WORKING PAPERS

N° 1493

December 2023

"Consumption, Wealth, and Income Inequality: A Tale of Tails"

Alexandre Gaillard, Christian Hellwig,
Philipp Wangner and Nicolas Werquin

Toulouse School of Economics

Consumption, Wealth, and Income Inequality: A Tale of Tails*

Alexandre Gaillard
Brown University

Christian Hellwig
Toulouse School of Economics, CEPR

Philipp Wangner
Columbia Business School

Nicolas Werquin
Federal Reserve Bank of Chicago, CEPR

December 5, 2023

Abstract

We provide evidence that the distributions of consumption, labor income, wealth, and capital income exhibit asymptotic power-law behavior with a strict ranking of upper tail inequality, in that order, from the least to the most unequal. We show analytically and quantitatively that the canonical heterogeneous-agent model cannot replicate the proper ranking and magnitudes of these four tails simultaneously. Mechanisms addressing the wealth concentration puzzle in these models through return heterogeneity lead to a mirror *consumption concentration puzzle*. We match the cross-sectional data on these four Pareto tails by positing a combination of non-homothetic, wealth-dependent preferences and scale-dependent returns to capital. We underscore the importance of these results by showing that all four dimensions of top inequality jointly determine the long-run elasticity that governs the revenue-maximizing capital tax rate.

*Gaillard: alexandre_gaillard@brown.edu. Hellwig: christian.hellwig@tse-fr.eu. Wangner: pw2608@columbia.edu. Werquin: nwerquin@gmail.com. First version: September 2023. An earlier version of the paper was circulated under the title "Inequality Accounting: A Tale of Tails". We thank Andy Atkeson, Gadi Barlevy, Susanto Basu, Richard Blundell, Pierre-Edouard Collignon, Jeanne Commault, Mike Golosov, Magnus Irie, Moritz Kuhn, Jaromir Nosal, Victor Rios-Rull, Joseph Stiglitz, Guillaume Sublet, Robert Ulbricht, Nathan Zorzi for useful comments. All remaining errors are ours. Opinions expressed in this article are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System. Christian Hellwig acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010).

1 Introduction

Over the past thirty years, heterogeneous-agent models pioneered by Bewley (1986); Imrohoroğlu (1989); Huggett (1993); Aiyagari (1994) have become the workhorse macroeconomic framework for studying income and wealth dynamics. These models offer an intuitive account of precautionary saving motives to buffer idiosyncratic income shocks within a rich aggregate economy. They easily lend themselves to extensions and calibrations that allow us to understand the forces that shape income and wealth concentration in advanced economies. Not surprisingly, they have thus been used to account for trends of rising inequality and to inform policy debates over the costs and benefits of redistributive income and capital taxation, among many other applications.

In this paper, we argue that despite its many appealing features, the baseline heterogeneous agent model fails to properly account for the consumption and investment decisions of the rich, i.e., the roughly top 10% of the population. Concretely, the canonical framework implies that the consumption and capital income of the richest households should be asymptotically proportional to their wealth. As a result, the Pareto coefficients of the upper tails of the consumption, capital income, and wealth distributions should theoretically all coincide. However, we evaluate empirically these measures of top tail inequality and conclude that they systematically differ—thus rejecting the canonical model. We propose instead a framework that generalizes the baseline model to replicate these cross-sectional moments from the data. To match all three Pareto tail coefficients, the model must be such that (i) on average, wealthier agents earn higher returns on capital; and (ii) they accumulate wealth for reasons other than funding their own future consumption. We argue that accounting for these mechanisms matters for the revenue-maximizing top capital tax rate.

We proceed in four steps. First, using the Panel Study of Income Dynamics (PSID) as our primary dataset, we establish a new empirical fact about the distributions of household-level consumption, labor income, wealth, and capital income: They all exhibit Pareto tails at the top, with a strict ranking going from thinnest to thickest, in that order. Specifically, the consumption tail is the most evenly distributed, with a Pareto coefficient above 3.0. Labor income is strictly more concentrated than consumption, with a Pareto coefficient around 2.2. Wealth has a strictly thicker tail than labor income, with a Pareto coefficient of 1.4. Finally, capital income is the most unequal distribution, with a Pareto coefficient around 1.2. Moreover, for all four variables, the best-fitting Pareto tail estimates closely match the data for the top 10% to top 5% of the population. These findings are robust to using different datasets (SCF, CEX, administrative IRS records) and adjusting for under-reporting and under-representation of the rich in survey data.

In the second part of our analysis, we set up a tractable, continuous-time environment with idiosyncratic labor productivity and investment return shocks to study theoretically which class of models is able to replicate these findings. The canonical Aiyagari (1994) model is notoriously unable to generate the high empirical levels of wealth inequality at the top of the distribution, giving rise to a *wealth concentration puzzle*. Modern versions of the heterogeneous-agent model were designed to address this puzzle, i.e., to simultaneously match the distributions of labor income and wealth, typically via a mechanism involving *type-dependent returns* to capital (e.g., heterogeneous portfolio investment skills or entrepreneurship). However, this class of models has two potential shortcomings.

First, type-dependent returns imply—at least in theory—that capital income is asymptotically linear in wealth. Thus, the Pareto coefficients of capital income and wealth converge to the same value as wealth grows arbitrarily large. That is, the model cannot decouple the two tail coefficients for the richest households. For the discrepancy between these two tails to persist at sufficiently high levels of wealth, the model must feature *scale-dependent returns*, i.e., a mechanism by which wealthier agents earn higher returns by virtue of being wealthy. Capital income then becomes convex in wealth at the top, rather than linear, thus rationalizing its higher concentration.

Second, and most importantly, allowing the canonical model to replicate the extreme concentration of wealth among top earners via type-dependent returns generates a far too high concentration of consumption at the top. In other words, because this modern class of models retains the fundamental structure whereby consumption is asymptotically linear in wealth, the mechanisms introduced to solve the wealth concentration puzzle mechanically create the opposite issue, a *consumption concentration puzzle*. What is more, introducing scale-dependent returns, as described above, to decouple the capital income tail from the wealth tail, aggravates this puzzle—it implies that consumption is convex in, and hence should be even more concentrated than, wealth. We show that introducing non-homothetic, wealth-dependent preferences allows us to decouple the consumption tail from the wealth tail. That is, we assume that for reasons that we do not seek to model explicitly, the rich draw utility directly from being wealthy. The non-homotheticity implies that the marginal utility is less steeply declining for wealth than for consumption, or in other words, wealth appears to be a luxury good.[1] Consumption then becomes concave in wealth at the top,

[1] One interpretation of these preferences is that households face a positive probability of death at each

which rationalizes its lower concentration.

The third part of our paper examines whether these departures from the canonical model of consumption-saving decisions are quantitatively important. While our theoretical results focus on the limit as wealth becomes arbitrarily large, Carroll and Kimball (1996) show that at finite levels of wealth, standard preferences over consumption only, and precautionary saving motives, are sufficient to make consumption concave in—and thus strictly less concentrated than—wealth. This raises the possibility that our finite sample estimates of Pareto tail coefficients for consumption are upward-biased, and thus that standard models could still be consistent with the ordering of estimated Pareto coefficients. Analogously, type-dependence might bias downwards the estimated Pareto coefficient for capital income in finite samples, due to a composition effect by which the share of high-return type agents increases along the wealth distribution. To gauge the importance of these concerns, we structurally estimate the key parameters of a model featuring a skewed labor income distribution, return heterogeneity—both type- and scale-dependence—and non-homothetic preferences, and compare its predictions against those of the canonical heterogeneous-agent model. Importantly, we target the Pareto tail coefficients by simulating model-based samples of the relevant variables in a manner that closely mirrors our empirical analysis.

On the one hand, our model-based estimation confirms that non-homothetic preferences are key in accounting for the empirical ranking and levels of the Pareto tail coefficients. Concretely, the concavity of consumption implied solely by precautionary saving motives is not sufficient to generate the relative magnitudes of consumption and wealth concentration at the top. Absent return heterogeneity, the consumption tail implied by the model—calibrated to match the concentration of top labor incomes—has the correct magnitude, but wealth is not concentrated enough. Introducing type- or scale-dependent returns generates the right magnitudes of income and wealth concentration, but consumption now becomes far too concentrated. By contrast, by relaxing the asymptotic linearity of consumption, non-homothetic preferences can account for the tail properties of all of these variables simultaneously.

On the other hand, we find that the composition effect generated by type-dependent returns is sufficiently strong to generate enough convexity of capital income in wealth at the finite levels of wealth observed in PSID, and hence to match quantitatively the ratio of these two Pareto coefficients estimated in survey data. However, using administrative records from Saez and Zucman (2016) that cover a much larger range of wealth and capital income quantiles allows us to disentangle the relative importance of type- vs. scale-dependent re-

instant and experience a warm glow utility from transferring their assets to their heirs.

turns. This is because the composition effect becomes negligible at such high wealth levels—implying that any remaining discrepancy between the two Pareto coefficients estimated at or above the top 1% can be attributed to scale-dependence. This identification procedure reveals a strictly positive degree of scale-dependence in the data.

As our fourth and final step, we illustrate the importance of matching all four Pareto tails by computing revenue-maximizing taxes on capital—i.e., the upper bound of the set of Pareto efficient top tax rates—under different model alternatives. Non-homothetic preferences, type- and scale-dependence of asset returns all introduce different additional motives for taxing or subsidizing capital—or equivalently, they all factor into the elasticity of savings to after-tax returns. The combined strength of these motives to tax capital is then identified from the different Pareto tail coefficients. Ignoring any of them thus inevitably biases conclusions about optimal taxation. If we were to abstract from matching the Pareto tails on consumption or capital income, we would conclude that the revenue-maximizing tax on capital was higher than in our full model. This result stems from two opposing forces. On the one hand, both scale-dependent returns and non-homothetic preferences lead to convex saving functions. As such, the Le Châtelier principle raises the long-run capital supply elasticity and calls for lower top marginal tax rates; see, e.g., Stiglitz (1969); Bourguignon (1981). On the other hand, non-homothetic preferences deliver additionally a characterization of wealth as a "luxury good", which offers a rationale for higher capital taxation in line with departures from Atkinson and Stiglitz (1976)'s uniform commodity taxation theorem. As a result, the optimal capital tax depends on the extent to which the convex savings function is generated by non-homothetic preferences vs. scale-dependent returns, when the model is calibrated to match the observed Pareto tails. In sum, the distributions of consumption and capital income inform us about the underlying mechanisms that drive the long-run elasticity of capital and, therefore, the optimal policy design.

Related Literature. Our paper contributes to an extensive literature that extends the Bewley-Imrohoroglu-Hugget-Aiyagari model to match the empirical wealth concentration: e.g., Carroll (1998); Krusell and Smith Jr. (1998); Quadrini (2000); Castaneda et al. (2003); Cagetti and De Nardi (2006); Benhabib et al. (2011, 2015); Aoki and Nirei (2017); Benhabib and Bisin (2018). While many recent papers use type-dependent returns to match the upper tail of the wealth distribution, De Nardi (2004); Straub (2019); Benhabib et al. (2019); Mian et al. (2021); Lee (2021); Michau et al. (2023) introduce, as we do, non-homothetic wealth-dependent preferences. Blanchet (2022) highlights the role of wealth-dependent savings rates for the evolution of top wealth inequality, while Hubmer et al. (2021) emphasizes the importance of scale-dependent returns.[2] In contrast to all of these papers, (i) we study jointly the distributions of all of the variables that appear in the budget constraint—consumption, labor income, wealth, and capital income—rather than a strict subset of these moments in isolation; and (ii) we focus on their Pareto tails, which delivers both straightforward cross-sectional estimation and stark theoretical identification of the required departure from the canonical model, i.e., the underlying degrees of non-homotheticity and scale-dependence.

Empirically, we are the first simultaneously and comprehensively estimate, using a single dataset, the tails of the consumption, labor income, wealth, and capital income distributions, and by identifying the model features that are necessary to match their ranking and magnitudes. Our results are consistent with prior research that explores the relative concentration of a subset of these distributions (e.g., Dynan et al. (2004); Krueger and Perri (2006); Toda and Walsh (2015); Blundell et al. (2016); Saez and Zucman (2016); Straub (2019); de Vries and Toda (2022); Garner et al. (2022); Buda et al. (2022); Meyer and Sullivan (2023)) or the relative importance of type- and scale-dependent returns to capital (e.g., Fagereng et al. (2020); Bach et al. (2020); Xavier (2021); Balloch and Richers (2021)). Methodologically, our estimation applies and extends the techniques of Clauset et al. (2009); Vermeulen (2016), and our theoretical analysis leverages the continuous-time techniques developed by Moll (2014); Gabaix et al. (2016); Achdou et al. (2022). Our application to capital taxation underscores the importance of accurately accounting for the distributions of consumption and capital income to discipline the savings elasticity, and builds on previous papers studying redistributive policies with non-homothetic preferences or heterogeneous returns: Atkinson (1971); Saez and Stantcheva (2018); Gerritsen et al. (2020); Gaillard and Wangner (2021); Schulz (2021); Hellwig and Werquin (2022); Yi (2022); Ferey et al. (2023); Morrison (2023); Guvenen et al. (2023); Boar and Midrigan (2023).

Structure of the Paper. Section 2 contains the empirical analysis. Section 3 outlines the analytical framework and provides theoretical insights on the underlying mechanisms that generate the accurate ranking of Pareto tails. We set up and estimate the quantitative framework in Section 4, and explore its implications for capital taxation in Section 5. Additional results and technical details are gathered in Appendices A to C.

[2] Ma and Toda (2021) propose an alternative potential justification for the thin consumption tail, by showing that homothetic preferences may be consistent with asymptotic MPC converging to zero for a particular combination of parameters, notably the coefficient of relative risk aversion and the stochastic process for the discount rate. We evaluate their mechanism quantitatively and show that it cannot match the four tail parameters in practice for reasonable parameter values.

2 Empirical Evidence

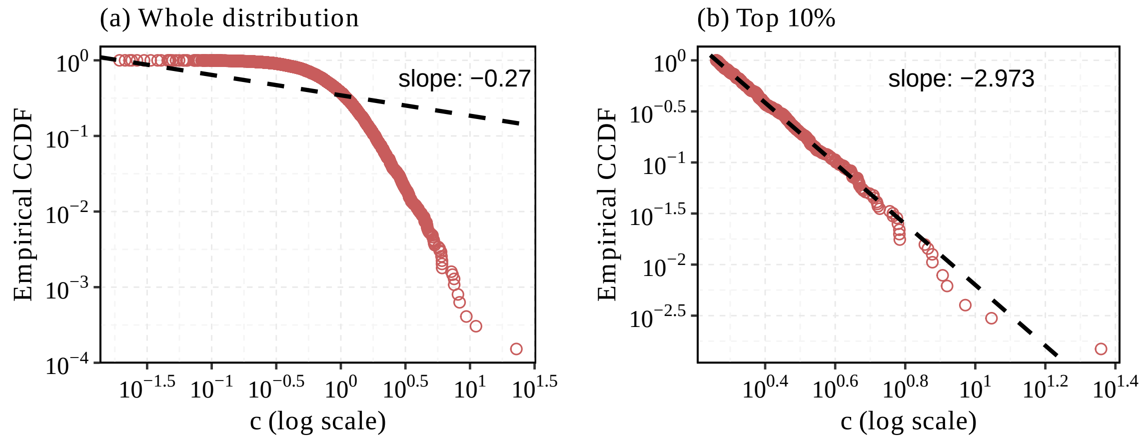
We begin by examining the upper tails of the distributions of consumption, labor income, wealth, and capital income in the data. The main results of this section are twofold: (i) all four tails are well-approximated by Pareto distributions; (ii) these tails obey a strict ranking from thinnest to thickest, in the order given in the previous sentence. Our analysis proceeds in several steps. First, in Section 2.1, we lay out the methodology that we use to estimate the Pareto tails. In Section 2.2, we describe our data and main variables. In Section 2.3, we propose a procedure to adjust for the potential under-representation and under-reporting in survey data. In Section 2.4, we present our main empirical results. Finally, we explore the robustness of our findings in Sections 2.5 and 2.6.

2.1 Methodology

Pareto Distributions. Consider a random variable X with cumulative distribution function (CDF) $F_X(x)$, complementary CDF (CCDF) $\bar{F}_X(x) = 1 - F_X(x)$, and probability density function (PDF) $f_X(x)$. We say that X has a Pareto (or power-law) distribution if there exist $\underline{x} > 0$ and $\zeta_X > 0$ such that its CCDF satisfies $\bar{F}_X(x) = (\underline{x}/x)^{\zeta_X}$ for all $x \geq \underline{x}$. We say that X is asymptotically Pareto distributed, or that is has a Pareto tail, if there exists $\zeta_X > 0$ such that $\bar{F}_X(x)/x^{-\zeta_X}$ converges to a positive constant as $x \to \infty$. We call the parameter ζ_X the *Pareto coefficient* of the tail of the distribution. A larger value of ζ_X indicates a thinner tail, or a more equal distribution.

The traditional approach to attest the existence of a Pareto tail in the data and measure its thickness is to verify whether the variable displays an approximately log-linear relationship between its CCDF and its log-values in the upper tail. Figure 1 represents this relationship for consumption using the first year of our PSID sample (2004), over an interval $c \geq \underline{c}$. The left panel displays the entire distribution ($\underline{c} = 0$), and the right panel focuses on the top 10 percentiles ($\underline{c} = \$78,000$ in 2004 dollars or, as a fraction of the mean consumption, $\underline{c} = 1.84$). While the relationship appears to be linear in the upper tail, two issues arise. First, it is *a priori* unclear where to locate the lower bound of the tail, $\underline{c}$. Second, the relationship departs from linearity at the very top of the sample—we thus need a method to assess whether such a departure should be interpreted as inevitable finite-sample noise or is informative about actual underlying changes in the shape of the distribution. To address these issues, we employ a purely data-driven procedure based on the methodology of Clauset

Figure 1. Pareto tail estimation of consumption in the 2004 PSID wave.



et al. (2009) to (i) identify a best-fitting Pareto tail for a finite data set; (ii) statistically test the null hypothesis that the distribution indeed displays a Pareto tail; and (iii) statistically test whether an alternative distribution—e.g., lognormal or exponential— gives a better fit.

Estimating the Asymptotic Pareto Distribution. Throughout this section, we consider a generic sample of survey data $\{x_i\}_{i=1}^n$ with sample weights $\{\omega_i\}_{i=1}^n$, where $\sum_{i=1}^n \omega_i = 1$. To assess whether the distribution of X has a Pareto tail and estimate the corresponding coefficient ζ_X, we proceed as follows. As mentioned above, this hypothesis implies an approximately linear relationship between $\ln \bar{F}(x_i)$ and $\ln x_i$ for large enough values of x_i, say $x_i \geq \underline{x}$. The slope of this linear relationship gives the tail coefficient $-\zeta_X$. For any given $\underline{x}$, we can thus estimate ζ_X via maximum-likelihood (MLE) or ordinary least squares (OLS) as follows:

$$\hat{\zeta}_X^{MLE}(\underline{x}) = \frac{\sum_{i=1}^n \omega_i \mathbb{1}_{\{x_i \geq \underline{x}\}}}{\sum_{i=1}^n \omega_i \ln(\underline{x}/x_i) \mathbb{1}_{\{x_i \geq \underline{x}\}}}, \quad \hat{\zeta}_X^{OLS}(\underline{x}) = \frac{\sum_{i=1}^n \omega_i \ln(\underline{x}/x_i) \ln \bar{F}(x_i) \mathbb{1}_{\{x_i \geq \underline{x}\}}}{\sum_{i=1}^n \omega_i (\ln(\underline{x}/x_i))^2 \mathbb{1}_{\{x_i \geq \underline{x}\}}}. \tag{1}$$

Notice that these estimators depend on the choice of $\underline{x}$. Choosing a value that is too low may result in fitting a power-law model to non-power-law data, while a too high value may discard valid data points. To find the optimal $\underline{x}$, we compute the Kolmogorov-Smirnov (KS) statistic:

$$KS(\underline{x}, \hat{\zeta}_X) = \sup_{x \geq \underline{x}} |\bar{F}(x) - \hat{F}(x)|,$$

where $\bar{F}(x)$ is the empirical CCDF for the observations with value at least as large as $\underline{x}$, and $\hat{F}(x)$ is the CCDF of the Pareto distribution that best fits the data—as described in the previous paragraph—on the sub-sample $x_i \geq \underline{x}$. We then choose the value of $\underline{x}$ that minimizes the KS statistic:

$$\hat{\underline{x}} = \underset{\underline{x} \in \mathbb{R}_+}{\arg\min} \;\; KS(\underline{x}, \hat{\zeta}_X). \tag{2}$$

This method makes the best-fit power-law distribution as close as possible—in the KS-statistic sense—to the empirical distribution above $\hat{\underline{x}}$.

Statistical Tests of the Pareto Hypothesis. To evaluate the goodness-of-fit of the Pareto distribution, we start by estimating the parameters $(\hat{\underline{x}}, \hat{\zeta}_X)$ that best fit the data and the resulting KS statistic $KS(\hat{\underline{x}}, \hat{\zeta}_X)$. We then generate semi-parametrically a large number of synthetic datasets, indexed by k, as follows. With probability $\pi = \sum_{i=1}^n \omega_i \mathbb{1}_{\{x_i \geq \hat{\underline{x}}\}}$, a random observation x_i is drawn from a Pareto distribution with parameters $(\hat{\underline{x}}, \hat{\zeta}_X)$ and sample weight π/n. With probability $1-\pi$, an observation below the threshold $\hat{\underline{x}}$ is randomly chosen from the initial sample. We then fit each synthetic dataset k individually to its own power-law model as described in equations (1)-(2), and denote the resulting parameters by $(\hat{\underline{x}}^{(k)}, \hat{\zeta}_X^{(k)})$ and the KS-statistic by $KS(\hat{\underline{x}}^{(k)}, \hat{\zeta}_X^{(k)})$. We reject the power-law hypothesis if the fraction p of synthetic data sets for which $KS(\hat{\underline{x}}^{(k)}, \hat{\zeta}_X^{(k)}) > KS(\hat{\underline{x}}, \hat{\zeta}_X)$ is small enough, say $p < 0.1$.

Finally, to compare the goodness-of-fit of the power-law model (with PDF $\hat{f}(x_i)$) to an alternative distribution with PDF $\hat{g}(x_i)$ for $x_i \geq \hat{\underline{x}}$, we compute the log-likelihood ratio $R = \sum_{i=1}^n \omega_i \ln(\hat{f}(x_i)/\hat{g}(x_i))\mathbb{1}_{\{x_i \geq \underline{x}\}}$. The power-law distribution outperforms the alternative distribution if R is statistically strictly positive.[3]

2.2 Data

2.2.1 Data Sources. Our main analysis uses the 2005–2021 waves of the Panel Study of Income Dynamics (PSID), since most consumption items are reported from 2005 onward. For robustness, we complement this analysis with results based on the 1989–2019 waves of the Survey of Consumer Finances (SCF) and the 2001–2017 waves of the Consumer Expenditure Survey (CEX). For consistency, we construct all of the variables at the household level.[4] We restrict our analysis to households whose head is between 19 and 80 years old, i.e., who make active consumption and saving decisions. For clearer comparison across surveys, we normalize all of the variables by their respective mean value.

[3] The p-value is $p = 1 - \text{erf}(|R|/(\sigma_R\sqrt{2n}))$, where $\text{erf}(\cdot)$ is the Gauss error function and the standard deviation is $\sigma_R^2 = \sum_{i=1}^n \omega_i[\ln(\hat{f}(x_i)/\hat{g}(x_i))]^2\mathbb{1}_{\{x_i \geq \underline{x}\}} - [\sum_{i=1}^n \omega_i \ln(\hat{f}(x_i)/\hat{g}(x_i))\mathbb{1}_{\{x_i \geq \underline{x}\}}]^2$. If p is small enough, say $p < 0.1$, then the measured sign of R is a reliable indicator of which model is the better fit to the data.

[4] In Section 2.5, we also report the results obtained by individualizing consumption expenditures.

Using the PSID as our primary dataset brings several advantages over our alternative datasets. First, it allows us to observe simultaneously the consumption, labor income, wealth, and capital income of each household, thus enabling a direct comparison of the thickness of all four tails using the same data. The SCF, known for oversampling at the higher end of the wealth distribution and having fewer missing observations than the PSID and CEX, unfortunately only includes data on wealth, income, and food consumption, lacking broader consumption data. By contrast, the CEX includes all those categories but is notoriously unreliable at the top end (see, e.g., Attanasio and Pistaferri (2014)). Second, anticipating our results, we find that the PSID yields estimates of the Pareto coefficients that are consistent with those obtained from the SCF or admistrative records from Saez and Zucman (2016) for the variables that they report.

2.2.2 Construction of the Variables. We define "wealth" as the net worth of households, calculated by subtracting liabilities (mortgages and home equity loans, student loans, medical debt, legal debt, family loans, credit card debt, and other credits) from the total value of assets comprising liquid assets (checking accounts, money market funds, certificates of deposit, government bonds, treasury bills) and illiquid assets (private equity businesses, real estate properties and vehicles, directly or indirectly held stocks, retirement accounts such as IRA and 401K, and "other savings or assets" such as cash value in a life insurance policy, a valuable collection for investment purposes, or rights in a trust or estate). The value of vehicles is depreciated at a 15% rate per year.

Part of our analysis requires disentangling liquid from illiquid wealth. We follow Kaplan et al. (2018) and construct "liquid wealth" in the PSID and the SCF as liquid assets net of revolving consumer credit card debt, and "illiquid wealth" as illiquid assets net of mortgages and home equity loans. Including stocks as part of liquid, rather than illiquid, wealth is inconsequential for our results.

"Labor income" is computed gross of taxes, benefits, and employee payroll deductions. It includes regular wages, bonuses, commissions, worker compensation, the part of business income that does not accrue to business assets, and unemployment insurance (social insurance benefits do not affect top income inequality). "Capital income" includes dividends, interests, the part of business income that accrues to business assets, rents, trusts and royalties and, when available, realized capital gains from the sale of assets. We refer to "total income" as the sum of labor income and capital income.

Finally, "consumption" refers to total expenditures, including various categories such as electricity, health, nondurables (such as food and gasoline), entertainment, trips, housing services (i.e., actual and imputed rents, where we follow Blundell et al. (2016) to measure the latter as 6% of the house value for homeowners), utilities (electricity, internet, water), transportation, clothing, furniture, vehicle repayments, leasing, maintenance, and repair. In addition, since 2003, the PSID reports detailed philanthropic donations, spanning various categories such as religion, combined purpose funds, basic needs assistance, healthcare and medical research, education, youth and family services, arts and culture, community improvement, environmental preservation, international aid, pandemic relief, and other purposes. Our baseline results include these charitable donations in the consumption variable. We remove observations for which households report not consuming any food. We explore the robustness of our results to alternative constructions of the consumption variable (such as including durable goods and individualizing consumption) in Section 2.5.

One might be concerned that the consumption of the richest households is imperfectly measured—we do not observe, for instance, direct political donations or extreme luxury goods such as jets, yachts, or artwork—thus possibly biasing our estimates of the Pareto coefficient upwards. Anticipating our empirical results, however, we find that the Pareto tails are already clearly apparent at relatively low quantiles—in particular, from the 88^{th} percentile onwards for consumption. This finding alleviates the need to precisely observe the behavior of the very richest agents. Nevertheless, the next section describes our procedure to adjust for under-reporting and under-representation at the top of the consumption, income, and wealth distributions.

2.3 Adjusting for Under-Reporting and Under-Representation

Methodology. Survey data often suffer from under-reporting of consumption, income, and wealth, as well as under-representation at the upper end of the distribution due to the limited survey participation of wealthy individuals.[5] One significant concern stems from the observed differences between survey estimates and their counterparts in national accounts. To address this concern, we employ an iterative adjustment approach.[6] Specifically, we align

[5] See, e.g., Koijen et al. (2014). In the Appendix, we argue that the alternative approach that consists of using administrative data to measure consumption as the residual of a budget constraint—if wealth, labor income, and capital income are observed—has its own pitfalls. In particular, measurement error in wealth (due to, e.g., offshoring) might lead to a spurious equality between the Pareto coefficients of measured consumption and measured wealth.

[6] Vermeulen (2016) employs a similar procedure for wealth using the SCF. Similarly, Jaravel and Lashkari (2023) adjust for under-reporting in CEX consumption using the BEA aggregates.

the aggregate survey values for the subcategories of each variable to their counterparts in the National Income and Product Accounts (NIPA) from the Bureau of Economic Analysis (BEA), and the Financial Accounts (FA) from the Board of Governors.

