

CAPITAL AND LABOR INCOME PARETO EXPONENTS ACROSS TIME AND SPACE

TJEERD DE VRIES* and ALEXIS AKIRA TODA

University of California San Diego

We estimate capital and labor income Pareto exponents across 475 country-year observations that span 52 countries over half a century (1967–2018). We document two stylized facts: (i) capital income is more unequally distributed than labor income in the tail; namely, the capital exponent (1–3, median 1.46) is smaller than labor (2–5, median 3.35), and (ii) capital and labor exponents are nearly uncorrelated. To explain these findings, we build an incomplete market model with job ladders and capital income risk that gives rise to a capital income Pareto exponent smaller than but nearly unrelated to the labor exponent. Our results suggest the importance of distinguishing income and wealth inequality.

JEL codes: C46, D15, D31, D52

Keywords: income fluctuation problem, inequality, power law

1. INTRODUCTION

The purpose of this paper is to estimate and document the Pareto exponents for capital and labor incomes separately for as many countries and years as possible. We say that a positive random variable X obeys a power law with Pareto exponent $\alpha > 0$ if the tail probability decays like a power function: $P(X > x) \sim x^{-\alpha}$ for large x .¹ In the context of the income distribution, the Pareto exponent characterizes the tail heaviness of high incomes and hence top tail inequality. We remain agnostic about the shape of the income distribution away from the tail. Our study is motivated by the following two observations. First, we are not aware of a comprehensive study that documents the capital and labor income Pareto exponents separately for many countries and years, despite their importance. Second, the Pareto exponent has desirable properties relative to other popular inequality measures such as the Gini coefficient or top income shares.

Consider the first point. Conceptually, capital and labor incomes are different entities. While the former is the return for providing capital (wealth), the latter is the return for providing labor services, and there is no particular reason to expect

Note: We thank Xavier Gabaix, Yannick Hoga, and Makoto Nirei for comments and suggestions. The replication files for this article are available at https://github.com/alexisakira/Pareto_time_space.

*Correspondence to: Tjeerd de Vries, Department of Economics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0021, USA (tjdevrie@ucsd.edu).

¹More precisely, we say that a random variable X has a Pareto upper tail with exponent $\alpha > 0$ if $P(X > x) = x^{-\alpha} \ell(x)$ for some slowly varying function ℓ . A function $\ell: (0, \infty) \rightarrow \mathbb{R}$ is said to be *slowly varying (at infinity)* if it is nonzero for sufficiently large x and $\lim_{x \rightarrow \infty} \ell(tx)/\ell(x) = 1$ for each $t > 0$. See Bingham *et al.* (1987) for a comprehensive treatment of the theory of regular variation.

a relation between the two. Although these two forms of income are conceptually distinct, it is often put together as just “income” and discussed in the context of inequality and related policies. If capital and labor incomes are quantitatively different, a policy design based on total income may be misleading. To give one example, consider the theory of optimal taxation (Saez, 2001), where the income Pareto exponent plays an important role. Saez and Stantcheva (2018) carefully distinguish capital and labor incomes and apply the theory of optimal taxation in the United States. They find that with an income elasticity of $e = 0.5$, the optimal top marginal tax rate is about 50 percent for labor and 60 percent for capital (see their Figure 5). This difference directly comes from the fact that capital and labor income Pareto exponents are distinct. Thus, distinguishing capital and labor income inequality is potentially important for policy designs.

Consider the second point about the desirable properties of Pareto exponents. In the applied literature such as Piketty (2003) and Piketty and Saez (2003), top income shares (such as the top 1 percent income share) are more commonly reported than the income Pareto exponent, perhaps because top shares are summary statistics that can be computed without specifying functional forms or can be understood by non-experts without special knowledge of statistics. However, Atkinson (2005) documents methodological problems regarding the cross-country comparison of top income shares, citing the differences in tax units (e.g., individual or household) and legislation (e.g., whether social security benefits are taxable). One of the reasons such issues arise is because it is not always clear how to define the population and measure small units.² Because common inequality measures such as the Gini coefficient and top income shares require the knowledge of the entire distribution, these quantities are greatly affected by the definition and measurement of small units. Using the Pareto exponents significantly alleviates these definition and measurement issues because the Pareto distribution is scale invariant (see Jessen and Mikosch, 2006 for a summary) and its exponent depends only on the tail behavior, not the entire distribution. For example, doubling the income of all households in the top 1 percent of the income distribution makes the top 1 percent income share (roughly) twice as large, but the Pareto exponent is unaffected. A similar comment applies to any inequality measure that depends on the entire distribution, such as the Gini coefficient. While we do not claim that the Pareto exponent is the only interesting inequality measure, it is certainly a robust (detail-independent) measure for top tail inequality. See Gabaix (2009, 2016) for more discussion on the robustness of the Pareto distribution.

In this paper, we use the harmonized *Luxembourg Income Study* database (hereafter LIS) to document the capital and labor income Pareto exponents across all available 475 country-year observations that span 52 countries over half a century. We document two empirical findings. First, we find that the capital income Pareto exponent is roughly in the range 1–3 (with median 1.46) and is smaller than the labor income Pareto exponent, which ranges between 2 and 5 (with median 3.35). This implies that capital income is more unequally distributed than labor income. This fact is unsurprising and well known for a specific country or year

²Imagine how to formally distinguish cities, towns, villages, and settlements; continents, islands, and islets; and inland seas, lakes, and ponds. How to define units and how to measure small units matter for the size distribution of population, land mass, and water surface area.

(see, e.g., the Lorenz curve in figure 1 of Saez and Stantcheva, 2018). However, we are not aware of a comprehensive study that systematically analyzes data sets from many countries and years, and therefore our finding suggests that capital income is generally more unequal than labor income. More specifically, using a statistical test recently developed by Hoga (2018), we formally test the equality of capital and labor income Pareto exponents and the null is rejected in 86 percent of samples. In every single case of rejection, the capital exponent is smaller than the labor exponent. Second, we find that the capital income Pareto exponent is nearly unrelated to the labor exponent. In particular, the correlation between the two exponents across countries is close to zero.

To explain our empirical findings, we build a simple incomplete market model with job ladders and capital income risk. In the model, agents get randomly promoted to the next job ladder. Because individual income follows a random multiplicative process, we obtain a Pareto-tailed labor income distribution. The agents also save assets and face idiosyncratic investment risk, which generates a Pareto-tailed wealth (hence capital income) distribution. Because the capital income Pareto exponent is mainly determined by the asset return distribution, while the labor income Pareto exponent is mainly determined by the income growth distribution, the relation between the two is weak. Furthermore, we analytically characterize the capital and labor income Pareto exponents and show that the former tends to be smaller than the latter for common parametrization. Our results suggest the importance of distinguishing income and wealth inequality.

1.1. Related Literature

The power law behavior of income was first recognized by Pareto (1895, 1896, 1897), who used tabulation data of tax returns in many European countries. More recent research that uses micro data include Reed (2001) for the US, Reed (2003) for the US, Canada, Sri Lanka, and Bohemia, Nirei and Souma (2007) for Japan, Toda (2011, 2012) for the US, Jenkins (2017) for the UK, and Ibragimov and Ibragimov (2018) for Russia. These papers all concern specific countries and years. Bandourian *et al.* (2003) estimate 11 parametric distributions (some of which exhibit Pareto tails) using 82 household labor income data sets from *Luxembourg Income Study* (LIS) as we do, though they neither focus on the Pareto exponent nor consider capital income. Gabaix (2009) mentions “[the] tail exponent of income seems to vary between 1.5 and 3,” citing Atkinson and Piketty (2007), though without providing specific details. Atkinson and Piketty (2010, Table 13A.23) document income Pareto exponents across many countries and years estimated from top income share data based on tax returns. However, these estimates are computed from total income, and because (as we document in Section 3.3) the capital income Pareto exponents tend to be smaller than labor exponents, their estimates are best understood as capital income (hence wealth) Pareto exponents. Benhabib *et al.* (2017) make the point that wealth is more skewed than income, citing a few Pareto exponent estimates from Badel *et al.* (2018) for income and Vermeulen (2018) for wealth. Using the 2013–2014 individual income tax data from Romania, Oancea *et al.* (2018)

document that the capital income Pareto exponent (1.44) is smaller than the labor exponent (2.53). Using time series regressions for 21 countries, Bengtsson and Waldenstrom (2018) document a positive correlation between the gross capital share in national accounts and the top 1 percent income shares. Their finding can be explained if top income earners tend to have higher capital income share, which is the case if the capital income Pareto exponent is smaller than the labor exponent as we document in this paper. As mentioned in the introduction, there seems to be no comprehensive studies that document the capital and labor income Pareto exponents separately for many countries and years.

