Where Did U.S. Tax Progressivity Go?
A Century of Income Tax Progression

Javier Cortés Orihuela*
Pablo Gutiérrez Cubillos†
Marcelo Montes Muñoz‡

June 26, 2026

Abstract

How has U.S. tax progressivity evolved since 1913? We recover residual progression, the elasticity of after-tax income with respect to pre-tax income, from grouped income shares, and show that our estimates closely track microdata benchmarks. Applied to WID and IRS tabulations, the method reveals opposing patterns. Top progressivity followed an inverted U, rising through World War II and declining after the 1970s. Below the top decile, progressivity is U-shaped, rising after the mid-1980s. The increase is strong in WID but muted in IRS, reflecting transfers and differences in income concepts. These findings reconcile competing accounts of declining and rising U.S. tax progressivity.

Keywords: Tax progressivity, residual income progression, income distribution, Lorenz curve, historical inequality.

JEL Classification: H20, H24, D31.

*Department of Economics, Yale University. Email: javier.cortesorihuela@yale.edu.

†Department of Management and Information Systems, Faculty of Economics and Business, University of Chile. Email: pgutiecu@fen.uchile.cl.

‡Department of Economics, Faculty of Economics and Business, University of Chile. Email: mmontesm@fen.uchile.cl.

1 Introduction

Tax progressivity, the degree to which the tax burden rises with income, is central to public finance and policy debates. Its case is both normative and economic: progressive taxation runs from the ability-to-pay principle (Mill, 1848) to modern theories of distributive justice (Rawls, 1971), while also compressing the inequality markets generate (Jalles and Mello, 2024) and insuring households against idiosyncratic labor-market shocks that private markets cannot (Heathcote *et al.*, 2017). However, how U.S. progressivity evolved from the introduction of the federal income tax to the modern microdata era remains largely unexplored.

The source of this gap is measurement. Modern empirical work measures progressivity through residual income progression: the elasticity of after-tax income with respect to pre-tax income (Jakobsson, 1976; Musgrave and Thin, 1948). This object requires observing how the tax-and-transfer system maps pre-tax into after-tax income, so existing estimates rely on matched pre- and after-tax microdata, which begin only in 1969 (Borella *et al.*, 2023). The first fifty-six years of the federal income tax have therefore been studied through statutory schedules, average effective tax rates, concentration curves and distributional tax tables rather than residual income progression (Auten and Splinter, 2024; Mathews, 2014; Musgrave and Thin, 1948; Pechman, 1985; Piketty and Saez, 2007; Saez and Zucman, 2019, 2026; Splinter, 2020).

In this paper we close this gap by constructing the first annual estimates of residual income progression for the United States from the introduction of the federal income tax in 1913 to 2021. Our method shows how to recover residual progressivity from grouped income data alone. In particular, under rank preservation and a common log-location-scale structure for pre- and after-tax income, the implied tax function is a power function—the benchmark specification through which modern microdata studies estimate residual income progression[1]—and its exponent, which captures the residual elasticity, equals the ratio of the standard deviations of log after-tax and log pre-tax income. Because log dispersions are scale-invariant, they are identified by complete Lorenz curves.

We apply this result to grouped World Inequality Database (WID) and Internal Revenue Service (IRS) tabulations, using a baseline Lorenz-based measure and semi-parametric and parametric alternatives.[2] The main patterns are robust across methods and validate well in the modern period: where matched microdata are available, the grouped-data estimates closely track microdata-based regression benchmarks in the full distribution, below the top decile, and at the top. The resulting series provides a long-run empirical counterpart to the power-tax parameters used in macro-public-finance models, extending their coverage from the late 1960s back to the origin of the federal income tax.

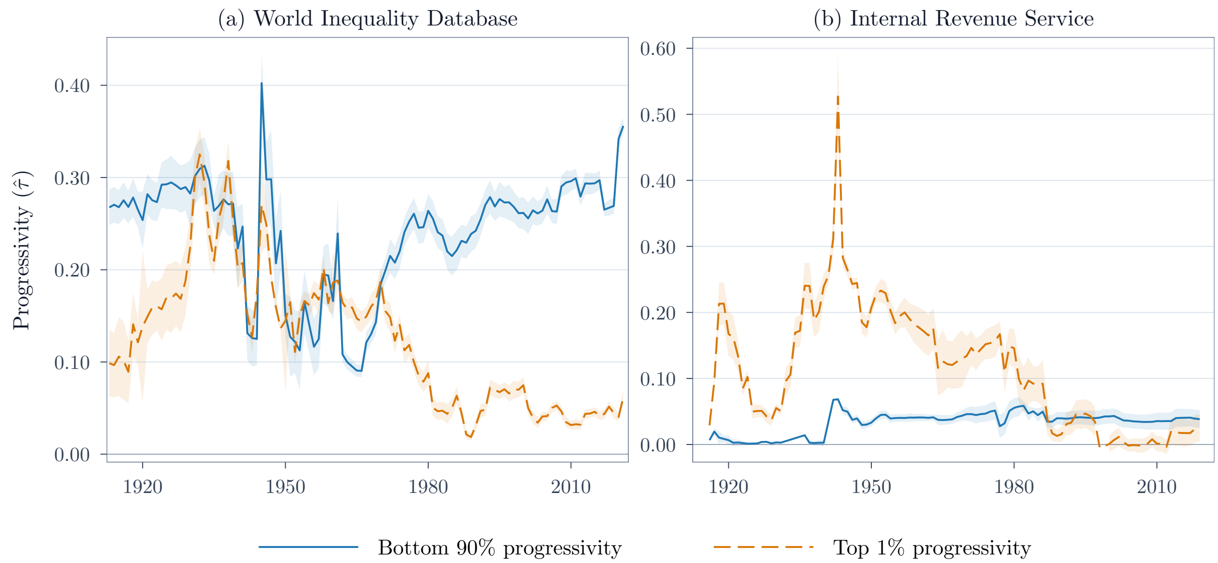
Our central finding is that U.S. progressivity did not simply rise or fall over the past century; it moved across the income distribution, as shown in Figure 1[3]. At the top 1, 5, and 10 percent, residual progressivity traces an inverted U: low under the early income tax, high through the mid-century high-rate regime, and low again after the 1980s. Below the top decile, the pattern is closer to its mirror image, a U shape: progressivity falls from the mid-century peak to a trough in the 1980s and rises thereafter. This heterogeneity matters for interpretation. When progressivity is not uniform across the income distribution, any single estimate of aggregate progressivity necessarily embeds an aggregation rule, whether explicit or implicit. Estimators that place more weight on the top may suggest declining progressivity after 1980, while estimators that place more weight below the top may suggest rising progressivity over the same period.

[1]Its popularity reflects four properties: a transparent one-parameter interpretation, analytical tractability inside structural macro-public-finance models (Bénabou, 2000; Boar and Midrigan, 2022; Castro Nofal *et al.*, 2024; Chang, 2018; Ferrière and Navarro, 2023; Heathcote *et al.*, 2017, 2020; Wu, 2021), a strong empirical fit with R2's above 90%, and direct estimability from microdata by OLS or PPML (Borella *et al.*, 2023; König, 2023).

[2]The baseline measure is an entropy-ratio estimate of relative log dispersion. The robustness checks include parametric Lorenz-curve specifications and a grouped PPML share-mapping estimator.

[3]We exclude the bottom 20 percent of WID share points from the baseline estimates. This restriction is motivated by the same concern emphasized by **?** for growth-incidence calculations, namely that income growth and related distributional objects are difficult to interpret at the very bottom of the distribution when pretax incomes are close to zero or negative.

Figure 1: Effective residual income progression at the bottom and top of the U.S. income distribution, 1913–2021



Notes: The figure reports Entropy Ratio estimates of progressivity, defined as one minus the ratio of after-tax to pre-tax log-income dispersion recovered from grouped Lorenz shares. Higher values indicate greater progressivity. Shaded areas are 95 percent delta-method confidence intervals accounting for covariance between the two dispersion estimates. For the World Inequality Database, the bottom and top ranges are percentiles 20–90 and 99–100. For the Internal Revenue Service, they are defined using the observed Lorenz cutpoints nearest percentiles 90 and 99 and therefore approximate the bottom 90 percent and top 1 percent. The income concepts are national income before and after taxes and transfers for the World Inequality Database and adjusted gross income before and after federal income tax for the Internal Revenue Service.

We therefore complement our global elasticity estimates with a local percentile-level measure of progressivity. This measure estimates progressivity separately across the distribution and then aggregates those local estimates, allowing us to ask what progressivity the average taxpayer faced rather than relying on the implicit weights of a single global estimator. The resulting equal-weighted local measure shows a strong post-1980 increase, consistent with the view that progressivity recovered below the top. This recovery is strong in WID but muted in IRS, a divergence that reflects income concepts. IRS tabulations remain close to fiscal income reported on tax returns, whereas WID post-tax national income incorporates transfers, retained earnings, in-kind benefits, and other components not captured in fiscal income Piketty *et al.* (2018). The recovery below the top therefore appears to be driven largely by the growing role of transfers and refundable credits. At the top, where transfers are quantitatively negligible but retained earnings remain relevant, WID and IRS yield much more similar patterns, and the inverted U is robust across both.

These findings contribute to the recent debate on whether U.S. tax progressivity has risen or fallen. Piketty and Saez (2007) and Saez and Zucman (2019) emphasize a long decline concentrated at the top, where effective tax rates fell sharply from their mid-century levels. By contrast, Splinter (2020) finds that, once refundable credits and transfers are included, the tax-and-transfer system became more progressive after the late 1970s, with a trough around 1986; Coleman and Weisbach (2023) reach a similar conclusion across multiple datasets.[4] Our estimates suggest that these accounts need not be inconsistent: they refer to different parts of the distribution, different income concepts, and different implicit weights. The decline is a statement about upper-tail income-tax progressivity; the rise is a statement about tax-and-transfer progressivity below the top. In this sense, the U.S. system did not simply become more or less progressive: progressivity moved down the distribution.

2 Residual progression from Lorenz curves

Tax progressivity is, fundamentally, a local and comparative property of the tax schedule. A schedule is progressive at income level y whenever the average tax rate rises with income, or equivalently whenever the marginal tax rate exceeds the average tax rate:

$$T'(y) > \frac{T(y)}{y}.$$

This condition identifies whether the tax burden rises locally with income, but it does not by itself provide a scalar measure of the degree of progressivity. We focus on residual income progression, which measures the elasticity of after-tax income with respect to pre-tax income and is therefore directly connected to the tax function itself.[5] Concretely, let y

[4]This question is informed by, but distinct from, the dispute over the *level* of top-income inequality (Auten and Splinter, 2024; Piketty *et al.*, 2018), which turns on the allocation of underreported and untaxed income. Our object is the residual progression of the tax schedule, not the concentration of income.

[5]Musgrave and Thin (1948) distinguished average rate progression, marginal rate progression, liability progression, and residual income progression. Other aggregate measures, such as the Kakwani, Suits, and Reynolds–Smolensky indices, summarize redistribution using concentration curves or changes in inequality indices.

denote pre-tax income, $T(y)$ tax liability, and $x = y - T(y)$ after-tax income. The residual income elasticity is

$$\epsilon(y) \equiv \frac{\partial \ln x}{\partial \ln y} = \frac{1 - T'(y)}{1 - T(y)/y}. \tag{1}$$

Progressivity at income level y is equivalent to $\epsilon(y) < 1$: after-tax income rises less than proportionally with pre-tax income, so the tax system compresses after-tax income differences. Since $\epsilon(y)$ can vary over the income distribution, empirical and quantitative work often summarizes the schedule with the power approximation

$$x(y) \approx \lambda y^{1-\tau}, \tag{2}$$

where $1 - \tau$ is the constant residual-income elasticity and τ is the corresponding progressivity parameter. With matched microdata, $1 - \tau$ is typically estimated from the relationship between individual or household pre-tax and after-tax income, using either log-linear specifications through OLS or level specifications via PPML.[6]

This matched-data requirement is restrictive for historical work. Before the late 1960s, U.S. income distributions are available primarily as grouped tabulations rather than matched pre-tax and after-tax income pairs, so modern regression-based estimators cannot be applied directly to the first five decades of the federal income tax. We therefore develop an identification result that recovers residual progression from grouped pre-tax and after-tax income shares. The central idea is that assumptions on the marginal distributions of pre-tax and after-tax income, together with rank preservation, imply a specific functional form for the effective tax schedule. We work under three standard assumptions:

Assumption 1 (Deterministic tax mapping)**.** *After-tax income satisfies $X = Y - T(Y)$, where $T(\cdot)$ is a deterministic, measurable tax function.*

Assumption 2 (Regular income distributions)**.** *The cumulative distribution functions F_x and F_y are strictly increasing and continuous on their supports.*

[6] Log-linear and level estimators impose different conditional-mean restrictions: the former restrict $\mathbb{E}[\log x_i \mid \log y_i]$, while the latter restrict $\mathbb{E}[x_i \mid y_i]$. They recover the same constant elasticity only under stronger restrictions, but both require matched observations of pre-tax and after-tax income. Other aggregate measures, such as the Kakwani, Suits, and Reynolds–Smolensky indices, summarize redistribution using concentration curves or changes in inequality indices.

Assumption 3 (Rank preservation)**.** *The function $Y \mapsto X(Y) = Y - T(Y)$ is strictly increasing.*

Assumption 3 is the key identifying assumption that allows the two marginal distributions to be interpreted as a single pre-tax-to-after-tax mapping. It requires the tax-and-transfer system not to reverse taxpayer ranks. When $T(\cdot)$ is differentiable, this condition implies marginal tax rates below 100 percent; more generally, it requires after-tax income to be strictly increasing in pre-tax income.

Lemma 1 (Implied Tax Function)**.** *Under Assumptions 1–3,*

$$T(Y) = Y - Q_x[F_y(Y)]\,,$$

where Q_x denotes the quantile function of X.[7]

Lemma 1 states that, under rank preservation, after-tax income is recovered by assigning each taxpayer their pre-tax rank and retrieving after-tax income at the same rank. We call $T(Y)$ the implied tax function of the distributional pair (F_y, F_x).[8] If after-tax income is a deterministic increasing function of pre-tax income, this implied function coincides with the structural tax schedule.

This structural interpretation is a benchmark. Deductions, credits, transfers, filing status, avoidance, reporting behavior, and income reclassification may generate different after-tax incomes among taxpayers with the same pre-tax income. In that case, marginals alone do not identify the structural schedule. The implied function should instead be read as the rank-preserving equivalent schedule: the effective schedule that rationalizes the observed transformation from the pre-tax to the after-tax marginal distribution without reranking. Thus rank preservation is required for a literal structural interpretation, but not for the

[7] A proof of Lemma 1 is provided in Appendix A1.

[8] Appendix A4 provides examples using well-known parametric families.

implied function to remain a well-defined summary of distributional compression.[9]

We now ask when this implied effective schedule takes the power form used in the modern progressivity literature. Under rank preservation, the tax schedule maps each pre-tax rank into the same after-tax rank. If, in addition, the two distributions differ only by location and scale in logs, this rank-preserving quantile map is log-affine, and the implied schedule is a power function.

