 1 - (1-p)b'(p)$$

Because the quantile function is increasing, the left hand side is nonnegative, which concludes the proof. ∎

A.2   Proof of proposition 2

From the proof of proposition 1, we have:

$$\frac{Q'(p)}{Q(p)} = \frac{1}{1-p} - \frac{1}{(1-p)b(p)} - \frac{b'(p)}{b(p)}$$

After integration:

$$Q(p) = Q(\bar{p}) \exp\left(\int_{\bar{p}}^p \frac{1}{1-u}\,\mathrm{d}u - \int_{\bar{p}}^p \frac{1}{(1-u)b(u)}\,\mathrm{d}u - \int_{\bar{p}}^p \frac{b'(u)}{b(u)}\,\mathrm{d}u\right)$$

$$= Q(\bar{p})\frac{(1-\bar{p})b(\bar{p})}{(1-p)b(p)} \exp\left(-\int_{\bar{p}}^p \frac{1}{(1-u)b(u)}\,\mathrm{d}u\right)$$

with $Q(\bar{p}) = \bar{x}$ by definition. ∎

A.3   Proof of proposition 3

The following representation of $b^*(x)$ will useful throughout the proofs.

Lemma 1.

$$b^*(x) = 1 + \frac{1}{x(1-F(x))}\int_x^{+\infty} 1 - F(z)\,\mathrm{d}z$$

Proof. Using integration by parts:

$$\int_x^{+\infty} z f(z)\, dz = \int_x^{+\infty} (-z)(-f(z))\, dz$$

$$= [-z(1 - F(z))]_{z=x}^{+\infty} + \int_x^{+\infty} 1 - F(z)\, dz$$

Because $\mathbb{E}[|X|] < +\infty$, Markov's inequality implies $1 - F(x) = o(1/x)$, so the bracketed term vanishes for $x \to +\infty$. Hence:

$$\int_x^{+\infty} z f(z)\, dz = x(1 - F(x)) + \int_x^{+\infty} 1 - F(z)\, dz$$

replacing in the expression of $b^*(x)$ yields the result. $\blacksquare$

First, note that since $\lim_{p \to 1} Q(p) = +\infty$, $\lim_{p \to 1} b(p) = \lim_{x \to +\infty} b^*(x)$. The assumption that L is slowly varying is equivalent to the assumption that $1 - F$ is regularly varying of index $-\alpha < -1$.

Direct half Applying the direct half of Karamata's theorem (Bingham, Goldie, and Teugels, 1989, 1.5.11, p. 28) to the representation of lemma 1, we have:

$$\lim_{x \to +\infty} \frac{1}{b^*(x)} = 1 + \frac{1}{\alpha - 1} = \frac{\alpha}{\alpha - 1}$$

Converse half We assume that $\lim_{p \to 1} b(p) = \alpha/(\alpha - 1)$. Hence:

$$\lim_{x \to +\infty} \frac{1}{b^*(x) - 1} = \alpha - 1$$

Then, we apply the converse half of Karamata's theorem (Bingham, Goldie, and Teugels, 1989, 1.6.1, p. 30) (with $\sigma = 0$) to the representation of lemma 1, proving that $1 - F$ is regularly varying of index $-\alpha$. $\blacksquare$

A.4   Proof of proposition 4

Direct half According to lemma 1, we have:

$$b^*(x) = 1 + \frac{1}{x(1 - F(x))} \int_x^{+\infty} 1 - F(z)\, dz$$

After a change of variable $z = tx$:

$$b^*(x) = 1 + \int_1^{+\infty} \frac{1 - F(tx)}{1 - F(x)}\, dt$$

$$= 1 + \int_1^K \frac{1 - F(tx)}{1 - F(x)}\, dt + \int_K^{+\infty} \frac{1 - F(tx)}{1 - F(x)}\, dt$$

for some $K > 1$. The function $t \mapsto (1 - F(xt))/(1 - F(x))$ is continuous over the compact interval $[1, K]$, so it is bounded. Therefore, Lebesgue's dominated convergence theorem implies:

$$\lim_{x \to +\infty} \int_1^K \left[\frac{1 - F(tx)}{1 - F(x)} \right] \mathrm{d}t = \int_1^K \left[\lim_{x \to +\infty} \frac{1 - F(tx)}{1 - F(x)} \right] \mathrm{d}t = 0$$

Moreover, we assumed that $1 - F$ is regularly varying. Therefore, using corollary 2.4.2 in Bingham, Goldie, and Teugels (1989, p. 85), the limit:

$$\lim_{x \to +\infty} \frac{1 - F(xt)}{1 - F(t)} = 0$$

holds uniformly for t over $[K, +\infty[$. The uniform convergence theorem implies:

$$\lim_{x \to +\infty} \int_K^{+\infty} \left[\frac{1 - F(tx)}{1 - F(x)} \right] \mathrm{d}t = \int_K^{+\infty} \left[\lim_{x \to +\infty} \frac{1 - F(tx)}{1 - F(x)} \right] \mathrm{d}t = 0$$

Therefore, we have $\lim_{x \to +\infty} b^*(x) = 1$. ∎

Converse half We assume that $\lim_{x \to +\infty} b^*(x) = 1$. Therefore:

$$\lim_{x \to +\infty} \int_1^{+\infty} \frac{1 - F(tx)}{1 - F(x)} \mathrm{d}t = 0$$

Let $\lambda > 1$ and $x \geq \bar{x}$. Because $t \mapsto (1 - F(xt))/(1 - F(x))$ is decreasing, we have for all $t < \lambda$:

$$\frac{1 - F(\lambda x)}{1 - F(x)} < \frac{1 - F(tx)}{1 - F(x)}$$

After integration with respect to t between 1 and λ:

$$\frac{1 - F(\lambda x)}{1 - F(x)} < \frac{1}{\lambda - 1} \int_1^\lambda \frac{1 - F(tx)}{1 - F(x)} \mathrm{d}t$$
$$< \frac{1}{\lambda - 1} \int_1^{+\infty} \frac{1 - F(tx)}{1 - F(x)} \mathrm{d}t$$

because $(1 - F(tx))/(1 - F(x)) \geq 0$ for all t. Since the inequality holds for all $x > \bar{x}$, and the left hand side is nonnegative, we have for all $\lambda > 1$:

$$\lim_{x \to +\infty} \frac{1 - F(\lambda x)}{1 - F(x)} = 0$$

Therefore, $1 - F$ is rapidly varying. ∎

B Detailed interpolation method

Recall that $\hat{\varphi}_k$ correspond to the quintic spline over the interval $[x_k, x_{k+1}]$ ($1 \leq k < K$). We parametrized the spline (ie. the polynomial of degree 5) with $(y_k, y_{k+1}, s_k, s_{k+1}, a_k, a_{k+1})$ so that:

$$\hat{\varphi}_k(x_k) = y_k \qquad\qquad \hat{\varphi}'_k(x_k) = s_k \qquad\qquad \hat{\varphi}''_k(x_k) = a_k$$
$$\hat{\varphi}_k(x_{k+1}) = y_{k+1} \qquad\qquad \hat{\varphi}'_{k+1}(x_{k+1}) = s_{k+1} \qquad\qquad \hat{\varphi}''_{k+1}(x_{k+1}) = a_{k+1}$$

The parameters $y_1, \ldots, y_K$ and $s_1, \ldots, s_K$ are directly given by the interpolation problem. But we still need to determine a_k, a_{k+1}. We first have $K - 2$ equations to ensure $\mathcal{C}^3$ continuity at the junctures:

$$\forall k \in \{2, \ldots, K-1\} \qquad \hat{\varphi}'''_{k-1}(x_k) = \hat{\varphi}'''_k(x_k)$$