2. DATA

In this section, we describe the data set that we use and discuss its limitations.

2.1. *The LIS Database*

We use the data from the *Luxembourg Income Study* LIS, which is a large, harmonized database of micro-level income data that covers over 50 countries worldwide and many years since the late 1960s. In many countries, the data derive from government surveys (e.g., the US data is based on the *Current Population Survey*). The LIS data are available at both individual and household levels. We focus on the household labor and capital income because (i) it is reasonable to assume that economic decisions such as financial planning are made at the household level, and (ii) incomes among couples are likely correlated due to assortative matching in the marriage market (Siow, 2015), which invalidates statistical estimation.³ The LIS defines *labor income* as “cash payments and value of goods and services received from dependent employment, as well as profits/losses and value of goods from self-employment, including own consumption.” *Capital income* is defined as “cash payments from property and capital (including financial and non-financial assets), including interest and dividends, rental income and royalties, and other capital income from investment in self-employment activity.” Together these two categories make up total *factor income*. See the LIS 2019 USER GUIDE⁴ for a detailed summary on how these data are retrieved and calculated.

2.2. *Data Limitations*

Our analysis draws upon data sets from many different countries that are harmonized into a common framework by the LIS. However, many details about the collection of data in the different countries are omitted. For example, we find evidence of top-coding in some countries and years, as the largest income order statistic is equal to the second largest.⁵ Top-coding induces an upward

³In our data, we find an average correlation of 0.22 between labor income of husband and wife, which underpins the conjectured dependency.

⁴<https://www.lisdatacenter.org/wp-content/uploads/files/data-lis-guide.pdf>

⁵Among all 475 country-year observations, the first- and second-order statistics are equal in 6 cases for labor income and 11 cases for capital income. Therefore, we conjecture that the top-coding issue is not severe.

bias in the estimation of the Pareto exponent. This issue is not necessarily resolved if, instead, one relies on administrative tax income data, for similar biases arise such as rich households trying to understate their taxable income (Atkinson *et al.*, 2011). Burkhauser *et al.* (2012) detail a method that can be used to overcome the bias due to top-coding; however, at the end of their paper they show that the results are robust even if estimates are based on the top-coded series. For these reasons, we treat the data sets as not being top-coded in our analysis.

Another limitation of the LIS database is that it is based on government surveys, and the measurement error may be larger compared to administrative data based on tax returns. The fact that the income distribution in administrative data is often reported as tabulations, not micro data, causes no problem for estimating Pareto exponents, as Toda and Wang (2021) provide an efficient estimation method for such data. In fact, Atkinson and Piketty (2010, Table 13A.23) document income Pareto exponents across countries and years estimated from top income share data. However, their table is based on total income, and because (as we document in Section 3.3) the capital income Pareto exponents tend to be smaller than labor exponents, the estimates in Atkinson and Piketty (2010) are best understood as capital income (hence wealth) Pareto exponents. Because we are not aware of a comprehensive income database that distinguishes capital and labor incomes, we decided to use the LIS database.

3. PARETO EXPONENTS ACROSS COUNTRIES AND YEARS

In this section we estimate the capital and labor income Pareto exponents for all countries and years that are available in the LIS database, which spans 52 countries over half a century (475 country-year observations in total). We then formally test for the equality of the Pareto exponents of capital and labor incomes.

3.1. Estimation Method

For each country and year, we suppose that the (capital or labor) income observations $\{X_n\}_{n=1}^N$ are independent and identically distributed (IID) with cumulative distribution function (CDF) $F(x) = P(X_n \leq x)$. The assumption that the upper tail of income obeys a power law with Pareto exponent $\alpha > 0$ translates into the regular variation condition

$$(1) \quad 1 - F(x) = x^{-\alpha} \ell(x)$$

for some slowly varying function ℓ (see footnote 1). Note that the assumption on ℓ involves only the limit as $x \rightarrow \infty$; we are assuming a power law behavior in the upper tail without taking a stance on the shape of the entire distribution.

We are interested in estimating the Pareto exponent α for each country and year. For this purpose, we use the Hill (1975) (maximum likelihood) estimator

$$(2) \quad \hat{\alpha}(k) := \frac{1}{k} \sum_{n=1}^k \log \left(\frac{X_{(n)}}{X_{(k)}} \right).$$

Here $X_{(n)}$ denotes the n -th largest order statistic from the sample $\{X_n\}_{n=1}^N$ and $k \in \{1, \dots, N\}$ denotes the number of tail observations used to estimate the Pareto exponent. Hall (1982) shows that the standard error of $\hat{\alpha}(k)$ is $\hat{\alpha}(k)/\sqrt{k}$ under flexible assumptions on the CDF (see (3) below). We use the Hill estimator because we are interested in formally testing the equality of the capital and labor income Pareto exponents using the method of Hoga (2018), for which the Hill estimator is required.⁶

When the population distribution is known to be exactly Pareto (so ℓ in (1) is zero below the minimum size x_{\min} and constant above this threshold), it is well known that the Hill estimator for the full sample ($k = N$) is consistent, asymptotically normal, and asymptotically efficient because it is a maximum likelihood estimator. In practice, the CDF is not exactly Pareto, and the researcher needs to select an appropriate value of k . For instance, if $F(x)$ satisfies

$$(3) \quad 1 - F(x) = Cx^{-\alpha}(1 + Dx^{-\beta} + o(x^{-\beta}))$$

with some $\beta > 0$, then Hall (1982) shows that choosing $k = o(N^{2\beta/(2\beta+\alpha)})$ together with $k \rightarrow \infty$ as $N \rightarrow \infty$ is sufficient for consistency and asymptotic normality (see also Embrechts and Kluppelberg, 2013).⁷ Notice that this choice puts a bound on the growth rate of k . The expansion (3) covers a wide range of distributions of interest, such as the t -distribution and the type II extreme value distribution (Danielsson and de Vries, 1997).

Despite these asymptotic results, it is notoriously difficult to pick k optimally in finite samples (Hall, 1990; Resnick and Stărică, 1997; Danielsson *et al.*, 2001). In practice, researchers often plot the Hill estimator (2) over a range of k to find a flat region or plot the log rank $\log 1, \dots, \log N$ against the log size $\log X_{(1)}, \dots, \log X_{(N)}$ to find a region that exhibits a straight line pattern and choose a size threshold to run the log-rank regression.⁸ Unfortunately, this graphical approach is not feasible in our setting because LIS does not allow researchers to download the micro data for confidentiality concerns (researchers are required to submit their execution files to conduct statistical analyses) and there is little scope for exploratory graphical data analysis. In this paper we simply use the largest 5 percent observations, so

$$(4) \quad k = \lfloor 0.05N \rfloor,$$

⁶If we are only interested in estimating the Pareto exponents, then there are many alternative methods available. Ivette Gomes and Guillou (2015) review 13 commonly used estimators. Fedotenkov (2020) reviews more than 100.

⁷Recall that we write $f(x) = o(g(x))$ if for all $\epsilon > 0$, we have $|f(x)| \leq \epsilon |g(x)|$ for large enough x .