Assumption 4 (Log-Location-Scale Incomes)**.** *Pretax income Y and after-tax income X belong to the same log-location-scale family: there exist constants $a \in \mathbb{R}$ and $b > 0$ such that*

$$Q_{\log X}(p) = a + b\, Q_{\log Y}(p) \quad \text{for all } p \in (0,1).$$

Canonical cases include the Pareto, lognormal, log-uniform, and Weibull distributions—all of which have been used extensively as models of the income distribution (Chotikapanich, 1993; Gabaix, 2009; Gibrat, 1931; Mirzaei *et al.*, 2019). Under Assumption 4, the implied tax function in Lemma 1 takes the following form:

Proposition 1 (Log-Location-Scale Identification of Residual Progression)**.** *Under Assumptions 1–3 and 4,*

$$T(Y) = Y - \exp\Big(\mu_{\log x} - \tfrac{\sigma_{\log x}}{\sigma_{\log y}}\, \mu_{\log y}\Big) Y^{\,\sigma_{\log x}/\sigma_{\log y}},$$

where $\mu_{\log z}$ and $\sigma_{\log z}$ denote the mean and standard deviation of $\log z$, for $z \in \{X, Y\}$. In particular, the tax function takes the power form (2) with

$$\lambda = \exp\left(\mu_{\log x} - \frac{\sigma_{\log x}}{\sigma_{\log y}} \mu_{\log y}\right), \qquad 1 - \tau = \frac{\sigma_{\log x}}{\sigma_{\log y}}.$$

Proposition 1 is the central theoretical result of the paper.[10] It shows that, under the

[9]This distinction is not specific to distributional data. Even with matched microdata, log-log regressions of after-tax income on pre-tax income generally estimate effective residual-progression objects rather than deterministic structural schedules when deductions, credits, transfers, filing status, avoidance, reporting responses, or income reclassification generate horizontal heterogeneity. Appendix A1.2 formalizes this point and relates the dispersion ratio of Proposition 1 to effective residual-progression objects based on the conditional expectation function.

[10]Proofs of Lemma 1 and Proposition 1 are provided in Appendix A1.

log-location-scale benchmark, residual progression is identified by the relative dispersion of log after-tax and pre-tax income. A more progressive tax-and-transfer system compresses the after-tax distribution more strongly relative to the pre-tax distribution, yielding a smaller value of $\frac{\sigma_{\log x}}{\sigma_{\log y}}$. The interpretation of this ratio follows the interpretation of the implied schedule in Lemma 1. Under deterministic rank preservation, it is the structural residual-income elasticity of the tax schedule. When rank preservation fails, this structural interpretation is no longer exact: the ratio need not equal the slope of an individual-level mapping from pre-tax to after-tax income. It remains, however, a well-defined distributional object. It measures how much the after-tax distribution is compressed in logs relative to the pre-tax distribution. Equivalently, it is the residual progression of the rank-preserving equivalent schedule: the effective schedule that would rationalize the observed transformation from the pre-tax marginal distribution to the after-tax marginal distribution without reranking. Appendix A1.2 shows that, with joint microdata, this total compression can be decomposed into average vertical progression, nonlinearity in the conditional expectation function, and horizontal heterogeneity.

Crucially, $\sigma_{\log x}$ and $\sigma_{\log y}$ are distributional parameters that can be estimated from income-share tabulations. To see why Lorenz information is sufficient for this purpose, let $L_z(p)$ denote the Lorenz curve of a positive income concept Z with quantile function $Q_z(p)$ and mean μ_z. Wherever the Lorenz curve is differentiable,

$$L'_z(p) = \frac{Q_z(p)}{\mu_z}.$$

Since adding the constant $\log \mu_z$ does not affect variance,

$$\text{Var}(\log Z) = \int_0^1 \left(\log L'_z(p) - \int_0^1 \log L'_z(u)\, du \right)^2 dp. \tag{3}$$

Thus the log-income variance, and hence the dispersion ratio in Proposition 1, is identified by the shape of the Lorenz curve alone. The estimation problem is therefore to recover this Lorenz shape smoothly from grouped income shares.

Several implementations follow from this result. Our baseline estimator uses the observed grouped Lorenz shares directly and chooses the log-dispersion contraction that makes the predicted after-tax shares closest to the observed after-tax tabulation in entropy-KL distance. As shown in the next section, this estimator can be written as a grouped PPML estimator and therefore links naturally to the PPML approach used with microdata. As parametric complements, we also estimate $\sigma_{\log z}$ from analytic Lorenz curves under Pareto, lognormal, and log-uniform specifications. These complementary estimates provide useful robustness checks while preserving the same identified object: the ratio of after-tax to pre-tax log-income dispersion. The next section describes the estimation procedure and validates the grouped-data estimators against microdata OLS and PPML benchmarks.

3 Estimation from grouped data and validation

Entropy dispersion ratio. Proposition 1 shows that the benchmark residual-progression parameter is the ratio of after-tax to pre-tax log-income dispersion. A complete Lorenz curve identifies this ratio because its derivative recovers normalized percentile income. Grouped tabulations, however, report only average Lorenz slopes within bins.

Let $\Delta p_j = p_j - p_{j-1}$ and $\Delta L_j^z = L_j^z - L_{j-1}^z$ ($z \in \{y, x\}$) denote the population and income-share mass of bin j. The ratio $z = \frac{\Delta L_j^z}{\Delta p_j}$ is the bin mean income relative to the overall mean, or equivalently the average Lorenz slope in the bin. Our baseline estimator replaces the Lorenz slope by this step function and computes the grouped log-income variance

$$\widehat{\sigma}^2_{\log z} = \sum_j \Delta p_j \left[\log\left(\frac{\Delta L_j^z}{\Delta p_j}\right) - \sum_\ell \Delta p_\ell \log\left(\frac{\Delta L_\ell^z}{\Delta p_\ell}\right) \right]^2. \tag{4}$$

The entropy dispersion-ratio estimator is therefore $\widehat{\tau} = 1 - \frac{\widehat{\sigma}_{\log x}}{\widehat{\sigma}_{\log y}}$. Notably, this estimator requires no interpolation within cells, tail restriction, or joint information on pre- and after-tax income. However, because it treats the Lorenz slope as constant within each reported cell, it does not recover within-bin log dispersion. The resulting coarsening error enters both the pre- and after-tax log variances, and may therefore partly cancel in the ratio, especially when the two marginal distributions have similar within-bin curvature.

As the partition refines, the estimator converges to the corresponding continuous Lorenz functional.

Parametric dispersion ratios. As complements, we fit analytic Lorenz curves that restore within-bin dispersion under standard parametric families. These estimates are not our baseline; they gauge sensitivity to tail shape and full-distribution functional form. Let $L(p;\sigma_{\log z})$ denote the Lorenz curve for income concept $z \in \{y, x\}$. We use

$$\text{Pareto:} \quad L(p;\sigma_{\log z}) = 1 - (1-p)^{1-\sigma_{\log z}}, \tag{5}$$

$$\text{Lognormal:} \quad L(p;\sigma_{\log z}) = \Phi\big(\Phi^{-1}(p) - \sigma_{\log z}\big)\,, \tag{6}$$

$$\text{Loguniform:} \quad L(p;\sigma_{\log z}) = \frac{\exp\big(\sqrt{12}\,\sigma_{\log z}\,p\big) - 1}{\exp\big(\sqrt{12}\,\sigma_{\log z}\big) - 1}, \tag{7}$$

where Φ is the standard normal CDF.[11] Let (p_k^z, L_k^z) be the empirical Lorenz coordinates. Imposing $\sigma_{\log x} = \epsilon\,\sigma_{\log y}$, we estimate $(\sigma_{\log y}, \epsilon)$ for each family by stacked nonlinear least squares:

$$(\hat{\sigma}_{\log y}, \hat{\epsilon}) = \arg\min_{\sigma_y, \epsilon} \left\{ \sum_k \left[L_k^y - L(p_k^y;\sigma_y)\right]^2 + \sum_k \left[L_k^x - L(p_k^x;\epsilon\sigma_y)\right]^2 \right\}. \tag{8}$$

Joint estimation lets the covariance between pre- and after-tax Lorenz moments enter the standard error of $\hat{\epsilon}$. Pareto is most natural for upper-tail exercises; lognormal and log-uniform provide full-distribution robustness checks.

Grouped PPML. As a separate check, we impose the power restriction directly on grouped income shares. If pre-tax cell j has relative mean $\Delta L_j^y / \Delta p_j$, the isoelastic share mapping predicts

$$\widehat{\Delta L}_j^x(\epsilon) = \frac{\Delta p_j (\Delta L_j^y / \Delta p_j)^\epsilon}{\sum_\ell \Delta p_\ell (q_\ell^y)^\epsilon}. \tag{9}$$

[11]For $Z \sim \text{Pa}(\eta_z, b_z)$, $\sigma_{\log z} = 1/\eta_z$, so the Pareto Lorenz curve can be written directly in $\sigma_{\log z}$ as in (5).

Equivalently, ϵ solves the grouped PPML model

$$\mathbb{E}[\Delta L_j^x \mid q_j^y, \Delta p_j] = \Delta p_j \exp\left(\alpha + \epsilon \log q_j^y\right), \tag{10}$$

with $\log \Delta p_j$ as an offset. The intercept normalizes fitted shares to sum to one, so the fitted values coincide with (9). The same estimator can be written as entropy-KL share fitting:

$$\widehat{\epsilon}^{PPML} = \arg\min_{\epsilon} \sum_j \Delta L_j^x \log\left(\frac{\Delta L_j^x}{\widehat{\Delta L_j^x}(\epsilon)}\right).$$

The PPML exercise differs from the dispersion-ratio estimators. Those estimators are functions of the two marginal Lorenz curves. PPML instead asks whether after-tax shares are well approximated by a common power transformation of pre-tax cell means. This restriction is implied by exact rank preservation, but it can also hold under weaker conditions: reranking may occur within cells, or cross-cell reranking may average out, as long as the grouped after-tax shares still satisfy the power share mapping. PPML therefore moves the identifying restriction from individual ranks to grouped shares. [12]

This flexibility comes at a cost. When the grouped share restriction is valid, PPML estimates the same benchmark elasticity and may be less sensitive to within-cell reranking. When it fails, the coefficient is only the best-fitting power share mapping. We therefore use PPML as a validation and robustness exercise rather than as the headline estimator.

No estimator dominates ex ante. The entropy ratio is transparent and nonparametric but subject to coarsening error. Parametric Lorenz curves restore within-cell dispersion at the cost of functional form. Grouped PPML can relax individual rank preservation, but imposes a share-mapping restriction. We compare these tradeoffs using microdata, where the full distribution is observed.

[12] See Appendix A1.5 for details.

3.1 Validation

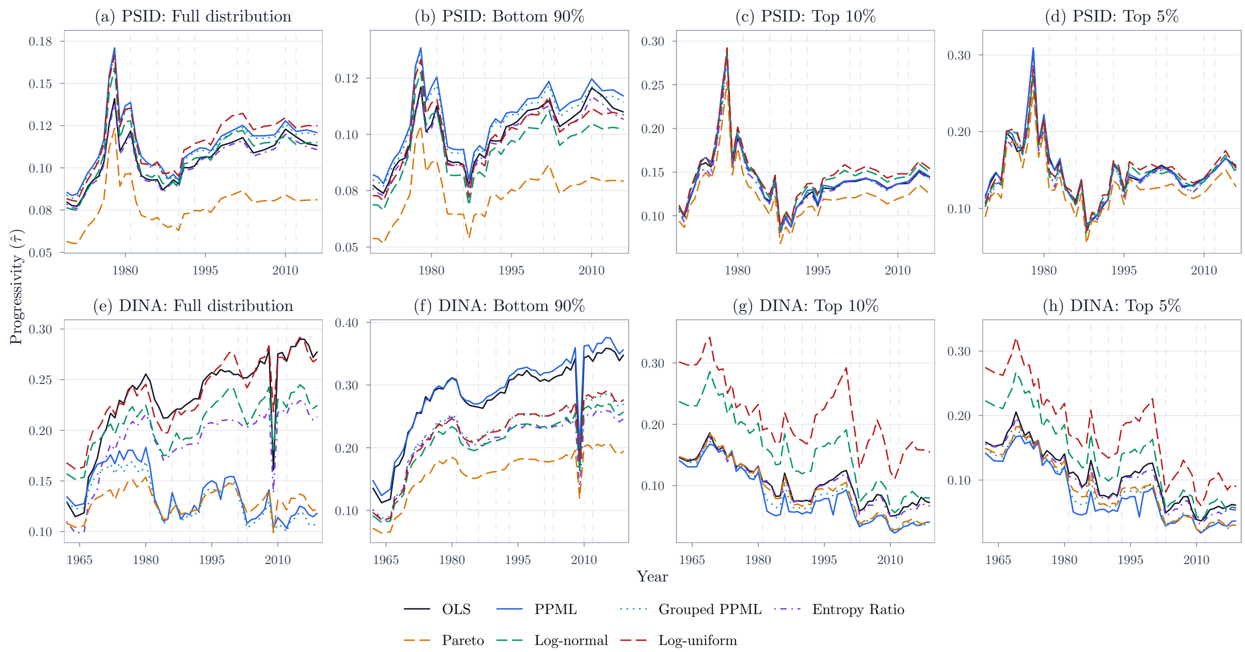
We validate the grouped-data estimators against OLS and PPML microdata benchmarks using two complementary U.S. sources. The first is the Panel Study of Income Dynamics (PSID). Our benchmark is Borella *et al*. (2023), who estimate residual progression by OLS for each wave from 1969 to 2016; we replicate their estimates using their replication code and augment them with PPML estimates following König (2023). The second is the Distributional National Accounts (DINA) synthetic microdata from Piketty *et al*. (2018), which cover 1962 to 2019, contain matched pre- and after-tax income, and are constructed to be closely aligned with the grouped WID tabulations used in our historical analysis.

As a first exercise, for each source we aggregate the microdata into synthetic income-share tabulations consisting of 100 Lorenz ordinates and re-estimate progressivity using grouped PPML, the Entropy Ratio, and the Pareto, log-normal, and log-uniform Lorenz estimators. We report $\tau = 1 - \epsilon$, so that higher values denote greater progressivity. For each estimator, we compare the grouped-data estimate with its OLS and PPML microdata counterparts for the full distribution, the bottom 90%, the top 10%, and the top 5%.

Two results hold across both sources and all four population ranges (full distribution, bottom 90, top 10, top 5); Figure 2 reports the PSID married-sample and DINA validations in a common eight-panel format. First, grouped PPML closely reproduces microdata PPML, while the Entropy Ratio lies nearer OLS. The grouped tabulations therefore recover not only the aggregate microdata estimates but their variation across the distribution, and the gap between OLS and PPML is not an artifact of grouping: each grouped estimator inherits the aggregation implicit in its microdata counterpart.

Second, the movements differ sharply by range, and DINA makes the consequence clearest. Progressivity for the bottom 90 percent rises substantially over the sample, apart from temporary reform-year reversals, whereas progressivity within the top 10 and top 5 percent falls from its late-1960s and 1970s levels to much lower post-1980 values, recovering only modestly. The full-distribution series combines these opposing movements, so its level and trend depend on how an estimator weights the bottom against the top: OLS and

Figure 2: PSID and DINA validation, 1962–2019



Notes: Higher values correspond to greater residual progressivity. OLS and PPML are estimated from DINA microdata (Piketty *et al.*, 2018). The remaining estimates use income-share tabulations constructed from the same sample, for the full distribution, bottom 90 percent, top 10 percent, and top 5 percent. Grouped PPML fits after-tax grouped Lorenz shares as a power tilt of before-tax shares. The Entropy Ratio separately estimates before- and after-tax log dispersions from entropy-step Lorenz densities and takes their ratio. Pareto, log-normal, and log-uniform estimates impose the corresponding parametric Lorenz curves. Each panel uses its own vertical scale.

the Entropy Ratio track the rising bottom, while PPML and grouped PPML weight high incomes more heavily and imply a lower, flatter series after 1980.

The parametric estimators reinforce this. Pareto is a poor approximation for the full distribution and bottom 90 percent but is informative in the upper tail; log-normal and log-uniform capture broader-distribution movements but differ in levels. Estimator choice is thus consequential precisely when progressivity moves in opposite directions across the distribution—which is why the historical analysis reports bottom and top progressivity separately rather than as a single full-distribution parameter.