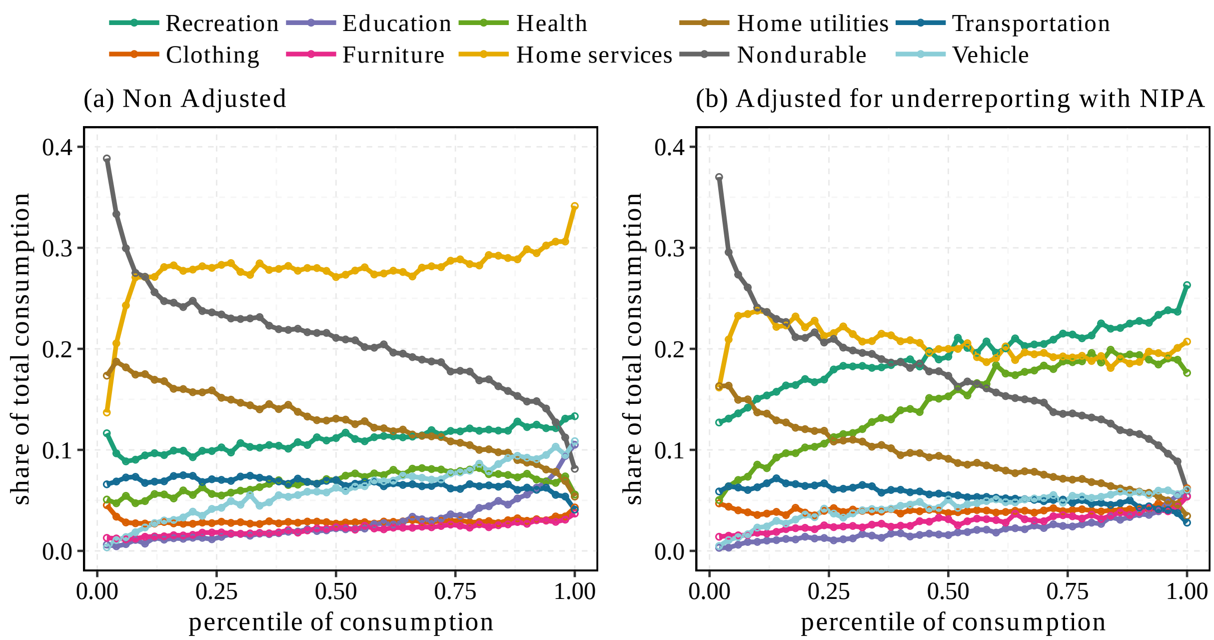
First, we create an *under-reporting-adjusted survey* by multiplying each subcategory of consumption (food, recreation, clothing, transportation, etc.), income (labor income, business income, dividends, interests, etc.), and wealth (real assets, financial assets, liabilities) by the category-specific ratio between national account and PSID aggregates. Appendix A.1 provides the correspondence tables that we use to merge survey categories with the NIPA and the FA. Although these adjustment factors are uniform across households, heterogeneous wealth portfolios, income sources, and consumption patterns lead to distinct adjustments across the distribution of household's total wealth, income, and consumption. For instance, if there is under-reporting of financial assets, the wealthiest households with a significant share of those assets would experience upward adjustments to their wealth in comparison to other households. Similarly, under-reported entertainment expenditures have a greater impact on high-consumption households, who allocate a larger fraction of their spending towards entertainment.

Second, we adjust for the potential under-representation of wealthy individuals—doing so reduces the risk that the previous procedure corrects too much for under-reporting in order to match aggregate quantities. For all of the variables, we reconstruct "missing data" at the very top based on the fitted Pareto distribution with parameters $\hat{\zeta}_X$ and $\hat{\underline{x}}$ estimated from the under-reporting-adjusted survey using equations (1) and (2). We can then compute aggregate consumption, income, and wealth by combining the under-reporting-adjusted survey for $x < \hat{\underline{x}}$ with the reconstructed Pareto distribution for $x \geq \hat{\underline{x}}$. We obtain estimates within each subcategory of consumption, income, and wealth in the Pareto tail segment by applying the average share of that subcategory in the overall category in the under-reporting-adjusted survey above $\hat{\underline{x}}$. We then compare the aggregate value of each subcategory of consumption, income, and wealth to its respective value in the NIPA and the FA. When discrepancies occur, we update the adjustment factors accordingly and re-estimate the Pareto coefficients. We iterate this procedure until the implied adjustment factors are invariant. As such, our data and the aggregate tables not only align in terms of overall quantities, but also across subcategories.[7]

[7]Note that this iterative procedure imposes that the Pareto distribution suitably represents the very top—beyond the values observed in our data set—of each distribution. Importantly, the under-representation adjustment is only meant to improve the accuracy of the corresponding Pareto coefficient estimate and does not affect the formal tests establishing the existence of such Pareto tails. Under-reporting adjustments can

Consumption Share Adjustments. For brevity, we report here the resulting adjustments to the distribution of consumption. We provide analogous results for the other variables in Appendix A.1.[8] Figure 2 illustrates how adjusting the consumption data in the PSID by applying factor adjustments to match these NIPA aggregates alters the consumption profile across the distribution. Particularly noteworthy is the considerable increase in the share of expenditure for health and recreation components compared to the raw PSID data. For example, at the 80^{th} percentile, the recreation share rises from 13.5% to 23%, while the health consumption share jumps from 8.5% to 20%. By contrast, there are reductions in the shares of nondurables and housing services, which decrease from 19% to 14% and 30% to 20%, respectively.

Figure 2. Consumption shares by category and percentile of consumption in the PSID.



2.4 Estimating the Pareto Coefficients

Throughout this section, we denote the parameters of the consumption, labor income, wealth, and capital income distributions with the subscripts c, y, a, ra, respectively. Our main results are based on the MLE estimates and are consistent with those obtain by OLS estimation.

alternatively be made without any under-representation adjustments, or using different distributions (e.g., lognormal). In practice, all of these adjustments yield very similar fixed point estimates.

[8] Table 5 gives the ratio of consumption in the PSID and in the NIPA tables for each subcategory. Four of them, namely clothing, entertainment, furniture, and health, are significantly distant from the NIPA estimates.

2.4.1 Graphical Evidence: Consumption. The left panel of Figure 3 plots the values of the consumption Pareto coefficient $\hat{\zeta}_c^{MLE}(\underline{c})$ obtained via MLE (equation (1)) in 2004 for values of the lower bound parameter $\underline{c}$ ranging from the 70^{th} to the 99^{th} percentiles. Intuitively, the best-fit Pareto coefficient $\hat{\zeta}_c^{MLE}(\hat{\underline{c}})$ and the optimal lower bound of the Pareto tail $\hat{\underline{c}}$ coincide with the values that they take in the region where $\hat{\zeta}_c^{MLE}(\underline{c})$ remains stable when varying the lower bound $\underline{c}$ of the truncated sample. To confirm this intuition, the middle panel of Figure 3 plots the KS-criterion (equation (2)) as a function of $\underline{c}$, along with the vertical line that selects the minimum of this KS statistic, obtained for $\underline{c} = \hat{\underline{c}}$. The corresponding estimate of the Pareto coefficient $\hat{\zeta}_c^{MLE}(\hat{\underline{c}})$ is represented by the horizontal line in the left panel. The resulting Pareto coefficient is equal to $\hat{\zeta}_c^{MLE} = 2.99$ and the Pareto tail starts at the 87.5^{th} percentile of the consumption distribution, corresponding to $\hat{\underline{c}} = 1.70$ (recall that we normalize all variables by their respective mean), or a nominal value for annual consumption of $\$78,000$ in 2004 dollars. Finally, as in Figure 1, the right panel of Figure 3 plots the relationship between $\ln \bar{F}_c(c)$ and $\ln c$ in 2004—but on an interval of values above the 87.5^{th} percentile of the distribution—along with the best linear trend (black dashed line) obtained from the MLE estimator $\hat{\zeta}_c^{MLE}(\hat{\underline{c}})$. The relationship is very close to linear. Above the 97^{th} percentile, the data becomes noisier, and the observations tend to lie below the best Pareto fit, indicating, if anything, that the tail of consumption becomes even thinner at the very top of the distribution.

Figure 3. Pareto tail estimation for consumption in the 2004 PSID wave.

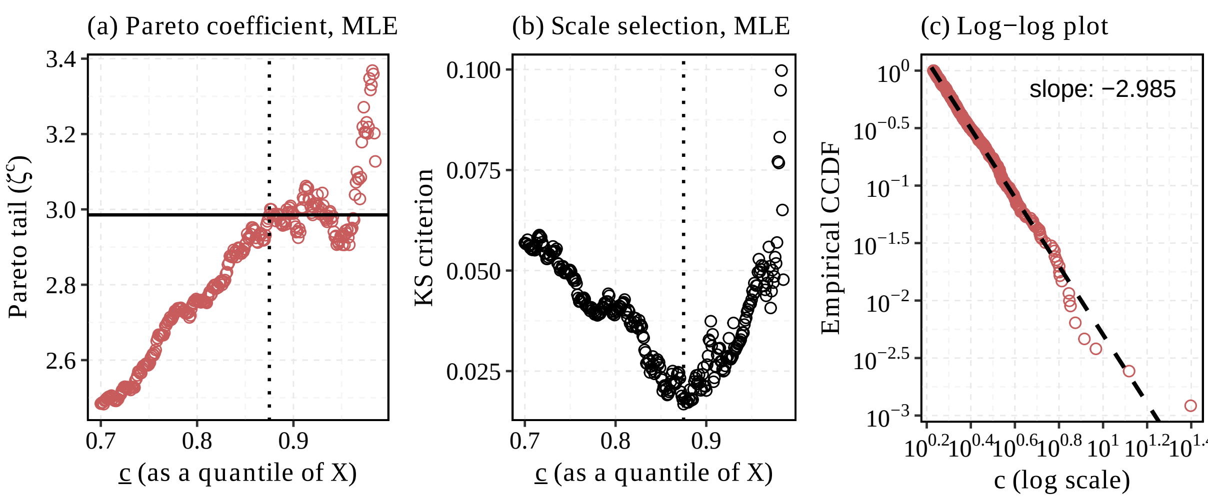


Figure 8 in Appendix A applies the same methodology to simulated data from several well-known parametric distributions: Pareto, lognormal, and double-Pareto-lognormal (DPLN). Such graphical evidence from distributions that possess or lack Pareto tails by construction serves as a useful benchmark for our empirical results. The similarity between the results of Figure 3 and those obtained with simulated DPLN distributions—in particular, the range of quantiles of $\underline{c}$ over which the estimated Pareto coefficient $\hat{\zeta}_c(\underline{c})$ remain stable—provides suggestive graphical evidence that consumption indeed has an upper Pareto tail with a well-identified slope and lower bound parameter. Finally, Figures 9 to 20 in the Appendix present graphs analogous to those of Figure 3 for every year in our sample and for the distributions of labor income, wealth, and capital income.

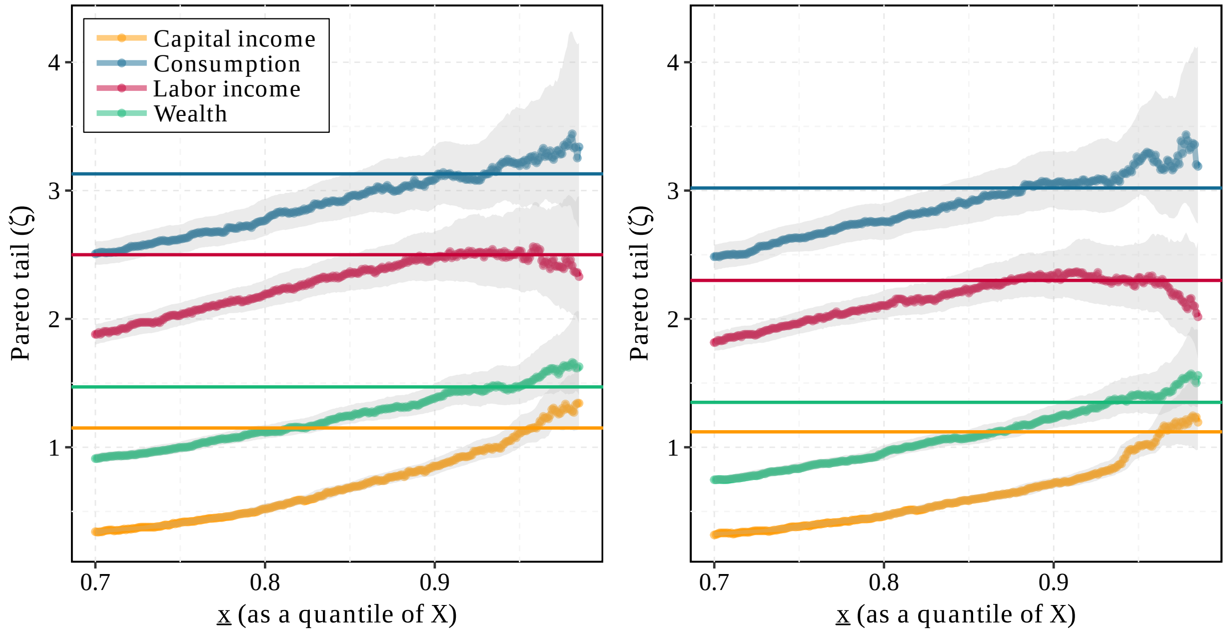
2.4.2 Main Result: A Strict Pareto Tail Ordering. Figure 4 presents our main empirical result. In both panels, the thick dotted lines indicate the median estimates, across all years in the PSID dataset, of the Pareto coefficients of the four variables (consumption, labor income, wealth, and capital income), as a function of the lower bound parameter $\underline{x}$. The four horizontal lines then give the median optimal Pareto coefficient for each variable, obtained by the procedure described in Section 2.1. The gray areas represent the median 90% confidence interval at each value of $\underline{x}$, obtained by bootstrapping estimates 1000 times for each wave.[9] In the left panel, the resulting Pareto tail estimates are shown without adjustments. In the right panel, we present the estimates obtained by applying our under-reporting and under-representation adjustments, as described in Section 2.3.

This figure shows that the Pareto tails of consumption, labor income, wealth, and capital income follow the strict order $\hat{\zeta}_c > \hat{\zeta}_y > \hat{\zeta}_a > \hat{\zeta}_{ra}$, with smaller coefficients corresponding to thicker tails, or more unequal distributions. While these coefficients are the medians across all years, the confidence intervals show furthermore that: (i) the estimates are remarkably similar across recent years (the intervals are tight around the median), and (ii) the differences in Pareto coefficients are significant (the intervals do not overlap, with the exception of wealth and capital income for very high values of $\underline{x}$). Recall furthermore that Figures 9 to 20 in Appendix A give the full set of estimated coefficients in all PSID waves: The respective ranges of estimates of any two different variables never overlap.

The median estimates for adjusted data across all years are $\hat{\zeta}_c = 3.02$, $\hat{\zeta}_y = 2.30$, $\hat{\zeta}_a = 1.35$, and $\hat{\zeta}_{ra} = 1.12$. The ranking and relative magnitudes of these coefficient estimates are broadly preserved by the under-reporting and under-representation adjustments. The median Pareto coefficient estimate for consumption declines slightly from 3.13 in the unadjusted data to 3.02 in the adjusted series. Similarly, that of labor income declines from 2.50 to 2.30, that of wealth declines from 1.42 to 1.35, and that of capital income declines from 1.24 to

[9]The bootstrap procedure generates replicate weights from the PSID using the relevant stratum and cluster variables.

Figure 4. Pareto tail estimates of the consumption, labor income, wealth, and capital income distributions for non-adjusted variables (left panel) and for under-reporting and under-representation adjusted variables (right panel).



1.16. In addition to slightly increasing the measured concentration of consumption, income, and wealth, these adjustments lead to more precise estimates.[10]

In sum, while it is well-known in the literature that the tail of the labor income distribution is strictly thinner than that of wealth, our main novel findings are that, using the same sample of households: (i) the distribution of consumption has a strictly thinner tail than labor income and, *a fortiori*, wealth; and (ii) the distribution of capital income has a thicker tail than wealth.[11,12]

Importantly, the series in Figure 4 show that these Pareto tail coefficients stabilize at relatively low levels of consumption ($82,000), labor income ($137,000), wealth ($785,000), and capital income ($9,000) (measured in 2016 dollars), corresponding roughly to the top 10 to 5 percent of households. This can also be seen graphically in Figure 5, which reports the average KS-statistic across years for each of the four variables—recall that its minimum

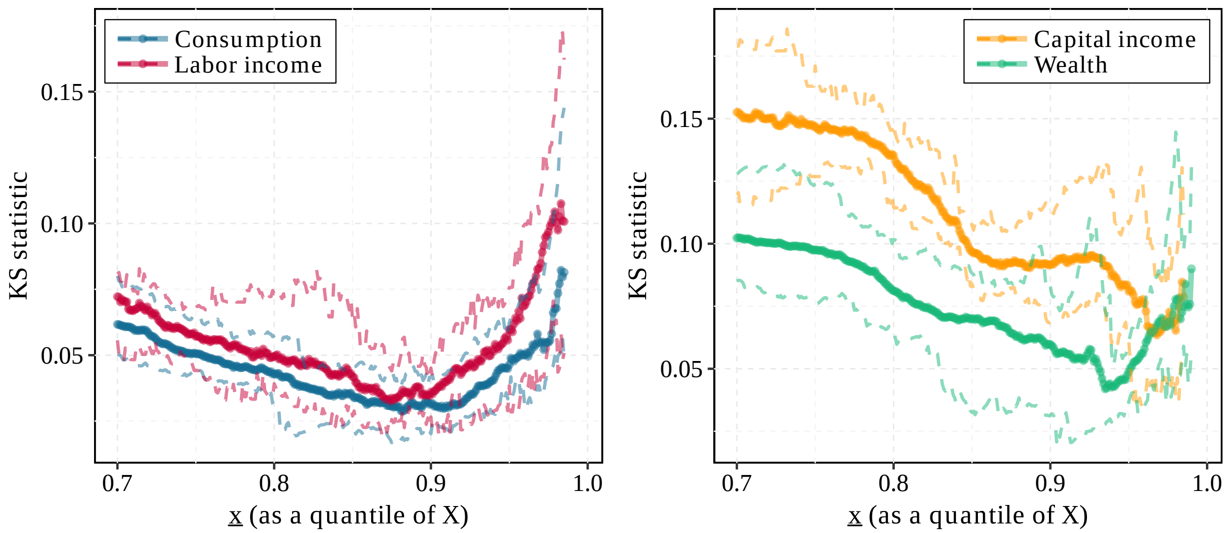
[10]Using nonlinear least squares (NLS) estimation leads to similar results; see Appendix A.3.

[11]Our preferred construction of the variables takes part of the durable goods (vehicle repayments and housing services) as a component of consumption. An alternative definition in which vehicle repayments and leases are not part of consumption expenditure yields a very similar MLE coefficient estimate, namely 3.05.

[12]Philanthropic donations exhibit significant concentration, reflected by a Pareto tail similar to that of wealth at 1.6. When we exclude them from our measurement of consumption, we obtain a slight increase in the average consumption Pareto tail estimate, which rises to 3.08. This minor change is due to the relatively small share of donations in total consumption, even at the upper end of the consumption distribution.

indicates the start of the Pareto tail. This finding implies that the PSID sufficiently covers rich households to observe and accurately measure the upper tails of these distributions.[13]

Figure 5. KS statistics for each of the four variables.



Note: The solid line refers to the average across PSID waves. The dashed lines refer to the minimum and maximum KS statistics for each variable across PSID waves.

2.4.3 Formal Tests of the Pareto Hypothesis. Table 6 in Appendix A reports the results of the statistical tests described in Section 2.1. It shows that for most of the variables—except perhaps labor income—the Pareto distribution accurately captures the data in the upper tail. Based on the MLE estimates, the power law hypothesis is rejected for wealth in two out of nine PSID waves, for labor income in five out of nine waves, for capital income in three out of nine waves, and for consumption in only one wave. Moreover, the power law outperforms the log-normal and exponential distributions for all variables and for most years. We cannot reject that the log-normal distribution performs at least as well as the Pareto distribution for three waves of only the consumption distribution.

2.5 Robustness and Alternative Assumptions

2.5.1 Evidence from the SCF and the CEX. For robustness, we apply the same empirical methodology to the SCF and the CEX. The SCF gives average Pareto tail coefficients

[13] This result is not spuriously driven by the small size of our data set. Drawing simulated data sets of much larger size gives exactly the same level and percentile for the bottom of the Pareto tail (estimated via the minimum of the KS-statistic) as in our smaller sample. The only difference is that with more observations, the estimated Pareto coefficients $\hat{\zeta}_X(\underline{x})$ become less noisy for values of $\underline{x}$ far above $\hat{\underline{x}}$, so that the coefficient estimates remain stable over a larger range of quantiles at the very top.

equal to 2.05 for labor income, 1.4 for wealth, and 1.2 for capital income. The SCF reports food consumption (at home and delivery) since 2004. The coefficient estimates (4.28 and 4.08 using OLS and MLE, respectively) are similar to those obtained in the PSID (4.40 and 4.26). Using the quarterly CEX data compiled by Meyer and Sullivan (2023) from 2001:I to 2017:IV, we find that consumption, total income, and wealth have Pareto tails with average MLE coefficients 3.4, 2.5, 1.6 respectively (this dataset does not distinguish between labor and capital income).[14] The PSID may thus slightly underestimate the thickness of the labor income tail, although to a lesser extent than the CEX.[15] Overall, however, these results from alternative data sets confirm the robustness of our coefficient estimates and the ranking of the four tails. This is because the Pareto tails emerge around or below the 90th percentile in all of these different surveys, i.e., far below the levels at which the PSID ceases to be representative. Table 1 summarizes the results obtained from all of the data sets.

2.5.2 Evidence from Tax Records. Administrative records from the Internal Revenue Service (IRS) provide valuable insights into the upper tails of labor income, wealth, and capital income, in the United States. Applying our method to the data compiled by Saez and Zucman (2016), we estimate via MLE a capital income tail coefficient of 1.35 (excluding capital gains) or 1.28 (including capital gains), both for the optimal KS-criterion and for the top 0.1% of the sample. The wealth tail, imputed from several capitalization methods, has a coefficient equal to 1.50 using the KS criterion, and 1.43 within the top 0.1%. The labor income tail coefficient is equal to 2.08 using the KS criterion, and 1.83 within the top 0.1%.[16] These estimates are naturally less noisy than in the PSID. They are shown graphically in Figures 14 and 17 in the Appendix.

2.5.3 Complementing the PSID with the Forbes Rich List. An alternative way of accounting for the under-reporting of wealth is to adopt the approach of Vermeulen (2016)

[14] Using unweighted CEX consumption data spanning the 1979-2004 period from Kocherlakota and Pistaferri (2009), Toda and Walsh (2015) estimate a Pareto tail coefficient of 3.38. Similarly, converting the top consumption shares $S(p), S(q)$ at various quantiles p, q estimated by Garner et al. (2022) for the years 2017-2019 into Pareto tail coefficients $\hat{\zeta}(p,q) = (1-\ln(S(p)/S(q))/\ln(p/q))^{-1}$, we obtain a stable implied coefficient of about 3.4 across years and quantile pairs above the 80^{th} percentile—thus supporting the existence of a power law at the top.

[15] Using the harmonized LWS data, de Vries and Toda (2022) present estimates of the labor income Pareto tail comparable to those we obtained in the PSID, ranging between 2.4-2.5, and estimates of the capital income Pareto tail ranging between 1.4-1.5, respectively. Updated series from Piketty and Saez (2003) indicate that the Pareto tail coefficient of individual wage income falls within the range of 2.0 to 2.1.

[16] The OLS estimates are 1.35 for the capital income tail excluding capital gains, 1.28 for capital income including capital gains, 1.51 for wealth using the KS criterion (or 1.40 when focusing on the top 0.1%), and 1.91 for wealth using the KS criterion (or 1.82 in the top 0.1%).

Table 1. Top consumption, income, and wealth inequality across surveys (raw data).[a]

Data	Variable	Best fit Pareto[b] $\hat{\underline{x}}^{OLS}$	$\hat{\zeta}^{OLS}$	$\hat{\underline{x}}^{MLE}$	$\hat{\zeta}^{MLE}$	Gini	$\sigma^2\vert_{\ln x}$	90/50 Ratio
PSID	Capital income	0.96	**1.22**	0.96	**1.21**	0.93	1.58	3073.2
		(0.02)	(0.15)	(0.02)	(0.14)	(0.01)	(0.20)	(2516.7)
	Wealth	0.93	**1.48**	0.92	**1.47**	0.80	0.55	11.42
		(0.03)	(0.09)	(0.03)	(0.09)	(0.02)	(0.04)	(2.27)
	Labor income	0.88	**2.42**	0.89	**2.50**	0.56	0.18	3.22
		(0.04)	(0.15)	(0.04)	(0.13)	(0.02)	(0.01)	(0.21)
	Consumption	0.89	**3.11**	0.90	**3.13**	0.36	0.11	2.21
		(0.04)	(0.28)	(0.04)	(0.20)	(0.01)	(0.01)	(0.06)
	Food consumption	0.93	**4.40**	0.93	**4.26**	0.38	0.06	2.18
		(0.05)	(0.33)	(0.05)	(0.43)	(0.01)	(0.01)	(0.17)
SCF	Capital income	0.96	**1.27**	0.95	**1.10**	0.95	1.67	∞
		(0.07)	(0.08)	(0.08)	(0.02)	(0.01)	(0.24)	–
	Wealth	0.92	**1.43**	0.92	**1.23**	0.82	0.76	9.44
		(0.03)	(0.10)	(0.03)	(0.13)	(0.03)	(0.10)	(2.45)
	Labor income	0.86	**2.01**	0.84	**2.08**	0.52	0.25	2.92
		(0.07)	(0.15)	(0.05)	(0.14)	(0.02)	(0.04)	(0.15)
	Food consumption	0.91	**4.28**	0.91	**4.08**	0.34	0.07	2.07
		(0.05)	(0.40)	(0.05)	(0.46)	(0.01)	(0.01)	(0.08)
CEX	Wealth	0.92	**1.69**	0.91	**1.62**	0.79	0.45	15.82
		(0.03)	(0.12)	(0.03)	(0.16)	(0.02)	(0.05)	(4.84)
	Total income	0.83	**2.66**	0.82	**2.50**	0.45	0.13	2.69
		(0.04)	(0.35)	(0.03)	(0.32)	(0.04)	(0.02)	(0.20)
	Consumption	0.88	**3.43**	0.88	**3.38**	0.31	0.09	2.02
		(0.05)	(0.25)	(0.05)	(0.26)	(0.01)	(0.01)	(0.04)
	Food consumption	0.88	**3.90**	0.87	**3.89**	0.31	0.07	1.91
		(0.05)	(0.44)	(0.05)	(0.39)	(0.01)	(0.01)	(0.05)

[a] The estimates represent the average value across all waves. Their standard deviation is reported within parentheses. The variance of $\ln x_i$ is calculated for x_i greater than the 80^{th} percentile.
[b] The threshold values $\hat{\underline{x}}$ are expressed in terms of quantiles of the underlying distribution, e.g., 0.84 refers to the value of x corresponding to the 84^{th} percentile of the respective distribution.

and supplement survey with data from the Forbes 500 rich list. Although this data is not necessarily consistent with our construction of wealth at the household level, it provides us with an additional source of information for the very top of the distribution. Using the raw PSID data, we find that this adjustment does not affect the mean Pareto tail coefficient for wealth estimated via MLE, while it raises the OLS estimate from 1.38 to 1.50. We provide the corresponding graphs in Figure 13 in the Appendix.

2.5.4 Alternative Measures of Inequality. The Pareto coefficients are a powerful and theoretically appealing metric of top inequality. Nevertheless, to confirm the strict ordering documented in Section 2.4, we report the Gini coefficient, the variance of logarithmic transformations of the variables above the 80^{th} percentile, and the ratio of the 90^{th} to the 50^{th}

percentiles of the distributions. Our results are collected in the last three columns of Table 1. All of these measures of inequality agree on the same strict ordering of the distributions. Capital income systematically appears to be the most unequally distributed variable, with a variance in the upper tail equal to 1.6 in the PSID and the SCF, followed by wealth (0.45-0.75), labor income (0.18-0.25) and consumption (0.1, unavailable in the SCF). The CEX seems to systematically underestimate inequality, with a variance of 0.45 for wealth, 0.13 for total income, and 0.09 for consumption.