⁸Gabaix *et al.* (2011) study the asymptotic behavior of log rank regression and show that the standard error is larger by a factor of $\sqrt{2}$ than the Hill estimator. However, they do not discuss how to select the threshold. In their empirical application, they consider the size distribution of population in the US metropolitan statistical areas, which are already far into the tail and hence the threshold selection is less of an issue. Ibragimov and Ibragimov (2018) apply the same methodology to Russian household income data and consider the top 5 percent and 10 percent thresholds.

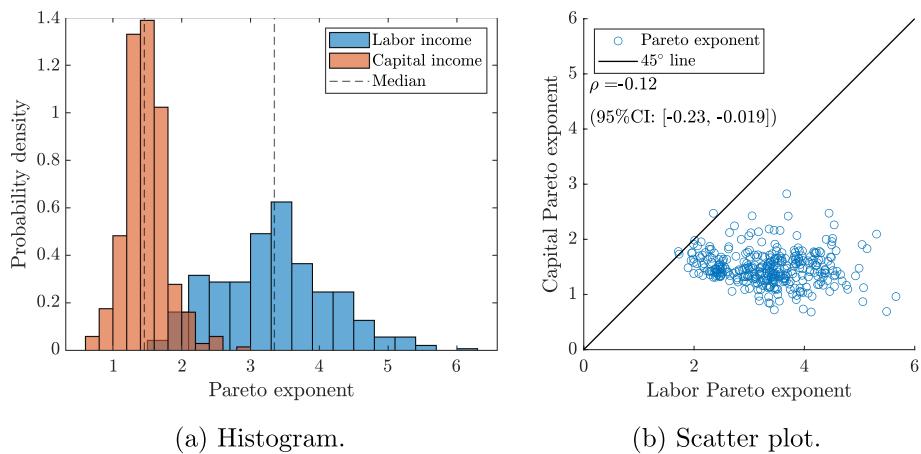


Figure 1. Histogram and Scatter Plot of Capital and Labor Income Pareto Exponents
[Colour figure can be viewed at wileyonlinelibrary.com]

which is standard in the literature.⁹ Unreported simulations show that our results are robust to using other thresholds such as the largest 1 percent and 10 percent observations, or using a data-driven procedure using the Kolmogorov–Smirnov distance as in Danielsson *et al.* (2016).

3.2. Capital and Labor Income Pareto Exponents

We estimate the capital and labor income Pareto exponents for all countries and years available in the LIS database. The database spans 52 countries across the years 1967–2018, with a total of 475 country-year observations. The point estimates of the capital and labor income Pareto exponents for each country and year as well as their standard errors can be found in Table 2 in the Online Appendix. To avoid small sample issues, we restrict our analysis to countries with at least 1,000 positive observations for income, resulting in (all) 475 country-year pairs for labor income and 342 for capital income. For visibility, Figure 1 shows the histogram and scatter plot of the Pareto exponents together with the 45° line. The confidence interval of the correlation coefficient ρ is computed assuming all country-year

Figure 1a shows the histogram of the capital and labor income Pareto exponents pooled across all available countries and years. The capital and labor income Pareto exponents are generally in the range 1–3 and 2–5 with medians 1.46 and 3.35, respectively. This suggests that (i) capital income is generally more unequally distributed than labor income, but (ii) there is significant heterogeneity in both capital and labor income inequality across countries and years. Figure 1b shows the scatter plot of the Pareto exponents together with the 45° line. The confidence interval of the correlation coefficient ρ is computed assuming all country-year

⁹An alternative approach is to estimate a parametric distribution F that admits a Pareto upper tail by maximum likelihood using the entire sample. The double Pareto-lognormal distribution proposed by Reed (2003) and Reed and Jorgensen (2004) often performs best. See Toda (2012) for a horse race across several parametric distributions in the context of the US labor income.

observations are independent and there is no sampling error in the point estimates of the Pareto exponents. Although it is not obvious how to account for these issues, doing so will only widen the confidence interval. Therefore the fact that the naïve confidence interval almost contains zero suggests that the correlation between the two Pareto exponents is weak. Furthermore, for the vast majority of countries and years, the capital income Pareto exponent is smaller than the labor exponent, again suggesting that capital income is more unequal than labor income.

How do the Pareto exponents evolve over time? Because many countries appear only sporadically in the LIS database, we only consider six countries for which the cross-sectional sample size is large and the time series is long, namely Canada, Germany, Switzerland, Taiwan, United Kingdom, and United States. [Figure 2](#) plots the capital and labor Pareto exponents and their 95 percent confidence intervals for these countries. The capital exponent tends to be stable at around 1–2 and is smaller than the labor exponent.¹⁰ Again, capital income appears to be more unequally distributed than the labor income.

3.3. Testing Equality of Capital and Labor Pareto Exponents

We now formally test whether the capital and labor Pareto exponents are equal. In particular our test is

$$H_0: \alpha_{\text{lab}} = \alpha_{\text{cap}} \quad \text{against} \quad H_1: \alpha_{\text{lab}} \neq \alpha_{\text{cap}},$$

where $\alpha_{\text{cap}}, \alpha_{\text{lab}}$ denote the capital and labor income Pareto exponents. Testing the null hypothesis H_0 is complicated by the fact that there is dependency between labor income $\{X_{\text{lab},n}\}_{n=1}^N$ and capital income $\{X_{\text{cap},n}\}_{n=1}^N$, because individuals who are rich (receive high labor income) tend to be wealthy and receive high capital income. Thus, we cannot use the 95 percent confidence intervals in [Figure 2](#) to test the equality of Pareto exponents. Instead, we apply the test recently developed by Hoga (2018), which allows for dependence in the data but assumes restrictions on the growth rate of the tail dependence (see Hoga, 2018, Assumption A2). The test is based on the inverse of the Hill estimator (2), which we denote by $\hat{\gamma} := 1/\hat{\alpha}$. The test statistic is defined by

$$(5) \quad T_N = \frac{(\hat{\gamma}_{\text{lab}}(1) - \hat{\gamma}_{\text{cap}}(1))^2}{\int_{t_0}^1 t^2 [(\hat{\gamma}_{\text{lab}}(t) - \hat{\gamma}_{\text{cap}}(t)) - (\hat{\gamma}_{\text{lab}}(1) - \hat{\gamma}_{\text{cap}}(1))]^2 dt},$$

where $t_0 \in (0, 1)$ is a tuning parameter and $\hat{\gamma}(t)$ is the inverse Hill estimator

$$(6) \quad \hat{\gamma}(t) := \frac{1}{[kt]} \sum_{n=1}^{[kt]} \log \left(\frac{X_{(n)}}{X_{([kt])}} \right).$$

¹⁰An exception may be Germany before 1983. However, this may be an artifact of the sudden change in sample size, which was over 40,000 until 1983 and about 5,000 since 1984. Thus, the 5 percent rule for capital income may be including too many observations in the body of the distribution before 1983.