Appendix Figures A2.1–A2.3 report three robustness versions of the same eight-panel exercise. The first uses concentration curves, cumulating after-tax income over before-tax ranks, instead of marginal after-tax Lorenz curves. These estimates are slightly closer to the microdata benchmarks, as expected: concentration curves preserve the before-tax ranking and therefore avoid the reranking problem that arises when only marginal after-tax Lorenz curves are observed. The second and third repeat the validation after coarsening the synthetic grouped tabulations to the resolution of historical IRS tabulations, once with marginal Lorenz curves and once with concentration curves. These harsher tests leave the broad DINA patterns largely intact but weaken the upper-tail PSID comparisons, consistent with the fact that IRS-style coarsening bites hardest when the validation sample is a survey with few upper-tail observations. Appendix Figures A2.4–A2.7 also repeat the PSID validation for single households, using the same four population ranges and the same marginal-versus-concentration and WID-versus-IRS coarseness distinctions.

Appendix Figures A2.8 and A2.9 provide a complementary Monte Carlo check. The simulations use grouped marginal Lorenz curves, matching the least-informative data environment of the main estimates. They show that the grouped estimators are well behaved when the power-tax structure is correct and remain informative under a nonlinear bracket tax schedule, where the single progressivity parameter is necessarily an approximation.

4 Historical progressivity trends, 1913–2021

4.1 Data and income concepts

We apply our estimators to annual U.S. data from the World Inequality Database (WID), covering before- and after-government income shares from 1913 to 2021. Following Piketty *et al.* (2018), pretax national income is factor plus pension income before personal taxes and transfers, and post-tax national income subtracts taxes and adds transfers. Full-distribution WID estimates therefore capture the combined effect of taxes and transfers; at the top, where transfers are negligible relative to income, they reflect tax progressivity alone.

Two caveats apply. First, the WID builds pre-1966 top shares from IRS tabulations by Pareto interpolationOur method inherits any such biases, so pre-1966 estimates are best read as trend indicators rather than precise levels. Second, we complement WID with IRS tabulations, which include or exclude transfers symmetrically across pre- and post-tax income and so stay closer to a fiscal income concept.[13] The IRS series is not a statutory-tax benchmark—it still reflects deductions, realizations, filing behavior, and reclassification—but relative to WID post-tax income it is far less affected by transfers, making the treatment of transfers the main distinction between the two sources.

4.2 Two histories: an inverted-U at the top, a U-shape below

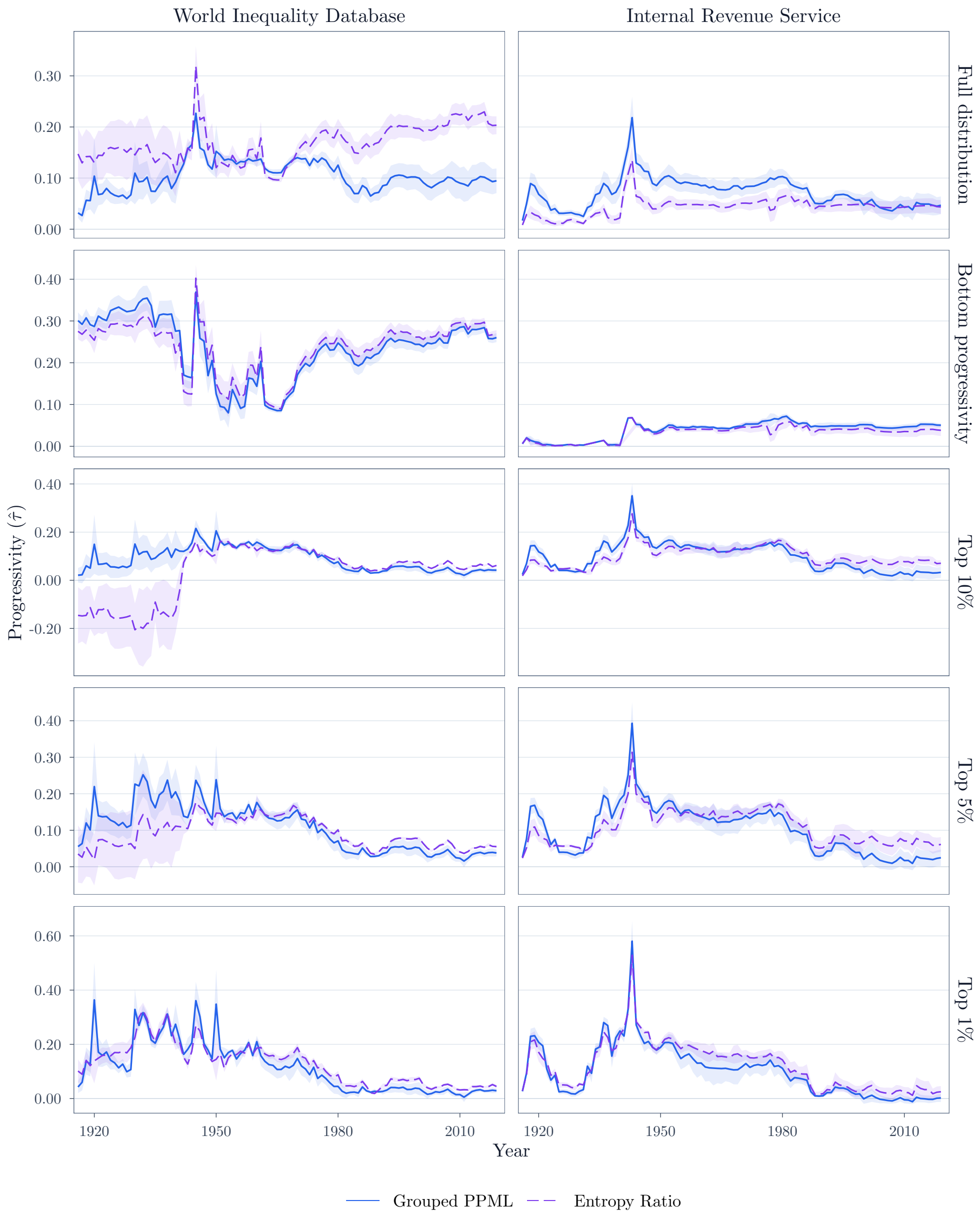
Figure 3 presents our main estimates for 1913–2021, by income range and for both WID and IRS. They tell two histories. At the top, progressivity traces an inverted-U: low under the early income tax, high through the mid-century high-rate regime, and low again after the 1980s. Below the top it follows a U-shape: high under the early income tax, peaking at mid-century with the wartime extension of the income tax to most households, falling to a trough around 1980, and rising again thereafter. The century thus records not a single rise or fall in progressivity but a reallocation of it down the distribution—the income tax builds and then dismantles progressivity at the top, while transfers and refundable credits increasingly carry redistribution below it.

The WID–IRS comparison serves as an internal diagnostic. Where transfers matter, as in the full distribution, the two sources diverge because they embed different income concepts; where transfers are negligible, as in the top 1, 5, and 10 percent, they converge. The divergence below the top is the signature of transfers; the convergence at the top confirms that upper-tail progressivity genuinely declined.

For most of the century the top and full-distribution histories coincide, because transfers

[13]WID data provide marginal Lorenz curves, whereas IRS tabulations provide before-tax income shares and concentration curves. That is, IRS after-tax shares are cumulated over groups ranked by before-tax income, not by after-tax income itself. Formally, if Y denotes before-tax income and X after-tax income, WID identifies the marginal Lorenz curve $L_X(p) = \mathbb{E}[X\mathbf{1}\{X \leq Q_X(p)\}]/\mathbb{E}[X]$, while IRS tabulations identify the concentration curve $C_X^Y(p) = \mathbb{E}[X\mathbf{1}\{Y \leq Q_Y(p)\}]/\mathbb{E}[X]$. These objects coincide only under rank preservation. When taxes or transfers induce reranking, the IRS object is therefore not the marginal Lorenz curve of after-tax income. In principle, however, concentration curves are more informative for estimating residual progressivity, since they retain information about how after-tax income varies along the before-tax income ranking. They do not identify the full joint distribution of before- and after-tax incomes, but they do identify the conditional mean of after-tax income along before-tax ranks. Appendix A1.3 develops this case. In the main text, we focus on marginal Lorenz curves because they represent the least informative data environment: they provide only the marginal distributions of before- and after-tax income and no direct information on how the two variables are paired.

Figure 3: Historical progressivity by income range, U.S., 1913–2021



Notes: Data come from the World Inequality Database (WID) and the Internal Revenue Service (IRS). Higher values correspond to greater residual progressivity. Grouped PPML and the Entropy Ratio are reported for every income range. Shaded areas are 95 percent confidence intervals. Grouped PPML intervals use percentile-bin robust sandwich standard errors. Entropy Ratio intervals stack the pre-tax and after-tax log-variance equations, estimate their percentile-bin sandwich covariance, and apply the delta method to their ratio. The WID full-distribution estimates exclude the bottom 20 percent and are computed over p20–p100; the IRS full-distribution estimates use the full observed IRS distribution. For WID, bottom progressivity is estimated over p20–p90. For IRS, bottom progressivity is estimated below the observed Lorenz cutpoint closest to p90 in each year, so this group is approximately, but not always exactly, the bottom 90 percent. Top-group estimates are computed within the indicated top 10, 5, and 1 percent ranges. WID uses pre-tax and post-tax national income; IRS uses AGI before and after federal income tax.

are small and full-distribution progressivity is essentially upper-tail income-tax progressivity. They separate only after 1980. Four phases trace this path.

One exception is informative. In the top 10 percent before 1941, the dispersion-ratio estimates imply negative progressivity. This is not a numerical anomaly but the consequence of crossing pre-tax and after-tax Lorenz curves within the top group: at some ranks the after-tax distribution is locally less compressed than the pre-tax distribution. The grouped PPML and parametric estimates do not reproduce this pattern because they impose a single-shape representation of the Lorenz curve, and hence a single residual elasticity over the whole range. The discrepancy is therefore diagnostic. It shows that, especially in the early income-tax period, a single residual elasticity may be an inadequate summary of progressivity within the top 10 percent. This motivates the local analysis in the next section, where progressivity is first estimated rank by rank and only then aggregated.

1913–1941: a nascent income tax. The Revenue Act of 1913 set a top rate of just 7 percent, and upper-tail progressivity was initially limited, consistent with near-proportional effective taxation despite a nominally progressive schedule. The War Revenue Act of 1917 raised the top rate to 67 percent, and both sources show a clear rise in top progressivity. The rate cuts of the 1920s (to 25 percent by 1925) produced a moderate erosion, consistent with Piketty and Saez (2007) on the role of statutory rates and with Splinter (2020), who notes that heavy sheltering muted the effect of high 1910s rates. The New Deal reversed course: the Revenue Acts of 1932 and 1935 raised the top rate to 63 and then 79 percent, and top progressivity rose substantially. Below the top, progressivity is already relatively high in this period, comparable to its modern level, so the U-shape's left arm begins from an elevated level rather than from a trough.

1941–1964: peak progressivity. The Revenue Act of 1942 raised the top rate to 88 percent and extended the income tax to most households; subsequent law held it at 91 percent. This is the peak of effective top progressivity in our sample: both sources show mid-century upper-tail compression far exceeding the early and post-1980 regimes. Progressivity below the top is also near its mid-century high, lifted by the same extension of the income tax to most households; top and bottom are elevated together, not yet apart. The pattern matches Piketty and Saez (2003) on the wartime compression of top shares and Borella *et al*. (2023), whose calibrated model finds these high-tax regimes materially affected labor supply and saving.

1964–1980: gradual erosion. The Kennedy–Johnson cut of 1964 lowered the top rate from 91 to 70 percent, producing a moderate decline—smaller than the rate cut implies, again because effective rates ran well below statutory ones (Splinter, 2020); **?** report a top-1% average federal income-tax rate of just 16 percent in 1962. The 1969 Alternative Minimum Tax and 1970s bracket creep roughly offset the cut, leaving top progressivity broadly stable into the late 1970s, even as it fell faster than full-distribution progressivity.

1980–2021: Reagan-era compression and the reversal. The Economic Recovery Tax Act of 1981 and TRA86 cut the top rate to 50 and then 28 percent while broadening the base; both sources show top progressivity falling sharply, to its lowest level since the 1920s—the post-WWII trough that Splinter (2020) dates to 1986 and that Borella *et al*. (2023) identify as the pivotal reform. Later reforms (1993, 2001–03, 2012, 2017) generate the partial recoveries and reversals visible at the top. This is where the two histories part: top progressivity stays far below its mid-century peak, while below the top the rising arm of the U takes over through transfers and refundable credits. For the first time in the sample, top and bottom move in opposite directions.

The top and bottom are therefore unambiguous across estimators; the full distribution is not. Because it aggregates two movements of opposite sign, its post-1980 trend is not a fact about the data but a consequence of weighting. Grouped PPML, fitting income-share moments, registers the top's decline and reports *falling* full-distribution progressivity after 1980; the dispersion-ratio estimator, summarizing compression in log Lorenz slopes, registers the bottom's rise and reports *growing* progressivity over the same period. No single full-distribution number resolves the disagreement. This maps directly onto the debate: the upper-tail estimates speak to the Piketty–Saez view of a mid-century rise and post-1970s decline at the top, while the bottom and post-1980 WID estimates speak to the Auten–Splinter emphasis on refundable credits and transfers after 1979. The tension dissolves once the two are read as claims about different parts of the distribution. The parametric robustness exercises in Appendix A3 lead to the same conclusion: estimator choice affects levels and the full-distribution post-1980 trend, but the inverted-U at the top and the post-1980 rise below the top are robust. The next section makes the weighting explicit.

4.3 What progressivity did the average taxpayer face?

The estimators above summarize the pre- to after-tax mapping with a single parameter, so their aggregation weights are implicit and need not agree when progressivity varies across the distribution. As a complement, we estimate local progressivity percentile by percentile and average it with equal population weight. Writing $q_z(p) = L'_z(p)$ for the Lorenz density of income concept $z \in \{y, x\}$, the local residual elasticity and progressivity of the rank-preserving schedule are

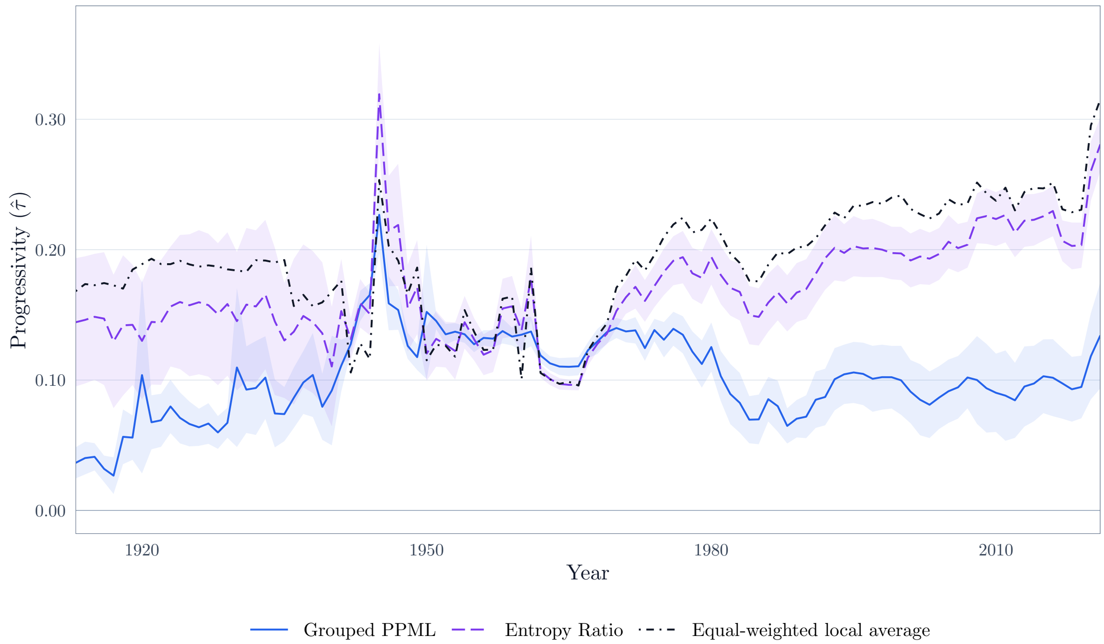
$$\epsilon(p) = \frac{d \log q_x(p)}{d \log q_y(p)}, \qquad \tau(p) = 1 - \epsilon(p),$$

which depend on distributional shape, not income levels. Because ranks are uniform, equal weighting across percentile cells is population weighting, so we report

$$\bar{\tau}^L_{20,100} = \frac{1}{0.80} \int_{0.20}^{1} \tau(p)\, dp,$$

the progressivity faced by the average rank cell above the 20th percentile. We estimate it from WID one-percentile tabulations, taking local derivatives of log normalized income by local linear regression over an 11-cell moving window; Appendix A1.6 gives the details.