2.5.5 Individualizing Variables. We can scale our data by family size to construct measures of individualized consumption, income, and wealth rather than household-level measures. Specifically, we apply an equivalence scale recommended by the National Academy of Sciences (NAS, see Citro et al. (1995)) to our measures, using the formula $(N_a + 0.7N_k)^{0.7}$, where N_a represents the number of adults in the family, and N_k represents the number of children. Applying this scaling does not affect the previous ranking and produces very similar estimates: The median MLE coefficients $(\zeta_c, \zeta_y, \zeta_a, \zeta_{ra})$ are $(3.05, 2.25, 1.34, 1.22)$ for individualized consumption, versus $(3.02, 2.30, 1.35, 1.12)$ in the unscaled data.

2.5.6 Age and Life-Cycle Dynamics. The ordering of Pareto tails that we documented may potentially be driven by the tight link between age, on the one hand, and the dynamics of consumption, income, and wealth, on the other hand. For instance, older households might accumulate large stocks of wealth, without any apparent reason to spend it, for bequest motives or as a buffer against medical hardships. To assess the relevance of these mechanisms, we remove the effects of age from the consumption, income, and wealth variables by regressing them on a fifth-order age polynomial and extracting the residuals. The tails of these residuals (re-centered around the mean of the corresponding variables) have coefficients equal to, in the above-mentioned order: 3.02, 2.20, 1.28, 1.12. Thus, we find that the part of consumption that is not explained by age has the same Pareto tail coefficient. That is, age does not appear to be a critical factor driving our results.

To further validate this observation, Table 8 in the Appendix presents the resulting tail estimates for two age groups: 19-45 and 46-80. We observe that for both age groups, the ranking and magnitude of the Pareto tail coefficients are preserved. Interestingly, we find that the Pareto tail coefficients for wealth and capital income tend to increase along the life cycle, while decreasing for consumption and labor income. Finally, we confirm that excluding health expenditures, known to substantially increase throughout the life-cycle (see De Nardi et al. (2017)), keeps the median consumption tail consistent at 3.01.

2.5.7 Time Series Trends. Appendix A.5 details the evolution of the estimated Pareto tail coefficient for wealth and consumption over time. Using the CEX data, we observe a decline in consumption tail estimates from 1970 to 1995, falling from above 4 to approximately 3.5-3.6. Since the 2000s, for the PSID, it has remained stable around 3-3.2, and for the CEX, between 3.2-3.5. In terms of wealth, the data show a decrease since 1980, dropping from 1.55 in 1980 to between 1.25-1.4 in 2020.

2.6 Pareto Tails of Other Variables

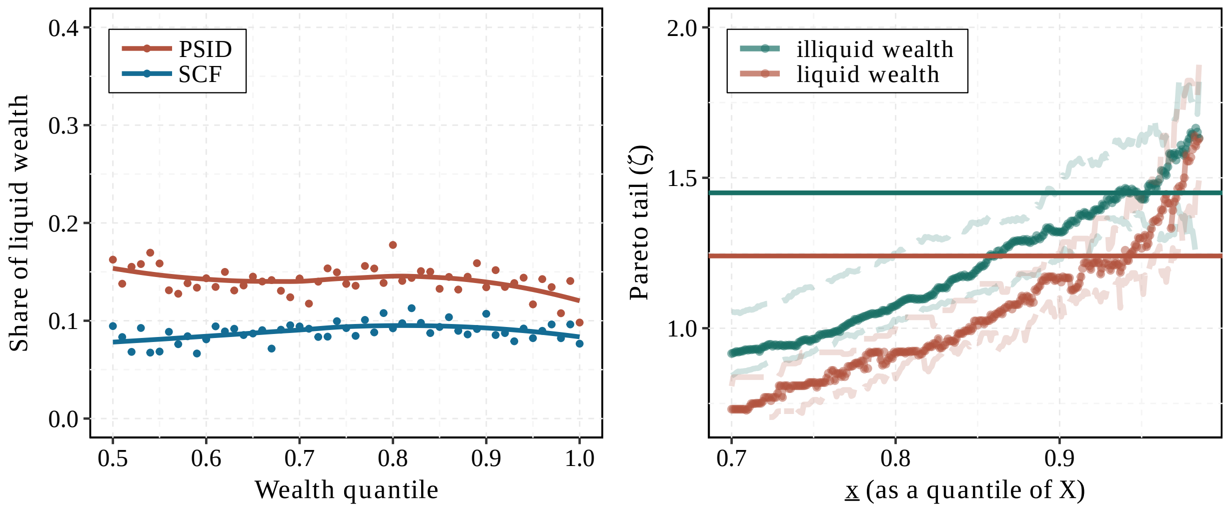
In this last section of our empirical analysis, we explore the Pareto tails of several other variables that will help us to further discipline the theoretical and quantitative analysis.

2.6.1 After-Tax Labor Income and Total Income. We first evaluate the role of the redistributive tax system in reducing top inequality. Using TAXSIM35 from the NBER, we construct measures of after-tax labor income and total income and compute the corresponding Pareto coefficients. The Pareto coefficient estimate for labor income is 2.52 (2.61) using OLS (MLE). The estimate for total income is 2.04 (2.10). While these estimates show that the tax system substantially reduces inequality, the after-tax labor income Pareto coefficient is still far below that of consumption.

2.6.2 Liquid and Illiquid Wealth. A potential explanation of the theoretically puzzling fact that consumption is strictly more evenly distributed than wealth might be that the distribution of consumption reflects that of *liquid* wealth only. This hypothesis would require liquid wealth to have a thinner tail than illiquid and, therefore, total wealth. This would imply in turn that the share of liquid (resp., illiquid) assets in total wealth should converge to zero (resp., one) for the richest households.

To test this hypothesis, the left panel of Figure 6 plots the share of liquid wealth by wealth percentile in the PSID and the SCF. The liquid wealth share converges to a strictly positive constant—about 10 percent—for the richest households. Moreover, this share is remarkably flat between the 50th and the 99th percentiles. Our baseline measure of liquid wealth excludes stocks. Including stocks into our measure of liquid wealth only reinforces this conclusion. These findings show that, on average, the wealthiest households keep a strictly positive and constant fraction of their wealth in liquid assets. The Pareto tail coefficients of liquid and illiquid wealth should therefore coincide. This is indeed confirmed by the right panel of Figure 6. If anything, liquid wealth has a slightly lower estimated coefficient relative to that of illiquid wealth, and both are approximately equal to that of total wealth—hence much lower than that of consumption. As a result, a model in which consumption follows the behavior of liquid wealth at the top of the distribution would be inconsistent with the data.

Figure 6. Pareto tail estimation for liquid and illiquid wealth.



Note: On the left panel, the solid lines are a smooth approximation of the underlying observations. The right panel performs the same exercises as in the middle panel of Figure 3 using the PSID. The shaded dashed lines indicate the estimates for the sample year that leads to the lowest and highest coefficients.

2.6.3 Financial vs. Real Wealth and Other Subcategories of Consumption. We can distinguish real wealth (illiquid wealth net of private equity, i.e., mostly real estate and durables) from financial wealth (liquid wealth plus private equity), using the Fed Board's Balance Sheet of Households table to define these categories. The former may provide value through flow utility, while the latter should only provide value through the budget constraint. Note that the elements that compose real wealth are already included in our consumption variable.[17] We report the Pareto tail coefficients of these two variables in Figure 15 in Appendix A. We obtain a coefficient of 1.9 for real assets, and a significantly smaller coefficient (a strictly thicker tail) of 1.3 for financial assets.

We can finally compute the Pareto coefficients of the most concentrated subcategories of consumption. We obtain 2.65 for healthcare, 2.65 for recreation, 2.3 for housing services (imputed rents). These tails are significantly thicker than that of food (with a coefficient

[17] In particular, recall from Figure 2 that imputed rents account for about 20% of household consumption at the top. Note also that the PSID has a "housing consumption imputation" variable, which has a Pareto coefficient of 2.8. Thus, if anything, our imputation is a lower bound to the Pareto tail of real wealth.

of above 4 both in the PSID and the SCF). Nevertheless, they are still far thinner than those of liquid or financial assets. This raises a further puzzle to the literature: Why do rich households hold large amounts of liquid or other financial assets with seemingly no desire to consume or invest in durables?

3 Theoretical Analysis

To replicate the empirical thickness of the wealth distribution, modern heterogeneous-agent incomplete-market models postulate the presence of heterogeneous returns to capital, reflecting for instance persistent idiosyncratic *type-dependent* portfolio investments or entrepreneurship (Krusell and Smith Jr., 1998; Cagetti and De Nardi, 2006) or *ex-post* investment luck (Angeletos and Calvet, 2006; Angeletos, 2007). The analytical results that we derive in this section, however, show that these models are unable to jointly generate the level and ranking of the consumption and capital income tails observed empirically. We propose instead a model that resolves this puzzle. Specifically, we incorporate into the canonical framework two additional features: *non-homothetic, wealth-dependent preferences* and *scale-dependent returns* to capital. The former decouples consumption from wealth at the tail, while the latter decouples capital income from wealth.

3.1 Analytical Model Environment

Time is continuous and indexed by t. There is a continuum of individuals indexed by i. Throughout this section, we focus on the stationary equilibrium of the economy, in which prices are time-invariant.

Preferences. At time t, individual i owns a stock of wealth a_{it}, has labor productivity e_{it} and investment productivity z_{it}, and consumes c_{it}. Her flow of utility is given by $u\left(c_{it}\right) + \kappa U\left(a_{it}\right)$, where $\kappa \geq 0$ and u, U are both strictly increasing and concave. Throughout our analysis, we assume the following parametric functional forms:

$$u\left(c\right) = \frac{1}{1-\gamma} c^{1-\gamma}, \qquad U\left(a\right) = \frac{1}{1-\nu} a^{1-\nu}, \tag{3}$$

with $\gamma > 1$ and $\nu \leq \gamma$. Preferences are *homothetic* whenever $\kappa = 0$ or $\nu = \gamma$; they are *non-homothetic* whenever $\kappa > 0$ and $\nu < \gamma$.

When $\kappa > 0$, wealth enters explicitly individual preferences (3). This reduced-form specification can be microfounded via, e.g., a warm glow for bequests: If agents die with some positive probability at every instant and value the amount of wealth that they pass on to their children, they receive a flow of utility from their stock of assets at each time t.[18] The critical element is that the utility of wealth $U(\cdot)$ can be strictly less concave than the utility of consumption $u(\cdot)$. Intuitively, if households have a strong enough desire to accumulate capital for its own sake rather than for future consumption, wealth can be strictly more concentrated than consumption at the top.

Labor Productivity Dynamics. The labor income of agent i at time t is $w \cdot y_{it}$, where w denotes an aggregate time-invariant wage rate. Labor productivity y_{it} is subject to idiosyncratic shocks. Specifically, we suppose that $y_{it} > 0$ evolves exogenously according to a geometric Brownian motion:

$$dy_{it} = \mu_y \, y_{it} \, dt + \sigma_y \, y_{it} \, dB_{it}, \tag{4}$$

with mean growth rate $\mu_y > 0$ and volatility $\sigma_y^2 > 0$. The assumption of a random growth process for labor productivity is the most natural way of generating a Pareto tail for the cross-sectional distribution of labor income.

Investment Productivity Dynamics. The expected return on capital of agent i at time t is given by $\mu_r R(z_{it}, a_{it})$, where the function R is specified below. Let $\mu_r \equiv (1 - \tau_K) r$, where τ_K is a constant capital income tax rate and r is an aggregate time-invariant interest rate.

Investment productivity z_{it} is subject to persistent idiosyncratic shocks. Specifically, we suppose that z_{it} evolves exogenously according to a continuous-time Markov chain with two states $z_H > z_L > 0$ and transition rates:[19]

$$q_{LH} \cdot R(z_L, a_{it}) \quad \text{and} \quad q_{HL} \cdot R(z_H, a_{it}) \ . \tag{5}$$

The parameter $q_{LH} > 0$ gives the instantaneous probability that the investment productivity jumps from the low state z_L to the high state z_H, and conversely for $q_{HL} > 0$.

Notice that the speed of transition from one productivity state to another rises propor-

[18] See, e.g., Atkinson (1971) or De Nardi (2004). Another possible interpretation of these preferences is a status-seeking motive, if we express alternatively the utility as a function of the individual's wealth relative to average wealth in the economy (as in the quantitative model of Section 4), a/A. Saez and Stantcheva (2018); Straub (2019); Mian et al. (2021) provide overviews of the various micro-foundations of wealth in the utility.

[19] In Appendix B, we extend our analysis to the case of general Markov chains for z_{it}.

tionally to $R(z_{it}, a_{it})$. That is, agents with higher expected returns also face more risk. In contrast to the *ex-post* investment risk introduced below, we refer to this source of uncertainty as *ex-ante* investment risk since the agent knows her current productivity type z_{it} at the time she makes her consumption and saving decisions.

Type- and Scale-Dependence. We suppose that $R(z_{it}, a_{it})$ is given by

$$R(z_{it}, a_{it}) = z_{it}\, S(a_{it}), \quad \text{where} \quad S(a_{it}) = 1 + \psi a_{it}^{\eta}, \tag{6}$$

with $\psi \geq 0$ and $\eta \geq 0$. The dependence of $R(z_{it}, a_{it})$ on the agent's idiosyncratic investment productivity z_{it} is called *type dependence*. This property, whereby some agents ("entrepreneurs") have access to higher-return technologies, on average, than others ("workers"), allows the model to replicate the empirical thickness of the Pareto tail of wealth.[20] In addition, $R(z_{it}, a_{it})$ is also directly increasing in the agent's wealth a_{it} whenever $\psi > 0$ and $\eta > 0$. This property is called *scale dependence* and plays a central role in our analysis to generate a strictly thicker tail for capital income than for wealth.[21]

Budget Constraint. The budget constraint of agent i at time t reads:

$$da_{it} = [wy_{it} - T(wy_{it}) - c_{it}]\, dt + a_{it} dr_{it}, \tag{7}$$

where $T(\cdot)$ is a time-invariant labor income tax schedule. The realized returns to capital evolve according to the following process:

$$dr_{it} = \mu_r \cdot R(z_{it}, a_{it})dt + \sigma_r \sqrt{R(z_{it}, a_{it})} dW_{it}, \tag{8}$$

where $\sigma_r \geq 0$ and W_{it} is a standard Brownian motion. The borrowing limit is $a_{it} \geq 0$.

If the volatility parameter σ_r is strictly positive, agents experience *ex-post* investment risk around their mean return $\mu_r R(z_{it}, a_{it})$.[22] Notice that the volatility of the return process (8) is proportional to the mean return; that is, agents with higher expected returns also face more capital income risk. Although not strictly necessary for our purposes, the assumption

[20]See, e.g., Quadrini (2000) and Cagetti and De Nardi (2006).

[21]See Fagereng et al. (2020), Bach et al. (2020) or Balloch and Richers (2021) for empirical evidence on the relative importance of type- and scale-dependence.

[22]This form of investment risk is introduced in, e.g., Angeletos and Calvet (2006); Angeletos (2007). By allowing for both *ex-ante* and *ex-post* capital income risk, our model combines features from this incomplete market growth literature and from that on entrepreneurship.

that both *ex-ante* and *ex-post* capital income risk scale up proportionally to the mean return ensures that scale dependence does not dwarf type dependence or capital income volatility for the wealthiest agents—thus allowing all three channels to jointly shape the Pareto distributions of wealth and capital income.

Death and Expropriation. Agents have finite lives. We suppose that each agent dies at a constant Poisson rate $\xi_1 > 0$. Upon death, she is replaced by a newborn ("child") with initial labor productivity $y_0 > 0$. The child inherits the wealth left by her deceased parent.

In addition, we introduce expropriation (or dissipation) shocks. Specifically, upon receiving such a shock, the wealth of an agent dissipates infinitely fast—although she survives and her labor income does not jump.[23] These shocks are independent of the death shocks and occur at a constant Poisson rate that is proportional to the agent's expected return, i.e., $\xi_2 \cdot R(z_{it}, a_{it})$ with $\xi_2 > 0$. Thus, the wealthier the agent and the more productive her investments, the more likely she is to be expropriated. The death and expropriation shocks ensure that the cross-sectional distributions of both labor income and wealth converge to stationary distributions.

Government. We suppose that the labor income tax $T(\cdot)$ has a parametric functional form with a constant rate of progressivity (CRP):

$$T(wy) = wy - \frac{1-\tau_0}{1-\tau_L}(wy)^{1-\tau_L}, \tag{9}$$

with $\tau_0 \in \mathbb{R}$ and $\tau_L \in (-\infty, 1)$. The rate of progressivity τ_L is equal to (minus) the elasticity of the retention rate $1 - T'(y)$ with respect to labor income y. If $\tau_L > 0$ (respectively, $\tau_L = 0$, $\tau_L < 0$), the tax schedule is progressive (resp., linear, regressive).[24]

Value Function and Stationary Distribution. Individuals maximize the present discounted value of their utility, subject to their budget constraint and the borrowing limit at each instant t. They discount the future at rate ξ_0. The death and expropriation shocks both amount to raising the agent's effective discount rate to $\xi(z,a) \equiv \xi_0 + \xi_1 + \xi_2 R(z,a)$. We denote by $V(y, z_n, a)$ the value function of an agent with labor productivity $y \in \mathbb{R}_+$, investment productivity $z_n \in \{z_L, z_H\}$, and wealth $a \in \mathbb{R}_+$. This value function satisfies the Hamilton-Jacobi-Bellman equation:

[23] See, e.g., Appendix C in Moll et al. (2022) for various micro-foundations of these shocks.

[24] The CRP tax code is a good representation of the U.S. income tax system (see, e.g., Heathcote et al. (2017)) and allows us to rationalize the ratio of after-tax to pre-tax labor incomes in the data.

$$\xi(z_n, a)V(y, z_n, a) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + \kappa\frac{a^{1-\nu}}{1-\nu} + \mu_y\, y\, \frac{\partial V}{\partial y}(y, z_n, a) + \frac{1}{2}\sigma_y^2\, y^2\, \frac{\partial^2 V}{\partial y^2}(y, z_n, a)$$
$$+ \left[wy - T(wy) + \mu_r R(z_n, a)\, a - c\right] \frac{\partial V}{\partial a}(y, z_n, a) + \frac{1}{2}\sigma_r^2\, R(z_n, a)a^2 \frac{\partial^2 V}{\partial a^2}(y, z_n, a) \quad (10)$$
$$+ q_{nm} R(z_n, a)\left[V(y, z_m, a) - V(y, z_n, a)\right],$$

for any y, z_n, a, where $m \neq n$. Let $c(y, z_n, a)$ denote the corresponding policy function. The resulting stationary joint distribution of labor productivity, investment productivity, and wealth, $f(y, z_n, a)$, satisfies a Kolmogorov-forward equation given in Appendix B.

3.2 The Case of Wealth-Independent Preferences

We start by characterizing the consumption and capital income of the wealthiest agents in the canonical model with wealth-independent, homothetic preferences.

Proposition 1 (**Wealth-Independent Preferences**). *Suppose that* $\kappa = 0$. *Let* $\xi_2 > (1-\chi)[\mu_r - \chi\sigma_r^2/2]$ *with* $\chi \equiv (1+\eta)\gamma$. *The consumption policy function satisfies* $c(y, z_n, a) \sim C_n^{-1/\gamma} \cdot a^{1+\eta}$ *for some constant* $C_n > 0$. *The distributions of wealth, consumption, and capital income have asymptotic Pareto tails with coefficients* $\zeta_c = \zeta_{ra} = \zeta_a/(1+\eta)$.

Consider first the case of scale-independent returns, i.e., $\eta = 0$. Proposition 1 shows that, under homothetic preferences, consumption is then asymptotically linear in wealth for any value of the labor and investment productivities.[25] Moreover, mean returns $\mu_r R(z_n, a)$ are independent of wealth, so that capital income is also asymptotically linear in wealth. This implies in turn that all three variables must be equally concentrated at the top: $\zeta_c = \zeta_{ra} = \zeta_a$.

Consider next the case of scale-dependent returns, $\eta > 0$. Capital income is then strictly convex in wealth by equation (6). This leads to a thicker tail of capital income than wealth in the stationary distribution: $\zeta_{ra} < \zeta_a$. However, consumption is then also strictly *convex* in wealth, with a cross-sectional elasticity $1 + \eta$. Intuitively, the higher returns earned by wealthier agents allow them to consume larger fractions of their wealth. That is, consumption inherits the concentration of the capital income tail, $\zeta_c = \zeta_{ra}$, and is therefore strictly more unevenly distributed than wealth. This is a manifestation of the fact that consumption is driven by *permanent income*, not wealth, which for the richest households corresponds to capital income. Thus, matching the concentration of the capital income tail does not help to resolve the consumption concentration puzzle—it makes it even worse.

[25] This linearity property is well-known and plays a key role in macroeconomics; see, e.g., Krusell and Smith Jr. (1998) or Achdou et al. (2022).

Note that the condition $\xi_2 > (1-\chi)[\mu_r - \chi\sigma_r^2/2]$ under which we stated Proposition 1 ensures that the constants C_L, C_H are positive. This restriction can be violated in the presence of *ex-post* capital income risk ($\sigma_r > 0$) if, for a given ξ_2, the risk-aversion coefficient γ is sufficiently large. Under scale-independent returns, Ma and Toda (2021) show that the asymptotic MPC is equal to zero in this case, so that $C_L = C_H = 0$. As a result, consumption might be strictly concave in, and therefore less concentrated than, wealth. In Section 4, however, we evaluate this argument quantitatively and show that matching the ratio of consumption and wealth tail parameters that we observe in the data would require an implausibly large risk-aversion coefficient under homothetic preferences.

3.3 General Preferences: Individual Behavior and Pareto Tails

We now study the general case of wealth-dependent preferences. We characterize analytically the individual policy functions in Proposition 2, and the tails of the stationary distributions of consumption, labor income, wealth, and capital income in Proposition 3.

Proposition 2 (**Asymptotic Policy Function**). *Suppose that $\kappa > 0$ and $\nu \notin (1-\eta, 1)$. Let $\xi_2 > (1-\chi)[\mu_r - \chi\sigma_r^2/2]$, with $\chi \equiv \nu + \eta$. The consumption policy function and the mean capital income satisfy*

$$c\,(y, z_n, a) \underset{a\to\infty}{\sim} C_n^{-1/\gamma} \cdot a^{(\nu+\eta)/\gamma} \qquad \textit{and} \qquad \mu_r R(z_n, a)a \underset{a\to\infty}{\sim} \mu_r\,\psi\,z_n\,a^{1+\eta} \tag{11}$$

where $C_L > C_H > 0$ are positive constants. Therefore, consumption is asymptotically linear in wealth iff $\nu + \eta = \gamma$, convex iff $\nu + \eta > \gamma$, and concave iff $\nu + \eta < \gamma$. Capital income is asymptotically convex in wealth iff $\eta > 0$.

To understand Proposition 2, consider first the case where preferences are homothetic, that is, $\nu = \gamma$. In this case, equation (11) implies that consumption is asymptotically either linear (if $\eta = 0$) or convex (if $\eta > 0$) in wealth. This result is similar to that of Proposition 1, except that the constants C_L, C_H are slightly different and the elasticity of consumption to wealth under scale-dependence is now $1 + \eta/\gamma$ rather than $1 + \eta$.

Suppose now that preferences are non-homothetic, $\nu < \gamma$. Equation (11) then shows that consumption can be asymptotically linear, convex, or concave in wealth, depending on the relative strength of the non-homotheticity and the scale-dependence elasticities. If $\eta = 0$, consumption is strictly *concave*, with cross-sectional elasticity $\nu/\gamma < 1$. Intuitively, the individual accumulates assets up to the point where the marginal utilities of consumption and wealth are of the same order, so that $c^{-\gamma} \propto a^{-\nu}$. The smaller the risk-aversion ν relative to γ, the less the individual consumes and the more she saves out of wealth. We show below that this concavity property is critical to yield a strictly thinner tail of consumption than wealth. However, absent scale-dependence in returns, capital income remains asymptotically linear in wealth. Letting $\eta > 0$ delivers both strictly concave consumption behavior *and* strictly convex capital income whenever $\nu + \eta < \gamma$.[26]

Our next proposition establishes that the convexity (respectively, concavity) of capital income (resp., consumption) in wealth translates into a strictly higher (lower) upper tail inequality than for wealth. Importantly, while intuitive, this result is not immediate. If there was no type-dependence, so that $z_L = z_H$, there would be a one-to-one map between capital income and wealth, and the ratio of their Pareto tails would be directly inherited from the linearity or the convexity of this relationship (equation (11)). However, under type-dependence, the share of high-return agents in a given wealth percentile increases with wealth. Thus, even with $\eta = 0$, capital income would grow faster than linearly along the wealth distribution due to this *composition effect*, and its distribution would be strictly more concentrated than that of wealth. Analogously, the constants C_L and C_H in equation (11) satisfy $C_L > C_H$, so that high-return types have a higher marginal propensity to consume than low-return types. Since the fraction of high types grows along the wealth distribution, again a composition effect might alter the resulting concavity of consumption along the wealth distribution. Nevertheless, Proposition 3 shows that, as wealth becomes arbitrarily large, these composition effects become negligible, because the share of high productivity-type agents converges to a constant that is independent of wealth.[27] Hence, the type-dependence of returns alone is unable to decouple the tails of capital income and consumption from that of wealth at the top. Any discrepancy between their Pareto coefficients must arise from the scale-dependence and the non-homotheticity channels that map into the power function in equation (11).[28]

[26]Note that the condition $\xi_2 > (1-\chi)[\mu_r - \chi\sigma_r^2/2]$ generalizes that of Proposition 1 to non-homothetic, wealth-dependent preferences and ensures that $C_L, C_H > 0$.

[27]Formally, the proof of Proposition 3 shows that the density function $f(y, z_n, a)$ becomes multiplicatively separable as $a \to \infty$. This property no longer holds at finite levels of wealth, and hence these composition effects play a key role in our quantitative analysis of Section 4.

[28]In our model, because agents switch stochastically from one investment productivity to another, the two conditional wealth distributions—for low-type and high-type agents—have the same tail thickness in the stationary equilibrium. We could instead model the returns process with an absorbing type, to allow the two conditional distributions to have different tail parameters. But then the distributions of capital income, consumption, and wealth would be driven by the behavior of the highest-type agents—so that at the top, the model would again reduce to one with a single investment type. Either way, we need scale-dependence to generate a discrepancy between the Pareto tails of capital income and wealth.

Proposition 3 **(Asymptotic Pareto Coefficients).** *Suppose that the assumptions of Proposition 2 are satisfied, and let* $\xi_2 > \mu_r$.

(a) Pre-tax and after-tax labor income are asymptotically Pareto distributed with respective coefficients ζ_y *and* $\zeta_y^{net} = (1 - \tau_L)\,\zeta_y$, *where*

$$\zeta_y = -\left(\frac{\mu_y}{\sigma_y^2} - \frac{1}{2}\right) + \sqrt{\left(\frac{\mu_y}{\sigma_y^2} - \frac{1}{2}\right)^2 + \frac{2\xi_1}{\sigma_y^2}}.$$

(b) Wealth is asymptotically Pareto distributed with coefficient $\zeta_a = \eta + \xi_2/\mu_r$ *if* $\sigma_r = 0$ *and, if* $\sigma_r > 0$,

$$\zeta_a = \eta - \left(\frac{\mu_r}{\sigma_r^2} - \frac{1}{2}\right) + \sqrt{\left(\frac{\mu_r}{\sigma_r^2} - \frac{1}{2}\right)^2 + \frac{2\xi_2}{\sigma_r^2}}.$$

(c) Consumption and capital income are asymptotically Pareto distributed with respective Pareto coefficients ζ_c *and* ζ_{ra} *given by:*

$$\zeta_c = \frac{\gamma}{\nu + \eta}\zeta_a \quad \text{and} \quad \zeta_{ra} = \frac{1}{1+\eta}\zeta_a. \tag{12}$$

Proposition 3 derives in closed form the Pareto coefficients of the four tails in the model with non-homothetic preferences and scale-dependent returns. Note first that the Pareto tails of before- and after-tax labor income depend on variables that are independent of those driving the other three tails: the drift and volatility of labor productivity, the progressivity of the labor income tax schedule, and the death rate. Intuitively, these parameters do not affect the other tails as labor income becomes irrelevant for the wealthiest individuals. Analogously, the Pareto coefficient of wealth depends on the volatility of the *ex-post* investment risk, the capital income tax rate, the expropriation rate, and the degree of scale-dependence. In particular, ζ_a increases—the wealth tail becomes thinner—when the capital tax rate τ_K increases or the volatility σ_r^2 falls.[29]

Equation (12) is the main result of this section. It shows that (i) the ratio of Pareto coefficients of consumption and wealth inherits the concavity of the consumption policy

[29] Interestingly, *ex-ante* entrepreneurial risk, i.e., the transition rates between types (q_{HL}, q_{LH}), leaves the Pareto coefficient unaffected: They only shift the tail of the wealth distribution uniformly. This result is due the particular scaling of the transition rates q_{LH} and q_{HL} by z_L and z_H, respectively. The joint density then behaves asymptotically as $f(y, z_n, a) \sim g(y)F_n a^{-\zeta_a - 1}$ for $a \to \infty$, with $F_n/F_m = (z_m q_{mn})/(z_n q_{nm})$ for $m, n \in \{L, H\}$. Under different modeling assumptions (e.g., uniform scaling), ζ_a would depend directly on the transition rates.

function, and (ii) the ratio of Pareto coefficients of capital income and wealth inherits the convexity of expected capital income. Hence, a positive degree of scale-dependence ($\eta > 0$) allows us to match the ratio of Pareto tail coefficients of wealth and capital income $\zeta_{ra}/\zeta_a < 1$ that we measured empirically in Section 2.4. Furthermore, preferences must be non-homothetic ($\nu < \gamma$) to rationalize the empirical ratio of the Pareto coefficients of consumption and wealth, $\zeta_c/\zeta_a > 1$. The underlying degree of non-homotheticity must be large enough to overcome the impact of scale-dependence, which tends to make consumption convex rather than concave; namely, we must have $\nu = \gamma\zeta_a/\zeta_c - \eta$.