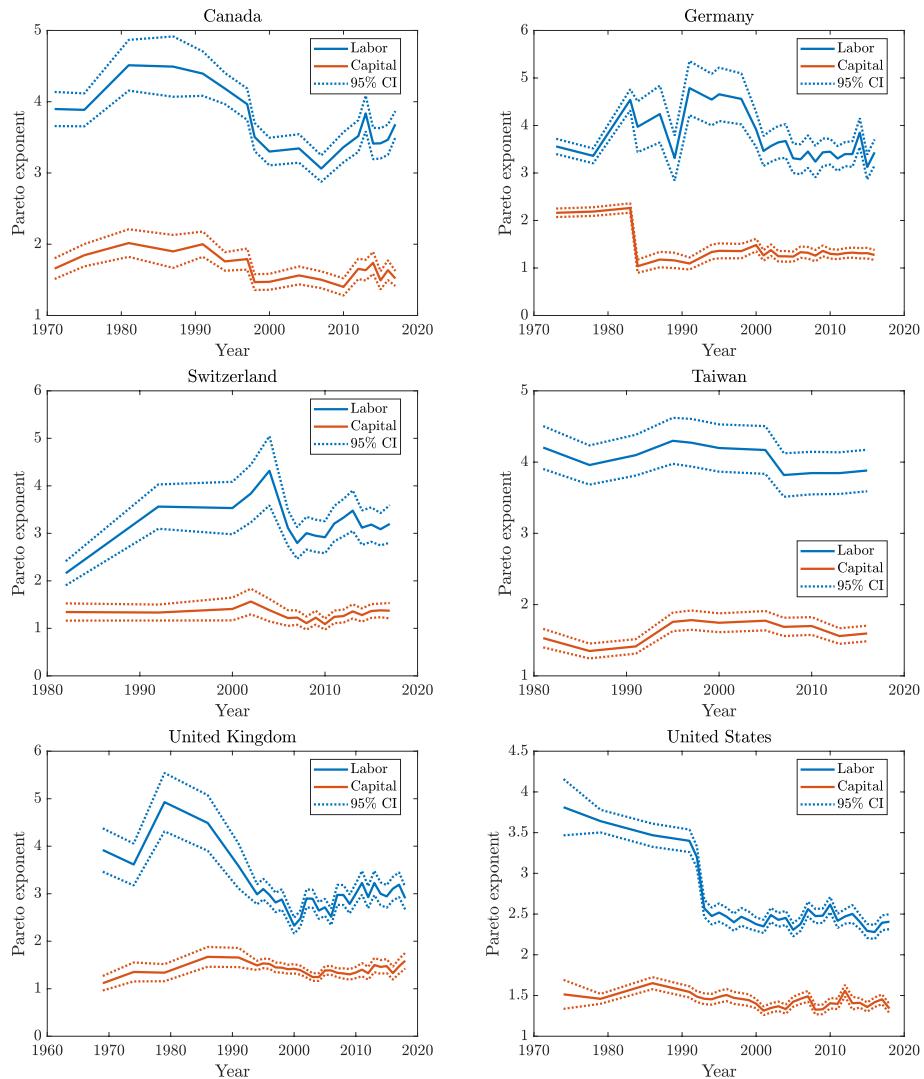


Figure 2. Time Evolution of Capital and Labor Income Pareto Exponents
[Colour figure can be viewed at wileyonlinelibrary.com]

Using the Hill estimator based on the subsample with only $[kt]$ observations leads to self-normalization of the test statistic T_N and renders a test that is asymptotically pivotal. The limiting distribution is

$$(7) \quad T_N \xrightarrow{d} \frac{W(1)^2}{\int_{t_0}^1 [W(t) - tW(1)]^2 dt},$$

where $W(t)$ is a standard Brownian motion. As the test statistic (5) can be computed using only the Hill estimator and conducting numerical integration, there is

no need to estimate the (potentially difficult) tail covariance. The tuning parameter t_0 affects the size of the test in finite samples: high values of t_0 make the integral in (5) based on too few differences of $\hat{\gamma}$, and low values of t_0 yield volatile $\hat{\gamma}$ in (6) when t is close to t_0 . Both of these effects may cause size distortions. Therefore, we set $t_0 = 0.2$ following the recommendation of Hoga (2018), who finds that this choice leads to favorable size properties.¹¹ We reject the null H_0 when the test statistic T_N is large. According to Table I of Hoga (2018), the 95 percentile of (7) for $t_0 = 0.2$ is 55.44, which we use as the critical value for testing H_0 at 5 percent significance level.

One issue with the test statistic (5) is that it requires the same number of tail observations k for both cross-sections of capital and labor incomes. Therefore, the 5 percent rule (4) discussed in Section 3.1 becomes problematic as the number of people with capital income in our data set is rather small. Many households do not hold liquid financial wealth and hence have no capital income. The resulting test is thus not feasible because k_{lab} based on our 5 percent rule could be wildly different from k_{cap} . To overcome this issue, we only test the equality of Pareto exponents for countries that have more than 1,000 positive capital income observations and set $k = \lfloor 0.05N_{\text{cap}} \rfloor$, where N_{cap} is the number of positive capital income observations (these households always have positive labor income), resulting in 342 country-year observations out of 475. In practice this means that we estimate the Pareto exponent of labor income further in the tail because the sample size of labor income N_{lab} tends to be larger than N_{cap} . However, this is acceptable because the Pareto approximation tends to fit better for smaller k . Figure 3 presents a scatter plot of the estimated labor income Pareto exponent $\hat{\alpha}_{\text{lab}}$ using 5 percent of the full sample (N_{lab}) and 5 percent of the sample with positive capital income (N_{cap}). The fact that most points are close to the 45° line supports our claim.

Table 3 in the Online Appendix shows the test results of the null hypothesis $H_0: \alpha_{\text{lab}} = \alpha_{\text{cap}}$. We reject the null in 294 country-year observations out of 342 (86%) that meet our sample selection criterion. In every single case of rejection, we have $\hat{\alpha}_{\text{lab}} > \hat{\alpha}_{\text{cap}}$, and therefore we formally confirm the observation in Section 3.2 that capital income is more unequally distributed than labor income.

4. MODEL OF CAPITAL AND LABOR PARETO EXPONENTS

Our empirical analysis in Section 3 suggests that (i) the capital income Pareto exponent is smaller than the labor one (i.e., capital income is more unequally distributed than labor income), and (ii) the correlation between capital and labor income Pareto exponents is weak. To explain these empirical findings, we present a simple dynamic model of consumption and savings, which builds on one of the authors' previous works (Ma *et al.*, 2020; Ma and Toda, 2021). Our model is more specialized but the characterizations are sharper.

¹¹The choice of t_0 in Hoga (2018) comes out using an automated selection procedure to choose k , which is different from our 5 percent rule. An earlier version of our paper uses the same automated procedure, which leads to very similar results.

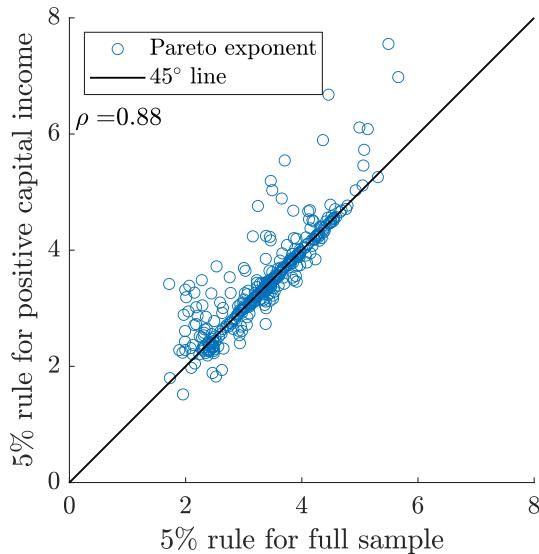


Figure 3. Labor Pareto Exponent with Full and Positive Capital Income Samples
[Colour figure can be viewed at wileyonlinelibrary.com]

4.1. Income Fluctuation Problem

Time is discrete and denoted by $t = 0, 1, 2, \dots$. Let a_t be the financial wealth of a typical agent at the beginning of period t including current income. The agent chooses consumption $c_t \geq 0$ and saves the remaining wealth $a_t - c_t$. The period utility function is $u: (0, \infty) \rightarrow \mathbb{R}$, the discount factor is $\beta > 0$, the gross return on wealth between time $t - 1$ and t is $R_t > 0$, and non-financial income at time t is $Y_t > 0$. Thus the agent solves

$$(8a) \quad \text{maximize} \quad E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$

$$(8b) \quad \text{subject to} \quad a_{t+1} = R_{t+1}(a_t - c_t) + Y_{t+1},$$

$$(8c) \quad 0 \leq c_t \leq a_t,$$

where the initial wealth $a_0 = a > 0$ is given, (8b) is the budget constraint, and (8c) implies that the agent cannot borrow (which is without loss of generality according to the discussion in Chamberlain and Wilson 2000). Throughout the rest of the paper, we maintain the following assumptions.