Figure 4: Progressivity of the average tax-payer, WID



Notes: The figure compares grouped PPML, the entropy-ratio estimator, and the equal-weighted local average of percentile-level progressivity. The local average is computed from WID one-percentile cells over $p20$–$p100$. For each year and percentile midpoint, local derivatives of log normalized income are estimated by local linear regressions using an 11-cell moving window. The years 1920, 1930, and 1940 are retained in the underlying data but omitted from the plot. Higher values indicate greater progressivity.

The local average lies above both global estimators, especially after the 1970s. The reason is weighting: grouped PPML loads on the parts of the distribution holding larger after-tax income shares, whereas the local average gives each percentile equal weight, so the post-1980 recovery of WID tax-and-transfer progressivity is more visible in it. The time path reinforces the paper's central result—progressivity rises around World War II, falls to the early 1980s, and rises again—but the post-1980 increase is stronger under equal weights, exactly as expected if redistribution increasingly operated below the top while upper-tail progressivity stayed far below its mid-century peak.[14] The statistic is local, derivative-

[14]We check the post-1962 trend against matched DINA microdata in Appendix Figure A3.2. The exercise estimates local quadratic log-log residual elasticities at the individual level and averages them using DINA weights. The resulting local-average series tracks OLS closely and shows the same post-1980 rise, while PPML remains flatter.

based, and noisier than grouped PPML, so it complements rather than replaces it: with each rank cell weighted equally, U.S. progressivity did not simply rise or fall—it changed location in the distribution.

5 Conclusion

U.S. progressivity did not simply rise or fall over the twentieth century; it changed shape. Upper-tail progressivity traces an inverted-U — limited under the early income tax, high through the mid-century high-rate regime, low again after the 1980s — and the pattern holds in both WID and IRS, where transfers are small. Below the top it is the mirror image: WID progressivity rises after the mid-1980s while the IRS series stays flat, which locates the recovery in transfers and refundable credits rather than in the income tax. The equal-weighted local average confirms the direction — with each percentile weighted equally, post-1980 progressivity rises even as top progressivity remains far below its postwar peak. Progressivity moved down the distribution. This reconciles the two standard accounts, a mid-century rise and post-1970s fall at the top and a post-1979 increase through credits and transfers below, as claims about different parts of the distribution.

The method applies wherever long-run grouped tabulations exist, and suggests asking not only whether progressivity rose or fell but where in the distribution it operated. Because the estimates identify rank-preserving effective schedules from marginal distributions, they are best read as effective distributional summaries rather than statutory schedules.

References

Auten, G. and Splinter, D. (2024), 'Income inequality in the united states: Using tax data to measure long-term trends', *Journal of Political Economy* **132**(7), 2179–2227. doi: 10.1086/728741.

Bénabou, R. (2000), 'Unequal societies: Income distribution and the social contract', *American Economic Review* **90**(1), 96–129. doi: 10.1257/aer.90.1.96.

Blanchet, T., Fournier, J. and Piketty, T. (2022), 'Generalized Pareto curves: Theory and applications', *Review of Income and Wealth* **68**(1), 263–288. doi: 10.1111/roiw.12510.

Boar, C. and Midrigan, V. (2022), 'Efficient redistribution', *Journal of Monetary Economics* **131**, 78–91. doi: 10.1016/j.jmoneco.2022.08.002.

Borella, M., De Nardi, M., Pak, M., Russo, N. and Yang, F. (2023), 'The importance of modeling income taxes over time: U.S. reforms and outcomes', *Journal of the European Economic Association* **21**(6), 2237–2286. doi: 10.1093/jeea/jvad053.

Castro Nofal, B., Gutiérrez Cubillos, P. and Micco, A. (2024), Optimal progressivity under informal consumption.

Chang, B. Y. (2018), 'Pareto weights in practice: A quantitative analysis across 32 OECD countries', *Review of Economic Dynamics* **28**, 181–204. doi: 10.1016/j.red.2017.08.006.

Charpentier, A. and Flachaire, E. (2022), 'Pareto models for top incomes and wealth', *Journal of Economic Inequality* **20**(1), 1–25. doi: 10.1007/s10888-021-09488-9.

Chotikapanich, D. (1993), 'A comparison of alternative functional forms for the Lorenz curve', *Economics Letters* **41**(2), 129–138. doi: 10.1016/0165-1765(93)90186-G.

Coleman, T. and Weisbach, D. A. (2023), How progressive is the U.S. tax system?, Working Paper 991, University of Chicago, Coase-Sandor Institute for Law and Economics.

Cortés, J. and Gutiérrez Cubillos, P. (2023), 'Chotikapanich meets Pareto: The log-uniform distribution as a bridge', *Economics Letters* **224**, 111009. doi: 10.1016/j.econlet.2023.111009.

D'Addario, R. (1974), 'Ricerche sulla curva dei redditi', *Giornale degli Economisti e Annali di Economia* **33**, 571–604.

Ferrière, A. and Navarro, G. (2023), 'The heterogeneous effects of government spending: It's all about taxes', *Review of Economic Studies* **90**(2), 849–895. doi: 10.1093/restud/rdac029.

Gabaix, X. (2009), 'Power laws in economics and finance', *Annual Review of Economics* **1**, 255–294. doi: 10.1146/annurev.economics.050708.142940.

Gibrat, R. (1931), *Les Inégalités Économiques*, Sirey, Paris.

Heathcote, J., Storesletten, K. and Violante, G. L. (2017), 'Optimal tax progressivity: An analytical framework', *Quarterly Journal of Economics* **132**(4), 1693–1754. doi: 10.1093/qje/qjx018.

Heathcote, J., Storesletten, K. and Violante, G. L. (2020), 'How should tax progressivity respond to rising income inequality?', *Journal of the European Economic Association* **18**(6), 2715–2754. doi: 10.1093/jeea/jvz052.

Jakobsson, U. (1976), 'On the measurement of the degree of progression', *Journal of Public Economics* **5**(1-2), 161–168. doi: 10.1016/0047-2727(76)90066-9.

Jalles, J. a. T. and Mello, L. d. (2024), 'Fiscal redistribution and income inequality: Evidence from a large panel of countries', *European Journal of Political Economy* **82**, 102484. doi: 10.1016/j.ejpoleco.2023.102484.

Jenkins, S. P. (2017), 'Pareto models, top incomes and recent trends in UK income inequality', *Economica* **84**(334), 261–289. doi: 10.1111/ecca.12217.

König, J. (2023), 'Bias in tax progressivity estimates', *National Tax Journal* **76**(2), 267–289. doi: 10.1086/724186.

Lebergott, S. (1959), 'The shape of the income distribution', *American Economic Review* **49**(3), 328–347.

Mathews, T. (2014), 'Historical trends in the degree of federal income tax progressivity in the united states', *The Social Science Journal* **51**(1), 90–99.

Mill, J. S. (1848), *Principles of Political Economy, with Some of Their Applications to Social Philosophy*, John W. Parker, London.

Mirzaei, M., Osati, M. and Davari, M. (2019), 'Applications of weibull distribution in income distribution', *International Journal of Statistical Distributions and Applications* **5**(2), 30–37. doi: 10.11648/j.ijsd.20190502.12.

Musgrave, R. A. and Thin, T. (1948), 'Income tax progression, 1929–48', *Journal of Political Economy* **56**(6), 498–514. doi: 10.1086/256742.

Pareto, V. (1895), 'La legge della domanda', *Giornale degli Economisti* **10**, 59–68.

Pechman, J. A. (1985), *Who Paid the Taxes, 1966–85?*, Brookings Institution Press, Washington, DC.

Piketty, T. and Saez, E. (2003), 'Income inequality in the United States, 1913–1998', *Quarterly Journal of Economics* **118**(1), 1–41. doi: 10.1162/00335530360535135.

Piketty, T. and Saez, E. (2007), 'How progressive is the U.S. federal tax system? A historical and international perspective', *Journal of Economic Perspectives* **21**(1), 3–24. doi: 10.1257/jep.21.1.3.

Piketty, T., Saez, E. and Zucman, G. (2018), 'Distributional national accounts: Methods and estimates for the United States', *Quarterly Journal of Economics* **133**(2), 553–609. doi: 10.1093/qje/qjx043.

Rawls, J. (1971), *A Theory of Justice*, Harvard University Press, Cambridge, MA.

Saez, E. and Zucman, G. (2019), *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*, W. W. Norton & Company, New York.

Saez, E. and Zucman, G. (2026), 'Distributional tax analysis in theory and practice: Harberger meets diamond-mirrlees', *Journal of Public Economics* **259**, 105675. doi: https://doi.org/10.1016/j.jpubeco.2026.105675.

Splinter, D. (2020), 'U.S. tax progressivity and redistribution', *National Tax Journal* **73**(4), 1005–1024. doi: 10.17310/ntj.2020.4.04.

Wu, C. (2021), 'More unequal income but less progressive taxation', *Journal of Monetary Economics* **117**, 1097–1108. doi: 10.1016/j.jmoneco.2020.09.004.

Online Appendix

Residual Income Progression in the Twentieth and Twenty-First Centuries

This Online Appendix is organized as follows. Appendix A1 collects the proofs and the additional identification results used in the paper, including the role of rank preservation, concentration curves, grouped PPML, local average progressivity, and the relation to global progressivity indexes. Appendix A2 reports the validation and robustness exercises: concentration-curve versions of the main validation, simulated IRS-coarseness exercises, PSID validations for single households, and Monte Carlo simulations based on marginal Lorenz curves. Appendix A3 reports historical trend robustness under alternative parametric assumptions. Appendix A4 gives parametric examples of implied tax functions.

A1 Proofs of the main propositions and additional remarks on identification and estimation

A1.1 Main proofs

Lemma 1

Proof. By Assumption 3, X is invertible. Since strictly increasing functions preserve probabilities, $F_x[x(y)] = F_y(y)$. By Assumption 2, Q_x exists, so $x(y) = Q_x[F_y(y)]$. The result follows from $T(y) = y - x(y)$. ■

Proposition 1

Proof. Let $\widetilde{X} = \log X$ and $\widetilde{Y} = \log Y$. Since the quantile function of e^Z satisfies $Q_{e^Z}(p) =$

$e^{Q_Z(p)}$ for any random variable Z, Lemma 1 gives

$$X(Y) \;=\; e^{Q_{\widetilde{x}}[F_y(Y)]}.$$

Assumption 4 implies that $\widetilde{X}$ and $\widetilde{Y}$ lie in the same location-scale family, so there exist constants a and $b > 0$ with $Q_{\widetilde{x}}(p) = a + b\, Q_{\widetilde{y}}(p)$. Hence

$$X(Y) \;=\; e^{a + b \log Y} \;=\; e^{a}\, Y^{b}.$$

To identify a and b, note that $\log X(Y)$ and $\log Y$ must share the same standardized score:

$$\frac{\log X - \mu_{\log x}}{\sigma_{\log x}} \;=\; \frac{\log Y - \mu_{\log y}}{\sigma_{\log y}},$$

which gives $b = \sigma_{\log x}/\sigma_{\log y}$ and $a = \mu_{\log x} - (\sigma_{\log x}/\sigma_{\log y})\, \mu_{\log y}$. ■

A1.2 What the dispersion ratio measures when rank-preservation and log-location-scale distributions assumptions fail

The benchmark interpretation of our estimator relies on two restrictions: rank preservation and log-location-scale structure. Under these conditions, the ratio $\frac{\sigma_{\log X}}{\sigma_{\log Y}}$ identifies the constant residual-income elasticity of the implied power tax schedule. This is the structural interpretation used in the main text.

If either rank preservation or log-location-scale structure fails, this structural interpretation is no longer valid. In particular, the ratio need not correspond to a constant residual-income elasticity, nor to the slope of a single power mapping from pre-tax to after-tax income. Assumption 3, rank preservation, may be violated when a tax reform induces income reclassification across tax forms. The most notable example is TRA86, which substantially reduced incentives to shelter income as corporate retained earnings and caused a large realization of previously deferred capital gains. **?** estimate that roughly one-third of the apparent increase in the top-1% fiscal income share between 1986 and 1989 reflects this reclassification rather than real pre-tax income growth.

Nevertheless, the ratio continues to measure a well-defined distributional object: the relative dispersion of after-tax log income to pre-tax log income. In this sense, $\frac{\sigma_{\log X}}{\sigma_{\log Y}}$ remains an interpretable measure of distributional compression even outside the benchmark model.

To clarify what is contained in this object, write the conditional expectation function of after-tax log income given pre-tax log income as

$$\mathbb{E}[\log X \mid \log Y = t].$$

Then the law of total variance implies

$$\left(\frac{\sigma_{\log X}}{\sigma_{\log Y}}\right)^2 = \frac{\mathrm{Var}\left(\mathbb{E}[\log X \mid \log Y]\right)}{\mathrm{Var}(\log Y)} + \frac{\mathbb{E}\left[\mathrm{Var}(\log X \mid \log Y)\right]}{\mathrm{Var}(\log Y)}. \tag{A1}$$

The first term is the vertical component: variation in expected after-tax log income across the pre-tax income distribution. The second term is the horizontal component: residual dispersion in after-tax log income among units with the same pre-tax log income.

The vertical component can be further decomposed when $\mathbb{E}[\log X \mid \log Y = t]$ is absolutely continuous in t. Suppose, for notational simplicity, that $\log Y$ has support $[a, b]$. Define the Yitzhaki weight

$$\omega_{\log Y}(t) = \frac{\mathrm{Cov}\left(\log Y, \mathbf{1}\{\log Y \geq t\}\right)}{\mathrm{Var}(\log Y)}.$$

Then

$$\int_a^b \frac{\partial}{\partial t}\mathbb{E}[\log X \mid \log Y = t]\, \omega_{\log Y}(t)\, dt = \frac{\mathrm{Cov}(\log X, \log Y)}{\mathrm{Var}(\log Y)}. \tag{A2}$$

Thus, the Yitzhaki-weighted average of local residual-income elasticities is the population OLS slope from regressing $\log X$ on $\log Y$.