Then, we impose the natural spline constraint at the first knot:

$$\hat{\varphi}'''_1(x_1) = 0$$

And we use a two points finite difference for the value of $\hat{\varphi}''_{K-1}(x_K)$:

$$\hat{\varphi}''_{K-1}(x_K) = \frac{s_K - s_{K-1}}{x_K - x_{K-1}}$$

That leads to a linear system of K equations for the K unknowns $a_1, \ldots, a_K$. We can put that system in matrix form to solve it numerically using standard methods. Define $\Delta_k = x_{k+1} - x_k$. Then $\boldsymbol{a} = \begin{bmatrix} a_1 & \cdots & a_K \end{bmatrix}'$ is given by $\boldsymbol{a} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{z}$, where:

$$\boldsymbol{z} = \begin{bmatrix} y_2 - y_1 & \cdots & y_K - y_{K-1} & s_1 & \cdots & s_K \end{bmatrix}'$$

$$\boldsymbol{B} = [\boldsymbol{B}_1 | \boldsymbol{B}_2]$$

$$\boldsymbol{B}_1 = \begin{bmatrix}
60/\Delta_1^3 & 0 & 0 & \cdots & 0 & 0 \\
-60/\Delta_1^3 & 60/\Delta_2^3 & 0 & \cdots & 0 & 0 \\
0 & -60/\Delta_2^3 & 60/\Delta_3^3 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 60/\Delta_{K-1}^3 & 0 \\
0 & 0 & 0 & \cdots & -60/\Delta_{K-1}^3 & 60/\Delta_{K-1}^3 \\
0 & 0 & 0 & \cdots & 0 & 0
\end{bmatrix}$$

$$\boldsymbol{B}_2 = \begin{bmatrix}
-36/\Delta_1^2 & -24/\Delta_1^2 & 0 & \cdots & 0 & 0 \\
24/\Delta_1^2 & 36/\Delta_1^2 - 36/\Delta_2^2 & -24/\Delta_2^2 & \cdots & 0 & 0 \\
0 & 24/\Delta_2^2 & 36/\Delta_2^2 - 36/\Delta_3^2 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 36/\Delta_{K-1}^2 - 36/\Delta_{K-1}^2 & -24/\Delta_{K-1}^2 \\
0 & 0 & 0 & \cdots & -1/\Delta_{K-1} & 1/\Delta_{K-1}
\end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix}
9/\Delta_1 & -3/\Delta_1 & 0 & \cdots & 0 & 0 \\
-3/\Delta_1 & 9/\Delta_1 + 9/\Delta_2 & -3/\Delta_2 & \cdots & 0 & 0 \\
0 & -3/\Delta_2 & 9/\Delta_2 + 9/\Delta_3 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 9/\Delta_{K-2} + 9/\Delta_{K-1} & -3/\Delta_{K-1} \\
0 & 0 & 0 & \cdots & 0 & 1
\end{bmatrix}$$

C Other comparisons of interpolation methods

We present here extended tables for the comparison of our new interpolation method with others. Those tables include a fourth interpolation method, described below, which was suggested by Cowell (2000, p. 158), yet virtually unused in the empirical literature. This method has a good pointwise performance, in many cases comparable to the generalized Pareto interpolation. However, it does not lead to a smooth quantile function or a continuous density.

We also include fiscal income in addition to pre-tax national income, as in the main text. Fiscal income tend to include a large fraction of individual with zero income, hence an important singularity near zero. To avoid that problem, we use a different tabulation in input, namely $p = 40\%, 70\%, 90\%, 99\%$.

Finally, we provide in table III the extrapolation results when the tabulation includes the top 10% and top 1%, and we seek the top 0.1%.

Method 4: piecewise Pareto distribution The method uses the Pareto distribution with information on both the thresholds and the means. It works by adjusting both the constant μ and the Pareto coefficient α of a Pareto distribution inside each bracket. The density over $[q_k, q_{k+1}]$ is:

$$f(x) = c_k x^{-\alpha_k - 1}$$

so that we get a nonlinear system of two equation with two unknowns (α_k and c_k), and two knowns $\xi_k = \int_{q_k}^{q_{k+1}} f(x)\,\mathrm{d}x$ and $\zeta_k = \int_{q_k}^{q_{k+1}} x f(x)\,\mathrm{d}x$. α_k is the solution of:

$$\frac{\alpha_k}{\alpha_k - 1} \frac{q_{k+1}^{1-\alpha_k} - q_k^{1-\alpha_k}}{q_{k+1}^{-\alpha_k} - q_k^{-\alpha_k}} = \zeta_k$$

which has no explicit solution but can be solved numerically. Then:

$$c_k = \frac{\alpha_k \xi_k}{q_{k+1}^{-\alpha_k} - q_k^{-\alpha_k}}$$

For $k = K$, so that $p_{K+1} = 1$ and $q_{K+1} = +\infty$, it becomes equivalent to method 1.

TABLE I: MEAN RELATIVE ERROR FOR DIFFERENT INTERPOLATION METHODS (FISCAL INCOME)

		mean percentage gap between estimated and observed values				
		M0	M1	M2	M3	M4
United States (1962–2014)	Top 50% share	0.042% (ref.)	0.59% (×14)	5.4% (×129)	0.035% (×0.83)	0.019% (×0.46)
	Top 20% share	0.037% (ref.)	0.34% (×9.3)	5.7% (×156)	0.021% (×0.56)	0.072% (×2)
	Top 5% share	0.11% (ref.)	1.3% (×11)	11% (×96)	0.54% (×4.8)	0.11% (×1)
	P50/average	0.57% (ref.)	14% (×25)	7.7% (×14)	0.39% (×0.68)	0.34% (×0.6)
	P80/average	0.13% (ref.)	2% (×16)	2.8% (×21)	1.2% (×9.2)	0.19% (×1.5)
	P95/average	0.42% (ref.)	6.9% (×16)	4.2% (×9.9)	1.6% (×3.7)	0.6% (×1.4)
France (1994–2012)	Top 50% share	0.055% (ref.)	0.42% (×7.6)	1.8% (×32)	0.019% (×0.34)	0.043% (×0.78)
	Top 20% share	0.032% (ref.)	0.35% (×11)	1.4% (×42)	0.02% (×0.63)	0.056% (×1.7)
	Top 5% share	0.05% (ref.)	0.35% (×6.8)	2.5% (×49)	0.43% (×8.5)	0.039% (×0.78)
	P50/average	0.48% (ref.)	7.3% (×15)	4.1% (×8.5)	0.31% (×0.65)	0.41% (×0.86)
	P80/average	0.058% (ref.)	2% (×35)	1.6% (×27)	1.1% (×18)	0.12% (×2)
	P95/average	0.11% (ref.)	1.4% (×13)	0.74% (×6.9)	2.1% (×20)	0.12% (×1.1)

Pre-tax national income. Sources: author's calculation from Piketty, Saez, and Zucman (2016) (United States) and Garbinti, Goupille-Lebret, and Piketty (2016) (France). The different interpolation methods are labeled as follows. M0: generalized Pareto interpolation. M1: constant Pareto coefficient. M2: log-linear interpolation. M3: mean-split histogram. M4: piecewise Pareto distribution. We applied them to a tabulation which includes the percentiles $p = 40\%$, $p = 70\%$, $p = 90\%$, and $p = 99\%$. We included the relative increase in the error compared to generalized Pareto interpolation in parentheses. We report the mean relative error, namely:

$$\frac{1}{\text{number of years}} \sum_{t=\text{first year}}^{\text{last year}} \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

where y is the quantity of interest (income threshold or top share), and $\hat{y}$ is its estimate using one of the interpolation methods. We calculated the results over the years 1962, 1964 and 1966–2014 in the United States, and years 1994–2012 in France.

TABLE II: MEAN RELATIVE ERROR FOR DIFFERENT INTERPOLATION METHODS
(PRE-TAX NATIONAL INCOME)

		mean percentage gap between estimated and observed values				
		M0	M1	M2	M3	M4
United States (1962–2014)	Top 70% share	0.059% (ref.)	2.3% (×38)	6.4% (×109)	0.054% (×0.92)	0.055% (×0.94)
	Top 25% share	0.093% (ref.)	3% (×32)	3.8% (×41)	0.54% (×5.8)	0.55% (×5.9)
	Top 5% share	0.058% (ref.)	0.84% (×14)	4.4% (×76)	0.83% (×14)	0.22% (×3.8)
	P30/average	0.43% (ref.)	55% (×125)	29% (×67)	1.4% (×3.3)	0.48% (×1.1)
	P75/average	0.32% (ref.)	11% (×35)	9.9% (×31)	5.8% (×18)	0.31% (×0.99)
	P95/average	0.3% (ref.)	4.4% (×15)	3.6% (×12)	1.3% (×4.5)	0.88% (×3)
France (1994–2012)	Top 70% share	0.55% (ref.)	4.2% (×7.7)	7.3% (×13)	0.14% (×0.25)	0.082% (×0.15)
	Top 25% share	0.75% (ref.)	1.8% (×2.4)	4.9% (×6.5)	0.37% (×0.49)	0.34% (×0.46)
	Top 5% share	0.29% (ref.)	1.1% (×3.9)	8.9% (×31)	0.49% (×1.7)	0.095% (×0.33)
	P30/average	1.5% (ref.)	59% (×40)	38% (×26)	2.6% (×1.8)	0.26% (×0.18)
	P75/average	1% (ref.)	5.2% (×5.1)	5.4% (×5.3)	4.7% (×4.6)	0.28% (×0.27)
	P95/average	0.58% (ref.)	5.6% (×9.6)	3.2% (×5.5)	1.8% (×3.2)	0.48% (×0.82)

Pre-tax national income. Sources: author's calculation from Piketty, Saez, and Zucman (2016) (United States) and Garbinti, Goupille-Lebret, and Piketty (2016) (France). The different interpolation methods are labeled as follows. M0: generalized Pareto interpolation. M1: constant Pareto coefficient. M2: log-linear interpolation. M3: mean-split histogram. M4: piecewise Pareto distribution. We applied them to a tabulation which includes the percentiles $p = 10\%$, $p = 50\%$, $p = 90\%$, and $p = 99\%$. We included the relative increase in the error compared to generalized Pareto interpolation in parentheses. We report the mean relative error, namely:

$$\frac{1}{\text{number of years}} \sum_{t=\text{first year}}^{\text{last year}} \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

where y is the quantity of interest (income threshold or top share), and $\hat{y}$ is its estimate using one of the interpolation methods. We calculated the results over the years 1962, 1964 and 1966–2014 in the United States, and years 1994–2012 in France.