Assumption 1. [CRRA utility] The utility function exhibits constant relative risk aversion (CRRA) with coefficient $\gamma > 0$, so $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ if $\gamma \neq 1$ and $u(c) = \log c$ if $\gamma = 1$.

Assumption 2. [IID shocks] Let $G_{t+1} := Y_{t+1}/Y_t$ be gross growth rate of income. The sequence $\{R_{t+1}, G_{t+1}\}_{t=0}^\infty$ is independent and identically distributed (IID).

These assumptions are similar to Carroll (2020), except that we allow for stochastic returns on savings. Note that the asset return R_{t+1} and income growth G_{t+1} are potentially mutually dependent. Due to the IID assumption, the state variables of the income fluctuation problem (8) are financial wealth $a_t > 0$ and current income $Y_t > 0$. Exploiting homotheticity (Assumption 1), we can reduce the number of state variables to just one, namely the wealth-income ratio (normalized wealth) $\tilde{a}_t := a_t/Y_t$. To see this, letting $\tilde{c}_t := c_t/Y_t$ be the consumption-income ratio (normalized consumption), dividing the borrowing constraint (8c) by Y_t , we obtain $0 \leq \tilde{c}_t \leq \tilde{a}_t$. Similarly, dividing the budget constraint (8b) by Y_{t+1} , we obtain

$$(9) \quad \begin{aligned} \tilde{a}_{t+1} &= a_{t+1}/Y_{t+1} = (R_{t+1}Y_t/Y_{t+1})(a_t/Y_t - c_t/Y_t) + 1 \\ &= (R_{t+1}/G_{t+1})(\tilde{a}_t - \tilde{c}_t) + 1 \\ &= \tilde{R}_{t+1}(\tilde{a}_t - \tilde{c}_t) + 1, \end{aligned}$$

where $\tilde{R}_{t+1} := R_{t+1}/G_{t+1}$ is the asset return relative to income growth. As for the utility function, because

$$c_t = Y_t \tilde{c}_t = Y_0 \left(\prod_{s=1}^t G_s \right) \tilde{c}_t$$

(here we interpret $\prod_{s=1}^0 \bullet = 1$), assuming $Y_0 = 1$ (which is without loss of generality) and $\gamma \neq 1$, it follows from (8a) that

$$(10) \quad \begin{aligned} E_0 \sum_{t=0}^{\infty} \beta^t u(c_t) &= E_0 \sum_{t=0}^{\infty} \left(\prod_{s=1}^t \beta G_s^{1-\gamma} \right) \frac{\tilde{c}_t^{1-\gamma}}{1-\gamma} \\ &= E_0 \sum_{t=0}^{\infty} \left(\prod_{s=1}^t \tilde{\beta}_s \right) \frac{\tilde{c}_t^{1-\gamma}}{1-\gamma}, \end{aligned}$$

where $\tilde{\beta}_t := \beta G_t^{1-\gamma}$. The discussion for $\gamma = 1$ is similar. Therefore, the problem reduces to an income fluctuation problem with CRRA utility, random discount factors $\{\tilde{\beta}_t\}_{t=1}^\infty$, stochastic returns $\{\tilde{R}_t\}_{t=1}^\infty$ on wealth, and constant income ($\tilde{Y}_t \equiv 1$). The general theory of income fluctuation problems with stochastic discounting, returns, and income in a Markovian setting was developed by Ma *et al.* (2020). Therefore, we immediately obtain the following result. In what follows, we drop the time subscript when no confusion arises.

Proposition 1. Suppose Assumptions 1 and 2 hold, and

$$(11) \quad \beta E[G^{1-\gamma}] < 1 \quad \text{and} \quad \beta E[RG^{-\gamma}] < 1.$$

Then the income fluctuation problem () has a unique solution. The consumption function can be expressed as $c(a, Y) = Y\tilde{c}(a/Y)$, where $\tilde{c} : (0, \infty) \rightarrow (0, \infty)$ is the

consumption function of the detrended problem (maximizing (10) subject to (9)), which can be computed by policy function iteration.¹²

Proof. Applying Theorem 2.2 of Ma *et al.* (2020) to the iid case, a sufficient condition for the existence of a solution to the detrended problem is $E[\tilde{\beta}] < 1$ and $E[\tilde{\beta} \tilde{R}] < 1$, which is equivalent to (11).

4.2. Tail Behavior of Income and Wealth

We now characterize the tail behavior of income and wealth in the context of the income fluctuation problem in Section 4.1.

To make the model stationary, suppose that agents survive to the next period with probability $v \in (0, 1)$ (perpetual youth model as in Yaari, 1965). Whenever agents die, they are replaced by newborn agents. For simplicity, assume that the discount factor β in (8a) already accounts for survival probability, and that there is no market for life insurance (allowing for life insurance only changes R to R/v and is thus mathematically equivalent after reparametrization). Without loss of generality, suppose that newborn agents start with income $Y_0 = 1$. Then the income of a randomly selected agent is Y_T , where T is a geometric random variable with mean $\frac{1}{1-v}$. By the assumption on income growth, the log income of a randomly selected agent

$$\log Y_T = \log(Y_T / Y_0) = \sum_{t=1}^T \log G_t$$

is a geometric sum of iid random variables, for which we can characterize the tail behavior as follows.

Proposition 2. [Income Pareto exponent] Suppose that $P(G > 1) > 0$ and $1 < vE[G^z] < \infty$ for some $z > 0$. Then the cross-sectional income distribution has a Pareto upper tail, whose exponent is the unique positive solution $z = \alpha_Y$ to

$$(12) \quad vE[G^z] = 1.$$

Proof. See Beare and Toda (2017).

To characterize the tail behavior of wealth, we first note that the normalized consumption function \tilde{c} in Proposition 1 is concave and asymptotically linear with a specific slope.¹³

Proposition 3. [Concavity and asymptotic linearity] Let everything be as in Proposition 1. Then \tilde{c} is concave and

¹²See Li and Stachurski (2014) and Ma *et al.* (2020) for details on policy function iteration.

¹³Carroll and Kimball (1996) showed the sufficiency of hyperbolic absolute risk aversion (HARA, which includes CRRA) for the concavity of the consumption function. Toda (2021) proved the necessity.

$$(13) \quad \lim_{a \rightarrow \infty} \frac{\tilde{c}(a)}{a} = \begin{cases} 1 - (E[\beta R^{1-\gamma}])^{1/\gamma} & \text{if } E[\beta R^{1-\gamma}] < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The concavity of \tilde{c} follows from Proposition 2.5 and Remark 2.1 of Ma *et al.* (2020). The asymptotic linearity of \tilde{c} follows from Ma and Toda (2021, Theorem 3). Noting that

$$E[\tilde{\beta} \tilde{R}^{1-\gamma}] = E[\beta G^{1-\gamma} (R/G)^{1-\gamma}] = E[\beta R^{1-\gamma}],$$

the limit (13) follows from their Example 2.

Using Proposition 3 and setting $\rho = \min \{ (E[\beta R^{1-\gamma}])^{1/\gamma}, 1 \}$, for high enough asset level, the detrended budget constraint (9) becomes approximately

$$\tilde{a}_{t+1} \approx \rho \tilde{R}_{t+1} \tilde{a}_t + 1,$$

which is a random multiplicative process (Kesten (1973) process). Under specific assumptions, Ma *et al.* (2020, Theorem 3.3) prove that the upper tail of the stationary distribution of normalized wealth \tilde{a} , has a Pareto lower bound. Although a sharp characterization of the tail behavior is generally difficult, in our setting it is possible to obtain an exact characterization due to concavity and the iid assumption.

Proposition 4. *Let $\rho = \min \{ (E[\beta R^{1-\gamma}])^{1/\gamma}, 1 \}$ and $H = \rho \tilde{R}$. Suppose that (i) R is thin-tailed (meaning $E[R^z] < \infty$ for all $z > 0$), (ii) $\log H$ is non-lattice (not supported on an evenly spaced grid), and (iii) $P(H > 1) > 0$ and $1 < vE[H^z] < \infty$ for some $z > 0$. Then the cross-sectional normalized wealth distribution is either bounded or has a Pareto upper tail, in which case the exponent is the unique positive solution $z = \tilde{\alpha}$ of*

$$(14) \quad vE[H^z] = 1.$$

The Pareto exponent for wealth and capital income is then $\alpha = \min \{ \tilde{\alpha}, \alpha_Y \}$.