To describe the remaining part of the vertical component, define the probability measure $W_{\log Y}$ on $[a, b]^2$ by

$$dW_{\log Y}(s, t) = \frac{\mathrm{Cov}\left(\mathbf{1}\{\log Y \geq s\}, \mathbf{1}\{\log Y \geq t\}\right)}{\mathrm{Var}(\log Y)}\, ds\, dt.$$

This measure has marginal density $\omega_{\log Y}(t)$. Therefore,

$$\frac{\operatorname{Var}\left(\mathbb{E}[\log X \mid \log Y]\right)}{\operatorname{Var}(\log Y)} = \left[\int_a^b \frac{\partial}{\partial t}\mathbb{E}[\log X \mid \log Y = t]\,\omega_{\log Y}(t)\,dt\right]^2 + \operatorname{Cov}_{W_{\log Y}}\left(\frac{\partial}{\partial s}\mathbb{E}[\log X \mid \log Y = s], \frac{\partial}{\partial t}\mathbb{E}[\log X \mid \log Y = t]\right).$$

Combining this identity with (A1) gives

$$\left(\frac{\sigma_{\log X}}{\sigma_{\log Y}}\right)^2 = \underbrace{\left[\int_a^b \frac{\partial}{\partial t}\mathbb{E}[\log X \mid \log Y = t]\,\omega_{\log Y}(t)\,dt\right]^2}_{\text{weighted effective average progression}} + \underbrace{\operatorname{Cov}_{W_{\log Y}}\left(\frac{\partial}{\partial s}\mathbb{E}[\log X \mid \log Y = s], \frac{\partial}{\partial t}\mathbb{E}[\log X \mid \log Y = t]\right)}_{\text{CEF nonlinearity}} + \underbrace{\frac{\mathbb{E}\left[\operatorname{Var}(\log X \mid \log Y)\right]}{\operatorname{Var}(\log Y)}}_{\text{horizontal heterogeneity}}. \tag{A3}$$

Equation (A3) shows that the squared dispersion ratio contains three components. The first is the square of the average local residual-income elasticity, with weights determined by the pre-tax income distribution. This component equals the square of the population OLS slope from regressing $\log X$ on $\log Y$. The second captures nonlinearity in the conditional expectation function: local elasticities may reinforce or offset each other across different parts of the income distribution. The third captures horizontal heterogeneity: dispersion in after-tax log income among units with the same pre-tax log income.

The power-tax benchmark is the special case in which $\frac{\partial}{\partial t}\mathbb{E}[\log X \mid \log Y = t]$ is constant and there is no horizontal heterogeneity. Then the nonlinearity term is zero, the horizontal component is zero, and the dispersion ratio collapses to the constant residual-income elasticity:

$$\frac{\sigma_{\log X}}{\sigma_{\log Y}} = \frac{\partial}{\partial t}\mathbb{E}[\log X \mid \log Y = t].$$

Outside this benchmark, the same ratio should not be interpreted as a structural constant elasticity. It should instead be interpreted as a measure of total distributional compression, combining average vertical progression, nonlinear progression, and horizontal heterogeneity. The ratio itself is identified from the marginal distributions of $\log X$ and $\log Y$; the decomposition in (A3), however, requires joint information on pre-tax and after-tax income.

A1.3 Effective residual progression from concentration curves

The main text studies the least informative case in which we observe only the marginal distributions of before- and after-tax income. In that environment, rank preservation is required to pair before-tax and after-tax quantiles. We now consider the more informative case in which after-tax income shares are observed over groups ranked by before-tax income. These data identify concentration curves.

Let Y denote before-tax income and X after-tax income. We allow after-tax income to depend not only on before-tax income, but also on other tax-relevant characteristics:

$$X = Y - T(Y, \xi),$$

where ξ collects deductions, credits, transfers, filing status, household composition, income composition, avoidance, reporting behavior, and other determinants of tax liability conditional on income. The relevant object is therefore not the realized individual tax schedule, but the conditional mean schedule

$$m(y) \equiv \mathbb{E}[X \mid Y = y].$$

Equivalently, define the effective tax function

$$\bar{T}(y) \equiv y - m(y) = \mathbb{E}[T(Y, \xi) \mid Y = y].$$

The corresponding effective residual-income elasticity is

$$\epsilon_C(y) \equiv \frac{\partial \ln m(y)}{\partial \ln y} = \frac{1 - \bar{T}'(y)}{1 - \bar{T}(y)/y}, \tag{A4}$$

whenever the derivative exists. Thus $\epsilon_C(y) < 1$ means that average after-tax income rises less than proportionally with before-tax income.

We work under the following assumptions.

Assumption A1 (Regular joint distribution)**.** *The pair (X, Y) has joint distribution $F_{X,Y}$, with $X > 0, Y > 0$, and $\mathbb{E}[X] < \infty$. The marginal distribution F_Y is strictly increasing and continuous on its support. The conditional mean $m(y) = \mathbb{E}[X \mid Y = y]$ exists, is positive, and is continuous.*

Assumption A2 (Monotone conditional mean)**.** *The conditional mean $m(y) = \mathbb{E}[X \mid Y = y]$ is strictly increasing in y.*

Assumption A2 replaces rank preservation of realized after-tax income. It allows taxpayers with the same before-tax income to have different after-tax incomes, and it allows realized reranking in the after-tax-income distribution. The restriction is only that average after-tax income is increasing in before-tax income.

Because $m(\cdot)$ is strictly increasing, the random variable $\mathbb{E}[X \mid Y]$ preserves the rank of Y. Hence

$$Q_{\mathbb{E}[X|Y]}(F_Y(Y)) = \mathbb{E}[X \mid Y],$$

where $Q_{\mathbb{E}[X|Y]}$ denotes the quantile function of the random variable $\mathbb{E}[X \mid Y]$.

Lemma 2 (Effective Tax Function)**.** *Under Assumptions A1–A2,*

$$\bar{T}(Y) = Y - Q_{\mathbb{E}[X|Y]}[F_Y(Y)]\,,$$

where $Q_{\mathbb{E}[X|Y]}$ denotes the quantile function of the random variable $\mathbb{E}[X \mid Y]$.

Lemma 2 states that, when average after-tax income is increasing in before-tax income, the effective after-tax schedule is recovered by assigning each taxpayer their before-tax rank and retrieving conditional-mean after-tax income at the same rank. We call $\bar{T}(Y)$ the effective tax function of the joint distribution $F_{X,Y}$. This is the concentration-curve analogue of Lemma 1. The difference is that the marginal case uses the quantile function of realized after-tax income, Q_X, whereas the concentration-curve case uses the quantile function of $\mathbb{E}[X \mid Y]$.

We now ask when this effective schedule takes the power form used in the modern progressivity literature. Under Assumption A2, the conditional mean schedule maps each pre-tax rank into the same rank of conditional-mean after-tax income. If, in addition, before-tax income and conditional-mean after-tax income differ only by location and scale in logs, this quantile map is log-affine, and the effective schedule is a power function.

Assumption A3 (Log-Location-Scale Conditional Mean)**.** *Pretax income Y and conditional-mean after-tax income $\mathbb{E}[X \mid Y]$ belong to the same log-location-scale family: there exist constants $a \in \mathbb{R}$ and $b > 0$ such that*

$$Q_{\log \mathbb{E}[X|Y]}(p) = a + b\, Q_{\log Y}(p) \quad \textit{for all } p \in (0,1),$$

where $Q_{\log \mathbb{E}[X|Y]}$ denotes the quantile function of the random variable $\log \mathbb{E}[X \mid Y]$.

Under Assumption A3, the effective tax function in Lemma 2 takes the following form:

Proposition A1 (Log-Location-Scale Identification of Effective Residual Progression)**.** *Under Assumptions A1–A3,*

$$\bar{T}(Y) = Y - \exp\left(\mathbb{E}[\log \mathbb{E}[X \mid Y]] - \frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}\mathbb{E}[\log Y]\right) Y^{\sigma_{\log \mathbb{E}[X|Y]}/\sigma_{\log Y}}.$$

In particular, the effective tax function takes the power form

$$\mathbb{E}[X \mid Y] = \lambda_C Y^{1-\tau_C}$$

with

$$\lambda_C = \exp\left(\mathbb{E}[\log \mathbb{E}[X \mid Y]] - \frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}\mathbb{E}[\log Y]\right), \qquad 1 - \tau_C = \frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}.$$

Proposition A1 is the concentration-curve analogue of Proposition 1. It shows that, under the log-location-scale benchmark, effective residual progression is identified by the relative dispersion of log conditional-mean after-tax income and log before-tax income.

A more progressive tax-and-transfer system compresses the conditional-mean after-tax schedule more strongly relative to the before-tax distribution, yielding a smaller value of

$$\frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}.$$

The interpretation of this ratio follows the interpretation of the effective tax function in Lemma 2. If $X = \mathbb{E}[X \mid Y]$ almost surely, then effective residual progression coincides with the structural residual-income elasticity of the tax schedule. More generally, the ratio is a conditional-mean object. It measures how much average after-tax income, evaluated along the before-tax income distribution, is compressed in logs relative to before-tax income. Horizontal heterogeneity around $\mathbb{E}[X \mid Y]$ reflects deductions, credits, transfers, filing status, income composition, avoidance, reporting behavior, and other factors that affect tax liability conditional on income.

It remains to connect this object to concentration-curve data. Let $C_X^Y(p)$ denote the concentration curve of after-tax income with respect to before-tax income:

$$C_X^Y(p) = \frac{\mathbb{E}[X\mathbf{1}\{Y \leq Q_Y(p)\}]}{\mathbb{E}[X]}.$$

By the law of iterated expectations,

$$C_X^Y(p) = \frac{\mathbb{E}[\mathbb{E}[X \mid Y]\mathbf{1}\{Y \leq Q_Y(p)\}]}{\mathbb{E}[\mathbb{E}[X \mid Y]]}.$$

Since $\mathbb{E}[X \mid Y]$ is strictly increasing in Y, this is exactly the Lorenz curve of the random variable $\mathbb{E}[X \mid Y]$:

$$C_X^Y(p) = L_{\mathbb{E}[X|Y]}(p).$$

Therefore, wherever the concentration curve is differentiable,

$$(C_X^Y)'(p) = L'_{\mathbb{E}[X|Y]}(p) = \frac{Q_{\mathbb{E}[X|Y]}(p)}{\mathbb{E}[\mathbb{E}[X \mid Y]]} = \frac{Q_{\mathbb{E}[X|Y]}(p)}{\mathbb{E}[X]}.$$

Since adding the constant $\log \mathbb{E}[X]$ does not affect variance,

$$\mathrm{Var}(\log \mathbb{E}[X \mid Y]) = \int_0^1 \left(\log(C_X^Y)'(p) - \int_0^1 \log(C_X^Y)'(u)\, du \right)^2 dp. \tag{A5}$$

Similarly, the before-tax Lorenz curve identifies

$$\mathrm{Var}(\log Y) = \int_0^1 \left(\log L_Y'(p) - \int_0^1 \log L_Y'(u)\, du \right)^2 dp.$$

Thus the effective residual-progression parameter is identified from the shape of the after-tax concentration curve and the before-tax Lorenz curve:

$$1 - \tau_C = \frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}. \tag{A6}$$

The means of before- and after-tax income are required to recover the effective tax function in levels, but not to recover the effective residual-progression ratio. The ratio depends only on the relative log dispersion of the conditional-mean after-tax schedule and the before-tax income distribution.

A1.4 What the concentration-curve ratio measures when log-location-scale structure fails

The concentration-curve estimator relies on a different object from the marginal-Lorenz estimator. Instead of comparing the marginal distributions of realized after-tax and pre-tax income, it compares pre-tax income to average after-tax income conditional on pre-tax income. The relevant object is therefore

$$\mathbb{E}[X \mid Y],$$

evaluated along the distribution of Y.

Under the log-location-scale benchmark for Y and $\mathbb{E}[X \mid Y]$, the ratio

$$\frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}$$

identifies the constant elasticity of the effective residual-income schedule

$$\mathbb{E}[X \mid Y = y] = \lambda_C y^{1-\tau_C}.$$

This is the effective residual-progression interpretation developed in Appendix A1.3.

If log-location-scale structure fails, this constant-elasticity interpretation is no longer exact. In particular, the ratio need not equal the slope of a single power mapping from pre-tax income to average after-tax income. It nevertheless remains a well-defined object: the relative dispersion of $\log \mathbb{E}[X \mid Y]$ to $\log Y$. Thus it measures how much the conditional mean after-tax schedule compresses the pre-tax income distribution in logs.

This object differs from the marginal-Lorenz ratio studied in Appendix A1.2. The marginal ratio $\sigma_{\log X}/\sigma_{\log Y}$ contains both vertical variation in expected after-tax log income and horizontal dispersion in realized after-tax log income among units with the same pre-tax income. By contrast, the concentration-curve ratio is based on $\mathbb{E}[X \mid Y]$. Since this object is already averaged within each pre-tax income level, horizontal variation in realized after-tax income is averaged out before the dispersion ratio is formed.

The ratio can be written as

$$\left(\frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}\right)^2 = \frac{\text{Var}\left(\log \mathbb{E}[X \mid Y]\right)}{\text{Var}(\log Y)}.$$

There is no law-of-total-variance residual term because $\log \mathbb{E}[X \mid Y]$ is a deterministic function of Y.

Suppose that $\log Y$ has support $[a, b]$, and that $\log \mathbb{E}[X \mid \log Y = t]$ is absolutely continuous in t. Define the Yitzhaki weight

$$\omega_{\log Y}(t) = \frac{\text{Cov}\left(\log Y, \mathbf{1}\{\log Y \geq t\}\right)}{\text{Var}(\log Y)}.$$

Then

$$\int_a^b \frac{\partial}{\partial t} \log \mathbb{E}[X \mid \log Y = t]\, \omega_{\log Y}(t)\, dt = \frac{\text{Cov}\left(\log \mathbb{E}[X \mid Y], \log Y\right)}{\text{Var}(\log Y)}. \tag{A7}$$

Thus, the Yitzhaki-weighted average of local effective residual-income elasticities is the population OLS slope from regressing $\log \mathbb{E}[X \mid Y]$ on $\log Y$.

To describe the remaining part of the concentration-curve ratio, define the probability measure $W_{\log Y}$ on $[a, b]^2$ by

$$dW_{\log Y}(s, t) = \frac{\text{Cov}\left(\mathbf{1}\{\log Y \geq s\}, \mathbf{1}\{\log Y \geq t\}\right)}{\text{Var}(\log Y)}\, ds\, dt.$$

This measure has marginal density $\omega_{\log Y}(t)$. Therefore,

$$\left(\frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}}\right)^2 = \underbrace{\left[\int_a^b \frac{\partial}{\partial t} \log \mathbb{E}[X \mid \log Y = t]\, \omega_{\log Y}(t)\, dt\right]^2}_{\text{weighted effective average progression}}$$
$$+ \underbrace{\text{Cov}_{W_{\log Y}}\left(\frac{\partial}{\partial s} \log \mathbb{E}[X \mid \log Y = s], \frac{\partial}{\partial t} \log \mathbb{E}[X \mid \log Y = t]\right)}_{\text{effective CEF nonlinearity}}. \tag{A8}$$

Equation (A8) shows that the squared concentration-curve ratio contains two components. The first is the square of the average local effective residual-income elasticity, with weights determined by the pre-tax income distribution. This component equals the square of the population OLS slope from regressing $\log \mathbb{E}[X \mid Y]$ on $\log Y$. The second captures nonlinearity in the effective conditional mean schedule: local elasticities may reinforce or offset each other across different parts of the income distribution.

The key difference from the marginal-Lorenz decomposition is the absence of a horizontal heterogeneity term. This does not mean that horizontal heterogeneity in realized after-tax income is absent. Rather, concentration curves identify the conditional mean schedule

$\mathbb{E}[X \mid Y = y]$, so variation in realized after-tax income around that schedule is not part of the identified ratio. The ratio is therefore a measure of effective vertical compression, plus any nonlinearity in the effective conditional mean schedule.