4 Quantitative Analysis

The theoretical analysis of Section 3 focuses on the limits as wealth becomes arbitrarily large. In practice, however, the values of wealth or consumption that we observe in the data are finite. Therefore, the standard model with homothetic preferences and scale-independent returns to capital *might* be able to match the empirical evidence from Section 2.4 at the empirically relevant levels of consumption, wealth, and capital income. In particular, Carroll and Kimball (1996) show that in a general class of heterogeneous agents models, consumption is concave in, and thus strictly less concentrated than, wealth. Moreover, the composition effects discussed in the previous section—caused by the increasing share of high-return type agents along the wealth distribution—could account for some of the observed concentration of capital income or consumption without the need to introduce non-homothetic, wealth-dependent preferences or scale-dependent returns. That is, our finite-sample estimates of Pareto tail coefficients for consumption (resp., capital income) may be upwards (resp., downwards) biased, so that the canonical model could still be consistent with the empirical ordering of Pareto coefficients. In this section, we evaluate these arguments quantitatively and study which class of models is able to match the data.

4.1 Quantitative Model Environment

In our quantitative analysis, we set up a discrete-time model that we use to sample distinct waves of a finite but large number of households. This approach is designed to mirror the PSID data collection process, thereby ensuring that our quantitative and empirical analysis closely align.

Households. In each period t, households consume c and save a', subject to the borrowing limit $a' \geq 0$.[30] They derive utility from consumption and wealth, as explained in Section 3.1, and discount future periods at rate $\beta \in (0,1)$. Additionally, households are subject to a death shock that occurs at a constant rate $\xi \in (0,1)$.

Analogous to Section 3, labor productivity is equal to $y = e \cdot h$, where e (resp., h) is the transitory (resp., persistent) component of the agent's idiosyncratic labor productivity. Capital returns are equal to $r(z,a) = r \cdot z \cdot S(a)$, where r is the interest rate, z is the idiosyncratic investment skill, and $S(a)$ is given by (6). We assume that h and z follow independent Markov processes described by the respective transition probabilities $P(h'|h)$, and $P(z'|z)$, while e is iid. We write $P(y'|y)$ the resulting markov chain for overall labor productivity. Upon death, children inherit their parents' wealth that is subject to an estate tax τ_B, and draw new realizations of h' and z' from their respective invariant distributions.

The household's maximization problem is given by

$$\begin{aligned} V(y,z,a) &= \max_{c,a' \geq \underline{a}} \frac{c^{1-\gamma}}{1-\gamma} + \kappa \frac{(a/A)^{1-\nu}}{1-\nu} + \beta(1-\xi) \sum_{y' \in \mathcal{Y}} \sum_{z' \in \mathcal{Z}} P(y'|y) P(z'|z) V(y',z',a') \\ &\text{s.t.} \quad c + a' = wy - T(wy) + (1-\tau_K) rzS(a)a + a \ , \end{aligned} \tag{13}$$

where the wage rate w and the interest rate r are determined in equilibrium. Note that we now assume that the utility of wealth arises relative to the average wealth A in the economy.[31] To ensure the stationarity of the equilibrium distributions, we assume that the government fully expropriates bequests, i.e., $\tau_B = 1$.[32]

Government. The government balances its budget at each instant:

$$\int_{(y,z,a)} \left[T\left(wy\right) + \tau_K rzS\left(a\right)a + \tau_B \xi a'(y,z,a) \right] dF\left(y,z,a\right) \;\; = \;\; G, \tag{14}$$

where $F(y,z,a)$ is the joint distribution of households over labor productivity, investment productivity, and wealth. In equilibrium, the level of expenditures G adjusts to ensure that (14) holds given the tax rates.

[30] If anything, the no-borrowing assumption makes consumption more concave in wealth and labor income than any other borrowing constraint, amplifying the Carroll and Kimball (1996) effect quantitatively.

[31] This assumption does not affect the results, but makes the model consistent with a balanced growth path.

[32] This assumption plays an analogous role as the expropriation (or dissipation) shocks in the theoretical model of Section 3. Alternatively, we could use random bequests to ensure the stationarity of the wealth distribution, as in De Nardi (2004) and Straub (2019).

Production. There is a representative firm that produces the final good using labor and physical capital as inputs according to a Cobb-Douglas production function:

$$Y = K^{\alpha} L^{1-\alpha}, \tag{15}$$

where K and L are aggregate efficiency units of capital and labor, respectively. The depreciation rate of capital is $\delta \in (0,1)$.

Equilibrium. We now formally define a stationary equilibrium of this economy.

Definition 1 (***Stationary Recursive Equilibrium (SRE)***)**.** *An SRE consists of a value function $V(y,z,a)$, a policy function $c(y,z,a)$, an invariant distribution of households $F(y,z,a)$, prices (w,r) and taxes (T,τ_K,τ_B), such that the following conditions hold:*

1. *The value function $V(y,z,a)$ and policy function $c(y,z,a)$ solve the household maximization problem (13), taking prices (w,r) as given.*

2. *The capital market clears, i.e., $K = A \equiv \int zS(a)a\,dF(y,z,a)$. The labor market clears, i.e., $L = \int y\,dF(y,z,a)$. Equilibrium prices are equal to the marginal products, that is, $r = \alpha K^{\alpha-1}L^{1-\alpha} - \delta$ and $w = (1-\alpha)K^{\alpha}L^{-\alpha}$.*

3. *The government balances its budget, i.e., (14) holds.*

4.2 Calibration

We use the theoretical insights of Section 3.3 to calibrate the model and replicate the tail concentration of consumption, labor income, wealth, and capital income. First, we fix a set of external parameters. Second, we estimate the remaining parameters internally by matching key targeted model-implied moments with their empirical counterparts using the method of simulated moments (MSM).

4.2.1 Externally Fixed Parameters. The model period is a year. We set $\beta = 0.90$ and $\xi = 0.02$. The inverse intertemporal elasticity of substitution is set to $\gamma = 2$, which is standard in the literature. Moreover, we set the capital income share of the Cobb-Douglas technology to $\alpha = 0.33$ and the depreciation rate to $\delta = 8\%$. Consistent with the current U.S. tax schedule, we impose a labor income tax progressivity of $\tau_L = 0.14$; this value matches almost exactly the gap between the after-tax and pre-tax labor income Pareto tails (see Proposition 3 (a)). The parameter τ_0 targets an average labor income tax rate in the economy of 30%. The linear capital income tax is set to $\tau_K = 0.15$.

We follow Hubmer et al. (2021) and assume that the labor income process is governed by a mixture of a log-normal and a Pareto distribution. Specifically, if the persistent labor productivity component h_t falls in the bottom q_h percentiles of its distribution, it satisfies

$$\ln h_{t+1} = \rho_h \ln h_t + \varepsilon_{ht} \ , \quad \text{with} \quad \varepsilon_{ht} \sim \mathcal{N}(0, \sigma_h^2) \ .$$

If instead h_t falls in the top $1 - q_h$ percentiles, it is distributed according to a power law, with a slope ζ_h set to match the labor income Pareto tail and a lower bound $\underline{x}_h$ set to ensure continuity at the ${q_h}^{\text{th}}$ percentile of the distribution. Following Storesletten et al. (2004), the persistent labor productivity process is parameterized by $\rho_h = 0.95$ and $\sigma_h^2 = 0.026$. We set the threshold of the Pareto distribution to the 90th percentile (i.e., $q_h = 0.9$), and the Pareto coefficient to $\zeta_h = 2.25$, following our estimates in Section 2.4. Finally, the transitory labor productivity process is parameterized with a variance of $\sigma_e^2 = 0.04$.

We model return heterogeneity following the worker–entrepreneur dichotomy as in Quadrini (2000) and Cagetti and De Nardi (2006). To do so, we assume that agents can have two types, $z \in \{z_L, z_H\}$. The low productivity type is associated with workers with no specific investment skill; thus, we let $z_L = 1$. By contrast, the high productivity type is associated with entrepreneurs; we calibrate z_H internally. Moreover, we calibrate the probability to switch from the entrepreneurial type to the worker type to the rate of exiting entrepreneurship in the PSID, which is $q_{HL} = 0.2$. Analogously, we calibrate the probability to become an entrepreneur to the entry rate into entrepreneurship in the PSID, i.e., $q_{LH} = 0.02$.

4.2.2 Internally Estimated Parameters. We endogenously estimate the remaining parameters. The MSM involves:[33]

(1) Estimating the two parameters governing the non-homotheticity of preferences, the high-return type, and the two parameters driving the scale-dependence of returns: that is, $\boldsymbol{\theta} = \{\kappa, \nu, z_H, \psi, \eta\}$;

(2.a) Targeting the Pareto tail coefficients of the three remaining distributions, evaluated at the lower bound $\underline{x}$ optimally selected by the KS criterion. Namely, we target a ratio of capital income to wealth tail coefficients $\zeta_{ra}/\zeta_a = 0.87$, a ratio of consumption

[33]This MSM procedure is standard. Its objective function minimizes the distance, given a matrix of weights, between the observed and the model-generated moments, for a given weighting matrix.

to wealth tail coefficients $\zeta_c/\zeta_a = 2.24$, and a ratio of wealth to labor income tail coefficients $\zeta_a/\zeta_y = 0.61$. We additionally target the ratio of the capital income to wealth tail coefficients evaluated within the top 1% of their respective distribution in the Saez and Zucman (2016) dataset, $\zeta_{ra}(0.99)/\zeta_a(0.99) = 0.90$.

(2.b) Targeting aggregate moments in order to be consistent with national accounts and the level of wealth concentration in the U.S.: the wealth-income ratio $K/Y = 3.8$ and the top 1% wealth share.[34]

Our baseline model is, therefore, over-identified: We internally estimate 5 parameters to target 6 moments.

To ensure a meaningful comparison between the model and the data, we estimate the Pareto tails within the model using the exact same methodology as in the empirical Section 2. Specifically, we first draw 10 model-generated data samples (PSID-like waves) of 10,000 households, each from the stationary distribution, in order to replicate the 10 waves of PSID data that we used empirically. While these draws are representative of the overall population, they are subject to noise due to their finite sample properties. Within each of these samples, we estimate the Pareto tail coefficient using both OLS and MLE estimators from equation (1), and we select the optimal lower bound parameter $\underline{x}$ using the KS-criterion defined in equation (2). We use the median value of these statistics for our quantitative analysis. Finally, when comparing the ratio of capital income to wealth tail coefficients within the top 1%, we use a finite sample of 55,000 observations, which corresponds to the annual number of observations in Saez and Zucman (2016).

4.2.3 Estimation Results Table 2 displays the MSM parameter estimates. The point estimates indicate significant degrees of non-homotheticity with $\nu = 0.76 < \gamma = 2$, of type-heterogeneity in capital returns with $z_H = 4.3 > z_L = 1$, and of scale-dependence in capital returns with $\eta = 0.24 > 0$.

Table 2. Parameter estimates: baseline model.

κ	ν	z_H	η	ψ
0.14	0.76	4.10	0.24	0.15

[34]In the model, κ controls the scale of the economy, akin to the role of β in models without wealth in the utility. Models with homothetic preferences and return heterogeneity exhibit similar properties whether we use β or κ to scale the economy. That is, using the model with $\kappa = 0$ or the model with $\kappa > 0$ and $\nu = \gamma$ as our benchmark homothetic environment is inconsequential for our results.

These estimates are consistent with a back-of-the-envelope calculation based on the theoretical results of Section 3. Using the closed-form expressions of Proposition 3, the parameter values of Table 2 imply asymptotic ratios of tail coefficients equal to $\zeta_c/\zeta_a = \gamma/(\nu+\eta) \approx 2.0$ and $\zeta_{ra}/\zeta_a = 1/(1+\eta) \approx 0.81$. These values are close to their empirical counterparts, $\zeta_c/\zeta_a = 2.24$ and $\zeta_{ra}/\zeta_a = 0.87$, respectively. Interestingly, our estimated degree of non-homotheticity is in line with those obtained in other papers using panel data on wealth mobility and individual consumption behavior (see, e.g., Straub (2019); Benhabib et al. (2019)), but without targeting the Pareto tails for their calibration. Moreover, the presence of both type- and scale-dependence echoes the recent empirical findings of Fagereng et al. (2020); Bach et al. (2020); Balloch and Richers (2021).

Figure 18 in Appendix A shows the coefficient estimates for consumption, wealth, and capital income obtained by simulating 10 PSID-like samples, in order to compare them with the corresponding empirical results in Figures 9 to 20. Three key observations stand out. First, the Pareto tails emerge, as in the data, around the 90% quantile. Second, these model-based estimates closely mirror our empirical findings. The point estimates are remarkably stable within the top 10%, followed by significant uncertainty at the highest percentiles. Third, applying the Pareto test of Section 2.1 to the model-generated samples confirms the validity of the power law, and reveals that the Pareto distribution consistently provides a better fit than both lognormal and exponential distributions.

Finally, the conditional distributions of the variables implied by the model are aligned with those in the data, as shown in Appendix C.2. The cross-sectional relationship between log-consumption and log-wealth is a straight line with a slope of 0.36 in the model, compared to 0.31 in the data. Additionally, the slope of the relationship between log-capital income and log-wealth is equal to 1.08 in the model, compared to 1.13 in the data.

4.3 Decomposing the Mechanisms

In this section, we evaluate the contribution of the key mechanisms of our model in replicating the magnitude and ranking of the four Pareto tails, and assess their identification power. We study six distinct counterfactual economies. Our baseline economy, (1), is a standard Aiyagari (1994) model with bequests and a thick-tailed earnings distribution, but with homothetic preferences and type- and scale-independent returns to capital. We then successively add to this model the following channels: (2) type-dependence in returns; (3) scale-dependence in returns; (4) type- and scale-dependence in returns; (5) non-homothetic preferences with type-dependent returns; (6) non-homothetic preferences with type- and scale-dependent returns. In all of these different models, we keep targeting the six moments mentioned above. The resulting parameter estimates are reported in Table 3, and the corresponding Pareto tails are presented in Table 4.

Table 3. Parameter estimates: counterfactual economies.

		κ	ν	z_H	η	ψ
Homothetic preferences						
(1)	Homogeneous returns	0.670	2.000	1.000	0.000	0.000
(2)	Type-dependence	0.007	2.000	12.10	0.000	0.000
(3)	Scale-dependence	0.000	2.000	0.000	0.190	1.290
(4)	Type- and Scale-dependence	0.001	2.000	4.900	0.600	0.110
Non-homothetic preferences						
(5)	Type-dependence	0.215	0.793	4.450	0.000	0.000
(6)	Type- and scale-dependence	0.140	0.763	4.100	0.240	0.150

Consider first the standard Aiyagari (1994) model with homothetic preferences ($\nu = \gamma$), no type-dependence ($z_H = z_L = 1$), no scale-dependence ($\eta = \psi = 0$), and a persistent–transitory process for labor income augmented with a Pareto distribution at the top. As shown in the first row of Table 4, this workhorse model is capable of replicating the proper ranking and magnitudes of the tails of consumption and labor income, i.e., $\zeta_c > \zeta_y$. This suggests in particular that the permanent income hypothesis is a powerful way to generate a less uneven distribution of consumption than current labor income. Moreover, consumption has a larger tail coefficient than wealth $\zeta_c = 3.04 > \zeta_a = 2.41$. This finding is consistent with Carroll and Kimball (1996) who show that under CRRA preferences, a borrowing constraint, and income uncertainty, consumption is generally concave in—and hence strictly more evenly distributed than—wealth. However, as is well known, this model generate far too little wealth concentration: $\zeta_a = 2.41$, for a targeted Pareto coefficient of 1.4. This is because, in this class of models, the wealth distribution inherits the tail of the *after-tax* labor income distribution. Furthermore, this model fails to significantly disentangle the capital income tail from the wealth tail ($\zeta_{ra}/\zeta_a = 1$).

In the second row of Table 4, we add heterogeneous—type-dependent—asset returns, captured by $z_H > z_L$. Unsurprisingly, this augmented model is now able to replicate the magnitude and ranking of the Pareto coefficients of both labor income and wealth. Interestingly, it is also able to reproduce the fact that the tail of capital income is strictly thicker than that of wealth. This is due to a composition effect: at finite levels of wealth, the fraction of high return-types is larger among wealthier agents, so that average returns conditional

Table 4. Counterfactual models and selected moments.

Data/Model	Pareto tails: mean MLE estimate[a] ζ_c	ζ_y	ζ_y^{net}	ζ_a	ζ_{ra}	Top 1% wealth	$\frac{\text{Wealth}}{\text{income}}$ ratio
Adjusted PSID (2005–2021)	3.06	2.25	2.57	1.38	1.20	0.35	3.8
Homothetic preferences							
(1) Homogeneous returns	3.04	2.25	2.57	2.42	2.42	0.09	3.8
(2) Type-dependence	2.65	2.25	2.57	1.32	1.02	0.29	3.7
(3) Scale-dependence	2.56	2.25	2.57	1.30	1.16	0.32	3.7
(4) Type- and scale-dependence	2.65	2.25	2.57	1.34	1.08	0.35	3.6
Non-homothetic preferences							
(5) Type-dependence	3.08	2.25	2.57	1.37	1.19	0.34	3.7
(6) Type- and scale-dependence	3.06	2.25	2.57	1.36	1.18	0.35	3.8

	Pareto tails, mean MLE estimate: fixed threshold $\underline{x}$ percentile[b]								
	ζ_c			ζ_a			ζ_{ra}		
	95^{th}	98^{th}	99^{th}	95^{th}	98^{th}	99^{th}	95^{th}	98^{th}	99^{th}
Adjusted PSID (2005–2021)	3.21	3.44	3.30	1.35	1.48	1.56	1.00	1.27	1.36
Saez and Zucman (2016)	–	–	–	1.50	1.49	1.55	1.27	1.28	1.28
Homothetic preferences									
(1) Homogeneous returns	3.08	2.97	2.84	2.47	2.34	2.35	2.47	2.34	2.35
(2) Type-dependence	2.65	2.50	2.32	1.33	1.44	1.67	0.99	1.17	1.22
(3) Scale-dependence	2.51	2.35	2.40	1.29	1.49	1.78	1.14	1.31	1.55
(4) Type- and scale-dependence	2.57	2.41	2.30	1.32	1.47	1.67	1.11	1.26	1.41
Non-homothetic preferences									
(5) Type-dependence	3.17	3.11	3.40	1.40	1.43	1.56	1.24	1.37	1.56
(6) Type- and scale-dependence	3.10	3.01	3.25	1.35	1.37	1.47	1.17	1.25	1.39

[a] The Pareto tails are estimated within the model using a sample of 10,000 observations to compare them with the tails obtained using the same method as in the PSID.
[b] The Pareto tails are estimated within the model using a sample of 55,000 observations to compare the upper tails with those obtained from the Saez and Zucman (2016) dataset.

on wealth are increasing and capital income is convex in wealth. This contrasts with our theoretical analysis of Section 3, which focused on the limits as wealth grew to infinity. In this case, type-dependence alone was unable to decouple these two tails at arbitrarily high levels of wealth, because the share of high return-types conditional on wealth converges to a constant at the limit. The composition effect was then negligible, resulting in a distribution of capital income that was just as concentrated as that of wealth.

Importantly, this quantitative model also *does* generate a strictly thinner tail for consumption ($\zeta_c = 2.65$) than for wealth ($\zeta_a = 1.32$) in finite samples, consistent with Carroll and Kimball (1996). However, and this is crucial for our purposes, the concavity of consumption implied by the model is far too small to replicate the thinness of the consumption tail that we observe in the data (3.09). In other words, the standard model correctly generates concavity in consumption, but simply not enough to rationalize the distribution of consumption among the top 10% richest households. This finding is of course driven by the same logic as that underlying the theoretical results of Section 3, according to which, as wealth grows to infinity, consumption should become asymptotically linear in wealth in this model. In fact, the convergence of the consumption Pareto coefficient to the same level as the wealth Pareto coefficient becomes evident in the second part of Table 4 as we estimate these tails in the quantitative model for several fixed levels of the lower bound parameter $\underline{x}$. The Pareto coefficient for consumption falls to 2.32 within the top 1%, while it rises to 1.67 for wealth, corresponding to a ratio ζ_c/ζ_a equal to 1.39 (compared to 2.10 in the PSID). In addition, the quantitative model generates an excessively thick capital income tail relative to that of wealth ($\zeta_{ra} = 1.02$). In other words, adding heterogeneous returns to the baseline Aiyagari framework solved the *wealth concentration puzzle*, but this came at the expense of introducing a mirror *consumption concentration puzzle* as well as a *capital income concentration puzzle*.

Next, consider in the third row of Table 4 the model with only scale-dependence ($\eta > 0$), assuming again homothetic preferences ($\nu = \gamma$). In this case, the model does a better job at replicating simultaneously the empirical magnitudes of the labor income, wealth, and capital income tails. However, it does not help generating a thin consumption tail: $\zeta_c = 2.56$ in the model. In fact, consistent with our theoretical findings of Section 3, this model leads to an even thicker consumption tail than the model with only type-dependence, since consumption is then strictly convex in wealth. Finally, incorporating scale-dependence into a model already featuring type-dependence (fourth row of Table 4) continues to fall short of accurately replicating the data.

These results imply that we must introduce an additional mechanism to decouple the upper tails of consumption and wealth, or equivalently, to generate much more concavity in the consumption function of the richest households. We do so in the last two rows of Table 4 by studying non-homothetic preferences. In the fifth row, we first assume that returns are only type-dependent. In this case, the model successfully replicates the thinness of the consumption tail observed in the data ($\zeta_c > 3$), and its Pareto coefficient does not converge to that of wealth at higher values of the lower bound parameter $\underline{x}$. This is again consistent with our theoretical results of Section 3, according to which non-homothetic preferences are able to generate consumption behavior that is asymptotically concave in wealth.

Interestingly, note that the model with solely type-dependence is already able to generate the correct wealth and capital income tails at the optimal lower bound $\underline{x}$ obtained using the

KS criterion. Adding scale-dependent returns in the sixth row of the table does not improve the ability of the model to match these moments. This suggests that, at the optimal KS criterion at which we estimated our Pareto coefficients, the tails of the four cross-sectional distributions do not carry sufficient information to allow us to disentangle type- from scale-dependence. Nevertheless, our theoretical results imply that data covering higher quantiles of wealth than the PSID would allow us to potentially identify these two mechanisms: Any discrepancy that remains between the capital income and wealth tails at very high levels of wealth—at which the composition effect due to type-dependence becomes negligible—can be attributed to, and precisely identifies, the scale-dependence elasticity η. Indeed, the results in the lower part of Table 4 reveal that, absent scale-dependence, the ratio of the Pareto coefficients of wealth and capital income converges to 1 within the top 1% in the non-homothetic model.[35] By contrast, incorporating scale-dependence into this model consistently generates a decoupling of the two tails ($\zeta_{ra} = 1.39 < \zeta_a = 1.47$). We view this identification of type- vs. scale-dependence, based on the ratio of tail coefficients at high levels of wealth, as a simple way to address the empirical debate regarding the respective importance of these two kinds of return heterogeneity in the data (see, e.g., Fagereng et al. (2020); Balloch and Richers (2021)).[36]

4.4 Robustness and Alternative Explanations

We conclude this section by briefly exploring potential alternative mechanisms that could rationalize the ranking and magnitudes of the four tail coefficients.

Labor Income Risk. First, following the logic of Carroll and Kimball (1996), we keep preferences homothetic and consider whether a higher amount of idiosyncratic income risk might lead to larger precautionary saving motives, and hence a thinner consumption tail. We explore this mechanism by raising the volatility of the persistent labor income shocks to $\sigma_h^2 = 0.25$ and reducing their persistence to $\rho_h = 0.90$. While this calibration requires greater

[35]Interestingly, this is not the case in the homothetic model with only type-dependence—this is because the degree of type-dependence z_H required to match wealth inequality is significantly lower in the non-homothetic model, accelerating the dissipation of the composition effect.

[36]While such identification may not be generalized to every return process, the composition effects should vanish at the very top. Therefore, using an even higher level of scale might be required. To this, the IRS data used by Saez and Stantcheva (2018) cover much higher levels of wealth. They show that the Pareto coefficient for capital income converges to a value of 1.38. The Forbes rich list we used in Section 2 implies a Pareto coefficient for wealth of 1.5. These values suggest a positive amount of scale dependence $\eta = \zeta_a/\zeta_{ra} - 1 = 0.087$. Using the Saez and Zucman (2016) dataset, we find a ratio of $\eta = \zeta_a/\zeta_{ra} - 1 = 1.43/1.28 - 1 = 0.12$ at the top 0.1%.

return heterogeneity to generate a tail coefficient for wealth of 1.4, it is still insufficient to account for the empirical magnitude of the consumption tail coefficient, which remains too low in the model at 2.60. Similarly, increasing the variance of the transitory component from $\sigma_e^2 = 0.04$ to $\sigma_e^2 = 0.25$ does not help in generating the disconnect between the consumption and wealth tails. Therefore, variables that we did not explicitly model and that would affect the budget constraint, such as health shocks, would not significantly affect our analysis. Finally, adding an unemployment state with a 2% probability and a 60% re-employment rate, consistent with the U.S. labor market data, has only minor effects on our results.

The Role of Taxes. Next, we examine the impact on the consumption tail of a progressive capital income tax schedule, by levying the CRP tax function (9) on total income. In that case, the necessary level of return heterogeneity is much higher. However, this does not lead to a thinner consumption tail under homothetic preferences, but it does decouple the wealth and capital income tails—the latter being lower than 0.9. Finally, note that the labor income CRP tax schedule plays an important role. Applying a constant top tax rate would seriously worsen the ability of the homothetic model to replicate the data. Absent return heterogeneity, the wealth tail would inherit the coefficient of the labor income tail (2.2). Under return heterogeneity, the consumption tail would be more concentrated, with a coefficient around 2.35 based on the KS-optimal subsample.

Higher Risk Aversion. As we discussed in Section 3, Ma and Toda (2021) argue that there exists a range of parameters under which the model with homothetic preferences yields a marginal propensity to consume that converges to 0 at the tail, and hence a strictly thinner tail for consumption than for wealth. To address this potential mechanism, we recalibrate our model with homothetic preferences and a higher coefficient of relative risk aversion γ in order to reproduce the ratio of Pareto coefficients ζ_c/ζ_a that we observe empirically. We find that, regardless of the persistence of the return-type process, we would need to assume $\gamma > 6$ to generate the thinness of the consumption tail in partial equilibrium. Such implausibly large values imply that the decay of asymptotic MPCs alone is far too weak to rationalize the empirical evidence on Pareto tails. This conclusion is reinforced by noting that raising the risk-aversion coefficient must be accompanied by a reduction in the discount factor β to match the economy's capital-income ratio K/Y, which further reduces the Pareto coefficient of consumption in the model.

Life-Cycle. Augmenting the model with overlapping generations (OLG) and age as a state variable does not alter the main conclusion of this paper. Return heterogeneity, which enables us to match the wealth Pareto tail, comes at the expense of excessive consumption concentration. Incorporating wealth into the utility function resolves this tension.