Proof. As \tilde{c} is concave by Proposition 3, it is in particular Lipschitz continuous. Under the maintained assumptions, we can apply Theorem 1.8 of Mirek (2011) to deduce that the normalized wealth \tilde{a} is either bounded or has a Pareto upper tail with exponent characterized as the solution to (14), where we have used the asymptotic linearity of \tilde{c} established in Proposition 3.

By accounting, capital income (excluding capital loss) is

$$Y_{\text{cap}} := \max \{ R - 1, 0 \} (a - c(a)) = \max \{ R - 1, 0 \} Y(\tilde{a} - \tilde{c}(\tilde{a})).$$

Using Proposition 3, this quantity is approximately equal to $\rho \max \{ R - 1, 0 \} Y \tilde{a}$. The claim $\alpha = \min \{ \tilde{\alpha}, \alpha_Y \}$ then follows because asset return R is thin-tailed and $a_t = Y_t \tilde{a}_t$ is the product of two (potentially dependent) random variables with Pareto upper tails, which inherits the smallest Pareto exponent by the result in Jessen and Mikosch (2006).

Proposition 4 is significant despite its simplicity. According to the model, we always have $\alpha \leq \alpha_Y$, typically with a strict inequality as we see in the numerical example below. This is in sharp contrast to canonical incomplete market general equilibrium models such as Aiyagari (1994), where agents can save using only a risk-free asset. In such models, the impossibility theorem of Stachurski and Toda (2019) implies that the tail behavior of income and wealth is the same, implying $\alpha = \alpha_Y$ in our setting.¹⁴ Therefore, unlike canonical incomplete market models, our model can explain the empirical fact that capital income is more unequal than labor income. The key assumption leading to this conclusion is the presence of stochastic returns.

We discuss an analytically solvable example to build intuition.

Example 1. Let $\Delta > 0$ be the length of time of one period and the discount factor be $\beta = e^{-\delta\Delta}$, where $\delta > 0$ is the discount rate. Suppose income grows at a constant rate $g > 0$, so $G = e^{g\Delta}$. Suppose asset return is risk-free, so $R = e^{r\Delta}$ with $r > 0$. Finally, let the survival probability be $v = e^{-\eta\Delta}$, where η is the death rate. Then (12) becomes

$$1 = e^{-\eta\Delta} e^{zg\Delta} \iff z = \eta/g,$$

so the income Pareto exponent is $\alpha_Y = \eta/g$. (This is the classical result of Wold and Whittle (1957) in discrete-time.) Suppose in addition that $-\eta + r(1 - \gamma) < 0$ so that $\beta R^{1-\gamma} < 1$. As

$$H = (E[\beta R^{1-\gamma}])^{1/\gamma} \tilde{R} = (\beta R)^{1/\gamma} / G = e^{(\frac{r-\eta}{\gamma} - g)\Delta},$$

solving (14) the normalized wealth Pareto exponent is

$$\tilde{\alpha} = \frac{\eta\gamma}{r - \eta - g\gamma}$$

assuming $r - \eta - g\gamma > 0$. Therefore

$$\tilde{\alpha} < \alpha_Y \iff \frac{\eta\gamma}{r - \eta - g\gamma} < \frac{\eta}{g} \iff r > \eta + 2g\gamma,$$

so the wealth (hence capital income) Pareto exponent is smaller than the labor income Pareto exponent if the return on wealth r is sufficiently large. In summary, we obtain the following result: suppose $-\eta + r(1 - \gamma) < 0$ and let $\alpha_{\text{cap}}, \alpha_{\text{lab}}$ be the capital and labor income Pareto exponents. Then

$$(15) \quad \begin{cases} \alpha_{\text{cap}} = \alpha_{\text{lab}} = \frac{\eta}{g} & \text{if } r \leq \eta + 2g\gamma, \\ \alpha_{\text{cap}} = \frac{\eta\gamma}{r - \eta - g\gamma} < \frac{\eta}{g} = \alpha_{\text{lab}} & \text{if } r > \eta + 2g\gamma. \end{cases}$$

¹⁴The original proof in Stachurski and Toda (2019) contained an error; it has been corrected in Stachurski and Toda (2020).

Note that the labor income Pareto exponent $\alpha_{\text{lab}} = \eta/g$ is highly sensitive to the income growth rate g . However, provided that $r > \eta + 2g\gamma$, the capital income Pareto exponent α_{cap} is not very sensitive to the value of g because the denominator is $r - \eta - g\gamma$. This example is consistent with our result in Section 3.2 that the capital Pareto exponent is smaller than the labor Pareto exponent but the two values are only weakly related.

4.3. Numerical Example

We further examine the tail behavior of income and wealth using a numerical example of the income fluctuation problem (8). Suppose that asset return is iid log-normal, so $\log R \sim N((\mu - \sigma^2/2)\Delta, \sigma^2\Delta)$, where $\Delta > 0$ is the length of one period, μ is the expected return, and σ is volatility. Suppose every period the agent is “promoted” with some probability, so the income growth rate is

$$G_{t+1} = Y_{t+1}/Y_t = \begin{cases} 1 & \text{with probability } 1-p, \\ e^g & \text{with probability } p, \end{cases}$$

where $p \in (0, 1)$ is the promotion probability and g is the log income growth rate conditional on promotion. We parametrize the promotion probability as $p = 1 - e^{-\Delta/L}$, where L is the expected length of time until a promotion. Using (12), the labor income Pareto exponent is determined such that

$$(16) \quad 1 = vE[G^{\alpha_Y}] = v(1 - p + pe^{g\alpha_Y}) \iff \alpha_Y = \frac{1}{g} \log \frac{1 - v + vp}{vp}.$$

We set the parameter values as in Table 1. One unit of time corresponds to a year and one period is a quarter, so $\Delta = 1/4$. The preference parameters (discount rate and risk aversion) are standard. The death rate of $\eta = 0.025$ implies an average (economically active) age of $1/\eta = 40$ years. The expected return and volatility roughly correspond to the stock market. We set the labor income Pareto exponent to $\alpha_Y = 3$, which is roughly the median value in Figure 1. Using the survival probability $v = e^{-\eta\Delta}$ and (16), the implied value of income growth upon promotion is $g = 0.0403$. The wealth Pareto exponent determined by (14) is then $\alpha = 1.201$.

To numerically solve the income fluctuation problem (8), we first detrend it as in (9) and (10). Suppressing the tildes, the detrended problem becomes

$$\begin{aligned} & \text{maximize } E_0 \sum_{t=0}^{\infty} \left(\prod_{s=1}^t \beta_s \right) \frac{c_t^{1-\gamma}}{1-\gamma} \\ & \text{subject to } a_{t+1} = R_{t+1}(a_t - c_t) + 1, \\ & \quad 0 \leq c_t \leq a_t, \end{aligned}$$

where $\{R_t, \beta_t\}_{t=1}^{\infty}$ is iid (though R_t and β_t are generally correlated). According to Ma *et al.* (2020), the Euler equation is

$$(17) \quad c_t^{-\gamma} = \max \{E_t[\beta_{t+1} R_{t+1} c_{t+1}^{-\gamma}], a_t^{-\gamma}\}.$$

TABLE 1
PARAMETER VALUES

Parameter	Symbol	Value
Length of one period	Δ	1/4
Discount rate	δ	0.04
Relative risk aversion	γ	2
Death rate	η	0.025
Expected return	μ	0.07
Volatility	σ	0.15
Expected time to promotion	L	5
Labor income Pareto exponent	α_Y	3

Let $c(a)$ be the consumption function. Taking the $-1/\gamma$ -th power of (17), we obtain

$$(18) \quad c(a) = \min \left\{ (E[\beta R c(a')^{-\gamma}])^{-1/\gamma}, a \right\},$$

where $a' = R(a - c(a)) + 1$. Therefore, we can compute the consumption function using the following variant of the policy function iteration algorithm:

- (i) Initialize the consumption function $c(a)$. For example, we can set $c(a) = \min \{a, 1 + ma\}$, where $m = \max \{1 - (E[\beta R^{1-\gamma}])^{1/\gamma}, 0\}$ is the theoretical asymptotic marginal propensity to consume according to (13).
- (ii) Update $c(a)$ by the right-hand side of (18), where $a' = R(a - c(a)) + 1$.
- (iii) Iterate the above step until $c(a)$ converges.