The power effective-tax benchmark is the special case in which

$$\frac{\partial}{\partial t} \log \mathbb{E}[X \mid \log Y = t]$$

is constant. Then the nonlinearity term is zero, and the concentration-curve ratio collapses to the constant effective residual-income elasticity:

$$\frac{\sigma_{\log \mathbb{E}[X|Y]}}{\sigma_{\log Y}} = \frac{\partial}{\partial t} \log \mathbb{E}[X \mid \log Y = t].$$

Outside this benchmark, the same ratio should not be interpreted as a structural constant elasticity. It should instead be interpreted as effective residual progression: the total log compression of the conditional-mean after-tax schedule relative to the pre-tax income distribution. The ratio itself is identified from the after-tax concentration curve and the before-tax Lorenz curve; the decomposition in (A8), however, requires recovering the local effective elasticity

$$\frac{\partial}{\partial t} \log \mathbb{E}[X \mid \log Y = t],$$

and therefore requires enough information to estimate the slope of the conditional mean schedule along the pre-tax income distribution.

A1.5 Grouped PPML and within-bin reranking

This appendix clarifies the conditions under which the grouped PPML share-mapping estimator has a structural interpretation when only marginal Lorenz tabulations are observed.

Let $U = F_Y(Y)$ denote the pre-tax rank and $V = F_X(X)$ the after-tax rank. For a reported bin $B_j = (p_{j-1}, p_j]$, define the after-tax income share of the pre-tax rank cell as

$$S_j^{X|Y} \equiv \frac{\mathbb{E}\left[X\mathbf{1}\{U \in B_j\}\right]}{\mathbb{E}[X]}.$$

This is the object that would be observed if after-tax income shares were tabulated by pre-tax income rank. By contrast, a marginal after-tax Lorenz tabulation reports

$$\Delta L_j^x = \frac{\mathbb{E}\left[X\mathbf{1}\{V \in B_j\}\right]}{\mathbb{E}[X]}.$$

These two objects coincide under rank preservation, but rank preservation is stronger than necessary. It is enough that reranking does not move units across reported bins:

$$\mathbf{1}\{U \in B_j\} = \mathbf{1}\{V \in B_j\} \quad \text{a.s. for all } j.$$

This condition allows arbitrary reranking within each reported bin. Since the grouped data identify only total income mass inside each cell, within-bin reranking is irrelevant for the grouped share-mapping estimator.

To see the role of cross-bin reranking, write the observed marginal after-tax share as

$$\Delta L_j^x = S_j^{X|Y} + e_j,$$

where e_j is the cross-bin rank-assignment error. Since both $\{\Delta L_j^x\}_j$ and $\{S_j^{X|Y}\}_j$ are income-share vectors, $\sum_j e_j = 0$.

The grouped PPML estimator uses the fitted share

$$\mu_j(\epsilon) = \frac{\Delta p_j \left(\frac{\Delta L_j^y}{\Delta p_j}\right)^{\epsilon}}{\sum_\ell \Delta p_\ell \left(\frac{\Delta L_\ell^y}{\Delta p_\ell}\right)^{\epsilon}}.$$

Equivalently, it solves the PPML score

$$\sum_j \left[\Delta L_j^x - \mu_j(\epsilon)\right] \log\left(\frac{\Delta L_j^y}{\Delta p_j}\right) = 0, \tag{A9}$$

with the intercept imposing $\sum_j \mu_j(\epsilon) = 1$.

If the structural grouped share mapping satisfies

$$S_j^{X|Y} = \mu_j(\epsilon_0) \quad \text{for all } j,$$

then the PPML score evaluated at ϵ_0 is

$$\sum_j \left[\Delta L_j^x - \mu_j(\epsilon_0)\right] \log\left(\frac{\Delta L_j^y}{\Delta p_j}\right) = \sum_j e_j \log\left(\frac{\Delta L_j^y}{\Delta p_j}\right).$$

The sign of this term is not arbitrary. Because ΔL_j^x is computed from the after-tax Lorenz ordering, for every cutoff p_m,

$$\sum_{j \le m} \Delta L_j^x = L_X(p_m) \le \frac{\mathbb{E}[X\mathbf{1}\{U \le p_m\}]}{\mathbb{E}[X]} = \sum_{j \le m} S_j^{X|Y}.$$

Thus, with

$$E_m = \sum_{j \le m} e_j,$$

we have $E_m \le 0$ for all $m < J$ and $E_J = 0$. Let

$$w_j = \log\left(\frac{\Delta L_j^y}{\Delta p_j}\right).$$

Since pre-tax cells are ordered by income, w_j is weakly increasing in j. Summation by parts gives

$$\sum_j e_j w_j = \sum_{m=1}^{J-1} E_m(w_m - w_{m+1}) \ge 0.$$

If w_j is strictly increasing across reported cells, this expression is zero only when $E_m = 0$ at every reported cutoff. Hence generic cross-bin reranking does not average out in the grouped PPML score. It moves the observed marginal after-tax Lorenz shares away from the pre-tax-rank-cell shares in a systematic direction.

The preceding discussion applies when only marginal after-tax Lorenz shares are observed. If instead the data provide concentration-curve shares, the rank-assignment problem disappears. Let

$$\Delta C_j^{X|Y} \equiv C_X^Y(p_j) - C_X^Y(p_{j-1}) = \frac{\mathbb{E}\left[X\mathbf{1}\{U \in B_j\}\right]}{\mathbb{E}[X]}.$$

Then

$$\Delta C_j^{X|Y} = S_j^{X|Y}$$

by construction. Hence the grouped PPML score based on concentration shares is

$$\sum_j \left[\Delta C_j^{X|Y} - \mu_j(\epsilon)\right] \log\left(\frac{\Delta L_j^y}{\Delta p_j}\right) = 0,$$

with no cross-bin rank-assignment error. Realized after-tax ranks may move across reported bins, but this reranking is irrelevant because the concentration tabulation already cumulates after-tax income over the pre-tax ranking. Thus, ignoring coarseness of the reported bins, concentration curves identify the correct grouped share-mapping object directly. No full rank-preservation or within-bin rank-preservation condition is needed to replace after-tax marginal shares with pre-tax-rank after-tax shares.

The remaining restrictions are therefore not rank-assignment restrictions but interpretation restrictions. If the pre-tax-rank share mapping satisfies the power form $S_j^{X|Y} = \mu_j(\epsilon_0)$, then grouped PPML recovers ϵ_0 from concentration shares. With heterogeneous taxes and transfers, this parameter should be interpreted as effective residual progression of the conditional-mean after-tax schedule, rather than as the elasticity of a deterministic individual tax schedule.

In conclusion, when only marginal after-tax Lorenz tabulations are observed, grouped PPML does not require full individual-level rank preservation, but it does require a weaker and specific condition: no reranking across reported bin cutoffs, or equivalently equality between after-tax marginal shares and after-tax shares tabulated by pre-tax rank at the reported cutoffs. Arbitrary reranking within reported cells is harmless, because grouped shares do not use within-cell ranks. Cross-bin reranking, however, generally changes the PPML score. By contrast, when concentration-curve shares are observed, the relevant pre-tax-rank after-tax shares are observed directly, so this rank-preservation requirement is unnecessary. In that case, grouped PPML targets the pre-tax-rank share mapping itself; its interpretation depends on whether that mapping is viewed as a structural tax schedule or as an effective conditional-mean schedule.

A1.6 Estimating the average local progressivity

We estimate $\bar{\tau}^{L}_{20,100}$ from WID one-percentile tabulations. For income concept z and cell j, the cell average relative to the overall mean is $\hat{q}^z_j = \bar{z}_j/\bar{z} = \Delta L^z_j/\Delta p_j$ with $\Delta p_j = 0.01$, evaluated at the midpoint $p_j = (j+0.5)/100$; let $s^z_j = \log \hat{q}^z_j$. For each year and midpoint we estimate $\partial_p \log q_z(p_j)$ for pre- and after-tax income by a local linear regression of s^z_ℓ on p_ℓ over an 11-cell moving window (interior cells use five cells on each side; endpoint cells use the nearest available 11). With estimated slopes $\hat{b}^y_j$ and $\hat{b}^x_j$, the plug-in estimates are $\hat{\epsilon}(p_j) = \hat{b}^x_j/\hat{b}^y_j$ and $\hat{\tau}(p_j) = 1 - \hat{b}^x_j/\hat{b}^y_j$, and the annual summary statistic is

$$\widehat{\bar{\tau}}^{L}_{20,100} = \frac{1}{N_t} \sum_{j:\, p_j \in [0.20,1),\, |\hat{b}^y_j| \geq 10^{-4}} \hat{\tau}(p_j),$$

where N_t is the number of valid percentile cells in year t.

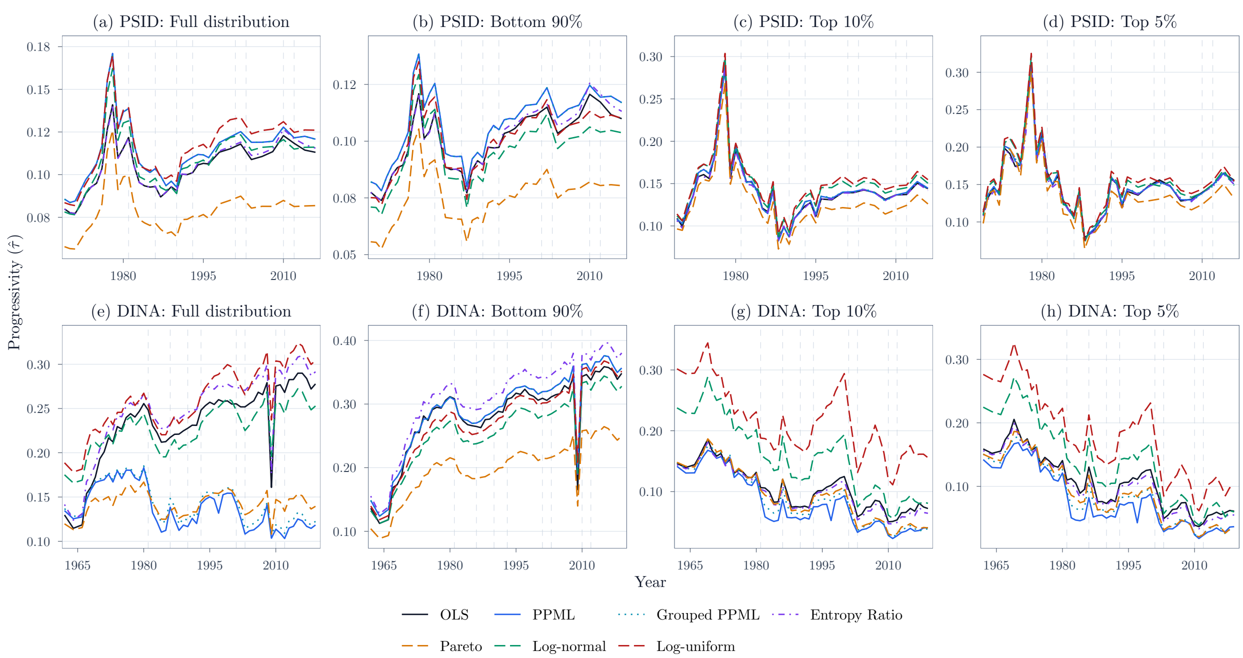
A1.7 Relationship to global progressivity indeces

Our residual elasticity $\epsilon = 1-\tau$ and the Kakwani index K both measure effective progressivity, but they are conceptually distinct. The Kakwani index is defined as $K = C_T - G_Y$, where C_T is the concentration coefficient of tax liabilities and G_Y is the Gini coefficient of pretax income. Under the power tax function, $K = G_Y(1-\epsilon)/(1-t_0)$ where t_0 is the average tax rate, so K is a scaled version of $1-\epsilon$ adjusted for the level of taxation. This means the Kakwani index and residual progression can move in opposite directions if the average tax rate changes. Specifically, if τ falls (progressivity declines) but t_0 also falls (lower tax level), K may still rise—which is consistent with the simultaneous decline in our top-income elasticity and the rise in Splinter (2020) Kakwani index: part of the increase in the Kakwani index reflects a declining tax level (numerator effect) rather than a change in the shape of the tax schedule.

A2 Validation Analysis: Additional Exercises

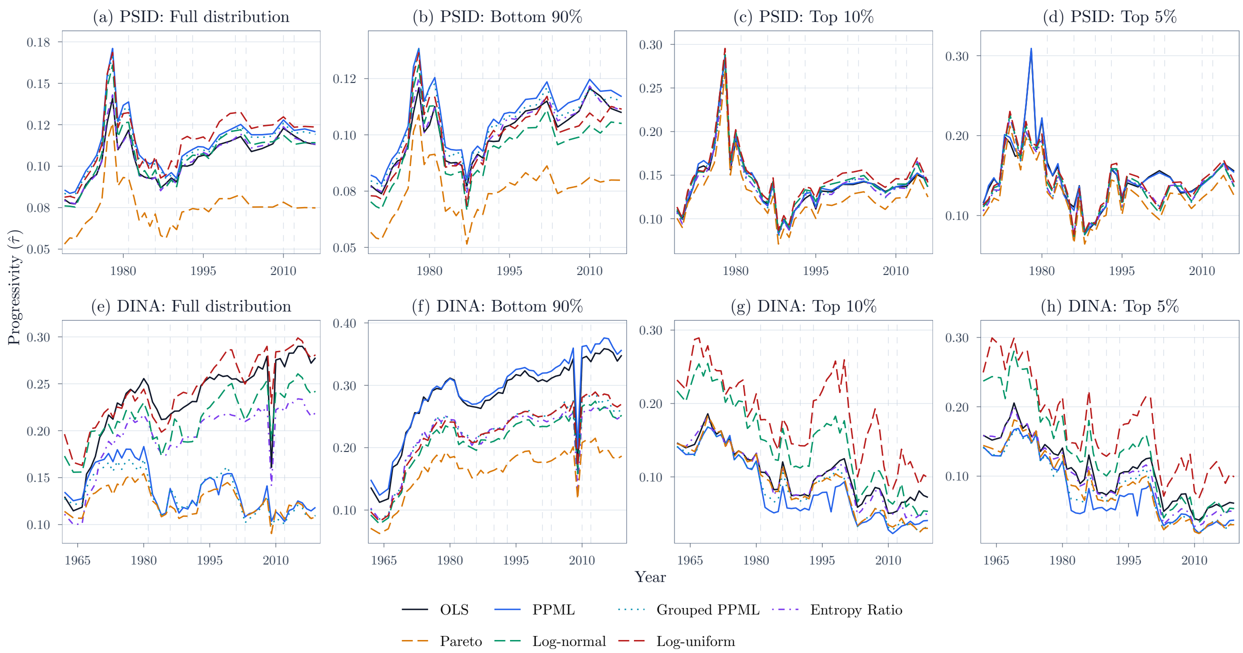
A2.1 Validation robustness

Figure A2.1: PSID and DINA validation using concentration curves



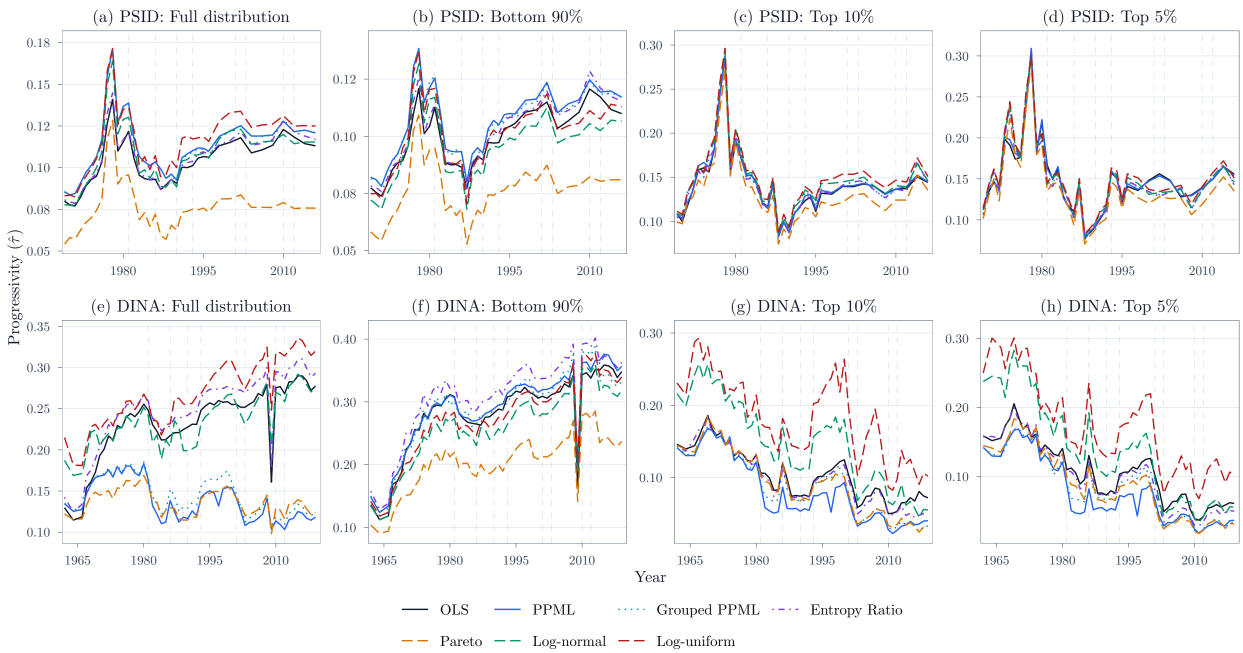
Notes: This figure repeats the main validation exercise using after-tax concentration curves rather than marginal after-tax Lorenz curves. After-tax grouped shares are cumulated over groups ranked by before-tax income. Higher values correspond to greater residual progressivity. OLS and PPML are estimated directly from microdata; the remaining estimates use grouped income-share tabulations constructed from the same samples. Each panel uses its own vertical scale.