TABLE III: MEAN RELATIVE ERROR ON THE TOP 0.1% FOR DIFFERENT EXTRAPOLATION METHODS, KNOWING THE TOP 10% AND THE TOP 1%

| | | mean percentage gap between estimated and observed values | | |
		M0	M1	M2
United States (1962–2014)	Top 0.1% share	3.4% (ref.)	4.2% (×1.2)	46% (×13)
	P99.9/average	5.5% (ref.)	4% (×0.72)	23% (×4.2)
France (1994–2012)	Top 0.1% share	1.4% (ref.)	4.5% (×3.2)	20% (×15)
	P99.9/average	1% (ref.)	2.1% (×2)	8.2% (×7.9)

Fiscal income. Sources: author's calculation from Piketty, Saez, and Zucman (2016) (United States) and Garbinti, Goupille-Lebret, and Piketty (2016) (France). The different extrapolation methods are labeled as follows. M0: generalized Pareto distribution. M1: constant Pareto coefficient. M2: log-linear interpolation. We applied them to a tabulation which includes the percentiles $p = 90\%$, and $p = 99\%$. We included the relative increase in the error compared to generalized Pareto interpolation in parentheses. We report the mean relative error, namely:

$$\frac{1}{\text{number of years}} \sum_{t=\text{first year}}^{\text{last year}} \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

where y is the quantity of interest (income threshold or top share), and $\hat{y}$ is its estimate using one of the interpolation methods. We calculated the results over the years 1962, 1964 and 1966–2014 in the United States, and years 1994–2012 in France.

D Error estimation

D.1 Decomposition of the error

Recall that the tabulation is based on $K \geq 3$ fractiles of the population $p_1, \ldots, p_K$ such that $0 \leq p_1 < \cdots < p_K < 1$. Let $k \in \{1, \ldots, K-1\}$ and $p \in [0,1]$ a fractile such that $p_k \leq p \leq p_{k+1}$. We also define $x = -\log(1-p)$.

Let n be the size of the population covered by the tabulation. Income or wealth are represented as a iid. copies $(X_1, \ldots, X_n)$ of the random variable X. The empirical quantile is $\hat{Q}_n(p) = X_{(\lfloor np \rfloor)}$ where $X_{(r)}$ is the r-th order statistic (i.e. the r-th largest value of the sample). We note $\overline{X}_n$ empirical average, and the empirical Lorenz curve is:

$$\hat{L}_n(p) = \frac{\sum_{i=1}^{\lfloor np \rfloor} X_{(i)}}{n \overline{X}_n}$$

Formally, we define the tabulation as the $(2K+1)$-tuple:

$$\boldsymbol{T}_n = [\overline{X}_n \quad \overline{X}_n \hat{L}_n(p_1) \quad \cdots \quad \overline{X}_n \hat{L}_n(p_K) \quad \hat{Q}_n(p_1) \quad \cdots \quad \hat{Q}_n(p_K)]$$

And its theoretical counterpart is:

$$\boldsymbol{T}_\infty = [\mathbb{E}[X] \quad \mathbb{E}[X]L(p_1) \quad \cdots \quad \mathbb{E}[X]L(p_K) \quad Q(p_1) \quad \cdots \quad Q(p_K)]$$

We define $\hat{\varphi}_n$ the function that we obtain through the procedure of section 3.1 on the tabulation $\boldsymbol{T}_n$. We also define $\hat{\varphi}_\infty$ the function that would be obtained with the same method on the tabulation $\boldsymbol{T}_\infty$. Then, we define $\varphi_n(x) = -\log((1 - \hat{L}_n(p))\overline{X}_n)$ the plug-in estimator of φ (hence, we may write $\varphi_\infty = \varphi$). We use analogous notations for φ'. The error at point x is:

$$\begin{aligned} e_n(x) &= \hat{\varphi}_n(x) - \varphi_n(x) \\ &= \underbrace{\hat{\varphi}_\infty(x) - \varphi_\infty(x)}_{\text{misspecification error}} + \underbrace{\hat{\varphi}_n(x) - \hat{\varphi}_\infty(x) + \varphi_\infty(x) - \varphi_n(x)}_{\text{sampling error}} \end{aligned}$$

We can set $u(x) = \hat{\varphi}_\infty(x) - \varphi_\infty(x)$ and $v_n(x) = \hat{\varphi}_n(x) - \hat{\varphi}_\infty(x) + \varphi_\infty(x) - \varphi_n(x)$, which proves the first part of theorem 5.

D.2 Misspecification error

The magnitude of the misspecification error depends on two features. First, the tightness of the tabulation in input: we can better estimate the true shape of the distribution if we have access to many percentiles of the data. Second, the "regularity" of the function we seek to

approximate: in loose terms, for the interpolation to work well, the function φ should not stray too far away from a polynomial of sufficiently low degree.

It is possible to express the misspecification error in a way that disentangle both effects. To that end, define an operator $\mathcal{E}$ which, to a function g over $[x_1, x_K]$, associates the interpolation error $\hat{g} - g$. It satisfies the three following properties:[1]

- **linearity:** $\mathcal{E}(f + \lambda g) = \mathcal{E}(f) + \lambda \mathcal{E}(g)$

- **inversion with integral sign:** $\mathcal{E}\left\{ \int_{x_1}^{x_K} f(\,\cdot\,, t) \, \mathrm{d}t \right\} = \int_{x_1}^{x_K} \mathcal{E}\{f(\,\cdot\,, t)\} \, \mathrm{d}t$

- **exact for polynomials of degree up to 2:** if $f \in \mathbb{R}[X]$, $\deg(f) \leq 2$, then $\mathcal{E}(f) = 0$

Under those conditions, the Peano kernel theorem gives a simple formula for the error term. Consider the Taylor expansion of the true function φ with integral remainder:

$$\varphi(x) = \varphi(x_1) + (x - x_1)\varphi'(x_1) + \frac{1}{2}(x - x_1)^2 \varphi''(x_1) + \frac{1}{2}\int_{x_1}^{x_K}(x - t)_+^2 \varphi'''(t) \, \mathrm{d}t$$

where $(x - t)_+ = \max\{x - t, 0\}$. Using the properties of $\mathcal{E}$, we have:

$$\hat{\varphi}_\infty(x) - \varphi_\infty(x) = \frac{1}{2}\int_{x_1}^{x_K} \mathcal{E}(K_t)(x)\varphi'''(t) \, \mathrm{d}t \tag{1}$$

where $K_t : x \mapsto (x - t)_+^2$, so that $\mathcal{E}(K_t)(x)$ is independent from φ. That last expression corresponds to the Peano kernel theorem. We get a similar expression for the first derivative $\varphi'(x)$. Therefore, setting $\varepsilon(x, t) = \mathcal{E}(K_t)(x)/2$ proves theorem 6.

The interpolation error at t can therefore be written as a scalar product between two functions. The first one, $t \mapsto \mathcal{E}(K_t)(x)$, depends only on the position of the brackets in terms the percentiles of the distribution. If the fractiles $p_1, \ldots, p_K$ included in the tabulation get more numerous and closer to each other, its value will get closer to zero. The other term, $t \mapsto \varphi'''(t)$, characterizes the regularity of the distribution. When $\varphi''' = 0$, the interpolated function is a polynomial of degree 2, so the interpolation error is zero. That is the case, in particular, of strict Pareto distributions, which the method can interpolate exactly. φ''' is best viewed as a "residual": it summarizes all the properties of the underlying distribution that are not properly captured by the functional form used in the interpolation.

We can obtain a first inequality on the absolute value of the error using the triangular inequality:

$$|\hat{\varphi}_\infty(x) - \varphi_\infty(x)| \leq \frac{||\varphi'''||_\infty}{2}\int_{x_1}^{x_K} |\mathcal{E}(K_t)(x)| \, \mathrm{d}t$$

---

[1] Many other interpolation methods would satisfy those three properties (possibly with different degrees for the polynomial), so that the results of this section could be extended to them with minimal changes.

where $||\varphi'''||_\infty = \sup\{|\varphi'''(t)| : x_1 \leq t \leq x_K\}$. That last formula is a conservative bound on the error, which would only be attained in the worst-case scenario where φ''' would frequently switch signs so as to systematically amplify the value of $\mathcal{E}(K_t)(x)$. Still, it remains interesting because we can evaluate it (using numerical integration), independently of φ, up to a multiplicative constant, and it gives insights on the shape of the error that will remain valid even after refinements. Figure 1 show this bound for a tabulation with fractiles 10%, 50%, 90% and 99%.