While the above algorithm has no guarantee to converge unlike the “true” policy function iteration algorithm discussed in Ma *et al.* (2020), it has the advantage of avoiding root-finding and therefore it is fast.

For computational purposes, we discretize the log asset return $\log R$ using a 7-point Gauss-Hermite quadrature and use a 100-point exponential grid for normalized wealth \tilde{a} that spans $[0, 10^4]$, with a median grid point of 10. The details on the exponential grid are discussed in Gouin-Bon enfant and Toda (2018). After solving the individual problem, we apply the Pareto extrapolation algorithm developed in Gouin-Bon enfant and Toda (2018) to accurately compute the stationary (normalized) wealth distribution. Finally, we also simulate an economy with 10^5 agents. Figure 4 shows the results.

Figure 4a shows the normalized consumption function $\tilde{c}(\tilde{a})$ in the range $\tilde{a} \in [0, 100]$. Consistent with Proposition 3, the consumption function is roughly linear for high asset level. Figure 4b shows the size distributions of income Y normalized wealth $\tilde{a} = a/Y$ in a log–log plot, both from the theoretical model and the simulation. The fact that the tail probability $P(X > x)$ exhibits a straight line pattern in a log–log plot suggests that the size distributions have Pareto upper tails, consistent with theory. Furthermore, the slope for income is steeper than that of normalized wealth, so wealth (hence capital income) is more unequally distributed than labor income.

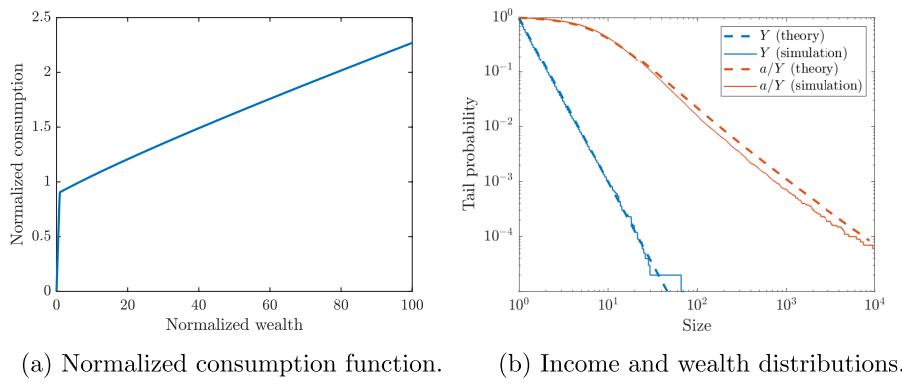


Figure 4. Solution to Income Fluctuation Problem
 [Colour figure can be viewed at wileyonlinelibrary.com]

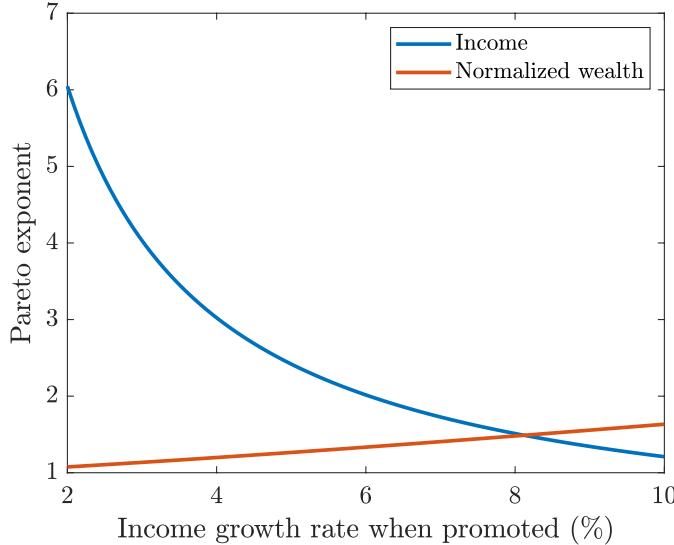


Figure 5. Dependence of Income and Wealth Pareto Exponents on g
 [Colour figure can be viewed at wileyonlinelibrary.com].

Finally, Figure 5 shows the income and wealth Pareto exponents when we change the income growth rate g in the range $g \in [0.02, 0.1]$, fixing other parameters. Because the income Pareto exponent is inversely proportional to income growth by (16), the labor income Pareto exponent is highly sensitive to income growth. On the contrary, the wealth (capital income) Pareto exponent does not depend much on income growth by the same intuition as in Example 1. Thus our model is consistent with our empirical findings in Section 3.2 that capital and labor income Pareto exponents are only weakly related.

5. CONCLUDING REMARKS

The Pareto exponent characterizes the tail heaviness of size distributions. This paper has systematically estimated and documented capital and labor income Pareto exponents for 52 countries over half a century (475 country-year observations). We have found that the capital income Pareto exponent (median 1.46) is substantially and robustly smaller than the labor exponent (median 3.35), suggesting that capital income (hence wealth) is more unequally distributed than labor income. This finding has two important implications. First, income inequality and wealth inequality are different entities and need to be distinguished in policy debates. Second, canonical Rao Aiyagari models that abstract from capital income risk—which necessarily imply the same tail behavior for income and wealth distributions as proved by Stachurski and Toda (2019)—are inadequate for studying wealth inequality. Our model in Section 4 provides a stylized example for alternative modeling.

REFERENCES

- Aiyagari, S. R., “Uninsured Idiosyncratic Risk and Aggregate Saving,” *Quarterly Journal of Economics*, 109(3), 659–84, 1994.
- Atkinson, A. B., “Comparing the Distribution of Top Incomes Across Countries,” *Journal of the European Economic Association*, 3(2–3), 393–401, 2005.
- Atkinson, A. B., and T. Piketty (eds), *Top Incomes over the Twentieth Century*. Oxford University Press, New York, NY, 2007.
- Atkinson, A. B., and T. Piketty (eds), *Top Incomes: A Global Perspective*. Oxford University Press, New York, 2010.
- Atkinson, A. B., T. Piketty, and E. Saez, “Top Incomes in the Long Run of History,” *Journal of Economic Literature*, 49(1), 3–71, 2011.
- Badel, A., M. Daly, M. Huggett, and M. Nybom, “Top Earners: Cross-Country Facts,” *Federal Reserve Bank of St. Louis Review*, 100(3), 237–57, 2018.
- Bandourian, R., J. B. McDonald, and R. S. Turley, “A Comparison of Parametric Models of Income Distributions Across Countries and Over Time,” *Estadística*, 55(164), 135–52, 2003. <https://ssrn.com/abstract=324900>
- Beare, B., and A. Toda, *Geometrically Stopped Markovian Random Growth Processes and Pareto Tails*. 2017. <https://arxiv.org/abs/1712.01431>
- Bengtsson, E., and D. Waldenström, “Capital Shares and Income Inequality: Evidence from the Long Run,” *Journal of Economic History*, 78(3), 712–43, 2018.
- Benhabib, J., A. Bisin, and M. Luo, “Earnings Inequality and Other Determinants of Wealth Inequality,” *American Economic Review: Papers and Proceedings*, 107(5), 593–7, 2017.
- Bingham, N. H., C. Goldie, and J. Teugels, “Regular Variation.” *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 1987.
- Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore, “Recent Trends in Top Income Shares in the United States: Reconciling Estimates from March CPS and IRS Tax Return Data,” *Review of Economics and Statistics*, 94(2), 371–88, 2012.
- Carroll, C., “Theoretical Foundations of Buffer Stock Saving,” *Quantitative Economics*, 2020. <http://qeconomics.org/ojs/forth/354/354-15.pdf>. Forthcoming.
- Carroll, C. D., and M. S. Kimball, “On the Concavity of the Consumption Function,” *Econometrica*, 64(4), 981–92, 1996.
- Chamberlain, G., and C. A. Wilson, “Optimal Intertemporal Consumption Under Uncertainty,” *Review of Economic Dynamics*, 3(3), 365–95, 2000.
- Danielsson, J., and C. G. de Vries, “Tail Index and Quantile Estimation with Very High Frequency Data,” *Journal of Empirical Finance*, 4(2–3), 241–57, 1997.
- Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries, “Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation,” *Journal of Multivariate Analysis*, 76(2), 226–48, 2001.
- Danielsson, J., L. M. Ergun, L. de Haan, and C. G. de Vries, “Tail Index Estimation,” *Quantile driven threshold selection*, 2016. <https://ssrn.com/abstract=2717478>