5 Revenue-Maximizing Top Capital Tax Rate

In this final section, we explore the implications of jointly matching the Pareto tail coefficients of consumption, labor income, wealth, and capital income for the revenue-maximizing long-run tax rate on capital—the top of the Laffer curve—levied on the top 1% richest households.[37] Our key result is that replicating only three (or fewer) of these tail coefficients delivers incorrect evaluations of the long-run elasticity of savings with respect to the tax rate, and hence of the revenue-maximizing top tax rate. In other words, the distributions of wealth and capital income are not sufficient statistics for the revenue-maximizing capital tax rate; matching the consumption distribution provides additional information that is necessary to properly design policy.

Non-Linear Capital Tax Schedule. We modify the baseline quantitative model as follows. The capital tax schedule is progressive and takes a piece-wise linear form:

$$t_a = \overline{\tau}_a (a - a^*) \mathbb{1}_{\{a \geq a^*\}}$$

for some $a^* > 0$ that represents the threshold of the top wealth bracket. We focus on a marginal tax rate levied on the richest households above a threshold a^* that coincides with the top 1%. At this threshold, the marginal tax rate jumps from 0 to $\overline{\tau}_a$.

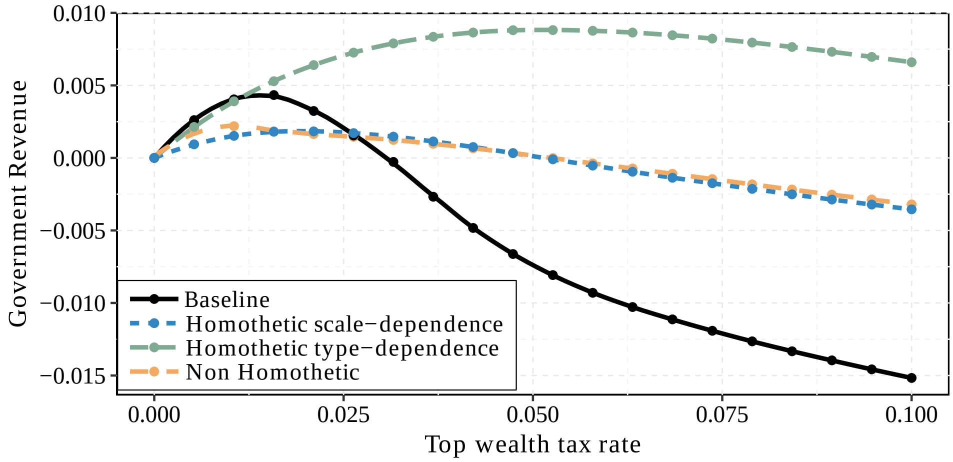
Laffer Curve for Top Wealth. As our main quantitative experiment, we compute the revenue-maximizing tax rate $\overline{\tau}_a^*$ and report the results in Figure 7.

Consider first the canonical model with skewed earnings and type-dependent returns—allowing us to match the Pareto tails of labor income and wealth—but homothetic preferences and scale-independent returns. The Laffer curve is represented by the green line in Figure 7. In this case, the revenue-maximizing top marginal tax rate is high, around 5%.

Second, consider the model with homothetic preferences and scale-dependent returns, allowing us to match the Pareto tails of labor income, wealth, and capital income, represented by the dashed blue line in the figure. In this setting, the revenue-maximizing top tax rate is

[37] This revenue maximization objective gives the upper bound of the set of Pareto efficient top tax rates, and avoids dealing with difficult issues of social welfare comparisons between models with standard vs. non-homothetic wealth-dependent preferences.

Figure 7. Laffer curves in alternative models.



much lower, with $\overline{\tau}_a^* \approx 1.2\%$. This is because, under scale-dependent returns, savings are a convex function of wealth. This convexity implies, following the Le Châtelier principle, that the elasticity of wealth to the tax rate is larger than in the canonical model that has a linear savings function. Intuitively, taxing capital reduces savings, which consequently yield lower returns, thus amplifying the initial fall in savings, and so on. This larger elasticity implies a smaller revenue-maximizing top tax rate compared to the canonical type-dependent model, aligning with theoretical and quantitative findings by Gerritsen et al. (2020); Schulz (2021), and Gaillard and Wangner (2021).

Third, consider the case of non-homothetic preferences with only type-dependent returns. In this case, the model can match the tail distributions of consumption, labor income, and wealth. Again, the revenue-maximizing top tax rate is lower than in the canonical model, with $\overline{\tau}_a^* = 0.8\%$. This is because non-homothetic preferences also convexify the savings policy function, leading again to a higher savings elasticity. This convexity is stronger than that implied by the estimated scale-dependent model.

In other words, while both scale-dependent returns and non-homothetic preferences are able to match the same wealth distribution, they predict different long-run capital elasticities, as illustrated in the right panel of Figure 7. The key difference between these two polar models is that the former (scale-dependent returns) implies a convex consumption function, while the latter (non-homothetic preferences) implies a concave consumption function. These two ways of generating a convex savings policy function predict distinct revenue-maximizing top capital tax rates.

The "correct" tax rate, at the peak of the black line in Figure 7, is obtained under the baseline model (black solid line), at $\tau_a^* \approx 1\%$. Both non-homotheticity and scale-dependence are operational in our setting and are essential for simultaneously aligning the Pareto coefficients of all four tails. However, the extent of scale-dependence is relatively moderate, and does not materialize in a significant shift of the revenue-maximizing tax rate to the left. Moreover, due to scale-dependence, the model assigns a lower weight to the wealth component in the utility function. This leads to a higher revenue-maximizing tax rate compared to the type-dependent non-homothetic model. In conclusion, matching only three of the four tails—and, most importantly, failing to properly target the consumption tail coefficient—would lead to incorrect policy recommendations. This is because the consumption and capital income tails discipline the strength of the underlying mechanisms—non-homotheticity of preferences and scale-dependence of returns—that generate a convex savings rate. Thus, both tails provide us with additional information that is critical to properly model the saving behavior and correctly evaluate the long-run capital supply elasticity.

6 Conclusion

This paper establishes that the cross-sectional distributions of consumption, labor income, wealth, and capital income follow asymptotic Pareto distributions in the upper tails with a strict ordering of their tail parameters. Specifically, consumption exhibits a thinner Pareto tail than labor income, which in turn has a thinner Pareto tail than wealth and capital income. We demonstrate that replicating the ranking and magnitude of these four tails constitutes a challenge for the canonical heterogeneous-agent models. Introducing non-homothetic preferences and scale-dependent asset returns allows the model to simultaneoously replicate the behavior of all four tails: The former (respectively, the latter) channel decouples the consumption (resp., capital income) tail from the wealth tail. The central take-away of our analysis is that the rich appear to save for reasons other than funding future consumption needs. Whether such behavior is driven by bequest motives, status-seeking concerns,[38] or other fundamental reasons that future research should explore and model, we believe that it should be more systematically incorporated into the workhorse heterogeneous-agent model in macroeconomics.

[38] As Adam Smith wrote in the *Theory of Moral Sentiments* (1759), "*The rich man glories in his riches because he feels that they naturally attract the world's attention to him ... He is fonder of his wealth on this account than for all the other advantages it brings him.*" For a modern treatment of this hypothesis and some of its implications, see, e.g., Roussanov (2010).

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll (2022): "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *The Review of Economic Studies*, Vol. 89, pp. 45–86.

Aiyagari, S. Rao (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, Vol. 109, pp. 659–684.

Angeletos, George-Marios (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, Vol. 10, pp. 1–30.

Angeletos, George-Marios and Laurent-Emmanuel Calvet (2006): "Idiosyncratic Production Risk, Growth and the Business Cycle," *Journal of Monetary Economics*, Vol. 53, pp. 1095–1115.

Aoki, Shuhei and Makoto Nirei (2017): "Zipf's law, Pareto's law, and the evolution of top incomes in the United States," *American Economic Journal: Macroeconomics*, Vol. 9, pp. 36–71.

Atkinson, Anthony B. (1971): "Capital Taxes, the Redistribution of Wealth and Individual Savings," *The Review of Economic Studies*, Vol. 38, pp. 209–227.

Atkinson, Anthony Barnes and Joseph E. Stiglitz (1976): "The Design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics*, Vol. 6, pp. 55–75.

Attanasio, Orazio and Luigi Pistaferri (2014): "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure," *The American Economic Review: Papers & Proceedings*, Vol. 104, pp. 122–126.

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini (2020): "Rich Pickings? Risk, Return, and Skill in Household Wealth," *The American Economic Review*, Vol. 110, pp. 2703–2747.

Balloch, Cynthia Mei and Julian Richers (2021): "Asset Allocation and Returns in the Portfolios of the Wealthy."

Benhabib, Jess and Alberto Bisin (2018): "Skewed Wealth Distributions: Theory and Empirics," *Journal of Economic Literature*, Vol. 56, pp. 1261–1291.

Benhabib, Jess, Alberto Bisin, and Mi Luo (2019): "Wealth Distribution and Social Mobility in the US: A Quantitative Approach," *The American Economic Review*, Vol. 109, pp. 1623–1647.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu (2011): "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents," *Econometrica*, Vol. 79, pp. 123–157.

——— (2015): "The Wealth Distribution in Bewley Economies with Capital Income Risk," *Journal of Economic Theory*, Vol. 159, pp. 489–515.

Bewley, Truman F. (1986): "Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers," *Contributions to Mathematical Economics in Honor of Gerard Debreu, Werner Hildenbrand and Andreu Mas-Colell, eds. (Amsterdam: North-Holland)*, Vol. 1, pp. 79–102.

Blanchet, Thomas (2022): "Uncovering the Dynamics of the Wealth Distribution," *Working Paper*.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten (2016): "Consumption Inequality and Family Labor Supply," *The American Economic Review*, Vol. 106, pp. 387–435.

Boar, Corina and Vigiliu Midrigan (2023): "Should We Tax Capital Income or Wealth," *The American Economic Review; Insights*, Vol. 5, pp. 259–274.

Bourguignon, François (1981): "Pareto Superiority of Unegalitarian Equilibria in Stiglitz' Model of Wealth Distribution with Convex Saving Function," *Econometrica*, Vol. 49, pp. 1469–1475.

Buda, Gergely, Vasco M. Carvalho, Stephen Hansen, Álvaro Ortiz, Tomasa Rodrigo, and José V. Rodríguez Mora (2022): "National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption," *CEPR Discussion Paper No. DP17519*.

Cagetti, Marco and Mariacristina De Nardi (2006): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, Vol. 114, pp. 835–870.

Carroll, Christopher D. (1998): "Why Do the Rich Save so Much?" *NBER Working Paper*.

Carroll, Christopher D. and Miles S. Kimball (1996): "On the Concavity of the Consumption Function," *Econometrica*, Vol. 64, pp. 981–992.

Castaneda, Ana, Javier Diaz-Gimenez, and Jose-Victor Rios-Rull (2003): "Accounting for the US Earnings and Wealth Inequality," *Journal of Political Economy*, Vol. 111, pp. 818–857.

Citro, Constance F., Robert T. Michael et al. (1995): *Measuring Poverty: A New Approach*: National Academy Press.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman (2009): "Power-Law Distributions in Empirical Data," *SIAM Review*, Vol. 51, pp. 661–703.

De Nardi, Mariacristina (2004): "Wealth Inequality and Intergenerational Links," *The Review of Economic Studies*, Vol. 71, pp. 743–768.

De Nardi, Mariacristina, Svetlana Pashchenko, and Ponpoje Porapakkarm

(2017): "The lifetime costs of bad health,"Technical report, National Bureau of Economic Research.

Dynan, Karen E., Jonathan Skinner, and Stephen P. Zeldes (2004): "Do the Rich Save More?" *Journal of Political Economy*, Vol. 112, pp. 397–444.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri (2020): "Heterogeneity and Persistence in Returns to Wealth," *Econometrica*, Vol. 88, pp. 115–170.

Ferey, Antoine, Benjamin B. Lockwood, and Dmitry Taubinsky (2023): "Sufficient Statistics for Nonlinear Tax Systems With General Across-Income Heterogeneity," *Working Paper*.

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

Gaillard, Alexandre and Philipp Wangner (2021): "Wealth, Returns, and Taxation: A Tale of Two Dependencies," *SSRN Working Paper*, Vol. 3966130.

Garner, Thesia I., Robert S. Martin, Brett Matsumoto, and Scott Curtin (2022): "Distribution of US Personal Consumption Expenditures for 2019: A Prototype Based on Consumer Expenditure Survey Data," *BLS Working Paper No. 557*.

Gerritsen, Aart, Bas Jacobs, Alexandra Victoria Rusu, and Kevin Spiritus (2020): "Optimal Taxation of Capital Income with Heterogeneous Rates of Return," *Working Paper*.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, and Daphne Chen (2023): "Use It or Lose It: Efficiency and Redistributional Effects of Wealth Taxation," *The Quarterly Journal of Economics*, Vol. 138, pp. 835—-894.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017): "Optimal Tax Progressivity: An Analytical Framework," *The Quarterly Journal of Economics*, Vol. 132, pp. 1693—-1754.

Hellwig, Christian and Nicolas Werquin (2022): "A Fair Day's Pay for a Fair Day's Work: Optimal Tax Design as Redistributional Arbitrage," *Working Paper*.

Hubmer, Joachim, Per Krusell, and Anthony A. Smith Jr. (2021): "Sources of US Wealth Inequality: Past, Present, and Future," *NBER Macroeconomics Annual*, Vol. 35, pp. 391–455.

Huggett, Mark (1993): "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, Vol. 17, pp. 953–969.

İmrohoroğlu, Ayşe (1989): "Cost of Business Cycles with Indivisibilities and Liquidity Constraints," *Journal of Political Economy*, Vol. 79, pp. 1364–1383.

Jaravel, Xavier and Danial Lashkari (2023): "Measuring Growth in Consumer Welfare with Income-Dependent Preferences: Nonparametric Methods and Estimates for the United States," *The Quarterly Journal of Economics (Forthcoming)*.

KAPLAN, GREG, BENJAMIN MOLL, AND GIOVANNI L. VIOLANTE (2018): "Monetary Policy According to HANK," *The American Economic Review*, Vol. 108, pp. 697–743.

KOCHERLAKOTA, NARAYANA AND LUIGI PISTAFERRI (2009): "Asset Pricing Implications of Pareto Optimality with Private Information," *Journal of Political Economy*, Vol. 117, pp. 555–590.

KOIJEN, RALPH, STIJN VAN NIEUWERBURGH, AND ROINE VESTMAN (2014): "Judging the quality of survey data by comparison with" truth" as measured by administrative records: Evidence from Sweden," in *Improving the measurement of consumer expenditures*: University of Chicago Press, pp. 308–346.

KRUEGER, DIRK AND FABRIZIO PERRI (2006): "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *The Review of Economic Studies*, Vol. 73, pp. 163–193.

KRUSELL, PER AND ANTHONY A. SMITH JR. (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, Vol. 106, pp. 867–896.

LEE, BYOUNGCHAN (2021): "Wealth Inequality, Aggregate Consumption, and Macroeconomic Trends under Incomplete Markets," *Working Paper*.

MA, QINGYIN AND ALEXIS AKIRA TODA (2021): "A Theory of the Saving Rate of the Rich," *Journal of Economic Theory*, Vol. 192.

MEYER, BRUCE D. AND JAMES X. SULLIVAN (2023): "Consumption and Income Inequality in the United States since the 1960s," *Journal of Political Economy*, Vol. 131, pp. 247–284.

MIAN, ATIF R., LUDWIG STRAUB, AND AMIR SUFI (2021): "Indebted Demand," *The Quarterly Journal of Economics*, Vol. 136, pp. 2243–2307.

MICHAU, JEAN-BAPTISTE, YOSHIYASU ONO, AND MATTHIAS SCHLEGL (2023): "The Preference for Wealth and Inequality: Towards a Piketty Theory of Wealth Inequality," *Working Paper*.

MOLL, BENJAMIN (2014): "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation," *The American Economic Review*, Vol. 104, pp. 3186–3221.

MOLL, BENJAMIN, LUKASZ RACHEL, AND PASCUAL RESTREPO (2022): "Uneven Growth: Automation's Impact on Income and Wealth Inequality," *Econometrica*, Vol. 90, pp. 2645–2683.

MORRISON, WENDY (2023): "Redistribution and Investment," *Working Paper*.

PIKETTY, THOMAS AND EMMANUEL SAEZ (2003): "Income Inequality in the United States, 1913–1998," *The Quarterly Journal of Economics*, Vol. 118, pp. 1–41.

Quadrini, Vincenzo (2000): "Entrepreneurship, Saving, and Social Mobility," *Review of Economic Dynamics*, Vol. 3, pp. 1–40.

Roussanov, Nikolai (2010): "Diversification and Its Discontents: Idiosyncratic and Entrepreneurial Risk in the Quest for Social Status," *The Journal of Finance*, Vol. 65, pp. 1755–1788.

Saez, Emmanuel and Stefanie Stantcheva (2018): "A Simpler Theory of Optimal Capital Taxation," *Journal of Public Economics*, Vol. 162, pp. 120–142.

Saez, Emmanuel and Gabriel Zucman (2016): "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *The Quarterly Journal of Economics*, Vol. 131, pp. 519–578.

Schulz, Karl (2021): "Redistribution of Return Inequality," *Working Paper*.

Stiglitz, Joseph E. (1969): "Distribution of Income and Wealth Among Individuals," *Econometrica: Journal of the Econometric Society*, pp. 382–397.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron (2004): "Cyclical Dynamics in Idiosyncratic Labor Market Risk," *Journal of Political Economy*, Vol. 112, pp. 695–717.

Straub, Ludwig (2019): "Consumption, Savings, and the Distribution of Permanent Income," *Working Paper*.

Toda, Alexis Akira and Kieran Walsh (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, Vol. 123, pp. 1177–1200.

Vermeulen, Philip (2016): "Estimating the Top Tail of the Wealth Distribution," *The American Economic Review*, Vol. 106, pp. 646–650.

de Vries, Tjeerd and Alexis Akira Toda (2022): "Capital and Labor Income Pareto Exponents Across Time and Space," *Review of Income and Wealth*, Vol. 68, pp. 1058–1078.

Xavier, Inês (2021): "Wealth Inequality in the US: The Role of Heterogeneous Returns," *SSRN Working Paper 3915439*.

Yi, Gang (2022): "Wealth-in-Utility and the Optimal Capital Taxation," *SSRN Working Paper*.

A Empirical Appendix

Figure 8. Simulated Data: Pareto tail coefficient estimates for three parametric distributions.

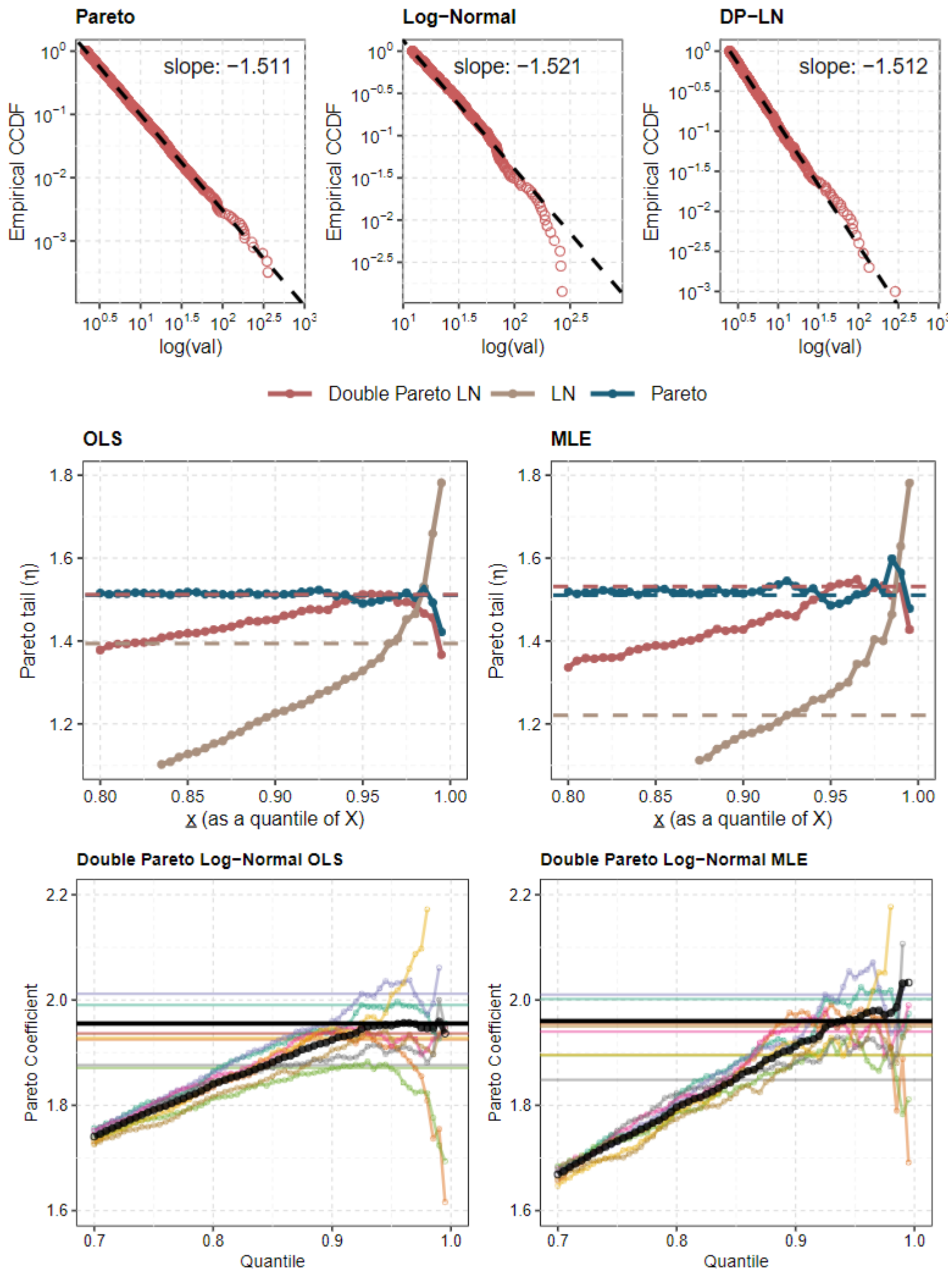


Figure 9. Consumption: log-log plots and Pareto tail coefficient estimates.

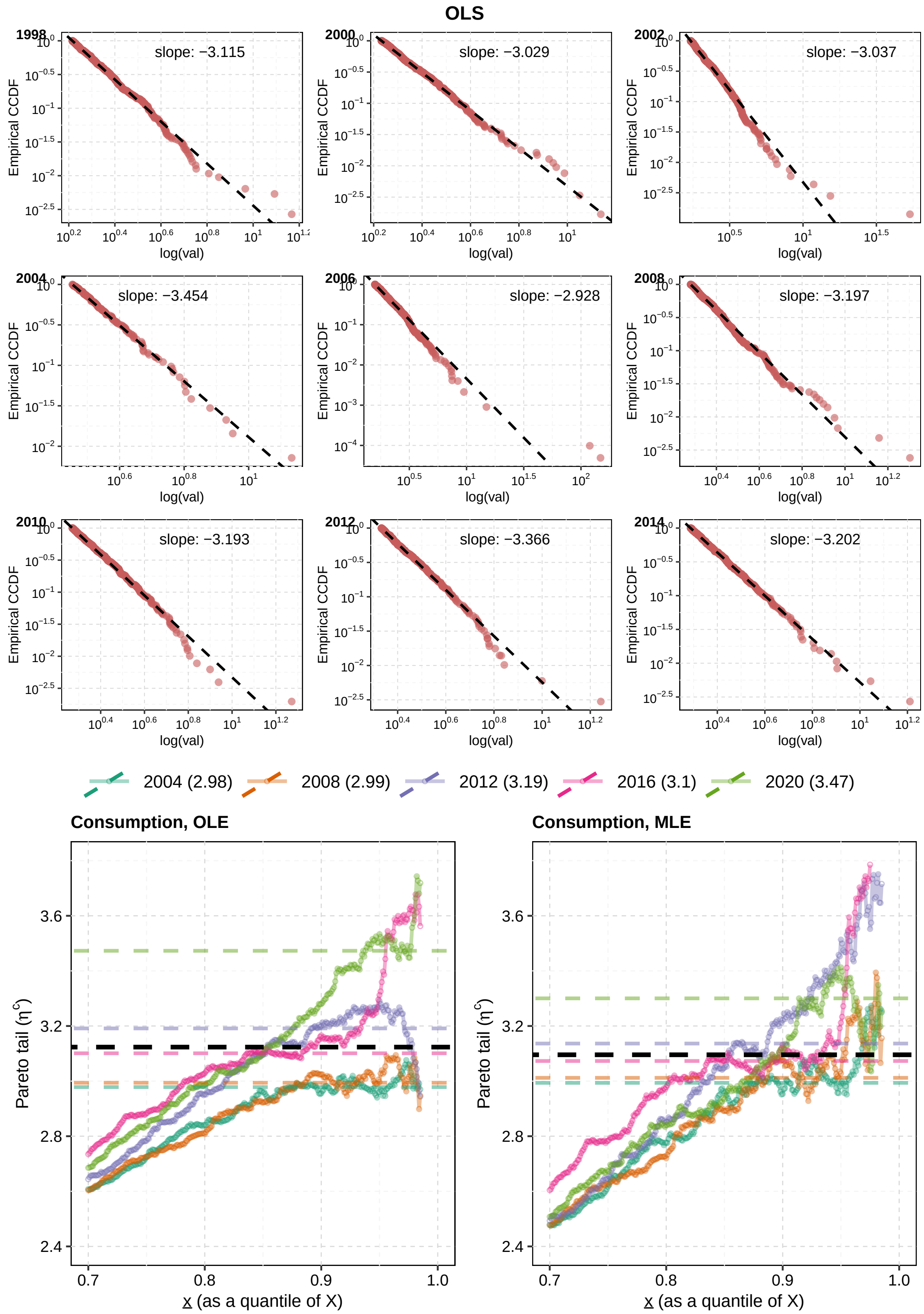


Figure 10. Labor income: log-log plots and Pareto tail coefficient estimates.

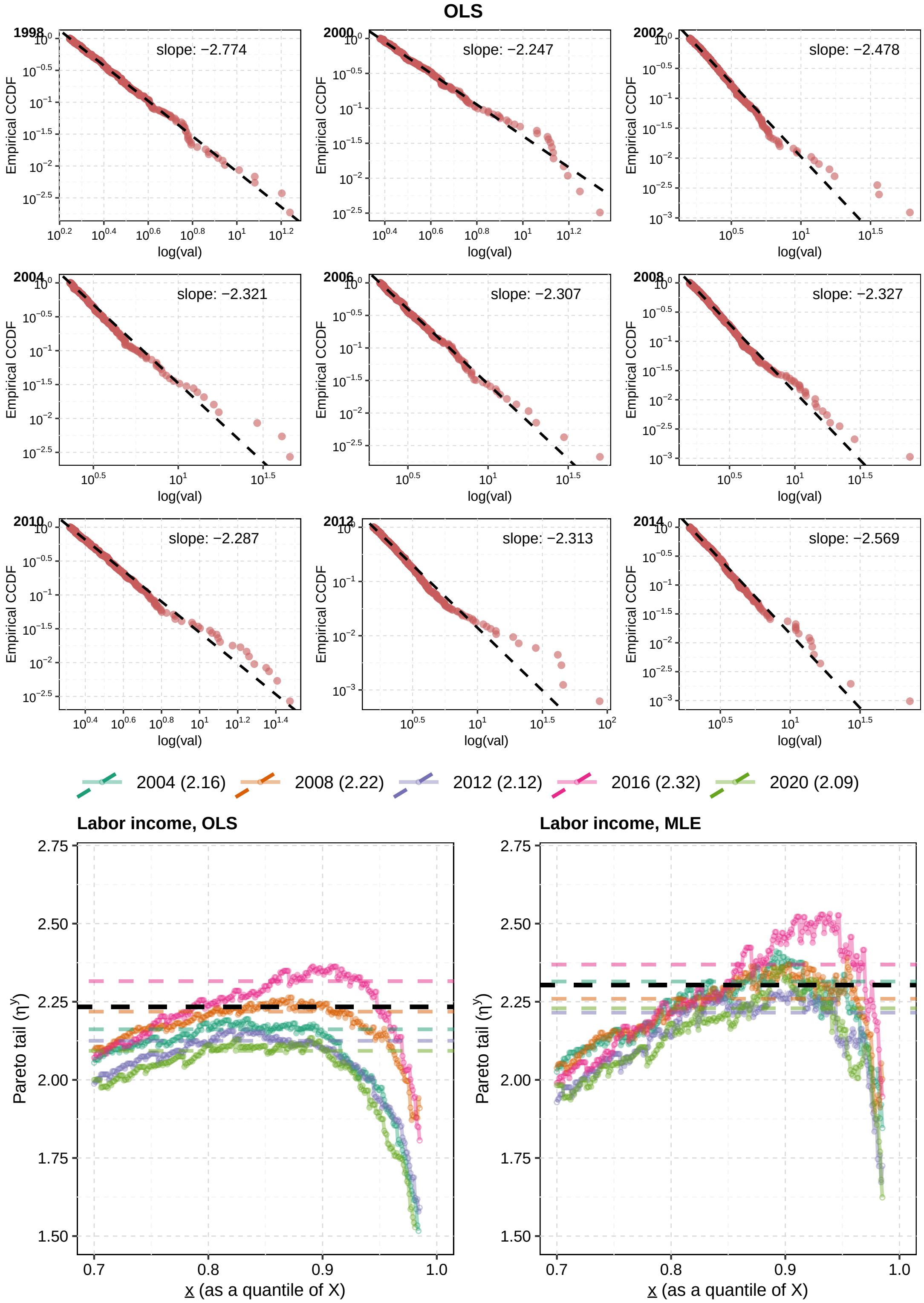


Figure 11. After-tax labor income: log-log plots and Pareto tail coefficient estimates.