Error bound on $\varphi(x)$

for a tabulation with $p = 10\%, 50\%, 90\%$ and 99%

Error bound on $\varphi'(x)$

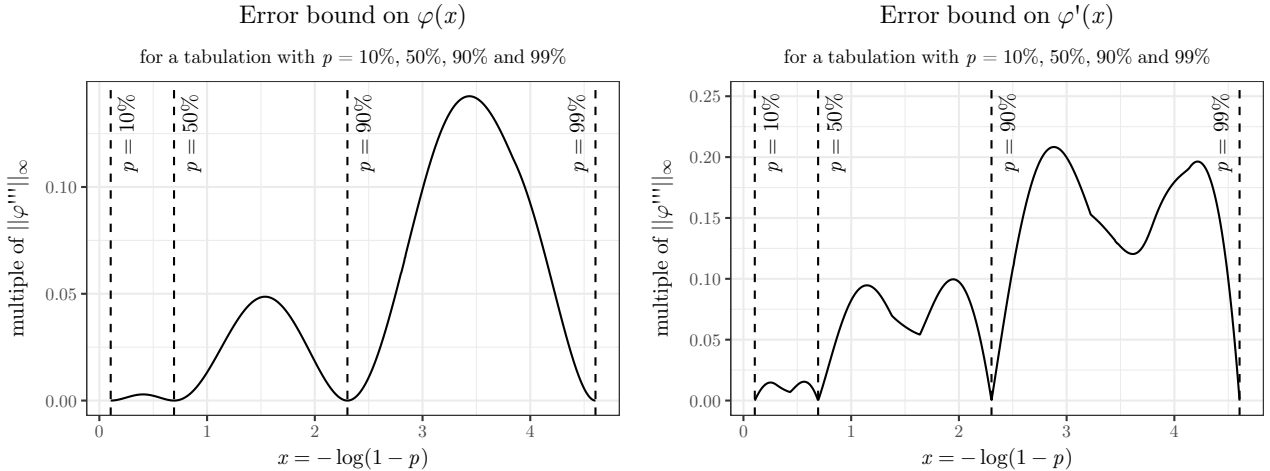for a tabulation with $p = 10\%, 50\%, 90\%$ and 99%

FIGURE 1: Bounds on the misspecification error term for φ and φ'

As expected, the error term is equal to zero for both φ and φ' at all the fractiles included in the tabulation. The error is also larger when the log-transformed bracket $[-\log(1-p_k), -\log(1-p_{k+1})]$ is wider. The overall shape of the error is quite different for φ and φ'. For φ, the error bound is bell-shaped within brackets, and its maximal value is attained near the middle of it. The error bound on φ' admits two peaks within each bracket, with the maximal error occuring somewhere near the 1/4th and the 3/4th of it. Estimates at the middle of each bracket are actually more precise than at those two values. That somewhat atypical profile is explained by the fact that the integral of φ' over $[x_k, x_{k+1}]$ is known and equal to $\varphi(x_{k+1}) - \varphi(x_k)$. Therefore, if we overestimate φ' in the first half of the bracket, we will have to underestimate it in the second half to compensate. By continuity, the error will have to equal to zero at some point, and that will happen somewhere near the middle of the interval.

Going back to the quantities that are truly of interest: the error on top shares follows the shape of the error for φ, while the error on quantiles follows the shape of the error for φ'. In

fact, for top shares, the error can be written as:

$$\left| \frac{e^{-\hat{\varphi}(x)} - e^{-\varphi(x)}}{\mathbb{E}[X]} \right| = \left| \frac{e^{-\varphi(x) - \varepsilon(x)} - e^{-\varphi(x)}}{\mathbb{E}[X]} \right|$$

$$= \left| \frac{e^{-\varphi(x)}}{\mathbb{E}[X]} (e^{-\varepsilon(x)} - 1) \right|$$

$$\approx \frac{e^{-\varphi(x)}}{\mathbb{E}[X]} |\varepsilon(x)|$$

where $\varepsilon(x)$ is the interpolation on φ. If it is small, then at the first order, the absolute error on φ corresponds to the relative error on top shares.

Because $\mathcal{E}(K_t)(x)$ only depends on known parameters, most of the work to use (1) in practice comes from φ'''. With sufficiently detailed tabulations, we can estimate it nonparametrically via local polynomial fitting on the empirical values of ϕ and ϕ'. Figure 2 shows the results, performed separately on the United States and France over all available years.
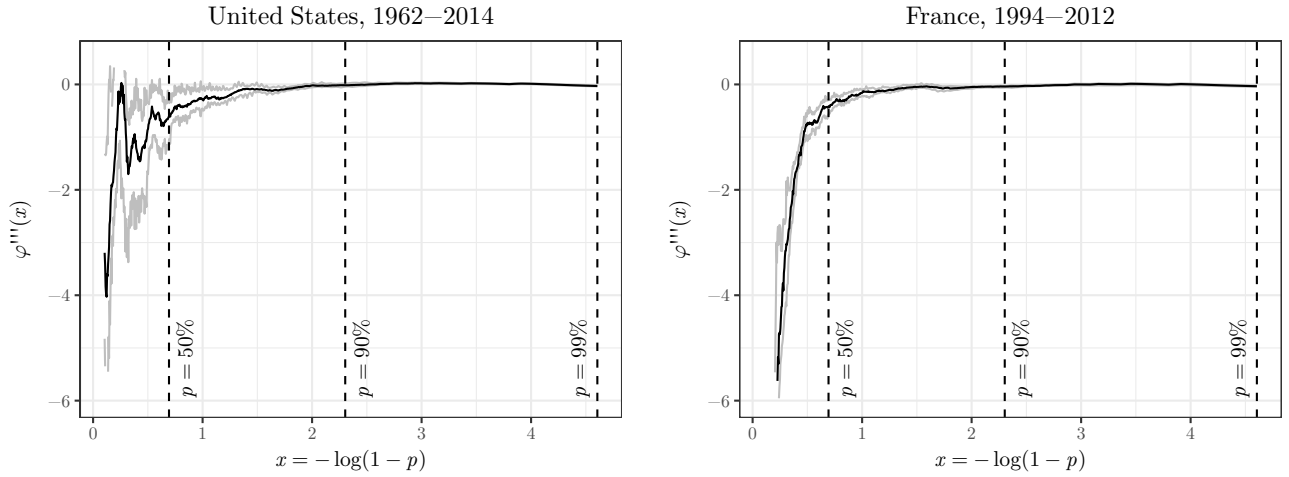
Pre-tax national income. Sources: author's computation from Piketty, Saez, and Zucman (2016) (for the United States) and Garbinti, Goupille-Lebret, and Piketty (2016) (for France). Median value over all years in black, first and tenth deciles in gray. Estimation by local polynomial fitting of degree 3 on both φ and φ' with Gaussian kernel and adaptive bandwidth so that 5% of observations are within one standard deviation of the Gaussian kernel.

FIGURE 2: Estimations of $\varphi'''(x)$

There is some similarity between both countries. The function φ''' can take relatively high values in the bottom half of the distribution, but then quickly converges to zero. Although it takes fairly different shapes in the bottom half, the shapes are actually very similar above $p = 50\%$. Within each country, there is also a certain stability over time (especially for France), as exhibited by the gray lines showing the first and the tenth decile of estimated values over all years.

D.3   Sampling error

We first prove that the sampling error converges to zero as the sample size increases (second part of theorem 5). This result is a natural consequence of the fact that the tabulation of the data used in input eventually converges to the theoretical values implied by the underlying distribution, so that the sampling error eventually vanishes.

Proof. Consider the function Θ:

$$\Theta : \quad \begin{matrix} \mathbb{R}^{2K+3} \\ \begin{bmatrix} \boldsymbol{T}_n \\ \overline{X}_n \hat{L}_n(p^*) \\ \hat{Q}_n(p^*) \end{bmatrix} \end{matrix} \quad \begin{matrix} \rightarrow \\ \\ \mapsto \end{matrix} \quad \begin{matrix} \mathbb{R}^2 \\ \begin{bmatrix} \hat{\varphi}_n(x^*) - \varphi_n(x^*) \\ \hat{\varphi}'_n(x^*) - \varphi'_n(x^*) \end{bmatrix} \end{matrix}$$

which is continuously differentiable since it is a combination of continuously differentiable functions. This function takes as input the tabulation $\boldsymbol{T}_n$, and the value of the Lorenz curve and the quantile at p^*. It returns the difference the estimated value of $\varphi_n(x^*)$ and its actual value. The sampling error, as we defined it, correspond to:

$$\Theta[\boldsymbol{T}_n, \overline{X}_n \hat{L}_n(p^*), \hat{Q}_n(p^*)] - \Theta[\boldsymbol{T}_\infty, \mathbb{E}[X]L(p^*), Q(p^*)]$$

The strong law of large number — alongside analogous results for the sample quantile — implies:

$$\boldsymbol{T}_n \xrightarrow{\text{a.s.}} \boldsymbol{T}_\infty$$
$$\overline{X}_n \hat{L}_n(p^*) \xrightarrow{\text{a.s.}} \mathbb{E}[X]L(p^*)$$
$$\hat{Q}_n(p^*) \xrightarrow{\text{a.s.}} Q(p^*)$$

Therefore, the continuous mapping theorem implies:

$$\Theta[\boldsymbol{T}_n, \overline{X}_n \hat{L}_n(p^*), \hat{Q}_n(p^*)] - \Theta[\boldsymbol{T}_\infty, \mathbb{E}[X]L(p^*), Q(p^*)] \xrightarrow{\text{a.s.}} 0$$

∎

To prove theorem 7 on the speed of convergence of the error, we need to make additional regularity assumptions on the distribution.