Figure A2.2: PSID and DINA validation under simulated IRS coarseness, marginal Lorenz curves



Notes: This figure repeats the main validation exercise after replacing the baseline validation grid with a synthetic IRS-style grouped tabulation. For each validation year, we construct the percentile cutoffs implied by the historical IRS tabulation grid and estimate progressivity using only the grouped income shares that remain available within each population range. After-tax grouped shares are treated as marginal Lorenz shares. Each panel uses its own vertical scale.

Figure A2.3: PSID and DINA validation under simulated IRS coarseness, concentration curves

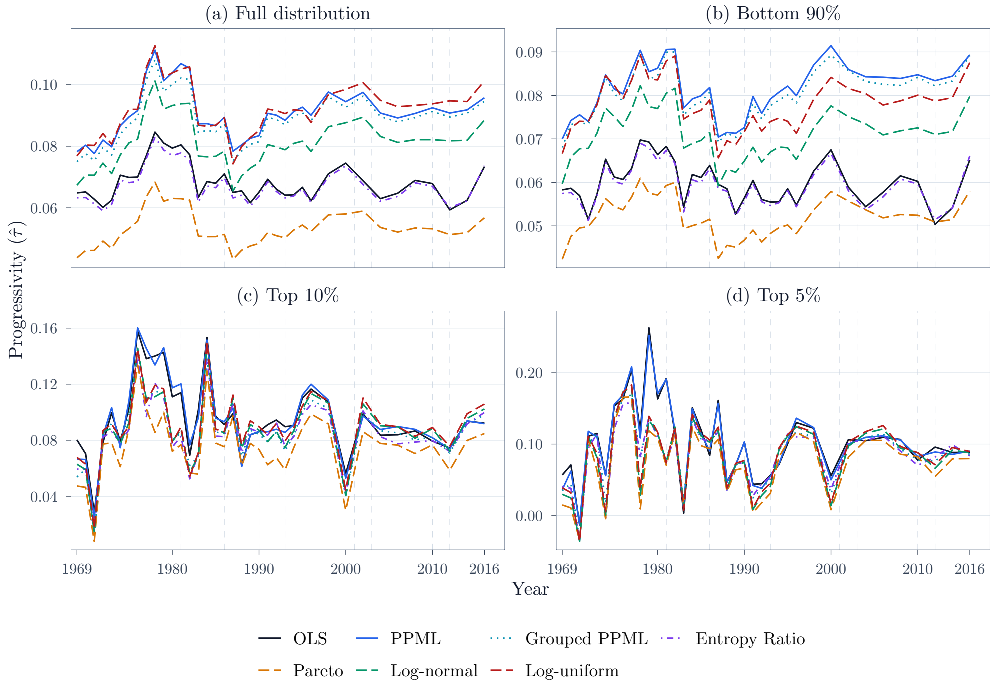


Notes: This figure repeats the simulated IRS-coarseness validation using concentration curves. After-tax grouped shares are cumulated over groups ranked by before-tax income, matching the information structure of IRS tabulations with before-tax ranking. Each panel uses its own vertical scale.

A2.2 PSID Validation for Singles

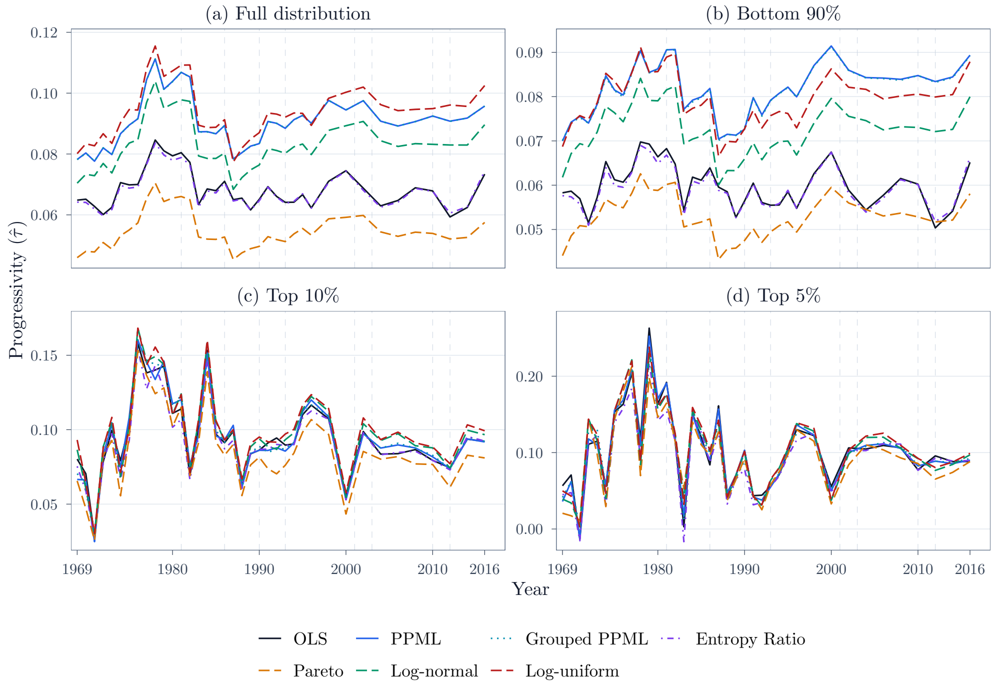
Figures A2.4–A2.7 repeat the validation exercise for single households. The four panels in each figure report the full distribution, bottom 90 percent, top 10 percent, and top 5 percent. The comparison separates two sources of robustness: the coarseness of the grouped tabulation, using either the WID-style validation grid or the historical IRS-style grid, and the information used to form after-tax shares, using either marginal Lorenz curves or concentration curves. As in the married-sample and DINA validations, concentration curves tend to track the microdata benchmarks more closely because they preserve before-tax ranking and therefore remove the reranking ambiguity present in marginal after-tax Lorenz curves. IRS-style coarsening is the stricter exercise and therefore produces larger deviations, especially in the upper-tail panels where the PSID sample has less support.

Figure A2.4: PSID singles validation, WID-style coarseness and marginal Lorenz curves



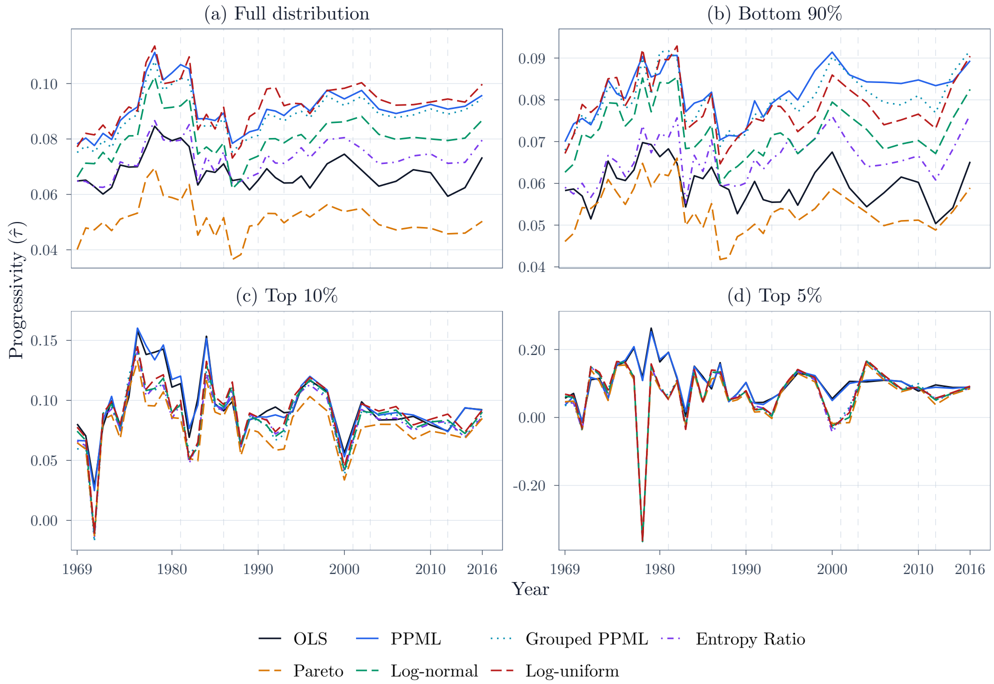
Notes: This figure repeats the PSID validation exercise for single households. Panels report the full distribution, bottom 90 percent, top 10 percent, and top 5 percent. Microdata OLS and PPML estimates are compared with grouped-data estimates constructed from WID-style synthetic income-share tabulations. After-tax grouped shares are treated as marginal Lorenz shares. Higher values correspond to greater residual progressivity. Each panel uses its own vertical scale.

Figure A2.5: PSID singles validation, WID-style coarseness and concentration curves



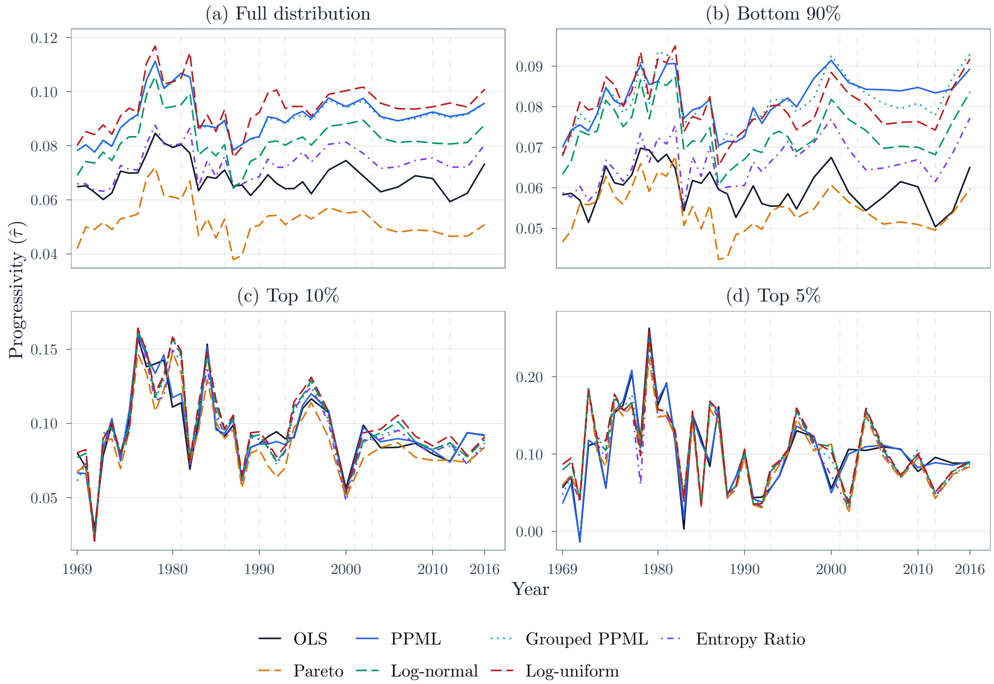
Notes: This figure repeats the PSID singles validation using WID-style synthetic grouped tabulations and concentration curves. After-tax income shares are cumulated over groups ranked by before-tax income, preserving the before-tax ranking. Higher values correspond to greater residual progressivity. Each panel uses its own vertical scale.

Figure A2.6: PSID singles validation, IRS-style coarseness and marginal Lorenz curves



Notes: This figure repeats the PSID singles validation after replacing the WID-style validation grid with a synthetic IRS-style grouped tabulation. After-tax grouped shares are treated as marginal Lorenz shares. Higher values correspond to greater residual progressivity. Each panel uses its own vertical scale.

Figure A2.7: PSID singles validation, IRS-style coarseness and concentration curves

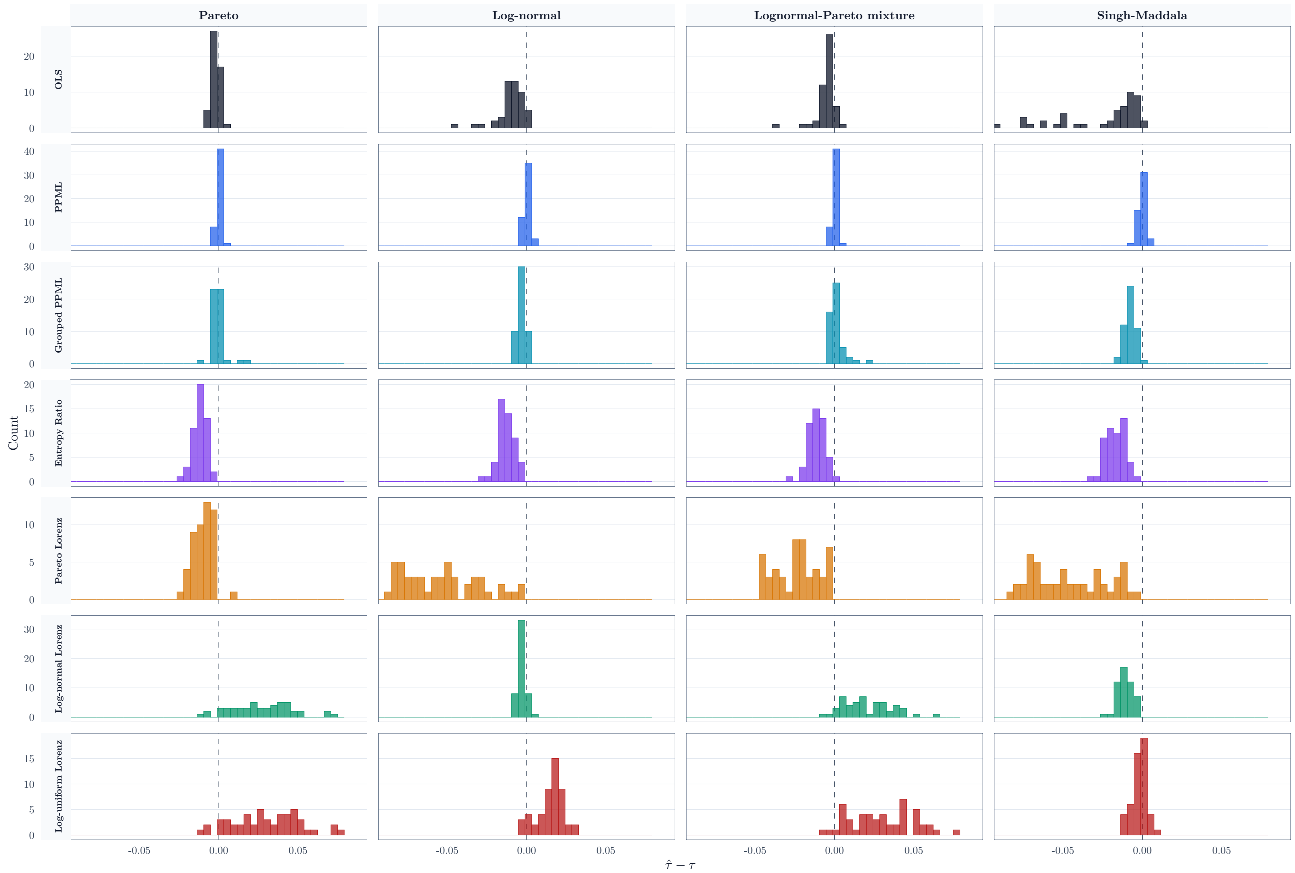


Notes: This figure repeats the PSID singles validation using synthetic IRS-style grouped tabulations and concentration curves. After-tax income shares are cumulated over groups ranked by before-tax income. Higher values correspond to greater residual progressivity. Each panel uses its own vertical scale.