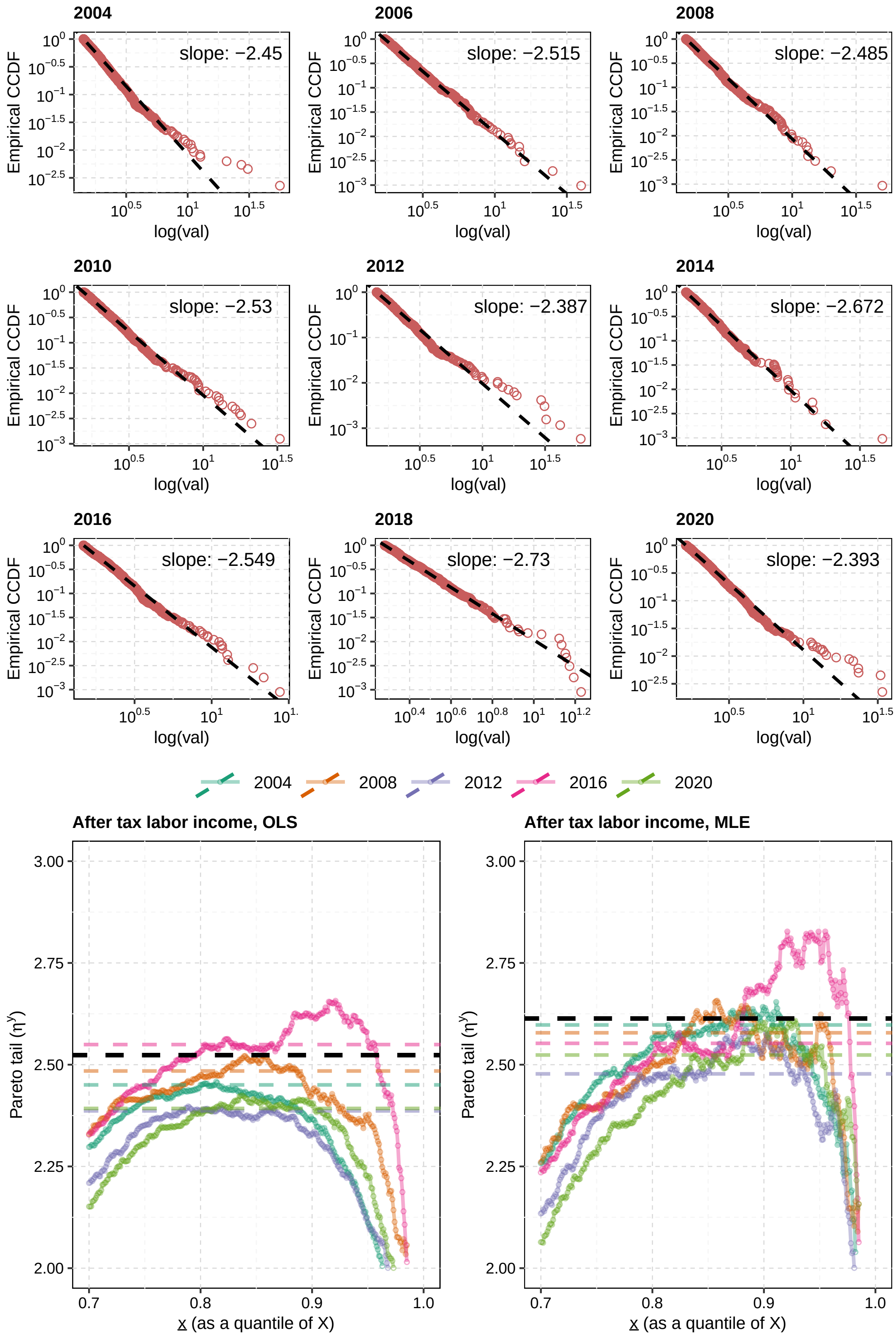
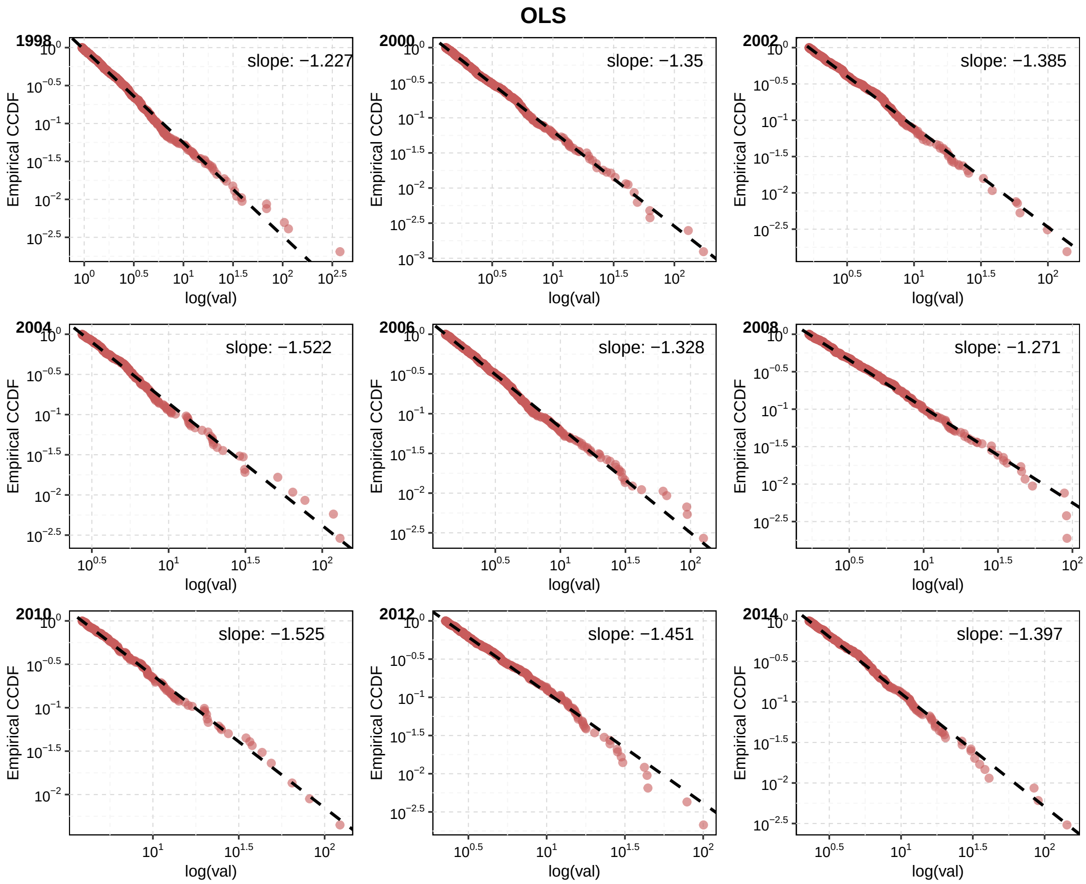


Figure 12. Wealth: log-log plots and Pareto tail coefficient estimates.



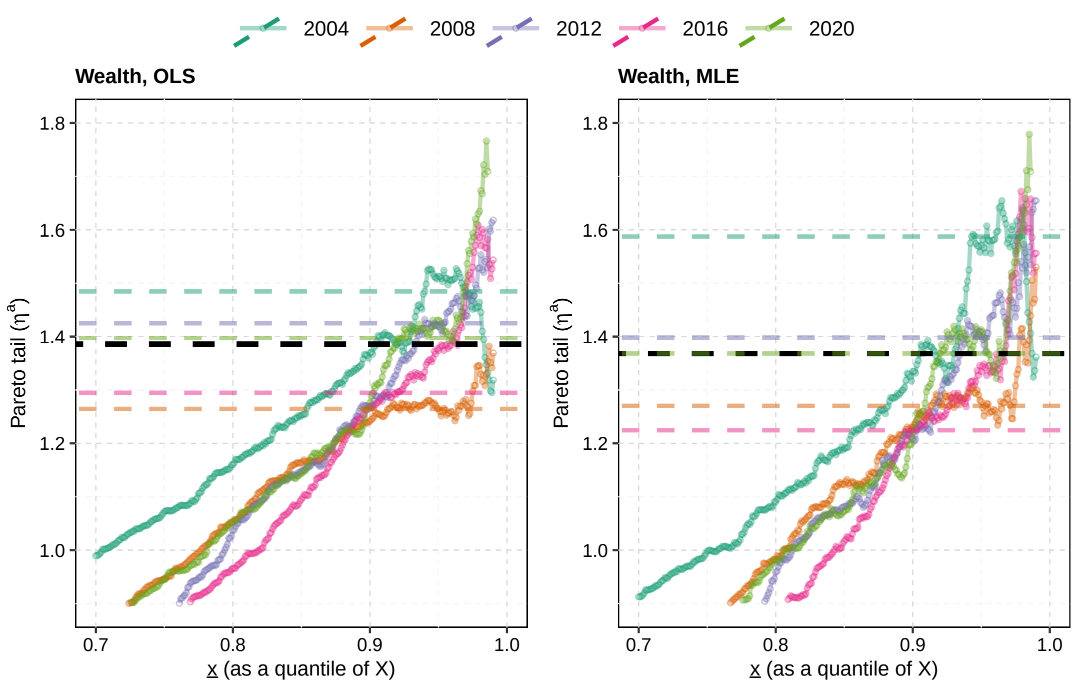


Figure 13. Wealth: log-log plots and Pareto tail coefficient estimates from the PSID augmented by the Forbes rich list.

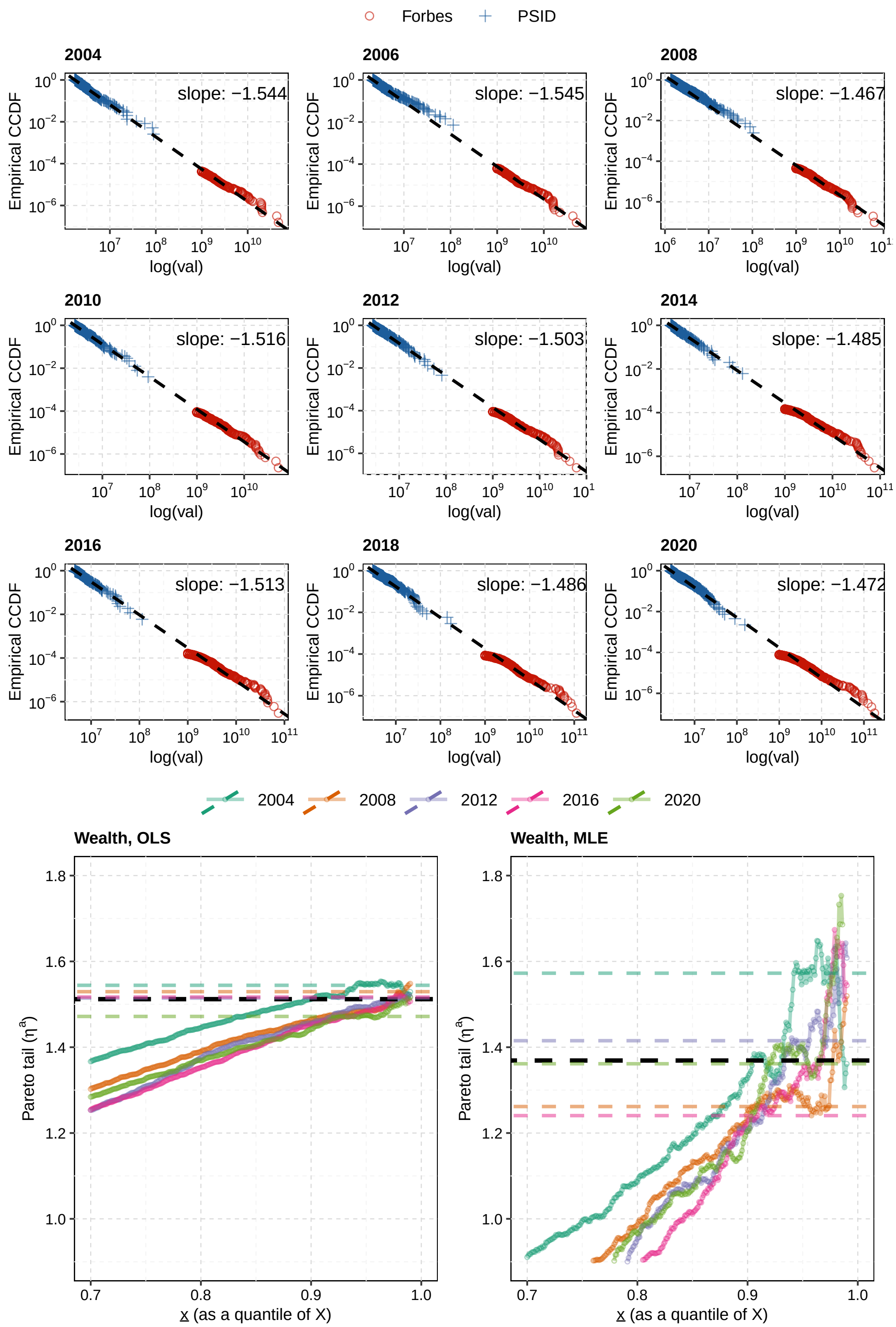


Figure 14. Labor income and wealth: data from Saez and Zucman (2016).

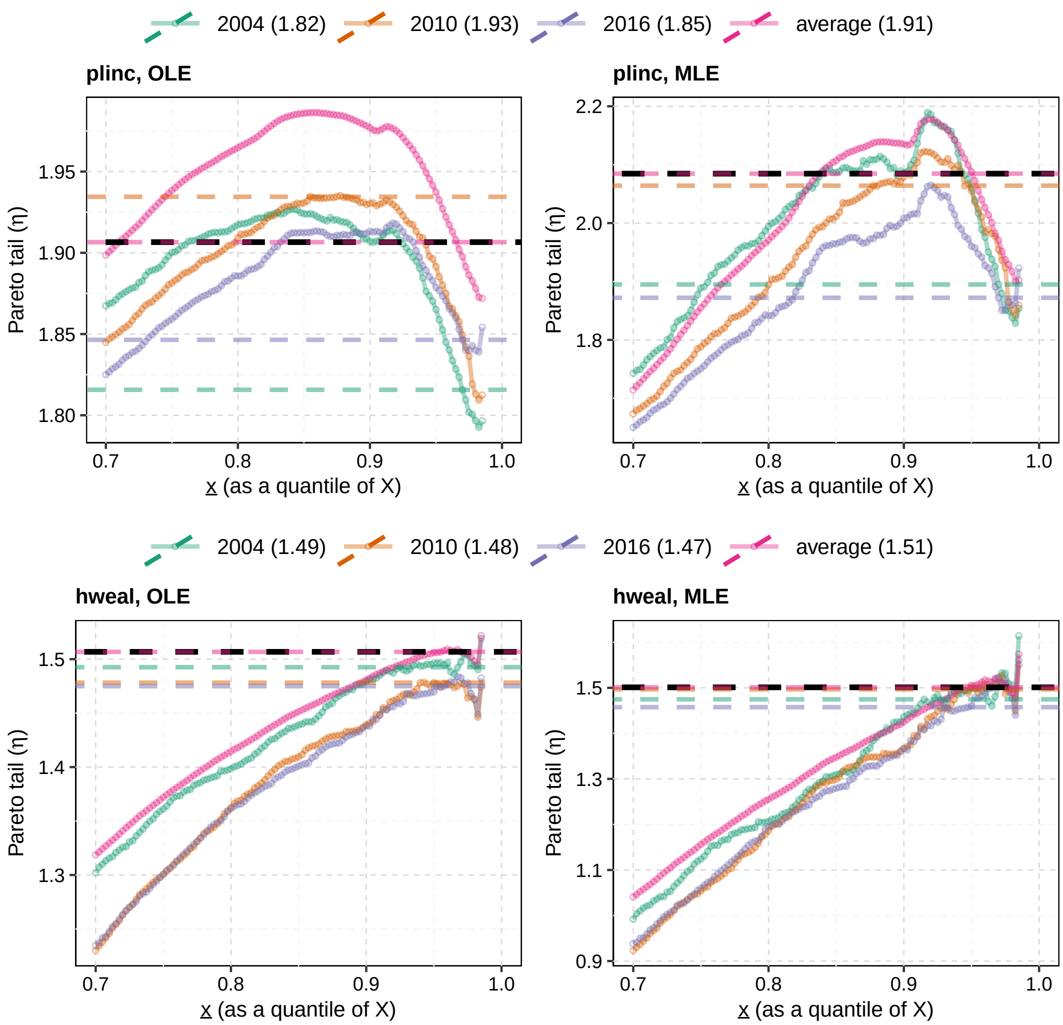


Figure 15. Real vs. financial wealth.

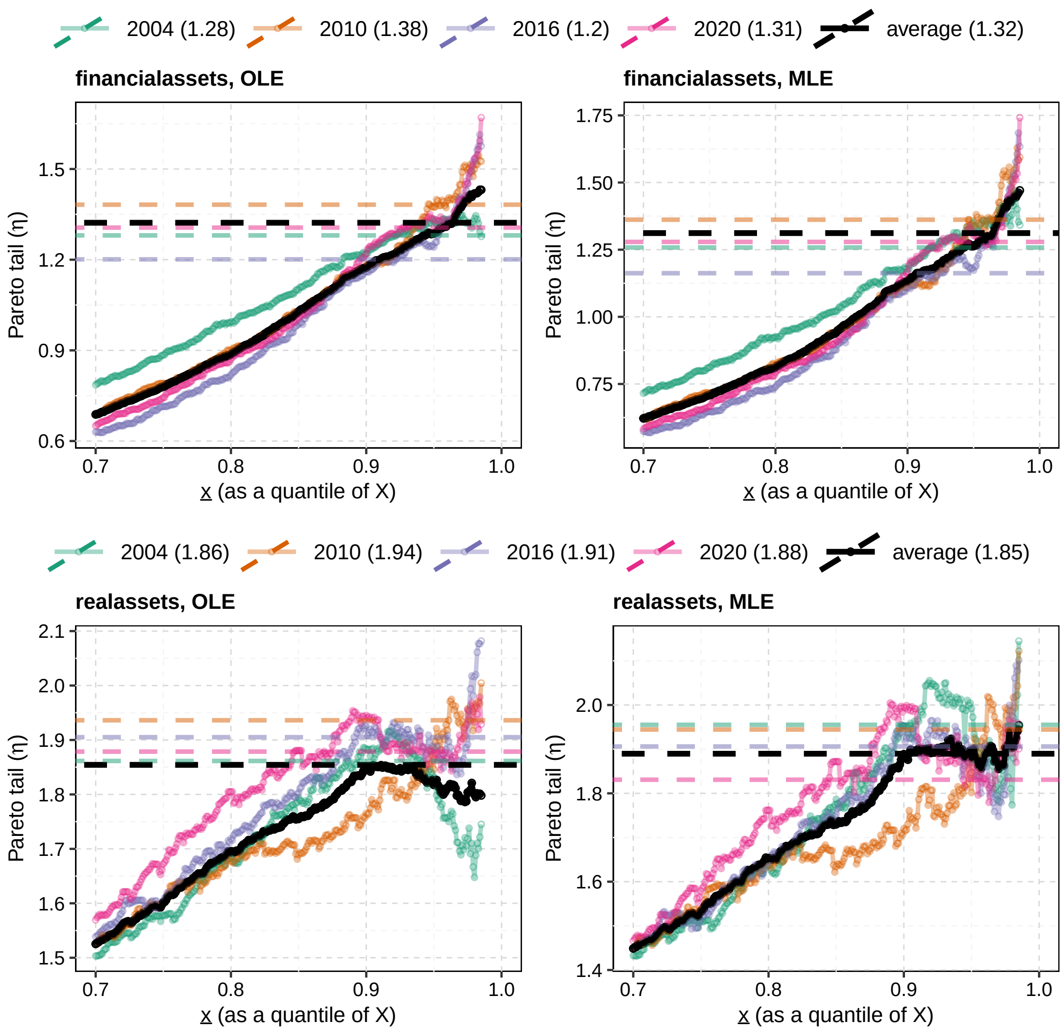


Figure 16. Capital income: log-log plots and Pareto tail coefficient estimates.

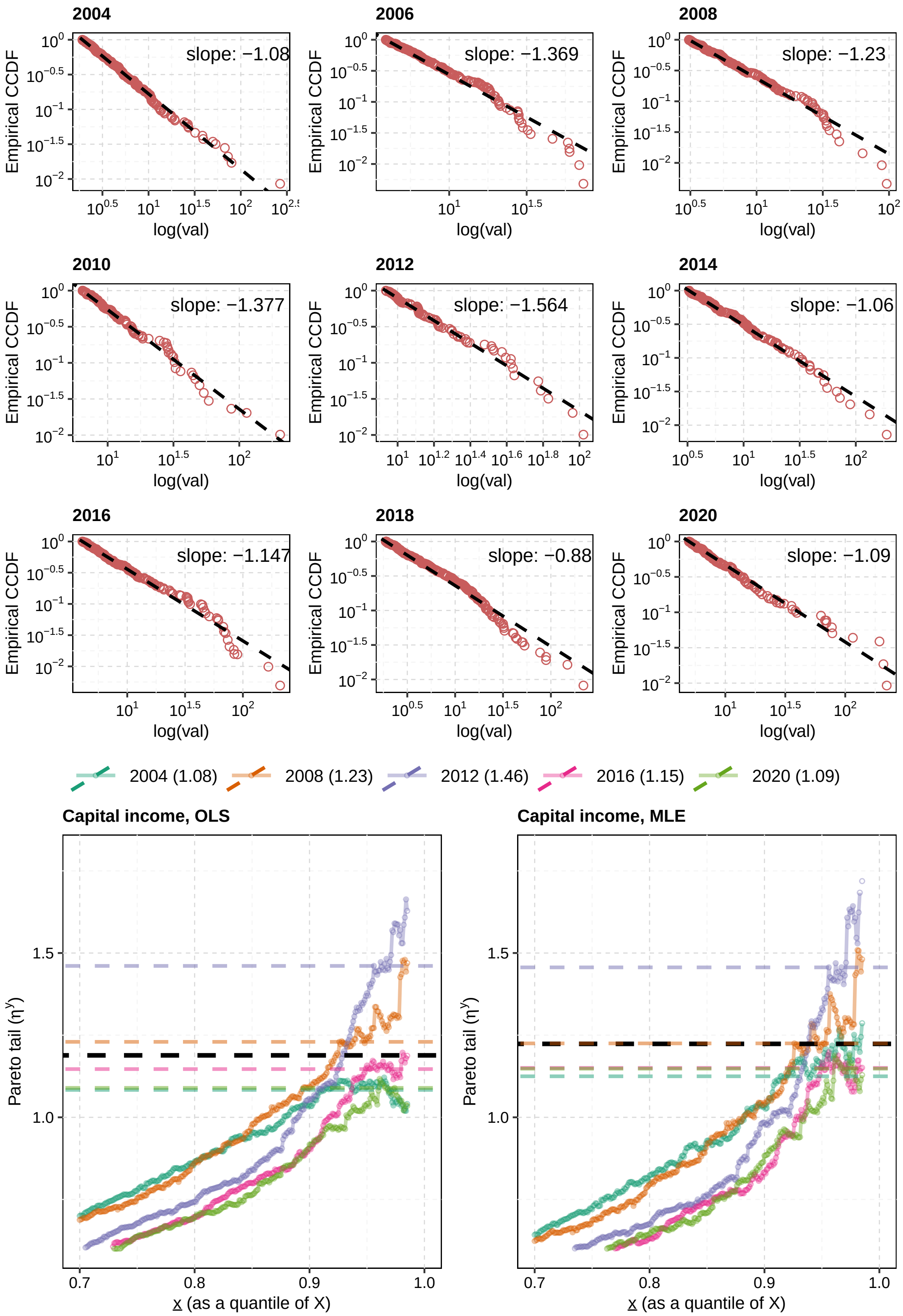


Figure 17. Capital income: data from Saez and Zucman (2016), excluding (top) or including (bottom) capital gains.