Assumption 1. *$f > 0$ and f' is bounded in a neighborhood of $[Q(p_1), Q(p_K)]$.*

Assumption 1 covers most relevant cases.[2] It allows the Bahadur (1966) representation of the sample quantile to hold, so that it has a regular asymptotic behavior. It also implies asymptotic normality of the trimmed mean (Stigler, 1973), and, by extension, of the Lorenz curve.

---

[2]There is one situation where that assumption may seem problematic, namely if $p_1 = 0$ and the distribution has a finite lower bound. However, in such cases, the value of $Q(p_1)$ is generally known *a priori* (typically, because we assumed income is nonnegative) and is therefore not subject to sampling variability. Hence, it will not affect the results.

Next, we distinguish two cases, depending on the tail behavior of the distribution. If it has a sufficiently thin upper tail, then the distribution will have finite variance. But it is common for the distribution of income (and *a fortiori* wealth) to have much fatter upper tails, leading to infinite variance. This distinction has important consequences for the asymptotic behavior of the sample mean, and by extension of our estimator.

D.3.1 The finite variance case

For a strict power law, finite variance corresponds to $b(p) > 2$. More generally, we can state the finite variance assumption using the second-order moment.

Assumption 2. *X has a finite second-order moment, i.e.* $\mathbb{E}[X^2] < +\infty$.

When variance is finite, the central limit theorem applies. Hence, we get the standard result of asymptotic normality and convergence rate $n^{-1/2}$ using the delta method.

Proof. We start by deriving the asymptotic joint distribution of all the quantiles and the means in the tabulation, which is multivariate normal. Then, theorem 7 for finite variance follows from the delta method applied to Θ.

For $k, k_1, k_2 \in \{1, \ldots, K+1\}$, define:

$$\mu_k = \mathbb{E}[U_i^k]$$
$$\sigma_k^2 = \mathrm{Var}(U_i^k)$$
$$\sigma_{k_1, k_2} = \mathrm{Cov}(U_i^{k_1}, U_i^{k_2})$$

We will also use the following matrix notations:

$$\boldsymbol{U}_i = [U_i^1 \quad U_i^2 \quad \cdots \quad U_i^{K+1}]'$$
$$\boldsymbol{V}_i = [V_i^1 \quad V_i^2 \quad \cdots \quad V_i^{K+1}]'$$
$$\boldsymbol{\mu} = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_{K+1}]'$$

We start with a lemma that gives the joint asymptotic distribution of $\overline{\boldsymbol{U}}_n$ and $\overline{\boldsymbol{V}}_n$.

Lemma 1.

$$\sqrt{n} \begin{bmatrix} \overline{\boldsymbol{U}}_n - \boldsymbol{\mu} \\ \overline{\boldsymbol{V}}_n - \boldsymbol{p} \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C}' & \boldsymbol{B} \end{bmatrix}\right)$$

where:

$$\boldsymbol{A} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,K+1} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K+1,1} & \sigma_{K+1,2} & \cdots & \sigma_{K+1}^2 \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} \tilde{p}_1(1-\tilde{p}_1) & \tilde{p}_1(1-\tilde{p}_2) & \cdots & \tilde{p}_1(1-\tilde{p}_{K+1}) \\ \tilde{p}_1(1-\tilde{p}_2) & \tilde{p}_2(1-\tilde{p}_2) & \cdots & \tilde{p}_2(1-\tilde{p}_{K+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_1(1-\tilde{p}_{K+1}) & \tilde{p}_2(1-\tilde{p}_{K+1}) & \cdots & \tilde{p}_{K+1}(1-\tilde{p}_{K+1}) \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} -\tilde{p}_1\mu_1 & \mu_1(1-\tilde{p}_2) & \cdots & \mu_1(1-\tilde{p}_K) & \mu_1(1-\tilde{p}_{K+1}) \\ -\tilde{p}_1\mu_2 & -\tilde{p}_2\mu_2 & \cdots & \mu_2(1-\tilde{p}_K) & \mu_2(1-\tilde{p}_{K+1}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\tilde{p}_1\mu_K & -\tilde{p}_2\mu_K & \cdots & -\tilde{p}_K\mu_K & \mu_K(1-\tilde{p}_{K+1}) \\ -\tilde{p}_1\mu_{K+1} & -\tilde{p}_2\mu_{K+1} & \cdots & -\tilde{p}_{K+1}\mu_{K+1} & -\tilde{p}_{K+1}\mu_{K+1} \end{bmatrix}$$

Proof. We have, for $i \in \{1, \ldots, n\}$ and $k_1, k_2 \in \{1, \ldots, K+1\}$ with $k_1 < k_2$:

$$\mathbb{E}[V_i^{k_1}] = p_{k_1}$$

$$\mathrm{Var}(V_i^{k_1}) = p_{k_1}(1 - p_{k_1})$$

$$\mathrm{Cov}(V_i^{k_1}, V_i^{k_2}) = p_{k_1}(1 - p_{k_2})$$

and for $k_1, k_2 \in \{1, \ldots, K+1\}$:

$$\mathrm{Cov}(U_i^{k_1}, V_i^{k_2}) = \begin{cases} -p_{k_2}\mu_{k_1} & \text{if} \quad k_2 \leq k_1 \\ \mu_{k_1}(1 - p_{k_2}) & \text{if} \quad k_2 > k_1 \end{cases}$$

Therefore, by the central limit theorem:

$$\sqrt{n} \begin{bmatrix} \overline{\boldsymbol{U}}_n - \boldsymbol{\mu} \\ \overline{\boldsymbol{V}}_n - \boldsymbol{p} \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C}' & \boldsymbol{B} \end{bmatrix}\right)$$

with $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ defined as in lemma 1. ∎

We know define for all $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, K\}$:

$$W_i^k = X_i \mathbb{1}\{\hat{Q}_n(p_k) < X_i \leq \hat{Q}_n(p_{k+1})\}$$

and for $k = K+1$:

$$W_i^{K+1} = X_i \mathbb{1}\{\hat{Q}_n(p_{K+1}) < X_i\}$$

and in matrix form:

$$\boldsymbol{W}_i = [W_i^1 \quad \cdots \quad W_i^{K+1}]$$

15

The definition of $\boldsymbol{W}_i$ is similar to that of $\boldsymbol{U}_i$, except that the quantile function was replaced by its empirical counterpart. We may now prove a second lemma, which correspond to the joint asymptotic distribution of $\overline{\boldsymbol{W}}_n$ and $\hat{\boldsymbol{Q}}_n$.

Lemma 2.

$$\sqrt{n}\begin{bmatrix}\overline{\boldsymbol{W}}_n - \boldsymbol{\mu} \\ \hat{\boldsymbol{Q}}_n - \boldsymbol{q}\end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix}\boldsymbol{A} + \boldsymbol{M}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{M}' + \boldsymbol{M}\boldsymbol{B}\boldsymbol{M}' & -\boldsymbol{C}\boldsymbol{N} - \boldsymbol{M}\boldsymbol{B}\boldsymbol{N} \\ -\boldsymbol{N}\boldsymbol{C}' - \boldsymbol{N}\boldsymbol{B}\boldsymbol{M}' & \boldsymbol{N}\boldsymbol{B}\boldsymbol{N}\end{bmatrix}\right)$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are defined as in lemma 1, $\boldsymbol{N} = \mathrm{diag}(1/f(\tilde{q}_1), \ldots, 1/f(\tilde{q}_{K+1}))$, and:

$$\boldsymbol{M} = \begin{bmatrix} \tilde{q}_1 & -\tilde{q}_2 & 0 & \cdots & 0 & 0 \\ 0 & \tilde{q}_2 & -\tilde{q}_3 & \cdots & 0 & 0 \\ 0 & 0 & \tilde{q}_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \tilde{q}_K & -\tilde{q}_{K+1} \\ 0 & 0 & 0 & \cdots & 0 & \tilde{q}_{K+1} \end{bmatrix}$$