- Embrechts, P., and C. Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events: For Insurance and Finance*. vol. 33. Springer Science & Business Media, 2013.
- Fedotenkov, I., "A Review of More than One Hundred Pareto-Tail Index Estimators," *Statistica*, 80(3), 245–99, 2020.
- Gabaix, X., "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–93, 2009.
- _____, "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30(1), 185–206, 2016.
- Gabaix, X., and R. Ibragimov, "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business and Economic Statistics*, 29(1), 24–39, 2011.
- Gouin-Bonfant, É., and A. A. Toda, *Pareto Extrapolation: An Analytical Framework for Studying Tail Inequality*. 2018. <https://ssrn.com/abstract=3260899>
- Ivette Gomes, M., and A. Guillou, "Extreme Value Theory and Statistics of Univariate Extremes: A Review," *International Statistical Review*, 83(2), 263–92, 2015.
- Hall, P., "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of the Royal Statistical Society, Series B*, 44(1), 37–42, 1982.
- _____, "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems," *Journal of Multivariate Analysis*, 32(2), 177–203, 1990.
- Hill, B. M., "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, 3(5), 1163–74, 1975.
- Hoga, Y., "Detecting Tail Risk Differences in Multivariate Time Series," *Journal of Time Series Analysis*, 39(5), 665–89, 2018.
- Ibragimov, M., and R. Ibragimov, "Heavy Tails and Upper-Tail Inequality: The Case of Russia," *Empirical Economics*, 54(2), 823–37, 2018.
- Jenkins, S. P., "Pareto Models, Top Incomes and Recent Trends in UK Income Inequality," *Economica*, 84(334), 261–89, 2017.
- Jessen, A. H., and T. Mikosch, "Regularly Varying Functions," *Publications de l'Institut Mathématique*, 80(94), 171–92, 2006.
- Kesten, H., "Random Difference Equations and Renewal Theory for Products of Random Matrices," *Acta Mathematica*, 131(1), 207–48, 1973.
- Li, H., and J. Stachurski, "Solving the Income Fluctuation Problem with Unbounded Rewards," *Journal of Economic Dynamics and Control*, 45, 353–65, 2014.
- Luxembourg Income Study (LIS) Database. <http://www.lisdatacenter.org> (Multiple Countries; September 2019–June 2021). Luxembourg: LIS 2021.
- Ma, Q., and A. A. Toda, "A Theory of the Saving Rate of the Rich," *Journal of Economic Theory*, 192, 105193, 2021.
- Ma, Q., J. Stachurski, and A. A. Toda, "The Income Fluctuation Problem and the Evolution of Wealth," *Journal of Economic Theory*, 187, 105003, 2020.
- Mirek, M., "Heavy Tail Phenomenon and Convergence to Stable Laws for Iterated Lipschitz Maps," *Probability Theory and Related Fields*, 151(3–4), 705–34, 2011.
- Nirei, M., and W. Souma, "A Two Factor Model of Income Distribution Dynamics," *Review of Income and Wealth*, 53(3), 440–59, 2007.
- Oancea, B., D. Pirjol, and T. Andrei, "A Pareto Upper Tail for Capital Income Distribution," *Physica A: Statistical Mechanics and its Applications*, 492, 403–17, 2018.
- Pareto, V., "La legge della demanda," *Giornale degli Economisti*, 10, 59–68, 1895.
- _____, *La Courbe de la Répartition de la Richesse*. Imprimerie Ch. Viret-Genton, Lausanne, 1896.
- _____, *Cours d'Économie Politique*. vol. 2. F. Rouge, Lausanne, 1897.
- Piketty, T., "Income Inequality in France, 1901–1998," *Journal of Political Economy*, 111(5), 1004–42, 2003.
- Piketty, T., and E. Saez, "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, 118(1), 1–41, 2003.
- Reed, W. J., "The Pareto, Zipf and Other Power Laws," *Economics Letters*, 74(1), 15–9, 2001. [https://doi.org/10.1016/S0165-1765\(01\)00524-9](https://doi.org/10.1016/S0165-1765(01)00524-9).
- _____, "The Pareto Law of Incomes—an Explanation and an Extension," *Physica A*, 319(1), 469–86, 2003.
- Reed, W. J., and M. Jorgensen, "The Double Pareto-Lognormal Distribution—a New Parametric Model for Size Distribution," *Communications in Statistics-Theory and Methods*, 33(8), 1733–53, 2004.
- Resnick, S., and C. Stărică, "Smoothing the Hill Estimator," *Advances in Applied Probability*, 29(1), 271–93, 1997.
- Saez, E., "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–29, 2001.
- Saez, E., and S. Stantcheva, "A Simpler Theory of Optimal Capital Taxation," *Journal of Public Economics*, 162, 120–42, 2018.

- Siow, A., "Testing Becker's Theory of Positive Assortative Matching," *Journal of Labor Economics*, 33(2), 409–41, 2015.
- Stachurski, J., and A. A. Toda, "An Impossibility Theorem for Wealth in Heterogeneous-agent Models with Limited Heterogeneity," *Journal of Economic Theory*, 182, 1–24, 2019.
- _____, "Corrigendum to 'An Impossibility Theorem for Wealth in Heterogeneous-Agent Models with Limited Heterogeneity' [Journal of Economic Theory 182 (2019) 1–24]," *Journal of Economic Theory*, 188, 105066, 2020.
- Toda, A. A., "Income Dynamics with a Stationary Double Pareto Distribution," *Physical Review E*, 83(4), 046122, 2011.
- _____, "The Double Power Law in Income Distribution: Explanations and Evidence," *Journal of Economic Behavior and Organization*, 84(1), 364–81, 2012.
- _____, "Necessity of Hyperbolic Absolute Risk Aversion for the Concavity of Consumption Functions," *Journal of Mathematical Economics*, 94, 102460, 2021.
- Toda, A. A., and Y. Wang, "Efficient Minimum Distance Estimation of Pareto Exponent from Top Income Shares," *Journal of Applied Econometrics*, 36(2), 228–43, 2021.
- Vermeulen, P., "How Fat is the Top Tail of the Wealth Distribution?" *Review of Income and Wealth*, 64(2), 357–87, 2018.
- Wold, H. O. A., and P. Whittle, "A Model Explaining the Pareto Distribution of Wealth," *Econometrica*, 25(4), 591–5, 1957.
- Yaari, M. E., "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *Review of Economic Studies*, 32(2), 137–50, 1965.

SUPPORTING INFORMATION

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

Table 2: Point estimates of income Pareto exponents across countries and years.

Table 3: Hypothesis testing of $\hat{\alpha}_{\text{lab}} = \hat{\alpha}_{\text{cap}}$.