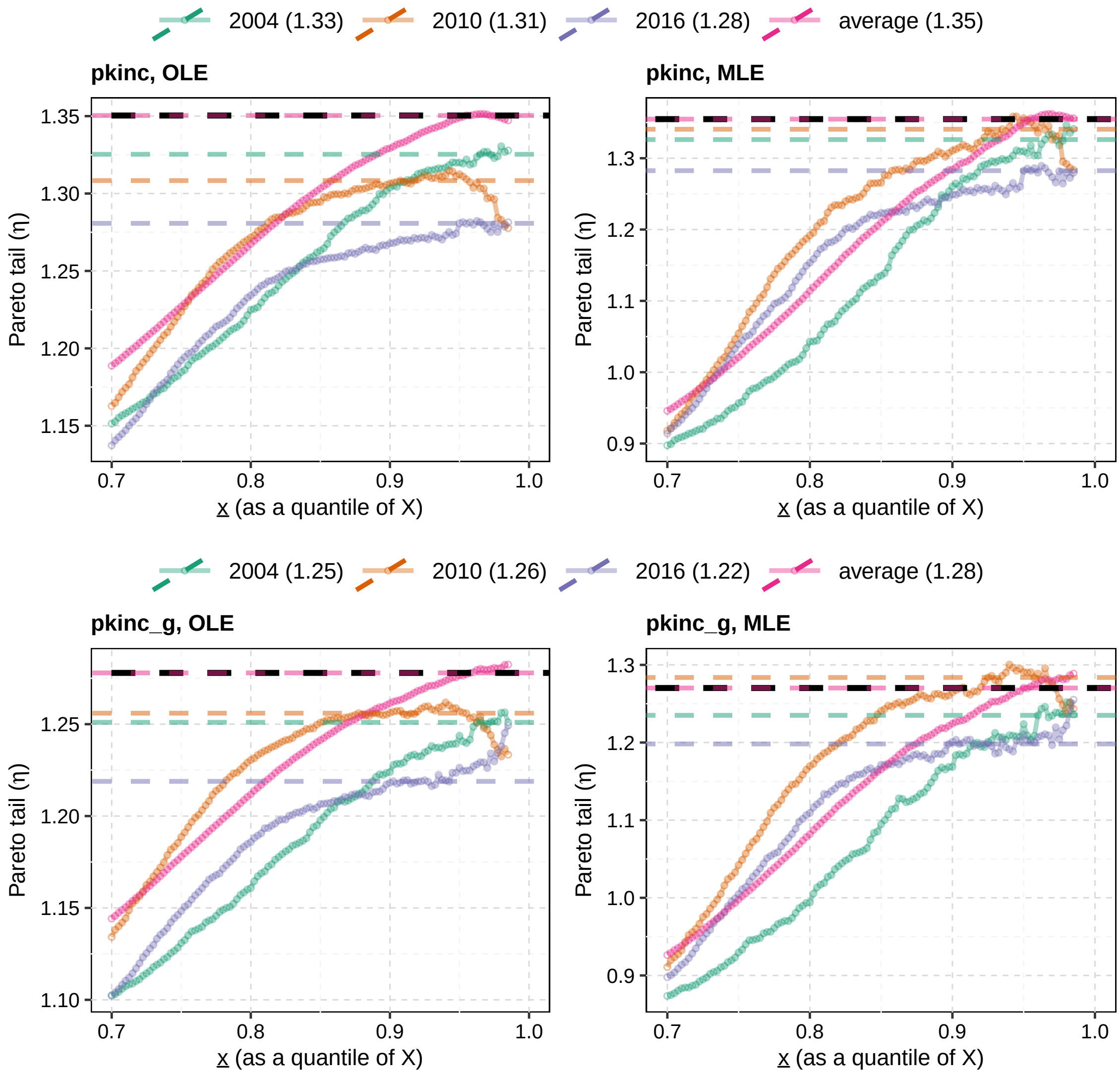
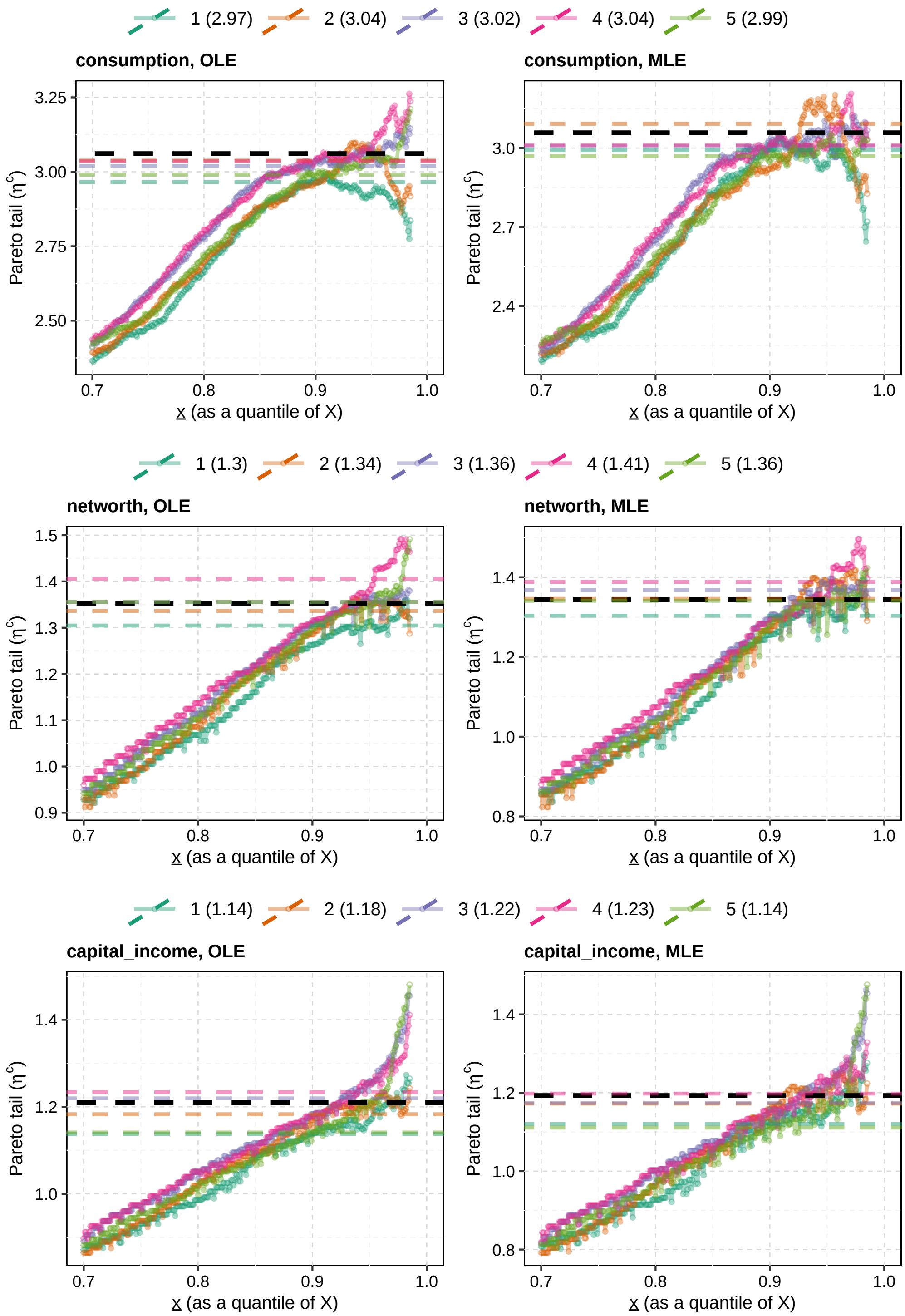


Figure 18. Pareto tail coefficient estimates generated by the quantitative model.



A.1 Under-Reporting and Under-Representation Adjustments

Table 5. Correspondence Table – PSID consumption and NIPA[a].

PSID subcategory	Corresponding NIPA subcategories[b]	Ratio (2019)
Nondurable goods: food (excl. outside food), food stamp, gasoline	Food and beverages purchased for off-premises consumption (71), Gasoline and other energy goods (111)	0.83
Clothing	Clothing and footwear (102), Clothing and footwear services (309)	0.40
Entertainment: recreation, trips, food outside	Recreation services (207), Recreational goods and vehicles (36), Recreational items (parts of 80, 92, and 93) (124), Magazines, newspapers, and stationery (part of 90) (140), Net expenditures abroad by U.S. residents (131) (143), Foreign travel by U.S. residents (129) (332), Food services and accommodations (232)	0.31
Transportation: car insurance, car repair, parking, bus fare, taxi fare, other transports	Transportation services (188), Net motor vehicle and other transportation insurance (116) (277)	0.63
Health: nurse, doctor, precription, health insurance	Net health insurance (112) (273), Health care (170) (excl. Hospitals (51) (181)), Pharmaceutical and other medical products (40 and 41) (119)	0.28
Furniture	Furnishings and durable household equipment	0.34
Education: tuition, other school expenditures, childcare	Education services (288), Educational books (96) (67)	0.78
Vehicle	Motor vehicles and parts (4)	0.94
Housing service: rent and imputed rent	Rental of tenant-occupied nonfarm housing (20) (152), Imputed rental of owner-occupied nonfarm housing (21) (158), Rental value of farm dwellings (22) (161), Group housing (23) (162)	0.88
Housing maintenance: water, heating, electric, property insurance, misc. utilities, home repair	Household utilities (163), Household supplies (parts of 32 and 36) (129), Communication (279), Household maintenance (parts of 31, 33, and 36) (325)	1.06
Total consumption	Personal consumption expenditures (1)	0.47
Total consumption (excl. housing services)	Personal consumption expenditures (1) (excl. housing service categories)	0.53

[a] The ratio is computed as the ratio of the total expenditure in the survey relative to the NIPA counterpart.
[b] We use the U.S. Bureau of Economic Analysis, "Table 2.4.5U. Personal Consumption Expenditures by Type of Product".

A.2 Testing the Pareto Hypothesis

Table 6. How plausible is the power-law hypothesis?

Data	Variable	Pareto rejected[a]		Pareto > $\mathcal{LN}$[b]		Pareto > Exp[b]	
		OLS	MLE	OLS	MLE	OLS	MLE
	Wealth	2/9	2/9	9—9/9	9—9/9	9—9/9	9—9/9
	Labor Income	6/9	5/9	9—9/9	9—9/9	9—9/9	9—9/9
PSID	Capital income	2/9	3/9	9—9/9	9—9/9	9—9/9	9—9/9
	Total income	3/9	2/9	9—9/9	9—9/9	9—9/9	9—9/9
	Consumption	1/9	1/9	9—6/9	9—6/9	9—9/9	9—9/9

[a] The power-law hypothesis is rejected with a p-value smaller than 0.1. The tests are performed with 500 bootstrap draws for each year.
[b] We present the statistical findings as $X|Y/Z$, where X is the number of occurrences with $R > 0$, Y counts the times where R is statistically different from zero, and Z is the total number of periods.

A.3 Non-linear Least Squares

Table 7. Mean KS-optimal Pareto coefficient: non-linear least squares.

Data	Consumption			Wealth			Labor income			Capital income		
	OLS	MLE	NLS	OLS	MLE	NLS	OLS	MLE	NLS	OLS	MLE	NLS
PSID	3.11	3.06	3.00	1.38	1.37	1.32	2.23	2.30	2.37	1.17	1.15	1.12

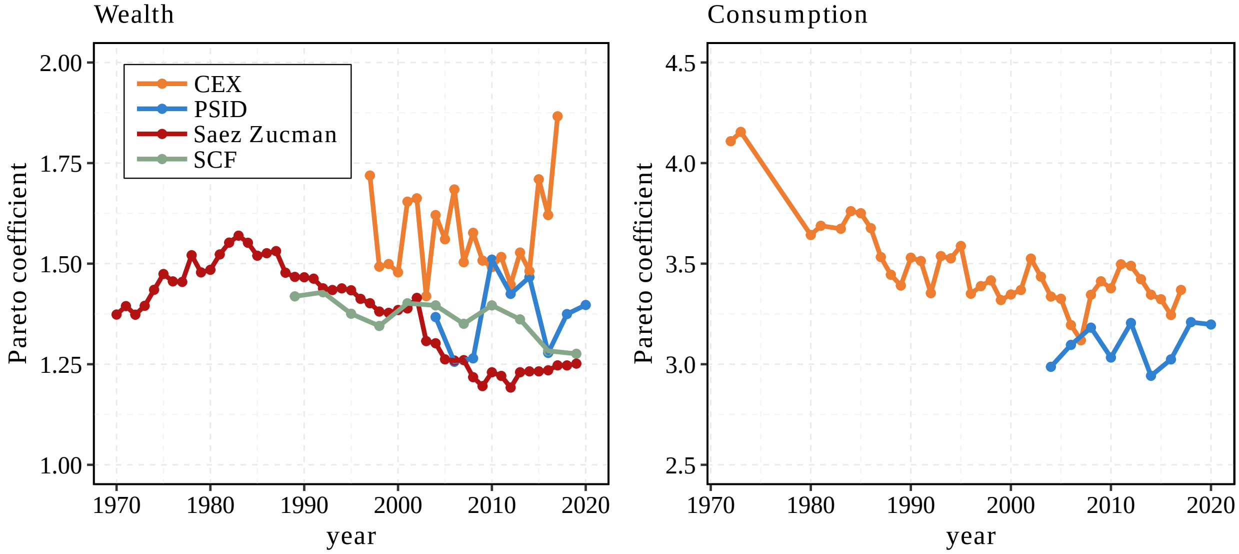
A.4 Pareto Tails along the Life-Cycle

Table 8. Pareto tails over the life-cycle in the PSID.

Head's age group	Mean KS-Optimal Pareto coefficient							
	Consumption		Wealth		Labor income		Capital income	
	OLS	MLE	OLS	MLE	OLS	MLE	OLS	MLE
All population (age $\in$ [19-80])	3.11	3.06	1.38	1.37	2.23	2.30	1.17	1.15
Age $\in$ [19-45]	3.35	3.30	1.26	1.25	2.35	2.41	0.88	0.89
Age $\in$]45-80]	2.96	2.97	1.40	1.41	2.17	2.27	1.26	1.26

A.5 Dynamics of the Pareto Tails

Figure 19. Evolution of Pareto tails over time.



A.6 Using Administrative Data to Measure Consumption

An alternative approach to measuring consumption is based on administrative panel data that are often advocated as being more accurate than survey data. The idea consists of backing out consumption as the residual from the budget constraint, if measures of wealth, labor income, and capital income are available—the main challenge being to evaluate the change in the valuation of assets to infer the change in the value of the stock. For the question we are interested in, however, this approach may not be as compelling as one might hope. To see this, suppose that the data reports labor and capital income correctly (call their sum $y_t + r_t a_t$), but that wealth is systematically under-reported by a factor $1/(1+\theta) < 1$ due to, e.g., tax evasion and offshoring. That is, for reported wealth $\hat{a}_t$, true wealth is $a_t = (1+\theta)\hat{a}_t$. Consumption is measured as $\hat{c}_t = y_t + r_t a_t - (\hat{a}_t - \hat{a}_{t-1})$, i.e., as total income net of the change in wealth. Simple algebra then shows that $\hat{c}_t/\hat{a}_t = c_t/a_t + \theta(y_t + r_t a_t)/a_t$. In other words, if there is systematic under-reporting of wealth, then there will be a systematic upwards bias in estimated consumption-to-wealth ratios from administrative data. Furthermore, if the income-to-wealth ratio does not converge to zero at the top (i.e., if the average return on wealth is strictly positive), then we would necessarily conclude that consumption-to-wealth ratios are bounded away from zero. Hence, we would mechanically conclude that measured consumption must have the same Pareto coefficient as measured wealth (or as capital income if returns are scale-dependent). That is, we would never be able to correctly detect the true Pareto tail for consumption.

B Theoretical Appendix

B.1 Proofs and Technical Details

Proof of Proposition 2. This proof, which follows Achdou et al. (2022), nests the derivation of the policy function given in Proposition 1. We proceed in two steps. First, we derive the solution to the HJB equation without labor income. Second, we show that this solution continues to characterize the asymptotic optimum in the general model with labor income because labor income becomes irrelevant relative to wealth as $a \to \infty$.

1. Model without labor income

Suppose that investment productivity follows a general Markov chain with transition matrix

$$\boldsymbol{Q} \cdot R\left(z_{it}, a_{it}\right), \quad \text{where} \quad \boldsymbol{Q} = \left(q_{nm}\right)_{1 \leq n,m \leq N} \quad \text{with} \quad q_{nn} = -\sum_{m \neq n} q_{nm}.$$

For any $n \in \{1, \ldots, N\}$, the parameter $q_{nn} \in \mathbb{R}_-$ gives (minus) the holding time in state z_n. That is, the probability that agent i's investment productivity jumps from z_n to another state at time t follows an exponential distribution with intensity $-q_{nn} > 0$. For any $m \neq n$, the probability that agent i's investment productivity jumps from z_n to z_m at time t follows an exponential distribution with intensity $q_{nm} > 0$.

The FOC for consumption in (10) reads $c\left(z_n, a\right) = \left(V_a\left(z_n, a\right)\right)^{-1/\gamma}$, where V_a denotes the derivative of the value function with respect to wealth a. The HJB equation in state z_n for $n \in \{1, \ldots, N\}$ can thus be rewritten as

$$\begin{aligned}
\left(\xi_0 + \xi_1 + \xi_2 R\left(z_n, a\right)\right) V\left(z_n, a\right) = & \frac{\gamma}{1-\gamma}\left(V_a\left(z_n, a\right)\right)^{\frac{\gamma-1}{\gamma}} + \frac{\kappa}{1-\nu} a^{1-\nu} \\
& + \sum_{m \neq n} q_{nm} R\left(z_n, a\right)\left[V\left(z_m, a\right) - V\left(z_n, a\right)\right] \\
& + \mu_r R\left(z_n, a\right) a V_a\left(z_n, a\right) + \frac{1}{2}\sigma_r^2 R(z_n, a) a^2 V_{aa}\left(z_n, a\right).
\end{aligned}$$

We guess and verify that the solution to this problem behaves asymptotically as

$$V\left(z_n, a\right) = \frac{C_n}{1-B} a^{1-B},$$

for some constant $C_n > 0$ and $B > 0$. Substituting this guess into the HJB leads to

$$\begin{aligned}
&\frac{C_n}{1-B}\left[\left(\xi_0+\xi_1+\xi_2 z_n\right)a^{1-B}+\xi_2 z_n\psi a^{1-B+\eta}\right] \\
&=\frac{\gamma}{1-\gamma}\left(C_n\right)^{\frac{\gamma-1}{\gamma}}a^{-B\frac{(\gamma-1)}{\gamma}}+\frac{\kappa}{1-\nu}a^{1-\nu}+\sum_{m\neq n}q_{nm}z_n\left(C_m-C_n\right)\frac{1}{1-B}(a^{1-B}+\psi a^{1-B+\eta}) \\
&\quad+\mu_r z_n C_n(a^{1-B}+\psi a^{1-B+\eta})-\frac{1}{2}\sigma_r^2 z_n C_n B(a^{1-B}+\psi a^{1-B+\eta}),
\end{aligned}$$

which can finally be rearranged as

$$\begin{aligned}
&\left[\left(\xi_0+\xi_1+\xi_2 z_n\right)C_n-(1-B)\left(\mu_r-\frac{1}{2}\sigma_r^2 B\right)z_n C_n-\sum_{m\neq n}q_{nm}z_n\left(C_m-C_n\right)\right]\frac{a^{1-B}}{1-B} \\
&+\left[\xi_2 z_n C_n-(1-B)\left(\mu_r-\frac{1}{2}\sigma_r^2 B\right)z_n C_n-\sum_{m\neq n}q_{nm}z_n\left(C_m-C_n\right)\right]\psi\frac{a^{1-B+\eta}}{1-B} \\
&=\frac{\gamma}{1-\gamma}\left(C_n\right)^{\frac{\gamma-1}{\gamma}}a^{-B\frac{(\gamma-1)}{\gamma}}+\frac{\kappa}{1-\nu}a^{1-\nu}. \qquad (16)
\end{aligned}$$

Recall that we impose $\kappa \geq 0$, $\gamma > 1$, $\nu \leq \gamma$, and $\nu \notin [1-\eta, 1]$ throughout the analysis. To solve for C_n, we consider several cases in turn.

(1a) *Homothetic preferences and scale-independent returns.* First, we consider the benchmark case of homothetic preferences, with either $\kappa = 0$ or $\nu = \gamma$, and scale-independent returns $\eta = 0$. In this case, we guess that $B = \gamma$, so that $\frac{1-\gamma}{\gamma}B = 1-\gamma$. Thus, all the terms in equation (16) are proportional to $a^{1-\gamma}$. Therefore, we must have

$$\left[\tfrac{\xi_0+\xi_1}{(1+\psi)z_n}+\xi_2-q_{nn}-(1-\gamma)\left(\mu_r-\tfrac{1}{2}\sigma_r^2\gamma\right)\right]C_n-\tfrac{\gamma}{(1+\psi)z_n}\left(C_n\right)^{\frac{\gamma-1}{\gamma}}-\sum_{m\neq n}q_{nm}C_m=\tfrac{\kappa}{(1+\psi)z_n}.$$

Consider in particular the case $N = 2$. This system of equations reads:

$$\left[\tfrac{\xi_0+\xi_1}{(1+\psi)z_L}+\xi_2+q_{LH}+(\gamma-1)\left(\mu_r-\tfrac{1}{2}\sigma_r^2\gamma\right)\right]C_L-\tfrac{\gamma}{(1+\psi)z_L}\left(C_L\right)^{\frac{\gamma-1}{\gamma}}-q_{LH}C_H=\tfrac{\kappa}{(1+\psi)z_L}$$

and

$$\left[\tfrac{\xi_0+\xi_1}{(1+\psi)z_H}+\xi_2+q_{HL}-(1-\gamma)\left(\mu_r-\tfrac{1}{2}\sigma_r^2\gamma\right)\right]C_H-\tfrac{\gamma}{(1+\psi)z_H}\left(C_H\right)^{\frac{\gamma-1}{\gamma}}-q_{HL}C_L=\tfrac{\kappa}{(1+\psi)z_H}.$$

Subtracting one equation from the other leads to

$$\begin{aligned}
&\left[\frac{\xi_0+\xi_1}{1+\psi}+\left\{\xi_2+q_{LH}+q_{HL}+(\gamma-1)\left(\mu_r-\frac{1}{2}\sigma_r^2\gamma\right)\right\}z_L\right]C_L-\frac{\gamma}{1+\psi}\left(C_L\right)^{\frac{\gamma-1}{\gamma}}\\
=&\left[\frac{\xi_0+\xi_1}{1+\psi}+\left\{\xi_2+q_{LH}+q_{HL}-(1-\gamma)\left(\mu_r-\frac{1}{2}\sigma_r^2\gamma\right)\right\}z_H\right]C_H-\frac{\gamma}{1+\psi}\left(C_H\right)^{\frac{\gamma-1}{\gamma}}.
\end{aligned}$$

Define

$$h\left(C,z\right)=\left(\alpha_0+\alpha_1 z\right)C-\frac{\gamma}{1+\psi}C^{\frac{\gamma-1}{\gamma}},$$

where $\alpha_0\equiv\left(\xi_0+\xi_1\right)/\left(1+\psi\right)$ and $\alpha_1\equiv\xi_2+q_{LH}+q_{HL}+\left(\gamma-1\right)\left(\mu_r-\sigma_r^2\gamma/2\right)$ are positive constants under the assumptions of Proposition 2. For a fixed z, the function $C\mapsto h\left(C,z\right)$ is U-shaped and crosses each positive value $A>0$ once. That is, the equation $h\left(C,z\right)=A$ has a unique solution $C\left(A,z\right)$ as long as $A>0$. Moreover, $C\left(A,z_L\right)>C\left(A,z_H\right)>0$ whenever $z_L<z_H$. Ignoring the trivial solution $C\left(0,z\right)=0$, we have

$$C\left(0,z\right)\;=\;\left[\frac{\gamma}{\left(1+\psi\right)\left(\alpha_0+\alpha_1 z\right)}\right]^{\gamma}.$$

Finally, the function $A\mapsto C\left(A,z\right)$ defined on $\mathbb{R}_+$ is continuous, strictly increasing, and unbounded in A.

Now, using the first equation of the original system, a solution $\left(C_L,C_H\right)\in\mathbb{R}_+^2$ must also satisfy

$$\begin{aligned}
&\left[\frac{\xi_0+\xi_1}{1+\psi}+\left\{\xi_2+q_{LH}+q_{HL}+(\gamma-1)\left(\mu_r-\frac{1}{2}\sigma_r^2\gamma\right)\right\}z_L\right]C_L-\frac{\gamma}{1+\psi}\left(C_L\right)^{\frac{\gamma-1}{\gamma}}\\
=&\frac{\kappa}{1+\psi}+\left(q_{HL}C_L+q_{LH}C_H\right)z_L\equiv A.
\end{aligned}$$

We must therefore verify that there exists $A>0$ such that

$$g\left(A\right)\;\equiv\;q_{HL}z_L C\left(A,z_L\right)+q_{LH}z_L C\left(A,z_H\right)-A+\frac{\kappa}{1+\psi}=0,$$

where $C\left(A,z_L\right),C\left(A,z_H\right)$ are as described above. Now, note first that $g\left(0\right)>0$ since $C\left(0,z_L\right),C\left(0,z_H\right)>0$, and $\kappa\geq 0$. Since g is continuous, we must therefore show that there exists $A>0$ such that $g\left(A\right)<0$, and that $g(\cdot)$ crosses 0 only once. To see this, apply the implicit function theorem to compute $\partial C(A,z)/\partial A$. Recall that

$$(\alpha_0+\alpha_1 z)\,C(A,z)-\frac{\gamma}{1+\psi}\left(C(A,z)\right)^{\frac{\gamma-1}{\gamma}}=A$$

so that

$$\frac{\partial C(A,z)}{\partial A}=\frac{1}{(\alpha_0+\alpha_1 z)-\frac{\gamma-1}{1+\psi}\left(C(A,z)\right)^{-\frac{1}{\gamma}}}\ .$$

We thus obtain

$$g'(A)=\frac{q_{HL}z_L}{(\alpha_0+\alpha_1 z_L)-\frac{\gamma-1}{1+\psi}\left(C(A,z_L)\right)^{-\frac{1}{\gamma}}}+\frac{q_{LH}z_L}{(\alpha_0+\alpha_1 z_H)-\frac{\gamma-1}{1+\psi}\left(C(A,z_H)\right)^{-\frac{1}{\gamma}}}-1\ .$$

Fix $\varepsilon>0$. As $A\to\infty$, we have $C(A,z_L),C(A,z_H)\to\infty$. Hence, there exists $\bar{A}(\varepsilon)>0$ such that, for any $A>\bar{A}(\varepsilon)$, $0<\frac{\gamma-1}{1+\psi}\left(C(A,z_H)\right)^{-\frac{1}{\gamma}}<\varepsilon$ and $0<\frac{\gamma-1}{1+\psi}\left(C(A,z_L)\right)^{-\frac{1}{\gamma}}<\varepsilon$. It follows that

$$g'(A)<\frac{q_{HL}z_L}{\alpha_0+\alpha_1 z_L-\varepsilon}+\frac{q_{LH}z_L}{\alpha_0+\alpha_1 z_H-\varepsilon}-1\ .$$

Since $z_L<z_H$, we can then write

$$g'(A)<\frac{(q_{HL}+q_{LH})z_L}{\alpha_0+\alpha_1 z_L-\varepsilon}-1\ =\ \frac{1}{1+D}-1\ ,$$

where

$$D\equiv\frac{\frac{\xi_0+\xi_1}{1+\psi}+\left\{\xi_2+(\gamma-1)\left(\mu_r-\frac{1}{2}\sigma_r^2\gamma\right)\right\}z_L-\varepsilon}{(q_{HL}+q_{LH})\,z_L}. \tag{17}$$

Now, for ε small enough, we have $D>0$ and bounded away from 0. Hence $g'(A)<0$ and $g(A)<0$ for A large enough. Finally, note that if $g'(A^*)<0$ for some $A^*>0$, then $g'(A)<0$ for all $A>A^*$, since $C(A,z)$ is monotonically increasing. Therefore, there exists a unique solution to the system. This concludes the proof.

(1b) *Homothetic preferences and scale-dependent returns.* Consider first the case of wealth-independent preferences, $\kappa=0$, and suppose that $\eta>0$. In this case, we guess that $B=(1+\eta)\gamma$, so that the dominant terms in the HJB equation (16) as $a\to\infty$ are those in $1-B+\eta=-B(\gamma-1)/\gamma=(1+\eta)(1-\gamma)$. We must therefore have

$$\left[\xi_2+|q_{nn}|-(1-(1+\eta)\gamma)\left(\mu_r-\tfrac{1}{2}\sigma_r^2(1+\eta)\gamma\right)\right]C_n-\tfrac{((1+\eta)\gamma-1)\gamma}{(\gamma-1)\psi z_n}\left(C_n\right)^{\frac{\gamma-1}{\gamma}}-\textstyle\sum_{m\neq n}q_{nm}C_m=0.$$

The proof then follows a reasoning analogous to that of Case (1a). Consider next the case of wealth-dependent preferences, $\kappa > 0$, with $\nu = \gamma$. In this case, we guess that $B = \gamma + \eta$, so that $1 - B + \eta = 1 - \gamma > -\frac{\gamma-1}{\gamma}B$, where the last inequality is equivalent to $\eta > 0$. Hence, the dominant term in the right-hand side of equation (16) as $a \to \infty$ is the term in $a^{1-\gamma}$. We must therefore have

$$\left[\xi_2 - (1-\gamma-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\gamma+\eta)\right) - q_{nn}\right] z_n C_n - \sum_{m \neq n} q_{nm} z_n C_m = \frac{(1-\gamma-\eta)\kappa}{(1-\gamma)\psi}$$

for all $n \in \{1, \ldots, N\}$. The proof then follows a reasoning identical to that of Case (1d) below, with $\nu = \gamma$.

(1c) *Non-homothetic preferences and scale-independent returns.* Second, consider the case $\nu < \gamma$ and $\eta = 0$. We guess that $B = \nu$. We then have $\frac{1-\gamma}{\gamma}B < 1 - \nu$, since this inequality is equivalent to $\nu < \gamma$. Hence, the dominant term in the right-hand side of equation (16) as $a \to \infty$ is the term in $a^{1-\nu}$. We must therefore have

$$\left[\frac{\xi_0 + \xi_1}{1+\psi} + \left\{\xi_2 - q_{nn} - (1-\nu)\left(\mu_r - \frac{1}{2}\sigma_r^2 \nu\right)\right\} z_n\right] C_n - \sum_{m \neq n} q_{nm} z_n C_m = \frac{\kappa}{1+\psi},$$

for all $n \in \{1, \ldots, N\}$. This is indeed a solution to the HJB equation as long as all of the constants C_n that solve this linear system are positive. In particular, in the case of $N = 2$, we obtain

$$C_L = \frac{\frac{\xi_0+\xi_1}{1+\psi} + \left[\xi_2 - (1-\nu)\left(\mu_r - \nu\sigma_r^2/2\right)\right] z_H + q_{LH} z_L + q_{HL} z_H}{\left[\frac{\xi_0+\xi_1}{1+\psi} + \left\{\xi_2 - (1-\nu)\left(\mu_r - \nu\sigma_r^2/2\right)\right\} z_L + q_{LH} z_L\right]\left[\frac{\xi_0+\xi_1}{1+\psi} + \left\{\xi_2 - (1-\nu)\left(\mu_r - \nu\sigma_r^2/2\right)\right\} z_H + q_{HL} z_H\right] - q_{LH} z_L q_{HL} z_H} \frac{\kappa}{1+\psi},$$

$$C_H = \frac{\frac{\xi_0+\xi_1}{1+\psi} + \left[\xi_2 - (1-\nu)\left(\mu_r - \nu\sigma_r^2/2\right)\right] z_L + q_{LH} z_L + q_{HL} z_H}{\left[\frac{\xi_0+\xi_1}{1+\psi} + \left\{\xi_2 - (1-\nu)\left(\mu_r - \nu\sigma_r^2/2\right)\right\} z_L + q_{LH} z_L\right]\left[\frac{\xi_0+\xi_1}{1+\psi} + \left\{\xi_2 - (1-\nu)\left(\mu_r - \nu\sigma_r^2/2\right)\right\} z_H + q_{HL} z_H\right] - q_{LH} z_L q_{HL} z_H} \frac{\kappa}{1+\psi}.$$

It is straightforward to see that a sufficient condition for C_L and C_H to be strictly positive is $\xi_2 > (1-\nu)\left(\mu_r - \nu\sigma_r^2/2\right)$.