Proof. Note that for $k \in \{1, \ldots, K\}$:

$$\overline{W}_n^k = \frac{1}{n}\sum_{i=\lfloor np_k \rfloor + 1}^{\lfloor np_{k+1} \rfloor} X_{(i)}$$

$$= \frac{1}{n}\left[\sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} X_{(i)} + \sum_{i=n\overline{V}_n^k + 1}^{n\overline{V}_n^{k+1}} X_{(i)} + \sum_{i=n\overline{V}_n^{k+1}+1}^{\lfloor np_{k+1} \rfloor} X_{(i)}\right]$$

$$= \overline{U}_n^k + \frac{1}{n}\left[\sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} X_{(i)} + \sum_{i=n\overline{V}_n^{k+1}+1}^{\lfloor np_{k+1} \rfloor} X_{(i)}\right]$$

Where $\sum_{i=a}^{b} x$ should be understood as $-\sum_{i=b}^{a} x$ if $a > b$. Therefore:

$$\sqrt{n}(\overline{W}_n^k - \mu_k) = \sqrt{n}(\overline{U}_n^k - \mu_k) + \frac{1}{\sqrt{n}}\left[\sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} X_{(i)} + \sum_{i=n\overline{V}_n^{k+1}+1}^{\lfloor np_{k+1} \rfloor} X_{(i)}\right]$$

We have:

$$\frac{1}{\sqrt{n}}\sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} X_{(i)} = \frac{1}{\sqrt{n}}\sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} (X_{(i)} - q_k + q_k)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} (X_{(i)} - q_k) + q_k\sqrt{n}\left(\overline{V}_n^k - \frac{\lfloor np_k \rfloor}{n}\right)$$

The first term converges in probability to zero because:

$$\left|\frac{1}{\sqrt{n}}\sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} (X_{(i)} - q_k)\right| \leq \left|\frac{n\overline{V}_n^k - \lfloor np_k \rfloor}{\sqrt{n}}\right| \max\{|X_{(\lfloor np_k \rfloor + 1)} - q_k|, |X_{(n\overline{V}_n^k)} - q_k|\}$$

16

where the first term is bounded in probability, $|X_{(\lfloor np_k \rfloor + 1)} - q_k| \xrightarrow{\mathbb{P}} 0$ and $|X_{(n\overline{V}_n^k)} - q_k| \xrightarrow{\mathbb{P}} 0$.

Hence:

$$\frac{1}{\sqrt{n}} \sum_{i=\lfloor np_k \rfloor + 1}^{n\overline{V}_n^k} X_{(i)} = q_k \sqrt{n}(\overline{V}_n^k - p_k) + \sqrt{n}\left(p_k - \frac{\lfloor np_k \rfloor}{n}\right) + o(1)$$

$$= q_k \sqrt{n}(\overline{V}_n^k - p_k) + o(1)$$

Similarly:

$$\frac{1}{\sqrt{n}} \sum_{i=n\overline{V}_n^{k+1}+1}^{\lfloor np_{k+1} \rfloor} X_{(i)} = -q_{k+1} \sqrt{n}(\overline{V}_n^{k+1} - p_{k+1}) + o(1)$$

Therefore:

$$\sqrt{n}(\overline{W}_n^k - \mu_k) = \sqrt{n}(\overline{U}_n^k - \mu_k) + q_k \sqrt{n}(\overline{V}_n^k - p_k) - q_{k+1} \sqrt{n}(\overline{V}_n^{k+1} - p_{k+1}) + o(1)$$

By similar arguments:

$$\sqrt{n}(\overline{W}_n^{K+1} - \mu_{K+1}) = \sqrt{n}(\overline{U}_n^{K+1} - \mu_{K+1}) + q_{K+1} \sqrt{n}(\overline{V}_n^{K+1} - p_{K+1}) + o(1)$$

Hence, in matrix notation:

$$\sqrt{n}(\boldsymbol{\overline{W}}_n - \boldsymbol{\mu}) = \sqrt{n}(\boldsymbol{\overline{U}}_n - \boldsymbol{\mu}) + \boldsymbol{M}\sqrt{n}(\boldsymbol{\overline{V}}_n - \boldsymbol{p}) + o(1)$$

The Bahadur (1966) representation of the quantile implies:

$$\boldsymbol{\hat{Q}}_n - \boldsymbol{q} = -\boldsymbol{N}(\boldsymbol{\overline{V}}_n - \boldsymbol{p}) + o(n^{-1/2})$$

Therefore, we have:

$$\sqrt{n} \begin{bmatrix} \boldsymbol{\overline{W}}_n - \boldsymbol{\mu} \\ \boldsymbol{\hat{Q}}_n - \boldsymbol{q} \end{bmatrix} = \sqrt{n} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{M} \\ \boldsymbol{0} & -\boldsymbol{N} \end{bmatrix} \begin{bmatrix} \boldsymbol{\overline{U}}_n - \boldsymbol{\mu} \\ \boldsymbol{\overline{V}}_n - \boldsymbol{p} \end{bmatrix} + o(1)$$

Using lemma 1, we get:

$$\sqrt{n} \begin{bmatrix} \boldsymbol{\overline{W}}_n - \boldsymbol{\mu} \\ \boldsymbol{\hat{Q}}_n - \boldsymbol{q} \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{A} + \boldsymbol{MC'} + \boldsymbol{CM'} + \boldsymbol{MBM'} & -\boldsymbol{CN} - \boldsymbol{MBN} \\ -\boldsymbol{NC'} - \boldsymbol{NBM'} & \boldsymbol{NBN} \end{bmatrix}\right)$$

with $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ defined as in lemma 1. ∎

Notice that:

$$\forall k \in \{1, \ldots, K+1\} \qquad \overline{X}_n \hat{L}_n(\tilde{p}_k) = \sum_{\ell=k}^{K+1} \overline{W}_n^\ell \qquad \text{and} \qquad \mathbb{E}[X]L(\tilde{p}_k) = \sum_{\ell=k}^{K+1} \mu_\ell$$

Therefore, we can write in matrix form that $\hat{\boldsymbol{L}}_n = \boldsymbol{P}\overline{\boldsymbol{W}}_n$ where:

$$\hat{\boldsymbol{L}}_n = [\overline{X}_n\hat{L}_n(\tilde{p}_1) \quad \cdots \quad \overline{X}_n\hat{L}_n(\tilde{p}_K)]$$

and $\boldsymbol{P}$ is the upper triangular matrix with only ones. Define $\nabla\Theta$ the gradient of Θ expressed at $[\boldsymbol{T}_\infty, \mathbb{E}[X]L(p^*), Q(p^*)]$, and:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}$$

Denote $\boldsymbol{S}$ the covariance matrix of lemma 2, and $\boldsymbol{\Sigma} = (\nabla\Theta)'\boldsymbol{R}'\boldsymbol{S}\boldsymbol{R}(\nabla\Theta)$. The delta method (van der Vaart, 2000, p. 25) then implies:

$$\sqrt{n}(\Theta[\boldsymbol{T}_n, \overline{X}_n\hat{L}_n(p^*), \hat{Q}_n(p^*)] - \Theta[\boldsymbol{T}_\infty, \mathbb{E}[X]L(p^*), Q(p^*)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\Sigma})$$

∎

D.3.2 The infinite variance case

When $\mathbb{E}[X^2] = +\infty$, the standard central limit theorem does not apply anymore. There is, however, a generalization due to Gnedenko and Kolmogorov (1968) that works with infinite variance. There are two main differences with the standard central limit theorem. The first one is that convergence operates more slowly than $n^{-1/2}$, the speed being determined by the asymptotic power law behavior of the distribution (the fatter the tail, the slower the convergence). The second one is that the limiting distribution belongs to a larger family than just the Gaussian, called *stable distributions*. With the exception of the Gaussian — which is an atypical member of the family of stable distributions — stable distributions exhibit fat tails and power law behavior. In most cases, their probability density function cannot be expressed analytically: only their characteristic function can. Although we designed our interpolation method with power laws in mind, we did not actually restrict the asymptotic behavior of the distribution, until now. But to apply the generalized central limit theorem, we need to make such an assumption explicitly (Uchaikin and Zolotarev, 1999, p. 62).

Assumption 3. $1 - F(x) \sim Cx^{-\alpha}$ *as* $x \to +\infty$ *for* $1 < \alpha \le 2$ *and* $C > 0$.

Assumption 3 implies that X is an asymptotic power law, but it is a little more restrictive than definition 2: instead of assuming that $x \mapsto L(x)$ in definition 2 is *slowly varying*, we make the stronger assumption that it *converges to a constant*. It still covers a vast majority of cases. We limit ourselves to situations where $1 < \alpha \le 2$, since when $\alpha > 2$ we are back to the finite variance case, and when $\alpha \le 1$ the mean is infinite.

Proof. We use the generalized central limit theorem of Gnedenko and Kolmogorov (1968), which gives the asymptotic distribution of the sample mean when $\mathbb{E}[X^2] = +\infty$. Once we apply this theorem, the rest of the proof becomes simpler than with finite variance. Indeed, with infinite variance, the sample mean in the last bracket converges slower than $1/\sqrt{n}$, while quantiles and means in other brackets still converge at the same speed. Therefore, asymptotically, there is only one source of statistical variability that eventually dominates all the other. Hence, we need not be concerned by, say, the joint distribution of the quantiles, because at the first order that distribution will be identically zero. This insight leads to the following lemma.

Lemma 3.

$$r_n \begin{bmatrix} \overline{\boldsymbol{U}}_n - \boldsymbol{\mu} \\ \overline{\boldsymbol{V}}_n - \boldsymbol{p} \end{bmatrix} \xrightarrow{\mathcal{D}} \boldsymbol{S}$$

where:

$$r_n = \begin{cases} n^{1-1/\alpha} & \text{if } 1 < \alpha < 2 \\ (n/\log n)^{1/2} & \text{if } \alpha = 2 \end{cases}$$

$$S_i = \begin{cases} \gamma Y & \text{if } i = K+1 \\ 0 & \text{otherwise} \end{cases}$$

and Y is a stable distribution with the characteristic function:

$$g(t) = \exp(-|t|^\alpha[1 - i\tan(\alpha\pi/2)\text{sign}(t)])$$

and:

$$\gamma = \begin{cases} \left(\frac{\pi C}{2\Gamma(\alpha)\sin(\alpha\pi/2)}\right)^{1/\alpha} & \text{if } 1 < \alpha < 2 \\ \sqrt{C} & \text{if } \alpha = 2 \end{cases}$$

Proof. Standard results on quantiles (David and Nagaraja, 2005) and the trimmed mean (Stigler, 1973) imply that quantiles and means in middle bracket converge in distribution at speed $1/\sqrt{n}$. Because $r_n = o(\sqrt{n})$, they converge to zero in probability when multiplied by r_n. Hence, the only nonzero term in $\boldsymbol{S}$ correspond to $\overline{U}_n^{K+1}$, which converges to γY according to the generalized central limit theorem (Uchaikin and Zolotarev, 1999, p. 62). ∎

We now move from the asymptotic distribution of $\overline{\boldsymbol{U}}_n$ and $\overline{\boldsymbol{V}}_n$ to the asymptotic distribution of $\overline{\boldsymbol{W}}_n$ and $\hat{\boldsymbol{Q}}_n$, as we did in the previous section. Except that now both distributions are the same, because the disturbances introduced by quantiles and middle bracket averages are asymptotically negligible.

Lemma 4.

$$r_n \begin{bmatrix} \overline{\boldsymbol{W}}_n - \boldsymbol{\mu} \\ \hat{\boldsymbol{Q}}_n - \boldsymbol{q} \end{bmatrix} \xrightarrow{\mathcal{D}} \boldsymbol{S}$$

with the same notations as in lemma 3.

Proof. Using the same method as in the proof of lemma 2, we get:

$$r_n(\overline{\boldsymbol{W}}_n - \boldsymbol{\mu}) = r_n(\overline{\boldsymbol{U}}_n - \boldsymbol{\mu}) + o(1)$$

Moreover, the Bahadur (1966) representation of the quantile implies:

$$r_n(\hat{\boldsymbol{Q}}_n - \boldsymbol{q}) = o(1)$$

Using lemma 3 give the result. ∎

We may now apply the delta method as we did in the previous section. We are in a somewhat non standard case because the convergence operates more slowly than $\sqrt{n}$, and the asymptotic distribution is not Gaussian, but the basic idea of the delta method applies nonetheless. We get:

$$r_n(\Theta[\boldsymbol{T}_n, \overline{X}_n \hat{L}_n(p^*), \hat{Q}_n(p^*)] - \Theta[\boldsymbol{T}_\infty, \mathbb{E}[X]L(p^*), Q(p^*)]) \xrightarrow{\mathcal{D}} (\nabla\Theta)\boldsymbol{R}\boldsymbol{S}$$

which proves the result of theorem 7 for infinite variance. ∎

The precise parameters of the stable distribution and the constants γ_1, γ_2 are given in appendix alongside the proof. For practical purposes, that theorem requires in particular the estimation of α and C. Using the generalized Pareto distribution model with parameters ξ, σ, μ as in section 3.3, we have:

$$\alpha = 1/\xi$$
$$C = (1-p)(\xi/\sigma)^{-1/\xi}$$

for $\xi \geq 1/2$ (if $\xi < 1/2$, variance is finite). If we have access to individual data, at the very top of the distribution, it is possible to use better estimates of α and C, using the large literature on the subject in extreme value theory (Haan and Ferreira, 2006, pp. 65–126).[3]

The infinite variance approximation is a rougher than the finite variance approximation for two reasons. First, because it relies on parameters, such as the asymptotic Pareto coefficient, which are harder to estimate than variances or covariances. Second, because it makes a first-order approximation which is less precise. That is, it assumes that all of the error comes from

---

[3]For example, wealth rankings such as the *Forbes 400* can give the wealth of a country's richest individuals. See Blanchet (2016).

mean of the top bracket (which converges at speed $1/r_n$), and none from the quantiles or the mean of the lower brackets (which converge at speed $n^{-1/2}$). Although that is asymptotically true, because r_n grows more slowly than $n^{1/2}$, it is possible that the second-order term is not entirely negligible for finite n. Still, it gives a good idea of the magnitude of the error.
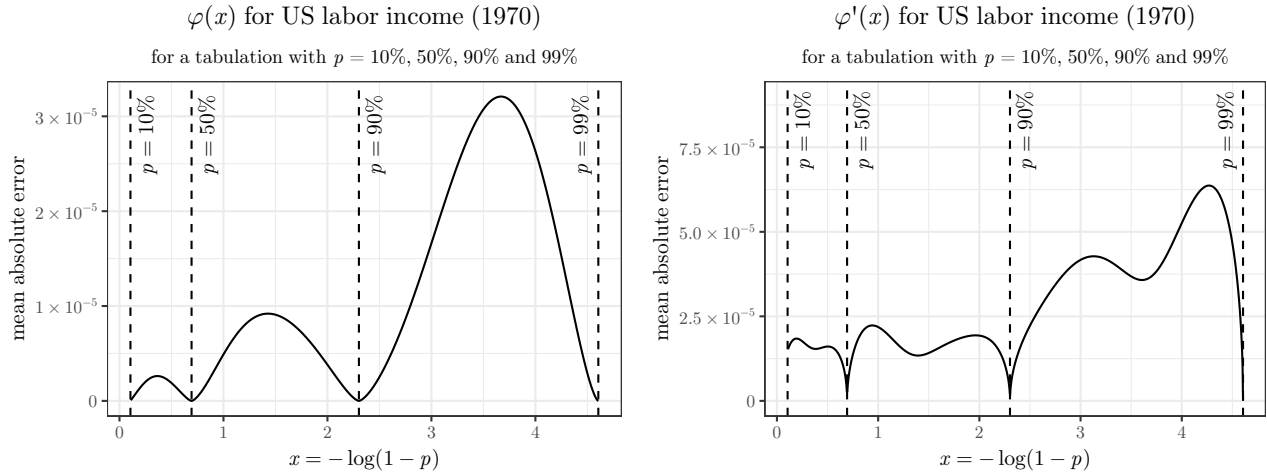
D.3.3 Comparison

FIGURE 3: Asymptotic mean absolute value of the sampling error with finite variance
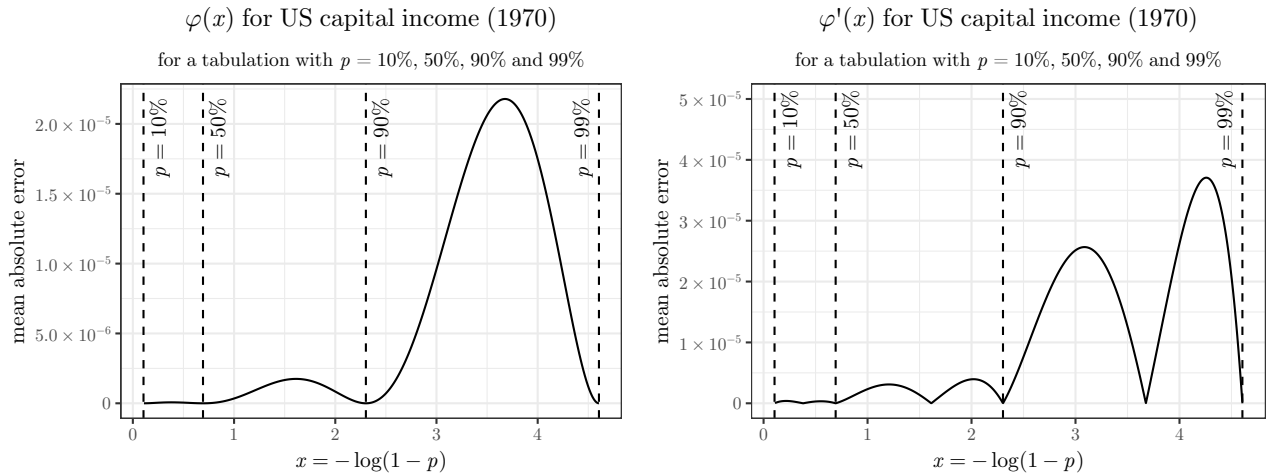

FIGURE 4: Asymptotic mean absolute value of the sampling error with infinite variance

To observe theorems 7 in practice, we turn to the distribution of labor and capital income in the United States in 1970. Back then, labor income inequality was low enough so that the asymptotic inverted Pareto coefficient was comfortably below 2 (somewhere between 1.4 and 1.6), which means that the distribution has finite variance. Capital income, on the other hand,

21

was as always more unequally distributed, so that its asymptotic inverted Pareto coefficient appeared to be above 2.3, which implies infinite variance.