(1d) *Non-homothetic preferences and scale-dependent returns.* Third, we consider the case of non-homothetic preferences, $\nu < \gamma$, and scale-dependent returns, $\eta > 0$. In this case, the dominant terms in the left-hand side of equation (16) as $a \to \infty$ are those in $a^{1-B+\eta}$. We guess that $B = \nu + \eta$, so that $1 - B + \eta = 1 - \nu > -\frac{\gamma-1}{\gamma}B$, where the last inequality is equivalent to $\nu < \gamma + (\gamma - 1)\eta$ and is thus satisfied as $\gamma > 1$. Thus, the dominant term in the right-hand side of equation (16) as $a \to \infty$ is the term in $a^{1-\nu}$.

We must therefore have

$$\left[\xi_2 - (1-\nu-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\nu+\eta)\right) - q_{nn}\right] z_n C_n - \sum_{m\neq n} q_{nm} z_n C_m = \frac{(1-\nu-\eta)\kappa}{(1-\nu)\,\psi}$$

for all $n \in \{1, \ldots, N\}$. This is indeed a solution to the HJB equation as long as all of the constants C_n are positive. In particular, in the case of $N = 2$, we obtain

$$\begin{aligned}
z_L z_H C_L &= \frac{1-\nu-\eta}{1-\nu}\frac{\kappa}{\psi}\frac{\left[\xi_2 - (1-\nu-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\nu+\eta)\right)\right] z_H + q_{LH} z_L + q_{HL} z_H}{\left[\xi_2 - (1-\nu-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\nu+\eta)\right) + q_{LH}\right]\left[\xi_2 - (1-\nu-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\nu+\eta)\right) + q_{HL}\right] - q_{LH}q_{HL}}, \\
z_L z_H C_H &= \frac{1-\nu-\eta}{1-\nu}\frac{\kappa}{\psi}\frac{\left[\xi_2 - (1-\nu-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\nu+\eta)\right)\right] z_L + q_{LH} z_L + q_{HL} z_H}{\left[\xi_2 - (1-\nu-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\nu+\eta)\right) + q_{LH}\right]\left[\xi_2 - (1-\nu-\eta)\left(\mu_r - \frac{1}{2}\sigma_r^2(\nu+\eta)\right) + q_{HL}\right] - q_{LH}q_{HL}}.
\end{aligned}$$

Note first that $(1-\nu-\eta)/(1-\nu) > 0$, since we assumed that either $\nu > 1$ or $\nu < 1-\eta$. We then easily obtain that C_L and C_H are strictly positive whenever $\xi_2 > (1-\nu-\eta)\left(\mu_r - (\nu+\eta)\,\sigma_r^2/2\right)$. Moreover, this condition also implies $C_L > C_H$.

2. Model with labor income

For any $\lambda > 0$, define an auxiliary value function V_λ by

$$V(y, z_n, a) = \lambda^{1-\nu-\eta} V_\lambda\left(y, z_n, \frac{a}{\lambda}\right).$$

We then have the following equalities:

$$\begin{aligned}
\frac{\partial V}{\partial a}(y, z_n, a) &= \lambda^{-(\nu+\eta)}\frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right), \quad & \frac{\partial^2 V}{\partial a^2}(y, z_n, a) &= \lambda^{-(\nu+\eta)-1}\frac{\partial^2 V_\lambda}{\partial a^2}\left(y, z_n, \frac{a}{\lambda}\right), \\
\frac{\partial V}{\partial y}(y, z_n, a) &= \lambda^{1-\nu-\eta}\frac{\partial V_\lambda}{\partial y}\left(y, z_n, \frac{a}{\lambda}\right), \quad & \frac{\partial^2 V}{\partial y^2}(y, z_n, a) &= \lambda^{1-\nu-\eta}\frac{\partial^2 V_\lambda}{\partial y^2}\left(y, z_n, \frac{a}{\lambda}\right).
\end{aligned}$$

Let us define $H(x) \equiv \max_c \frac{1}{1-\gamma} c^{1-\gamma} - x \cdot c = \frac{\gamma}{1-\gamma} x^{\frac{\gamma-1}{\gamma}}$. We can then write

$$H\left(\frac{\partial V}{\partial a}(y, z_n, a)\right) = \lambda^{-(\nu+\eta)\frac{\gamma-1}{\gamma}} H\left(\frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right)\right).$$

Substituting into equation (10) and dividing through by $\lambda^{1-\nu}$, we obtain that V_λ satisfies the following HJB equation:

$$
\begin{aligned}
&\left[\frac{\xi_0+\xi_1+\xi_2 z_n}{\lambda^\eta}+\xi_2 z_n \psi\left(\frac{a}{\lambda}\right)^\eta\right] V_\lambda\left(y, z_n, \frac{a}{\lambda}\right)=\max_c \ \lambda^{-(1-\nu)} \frac{c^{1-\gamma}}{1-\gamma}-\lambda^{-(1+\eta)} c \frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right) \\
&+\kappa \frac{\left(\frac{a}{\lambda}\right)^{1-\nu}}{1-\nu}+\left[\frac{w y-T(w y)}{\lambda^{1+\eta}}+\mu_r z_n \frac{a}{\lambda} \frac{1}{\lambda^\eta}+\mu_r z_n \psi\left(\frac{a}{\lambda}\right)^{1+\eta}\right] \frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right) \\
&+\frac{1}{2} \sigma_r^2 z_n\left[\left(\frac{a}{\lambda}\right)^2 \frac{1}{\lambda^\eta}+\psi\left(\frac{a}{\lambda}\right)^{2+\eta}\right] \frac{\partial^2 V_\lambda}{\partial a^2}\left(y, z_n, \frac{a}{\lambda}\right)+\frac{\mu_y y}{\lambda^\eta} \frac{\partial V_\lambda}{\partial y}\left(y, z_n, \frac{a}{\lambda}\right)+\frac{1}{2} \frac{\sigma_y^2 y^2}{\lambda^\eta} \frac{\partial^2 V_\lambda}{\partial y^2}\left(y, z_n, \frac{a}{\lambda}\right) \\
&+\frac{q_{nm} z_n}{\lambda^\eta}\left[V_\lambda\left(y, z_m, \frac{a}{\lambda}\right)-V_\lambda\left(y, z_n, \frac{a}{\lambda}\right)\right]+q_{nm} z_n \psi\left(\frac{a}{\lambda}\right)^\eta\left[V_\lambda\left(y, z_m, \frac{a}{\lambda}\right)-V_\lambda\left(y, z_n, \frac{a}{\lambda}\right)\right].
\end{aligned}
$$

The FOC for consumption in this problem reads:

$$
c_\lambda\left(y, z_n, \frac{a}{\lambda}\right)=\lambda^{\frac{\nu+\eta}{\gamma}}\left(\frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right)\right)^{-\frac{1}{\gamma}}.
$$

Hence, the HJB equation can be rewritten as:

$$
\begin{aligned}
&\left[\frac{\xi_0+\xi_1+\xi_2 z_n}{\lambda^\eta}+\xi_2 z_n \psi\left(\frac{a}{\lambda}\right)^\eta\right] V_\lambda\left(y, z_n, \frac{a}{\lambda}\right)=\lambda^{\frac{\eta+\nu}{\gamma}-(1+\eta)} H\left(\frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right)\right) \\
&+\kappa \frac{\left(\frac{a}{\lambda}\right)^{1-\nu}}{1-\nu}+\left[\frac{w y-T(w y)}{\lambda^{1+\eta}}+\mu_r z_n \frac{a}{\lambda} \frac{1}{\lambda^\eta}+\mu_r z_n \psi\left(\frac{a}{\lambda}\right)^{1+\eta}\right] \frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right) \\
&+\frac{1}{2} \sigma_r^2 z_n\left[\left(\frac{a}{\lambda}\right)^2 \frac{1}{\lambda^\eta}+\psi\left(\frac{a}{\lambda}\right)^{2+\eta}\right] \frac{\partial^2 V_\lambda}{\partial a^2}\left(y, z_n, \frac{a}{\lambda}\right)+\frac{\mu_y y}{\lambda^\eta} \frac{\partial V_\lambda}{\partial y}\left(y, z_n, \frac{a}{\lambda}\right)+\frac{1}{2} \frac{\sigma_y^2 y^2}{\lambda^\eta} \frac{\partial^2 V_\lambda}{\partial y^2}\left(y, z_n, \frac{a}{\lambda}\right) \\
&+\frac{q_{nm} z_n}{\lambda^\eta}\left[V_\lambda\left(y, z_m, \frac{a}{\lambda}\right)-V_\lambda\left(y, z_n, \frac{a}{\lambda}\right)\right]+q_{nm} z_n \psi\left(\frac{a}{\lambda}\right)^\eta\left[V_\lambda\left(y, z_m, \frac{a}{\lambda}\right)-V_\lambda\left(y, z_n, \frac{a}{\lambda}\right)\right].
\end{aligned}
$$

The policy function of the original problem $c(y, z_n, a)$ satisfies:

$$
c(y, z_n, a)=\left(\frac{\partial V}{\partial a}(y, z_n, a)\right)^{-1/\gamma}=\lambda^{\frac{\nu+\eta}{\gamma}}\left(\frac{\partial V_\lambda}{\partial a}\left(y, z_n, \frac{a}{\lambda}\right)\right)^{-\frac{1}{\gamma}}=c_\lambda\left(y, z_n, \frac{a}{\lambda}\right).
$$

In particular, for $\lambda=a$, we get

$$
\frac{c(y, z_n, a)}{a^{(\nu+\eta)/\gamma}}=\left(\frac{\partial V_\lambda}{\partial a}(y, z_n, 1)\right)^{-\frac{1}{\gamma}}.
$$

Hence, letting $a \rightarrow \infty$ leads to

$$
\lim_{a \rightarrow \infty} \frac{c(y, z_n, a)}{a^{(\nu+\eta)/\gamma}}=\lim_{\lambda \rightarrow \infty}\left(\frac{\partial V_\lambda}{\partial a}(y, z_n, 1)\right)^{-\frac{1}{\gamma}}.
$$

We now show that as $\lambda \rightarrow \infty$, $(\partial V_\lambda(y, z_n, 1)/\partial a)^{-1/\gamma}$ converges to a constant that is independent of labor productivity y. Letting $\lambda \to \infty$ in the previous HJB equation and noting that $(\eta + \nu) < (1 + \eta)\gamma$, we obtain that the solution V_λ to this problem is given by the function V that satisfies, for any $\tilde{a} \equiv a/\lambda$:

$$\xi_2 z_n \psi \tilde{a}^\eta V(z_n, \tilde{a}) = \frac{\kappa}{1-\nu}\tilde{a}^{1-\nu} + \mu_r z_n \psi \tilde{a}^{1+\eta} V_a(z_n, \tilde{a}) + \frac{1}{2}\sigma_r^2 z_n \psi \tilde{a}^{2+\eta} V_{aa}(z_n, \tilde{a})$$
$$+ q_{nm} z_n \psi \tilde{a}^\eta \left[V(z_m, \tilde{a}) - V(z_n, \tilde{a})\right].$$

We already saw that the solution to this equation is given by $V(z_n, \tilde{a}) = \frac{C_n}{1-\nu-\eta}\tilde{a}^{1-\nu-\eta}$ for a constant $C_n > 0$ derived above in the case $N = 2$. As a result, we obtain

$$\lim_{\lambda \to \infty} \left(\frac{\partial V_\lambda}{\partial a}(y, z_n, 1)\right)^{-\frac{1}{\gamma}} = (C_n)^{-\frac{1}{\gamma}}.$$

This concludes the proof. $\square$

Proof of Proposition 3.

<u>1. Pareto Coefficient of the Labor Income Tail</u>

The law of motion of labor productivity is exogenous and independent of wealth. The stationary Kolmogorov-forward equation (KFE) for its marginal distribution reads

$$0 = -\frac{\partial}{\partial y}\left[\mu_y y f_y(y)\right] + \frac{\partial^2}{\partial y^2}\left[\frac{1}{2}\sigma_y^2 y^2 f_y(y)\right] - \xi_1 f_y(y),$$

for all $y \neq y_0$. We guess and verify that $f_y(\cdot)$ is Pareto distributed on (y_0, ∞), i.e.,

$$f_y(y) = G y^{-\zeta_y - 1}$$

for some constants $G > 0$ and $\zeta_y > 1$. Substituting into the KFE leads to

$$0 = \frac{1}{2}\sigma_y \zeta_y^2 + \left(\mu_y - \frac{1}{2}\sigma_y^2\right)\zeta_y - \xi_1.$$

This is indeed a solution if

$$\zeta_y = -\left(\frac{\mu_y}{\sigma_y^2} - \frac{1}{2}\right) + \sqrt{\left(\frac{\mu_y}{\sigma_y^2} - \frac{1}{2}\right)^2 + \frac{2\xi_1}{\sigma_y^2}}.$$

Note that $\zeta_y > 1$ whenever $\xi_1 > \mu_y$. Since w is constant, the distribution of labor income is also Pareto with corresponding tail ζ_y. Moreover, the distribution of after-tax labor income then follows by a change of variable, i.e., $y^{net} = \frac{1-\tau_0}{1-\tau_L} y^{1-\tau_L}$ such that its corresponding Pareto tail is $\zeta_y^{net} = (1-\tau_L)\zeta_y$.

2. Pareto Coefficient of the Wealth Tail

The stationary joint distribution of labor productivity, investment productivity, and wealth, $f(y, z_n, a)$, if it exists, satisfies the following KFE:

$$\begin{aligned}
0 = &-\frac{\partial}{\partial a}\left[\big(wy - T(wy) + \mu_r R(z_n,a)a - c(y,z_n,a)\big) f(y,z_n,a)\right] \\
&+ \frac{1}{2}\sigma_r^2 \frac{\partial^2}{\partial a^2}\left[R(z_n,a)a^2 f(y,z_n,a)\right] - \frac{\partial}{\partial y}\left[\mu_y y f(y,z_n,a)\right] + \frac{1}{2}\sigma_y^2 \frac{\partial^2}{\partial y^2}\left[y^2 f(y,z_n,a)\right] \\
&+ \sum_{m\neq n} q_{mn} R(z_m,a) f(y,z_m,a) - \big(\xi_1 + \left[|q_{nn}| + \xi_2\right] R(z_n,a)\big) f(y,z_n,a).
\end{aligned}$$

We guess and verify that wealth is asymptotically Pareto distributed and independent of labor productivity and investment productivity as $a \to \infty$, that is,

$$f(y,z_n,a) \underset{a\to\infty}{\sim} f_y(y) F_n \, a^{-\zeta_a - 1}$$

for some constants $F_n > 0$ and $\zeta_a > 1$. Substituting this expression along with the asymptotic policy function $c(y,z_n,a) = C_n^{-1/\gamma} a^{(\nu+\eta)/\gamma}$ if $\kappa > 0$ into the KFE gives

$$\begin{aligned}
0 = &f_y(y) F_n \left\{ -\frac{\partial}{\partial a}\left[\mu_r z_n \left(a^{-\zeta_a} + \psi a^{-\zeta_a+\eta}\right) - C_n^{-1/\gamma} a^{\frac{\nu+\eta}{\gamma} - \zeta_a - 1}\right] + \frac{1}{2}\sigma_r^2 z_n \frac{\partial^2}{\partial a^2}\left[a^{-\zeta_a+1} + \psi a^{-\zeta_a+1+\eta}\right]\right\} \\
&+ \left\{ -\frac{\partial}{\partial y}\left[\mu_y y f_y(y)\right] + \frac{1}{2}\sigma_y^2 \frac{\partial^2}{\partial y^2}\left[y^2 f_y(y)\right] - \xi_1 f_y(y)\right\} F_n a^{-\zeta_a - 1} \\
&+ f_y(y)\left[\sum_{m\neq n} q_{mn} z_m F_m\right]\left(a^{-\zeta_a-1} + \psi a^{-\zeta_a-1+\eta}\right) - f_y(y) F_n \left(|q_{nn}| + \xi_2\right) z_n \left(a^{-\zeta_a-1} + \psi a^{-\zeta_a-1+\eta}\right).
\end{aligned}$$

Using the KFE for $f_y(y)$, dividing through by $z_n F_n a^{-\zeta_a+\eta-1}$ and letting $a \to \infty$ leads to

$$0 = -\left(-\zeta_a + \eta\right)\mu_r + \frac{1}{2}\sigma_r^2 \left(-\zeta_a + 1 + \eta\right)\left(-\zeta_a + \eta\right) - \left(|q_{nn}| + \xi_2\right) + \left[\sum_{m\neq n} q_{mn} \frac{z_m}{z_n}\frac{F_m}{F_n}\right].$$

where we used that $\nu + \eta < (1+\eta)\gamma$. Let us now define $\tilde{\zeta}_a \equiv \zeta_a - \eta$. The previous equation can be rewritten as

$$0 = \frac{1}{2}\sigma_r^2\tilde{\zeta}_a^2 + \left(\mu_r - \frac{1}{2}\sigma_r^2\right)\tilde{\zeta}_a + \left[\sum_{m \neq n} q_{mn}\frac{z_m}{z_n}\frac{F_m}{F_n} - |q_{nn}| - \xi_2\right]. \tag{18}$$

Let $\{F_n\}_{n=1,\ldots,N}$ be such that $\sum_{m \neq n} q_{mn} z_m F_m = z_n F_n \sum_{m \neq n} q_{nm} = z_n F_n |q_{nn}|$. For $N = 2$, this implies $F_H/F_L = (z_L q_{LH})/(z_H q_{HL})$. The previous equation then has the solution $\tilde{\zeta}_a = \xi_2/\mu_r$ if $\sigma_r = 0$, and

$$\tilde{\zeta}_a = -\left(\frac{\mu_r}{\sigma_r^2} - \frac{1}{2}\right) + \sqrt{\left(\frac{\mu_r}{\sigma_r^2} - \frac{1}{2}\right)^2 + \frac{2\xi_2}{\sigma_r^2}}$$

if $\sigma_r > 0$. It is straightforward to show that $\tilde{\zeta}_a > 1$ iff $\xi_2 > \mu_r$. Now consider the case $\kappa = 0$, and use instead the policy function $c(y, z_n, a) = C_n^{-1/\gamma} a^{1+\eta}$. We obtain, if $\eta > 0$, the same equation as (18) except that μ_r in the second term in the right-hand side is now replaced by $\mu_r - C_n^{-1/\gamma}/(z_n\psi)$. The case $\eta = 0$ can be treated analogously.

3. Pareto Coefficients of the Consumption and Capital Income Tails

Let $\pi(z_n)$ denotes the stationary mass of investment productivity z_n and $\pi(z_n|a)$ the conditional distribution at a given wealth level a. The derivations of the Pareto coefficients for consumption in the cases $\kappa = 0$ and $\kappa > 0$ follow similar steps, using the relevant policy functions; for simplicity we only focus on the case $\kappa > 0$ here. For any $C > 0$, we have

$$\begin{aligned}
\mathbb{P}(c \geq C) &= \int_0^\infty \left[\sum_{n=1}^N \pi(z_n|a)\,\mathbb{1}_{\{c_n(a) \geq C\}}\right] f(a)\,da \\
&= \int_0^\infty \left[\sum_{n=1}^N \pi(z_n|a)\,\mathbb{1}_{\{a \geq C_n^{1/(\nu+\eta)} C^{\gamma/(\nu+\eta)}\}}\right] f(a)\,da \\
&= \sum_{n=1}^N \int_{C_n^{\frac{1}{\nu+\eta}} C^{\frac{\gamma}{\nu+\eta}}}^\infty \pi(z_n|a)\, f(a)\,da \\
&= \sum_{n=1}^N \int_{C_n^{\frac{1}{\nu+\eta}} C^{\frac{\gamma}{\nu+\eta}}}^\infty \pi(z_n)\, f(a|z_n)\,da,
\end{aligned}$$

where the last equality uses Bayes' rule. Letting $C \to \infty$, one can rewrite this expression as

$$\begin{aligned} \mathbb{P}(c \geq C) &= \sum_{n=1}^{N} \pi(z_n) \int_{C_n^{\frac{1}{\nu+\eta}} C^{\frac{\gamma}{\nu+\eta}}}^{\infty} F_n a^{-\zeta_a - 1} da \\ &= \frac{1}{\zeta_a} \sum_{n=1}^{N} \pi(z_n) F_n \left[C_n^{\frac{1}{\nu+\eta}} C^{\frac{\gamma}{\nu+\eta}} \right]^{-\zeta_a} \\ &= \left[\frac{1}{\zeta_a} \sum_{n=1}^{N} \pi(z_n) F_n C_n^{-\zeta_a \frac{1}{\nu+\eta}} \right] C^{-\zeta_a \frac{\gamma}{\nu+\eta}}. \end{aligned}$$

Similarly, the distribution of pre-tax capital income satisfies, for any $R > 0$,

$$\mathbb{P}(r \geq R) = \int_0^{\infty} \left[\sum_{n=1}^{N} \pi(z_n | a) \mathbb{1}_{\{a + \psi a^{1+\eta} \geq R/(r z_n)\}} \right] f(a)\, da.$$

As $a \to \infty$, this expression can be rewritten as

$$\begin{aligned} \mathbb{P}(r \geq R) &= \int_0^{\infty} \sum_{n=1}^{N} \left[\pi(z_n | a) \mathbb{1}_{\{a \geq [R/(\psi r z_n)]^{1/(1+\eta)}\}} \right] f(a)\, da \\ &= \sum_{n=1}^{N} \int_{[R/(\psi r z_n)]^{1/(1+\eta)}}^{\infty} \pi(z_n) f(a | z_n)\, da \\ &= \sum_{n=1}^{N} \int_{[R/(\psi r z_n)]^{1/(1+\eta)}}^{\infty} \pi(z_n) F_n a^{-\zeta_a - 1} da \\ &= \left[\frac{1}{\zeta_a} \sum_{n=1}^{N} \pi(z_n) F_n (\psi r z_n)^{\zeta_a/(1+\eta)} \right] R^{-\zeta_a \frac{1}{1+\eta}}, \end{aligned}$$

which concludes the proof. □

B.2 Simulations Capital Supply Elasticity

In this section, we use the analytical model to simulate the time path of the average capital supply elasticity to an unanticipated capital tax reform at time $s = 0$ under four model variants: (a) homothetic, wealth-independent preferences, (b) non-homothetic preferences, (c) homothetic, wealth dependent preferences and scale-dependent returns, and (d) non-homohtetic preferences and scale-dependent returns. Specifically, we draw a sample of 100000 observations from a Pareto distribution for wealth that starts at the scale $\underline{a}$ of one million. We discretize the continuous time stochastic differential equation and simulate the economy for 20 years over 10000 time increments. For parsimony reasons, we abstract from type-dependent ex-ante return heterogeneity. The calibration across different model variants is shown in Table 9.

Table 9. Capital supply elasticity: calibrated parameters for various models.

Environment	Calibration				
	γ	κ	ν	ψ	η
a. Homothetic Preferences	2	0	0	0.0025	0
b. Non-Homothetic Preferences	2	10	0.93	0.0025	0
c. WIU Homothetic Preferences + Scale	2	10	2.00	0.0025	0.22
d. Non-Homothetic Preferences + Scale	2	10	0.72	0.0025	0.22

Across all model variants, we assume a standard time discount rate of $\xi_0 = 0.04$ and an inverse EIS of $\gamma = 2$. Regarding the labor productivity process, we assume that $w = 1$, $\sigma_y^2 = 0.02$, and $\mu_y = 0.0001$. To match a labor income tail of $\hat{\zeta}_y = 2.2$, we choose an implied death rate of $\xi_1 = 0.0266$. Moreover, we assume that the volatility of ex-post idiosyncratic investment risk is $\sigma_r = 0.175$ and the scale-dependent intercept equals $\psi = 0.0025$. For a baseline return of $\mu_r = 0.06$, the latter implies an additional 0.028 mean return difference between households with wealth close to the scale lower bound and the richest ones in the sample. For all model variants, we choose the expropriation risk ξ_2 to match a wealth tail of $\hat{\zeta}_a = 1.4$. In the other model variants, η and ν are chosen to match a consumption tail of $\hat{\zeta}_c = 3.0$ and a capital income tail of $\hat{\zeta}_{ra} = 1.15$.

In Table 10, we report the time path of the average capital supply elasticity across our sample of households to an unanticipated tax reform at period $s = 0$. The benchmark economy with homothetic preferences yields the lowest elasticity at every time instant, which is roughly half of the one of the non-homothetic economy as $\gamma = 2$. That is, the non-homothetic model yields a higher average capital supply elasticity than the homothetic model if and only if $\gamma > 1$. Interestingly, the model with wealth-dependent, homothetic preferences and scale-dependent capital returns provides a comparable time path of capital supply elasticities relative to the non-homothetic model. This finding depends on the choice of the wealth utility shifter κ and the degree of scale-dependence as captured by ψ. Higher values of κ and ψ both increase the elasticity in models with scale-dependent returns relative to the pure non-homothetic preference economy. Finally, a model that features both non-homothetic preferences and scale-dependent capital returns generates the highest capital supply elasticities at every instant. In particular, as time goes by, there arises a strong complementary between non-homothetic preferences and scale-dependence: A convex saving function that jointly arises from a convex capital income and a concave consumption function yields a capital supply elasticity that is roughly 5 times larger than a model that has only a convex capital income function or a concave consumption function. In other words, such a model allows the unbounded (and nonlinear) effects of scale-dependence to take off at much earlier stages of time.

Table 10. Capital supply elasticity: time path to tax reform at period $s = 0$ for various models.

Environment	Elasticity			
	$\varepsilon_{a,\mu_r}\|_{1\ years}$	$\varepsilon_{a,\mu_r}\|_{5\ years}$	$\varepsilon_{a,\mu_r}\|_{10\ years}$	$\varepsilon_{a,\mu_r}\|_{20\ years}$
Unanticipated Tax Reform				
a. Homothetic Preferences	0.12	0.60	1.21	2.43
b. Non-Homothetic Preferences	0.24	1.21	2.43	4.93
c. WIU Homothetic Preferences + Scale	0.23	1.11	2.22	4.30
d. Non-Homothetic Preferences + Scale	0.28	1.41	3.70	27.47

C Quantitative Appendix

C.1 Estimation

Method of Simulated Moments Formally, consider a set of $n = 1, \ldots, N$ moments. Let $\boldsymbol{\theta}$ be the parameter vector to be estimated. Moreover, let m_n denote a generic empirical moment and $\hat{m}_n(\boldsymbol{\theta})$ the corresponding model-generated moment for a given parameter vector $\boldsymbol{\theta}$. The MSM minimizes the difference between each empirical moment and its corresponding simulated model-implied moment. We define $G_n(\boldsymbol{\theta}) \equiv \hat{m}_n(\boldsymbol{\theta}) - m_n$ for all n. The MSM objective is:

$$\arg\min_{\boldsymbol{\theta}} \mathbf{G}(\boldsymbol{\theta})' \boldsymbol{W} \mathbf{G}(\boldsymbol{\theta}), \tag{19}$$

where $\boldsymbol{W}$ is a weighting matrix and $\mathbf{G}(\boldsymbol{\theta})$ is a column vector in which all moment conditions are stacked, i.e., $\mathbf{G}(\boldsymbol{\theta}) = [G_1(\boldsymbol{\theta}), \ldots, G_N(\boldsymbol{\theta})]^T$. In our baseline estimation, we assume that $\boldsymbol{W}$ is a diagonal matrix with uniform weights for all moments.[39]

C.2 Model Validation: Conditional Distributions

[39]In practice, the objective function is highly non-linear. We therefore utilize a global control random search algorithm in the estimation process.

Figure 20. Conditional distribution: data (top panels) and model (bottom panels).