Figures 3 and 4 apply theorem 7 to the distribution of labor and capital income in the United States. The patterns are reminiscent of what we observed for the misspecification error: a bell-shaped, single-peaked error in each bracket for φ, and a double-peaked error for φ'.

D.4   Comparing Misspecification with Sampling Error
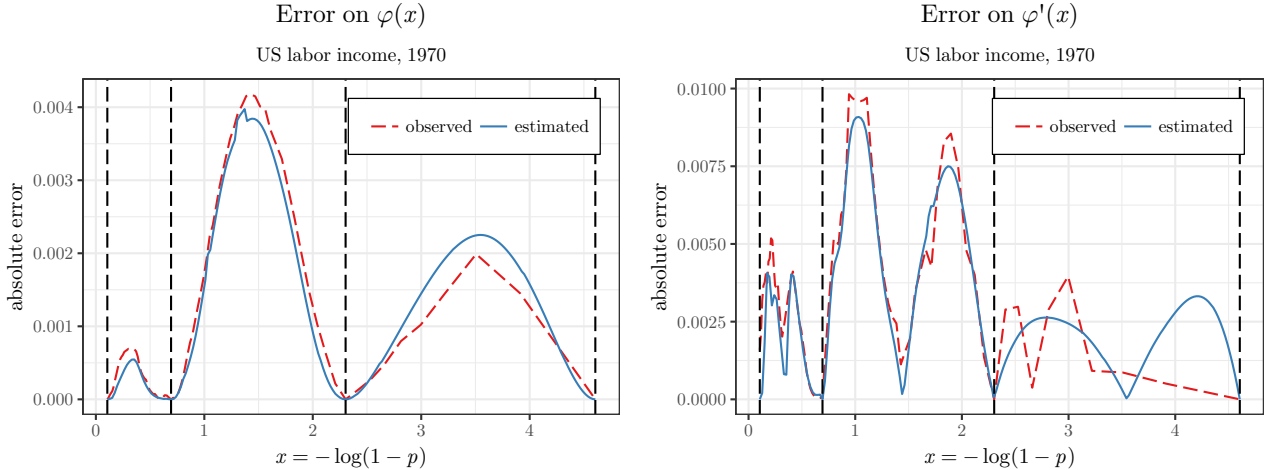
Pre-tax national income. Sources: author's computations from Piketty, Saez, and Zucman (2016). The solid blue line correspond to the misspecification error estimated with formula (1) and a nonparametric estimate of φ'''. The dashed red line correspond to the actual, observed error. We smoothed the observed error for φ' using the Nadaraya-Watson kernel estimator to remove excessive variability due to rounding.

FIGURE 5: Actual error and estimated misspecification error

The misspecification error largely dominates the sampling error given the sample sizes that are typical of tax tabulations. To see this, we may go back to the previous example of the US distribution of labor income in 1970. Figure 5 shows the misspecification error is this case estimated using formula (1) and a nonparametric estimate of φ''', alongside the actual, observed error. There is some discrepancy between both figures, largely due to the fact that φ''' cannot be estimated perfectly. Yet the estimated misspecification error appear to be a fairly good estimate of the actual error overall.

We may then look at figure 3 to see how the sampling error compares. At its highest, it reaches to 3.5×10^{-5} for φ and 7.5×10^{-5} for φ'. The misspecification error is several orders of magnitude higher, around 10^{-3}. Even if the population was 100 times smaller, the magnitude of the mean absolute deviation of the error would be multiplied by $\sqrt{100} = 10$, so it would remain

an order of magnitude lower. The figures would be similar for other years, countries or income concept. In practice, we can confidently neglect the sampling error.

E   Dynamic Model of Income Growth an Wealth Accumulation

We prove simultaneously the theorems 8 and 9. We assume that $\mu(x) \to \mu$, $\sigma^2(x) \to \sigma^2$. Recall that the process X_t follows the stochastic differential equation:

$$\frac{\mathrm{d}X_t}{X_t} = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$$

which means that the evolution of its density $f(x,t)$ is described by the Fokker-Planck equation:

$$\frac{\partial}{\partial t}f(x,t) = -\frac{\partial}{\partial x}[x\mu(x)f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[x^2\sigma^2(x)f(x,t)] \tag{2}$$

We also write:

$$\zeta(x) = 1 - \frac{2\mu(x)}{\sigma^2(x)}$$

and $\zeta = \lim_{x \to +\infty} \zeta(x) = 1 - 2\mu/\sigma^2$. For the stationary distribution $f(x)$, we have $\frac{\partial}{\partial t}f(x) = 0$, so equation (2) implies:

$$0 = -\frac{\partial}{\partial x}[x\mu(x)f(x)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[x^2\sigma^2(x)f(x)]$$

We can integrate that equation into:

$$x\mu(x)f(x) = \frac{1}{2}\frac{\partial}{\partial x}[x^2\sigma^2(x)f(x)]$$

$$= [x\sigma^2(x) + x^2\sigma(x)\sigma'(x)]f(x) + \frac{1}{2}x^2\sigma^2(x)f'(x)$$

Reordering terms, we get:

$$\frac{f'(x)}{f(x)} = -\frac{\zeta+1}{x} - \frac{2\sigma'(x)}{\sigma(x)} - \frac{\zeta(x)-\zeta}{x}$$

And after integration:

$$f(x) \propto x^{-\zeta-1} \exp\left(-\log(\sigma^2(x)) - \int_1^x \frac{\zeta(t)-\zeta}{t}\,\mathrm{d}t\right)$$

Rewrite that expression as $f(x) = L(x)x^{-\zeta-1}$. Because $x \mapsto \zeta(x) - \zeta$ converges to zero and $x \mapsto \sigma^2(x)$ converges to a positive constant, Karamata's (1930) representation theorem (Bingham, Goldie, and Teugels, 1989, p. 12) implies that L is slowly varying. Then, we can use the following property of slowly varying functions:

$$1 - F(x) = \int_x^{+\infty} L(t)t^{-\zeta-1}\,\mathrm{d}t \sim L(x)\frac{x^{-\zeta}}{-\zeta}$$

to see that the stationary distribution is in fact an asymptotic power law.   ∎

References

Bahadur, R. R. (1966). "A Note on Quantiles in Large Samples". In: *The Annals of Mathematical Statistics* 37.3, pp. 577–580. URL: http://dx.doi.org/10.1214/aoms/1177699450.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press. ISBN: 9780521379434.

Blanchet, Thomas (2016). "Wealth inequality in Europe and the United States: estimates from surveys, national accounts and wealth rankings". MA thesis. Paris School of Economics.

Cowell, Frank A. (2000). *Measuring Inequality*. LSE Economic Series. Oxford University Press.

David, H. A. and H. N. Nagaraja (2005). *Order Statistics*. John Wiley & Sons, Inc. ISBN: 9780471722168. DOI: 10.1002/0471722162.ch10.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2016). "Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)". In: WID.world Working Paper. URL: http://wid.world/document/b-garbinti-j-goupille-and-t-piketty-inequality-dynamics-in-france-1900-2014-evidence-from-distributional-national-accounts-2016/.

Gnedenko, B.V. and A.N. Kolmogorov (1968). *Limit distributions for sums of independent random variables*. Addison-Wesley series in statistics. Addison-Wesley.

Haan, L. de and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer New York. ISBN: 9780387239460. URL: http://www.springer.com/us/book/9780387239460.

Karamata, J. (1930). "Sur un mode de croissance regulière des fonctions". In: *Mathematica* 4.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2016). "Distributional National Accounts: Methods and Estimates for the United States". In: Working Paper Series 22945. DOI: 10.3386/w22945. URL: http://www.nber.org/papers/w22945.

Stigler, Stephen M. (1973). "The Asymptotic Distribution of the Trimmed Mean". In: *The Annals of Statistics* 1.3, pp. 472–477. URL: http://dx.doi.org/10.1214/aos/1176342412.

Uchaikin, V.V. and V.M. Zolotarev (1999). *Chance and Stability: Stable Distributions and their Applications*. Modern Probability and Statistics. De Gruyter. ISBN: 9783110935974. URL: http://staff.ulsu.ru/uchaikin/uchzol.pdf.

van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN: 9780521784504.