A2.3 Monte Carlo simulations

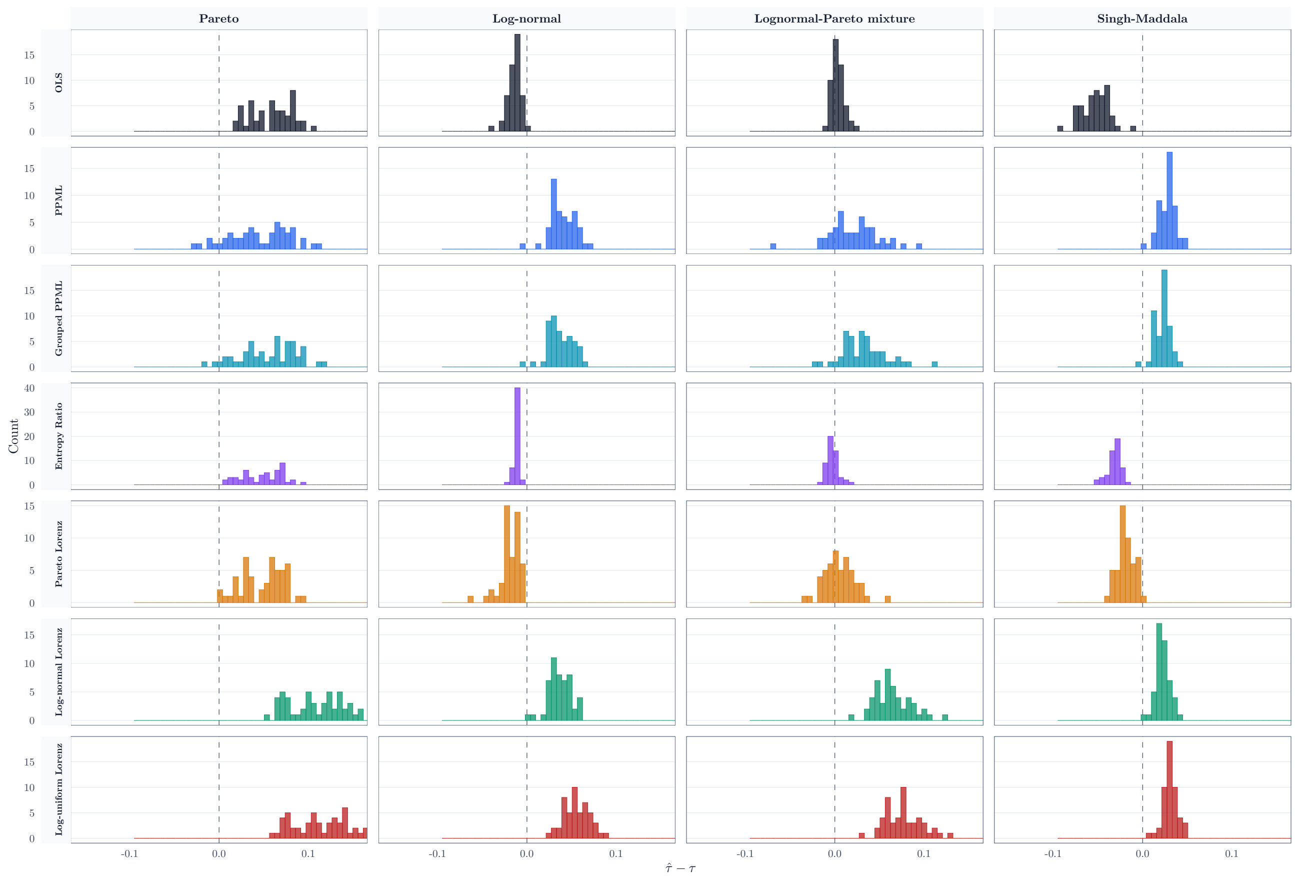
Figures A2.8 and A2.9 evaluate estimator performance in simulated data. Each replication draws a paired before-tax and after-tax microdata sample, constructs grouped marginal Lorenz curves for before-tax and after-tax income separately, and compares the resulting estimates with the simulation truth. The exercise therefore matches the least-informative data environment emphasized in the main text: only the two marginal distributions are used, not concentration curves. Under the power-tax design, the estimand is exactly aligned with the residual progression parameter. Under the bracket-tax design, the true object is the average progressivity implied by the nonlinear bracket schedule, so the exercise measures how the estimators behave when the power-tax approximation is misspecified.

Figure A2.8: Monte Carlo simulations under a true power tax function



Notes: The figure reports the distribution of estimation errors, $\hat{\tau} - \tau$, under a data-generating process in which after-tax income is generated by a power tax function. Columns distinguish the before-tax income distribution used in the simulation. Rows distinguish estimators. Grouped estimators use separately sorted before-tax and after-tax marginal Lorenz curves. The vertical dashed line marks zero error.

Figure A2.9: Monte Carlo simulations under a bracketed tax function

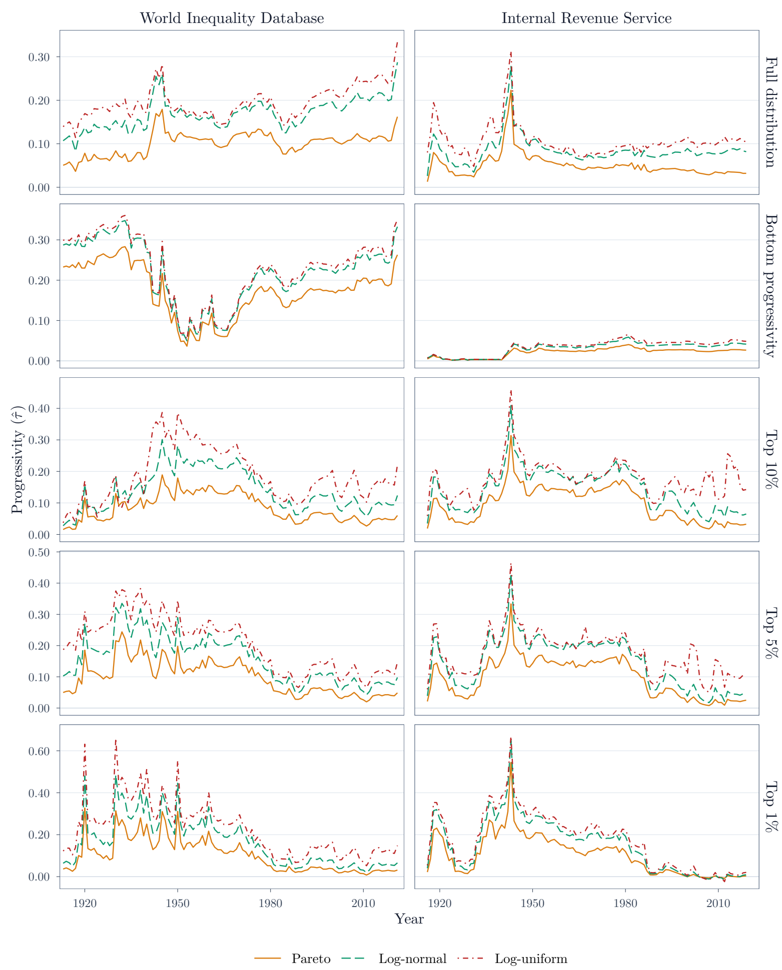


Notes: The figure reports the distribution of estimation errors, $\hat{\tau} - \tau$, when after-tax income is generated by a nonlinear bracket tax schedule rather than a power tax function. Columns distinguish the before-tax income distribution used in the simulation. Rows distinguish estimators. Grouped estimators use separately sorted before-tax and after-tax marginal Lorenz curves. The vertical dashed line marks zero error.

A3 Robustness checks

A3.1 Historical progressivity trends under parametric assumptions

Figure A3.1: Parametric Robustness for the Historical Progressivity Estimates

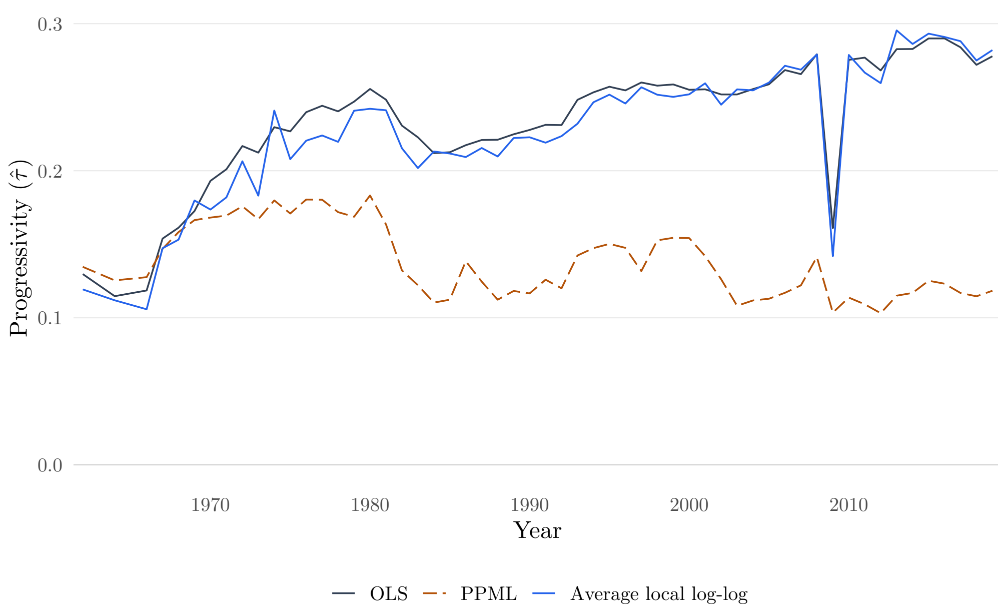


Notes: The figure reproduces the historical progressivity analysis using parametric Lorenz-curve estimators instead of the baseline nonparametric estimators. Columns distinguish the World Inequality Database and Internal Revenue Service income concepts. Rows report estimates for the full distribution, bottom progressivity, top 10 percent, top 5 percent, and top 1 percent. The full-distribution World Inequality Database estimates use percentiles 20–100, while bottom progressivity uses the bottom 90 percent of the relevant observed population range. Lines correspond to Pareto, log-normal, and log-uniform specifications. Lower values indicate lower effective tax progressivity.

A3.2 DINA Local Log-Log Elasticity

Figure A3.2 uses the matched DINA microdata to estimate progressivity without imposing a single global log-log slope. For each year, we run local quadratic regressions of log after-tax income on log before-tax income, evaluate the implied residual elasticity at each individual, and average these elasticities using DINA weights. The resulting local-average progressivity series is close to the OLS benchmark and rises after 1980. PPML remains lower and flatter, consistent with the main validation results: because PPML fits levels rather than logs, it assigns more influence to high-income observations, where post-1980 progressivity falls rather than rises.

Figure A3.2: DINA local log-log progressivity compared with OLS and PPML



Notes: The figure compares three DINA microdata estimates for the full distribution. OLS estimates a single weighted regression of log after-tax income on log before-tax income. PPML estimates after-tax income in levels with a log before-tax-income index. The local log-log series estimates a local quadratic regression of log after-tax income on log before-tax income, evaluates the derivative at each observation, and averages those individual elasticities using DINA weights. The figure reports progressivity, $\tau = 1 - \epsilon$, so higher values denote greater residual progressivity.

A4 Parametric Examples of Implied Tax Functions

This section derives the implied tax function for six distributional families: Pareto, log-normal, log-uniform, Weibull, normal, shifted exponential, and generalized Pareto. All derivations proceed by applying Proposition 1 (main text) to the relevant quantile functions.

A4.1 Case 1: Pareto Distribution

The Pareto distribution is the canonical model of the upper tail of the income distribution, supported by extensive empirical evidence (Blanchet *et al.*, 2022; Gabaix, 2009; Pareto, 1895). Formally, a random variable Z is Pareto distributed with shape $\eta > 1$ and scale $b > 0$, written $Z \sim \text{Pa}(\eta, b)$, if its CDF is

$$F_Z(z) \;=\; 1 - \left(\frac{b}{z}\right)^{\eta}, \quad z \geq b.$$

The quantile function is $Q_Z(p) = b\,(1-p)^{-1/\eta}$, the mean is $\mu_Z = b\eta/(\eta-1)$ (for $\eta > 1$), and $\sigma_{\log Z} = 1/\eta$.

Proposition A2. *If $Y \sim \text{Pa}(\eta_y, b_y)$ and $X \sim \text{Pa}(\eta_x, b_x)$, the implied tax function under Assumptions 1–3 (main text) is*

$$T(Y) \;=\; Y \;-\; \left(\frac{b_x^{\eta_x}}{b_y^{\eta_y}}\right)^{1/\eta_x} Y^{\,\eta_y/\eta_x},$$

a power tax function with residual elasticity $\epsilon = \eta_y/\eta_x$.

Proof. From Proposition 1 (main text),

$$X(Y) = Q_x[F_y(Y)] = b_x\big[1 - F_y(Y)\big]^{-1/\eta_x} = b_x\left(\frac{b_y}{Y}\right)^{\eta_y/\eta_x} \cdot \frac{1}{b_x^0}$$

Working through:

$$1 - F_y(Y) = \left(\frac{b_y}{Y}\right)^{\eta_y},$$

so

$$X(Y) = b_x \left(\frac{b_y}{Y}\right)^{-\eta_y/\eta_x} = b_x\, b_y^{-\eta_y/\eta_x}\, Y^{\,\eta_y/\eta_x} = \left(\frac{b_x^{\eta_x}}{b_y^{\eta_y}}\right)^{1/\eta_x} Y^{\,\eta_y/\eta_x}.$$

Hence $T(Y) = Y - X(Y)$ takes the stated form. The residual elasticity is the exponent of Y in $X(Y)$, which equals η_y/η_x. ■

Remark 1. *The ratio η_y/η_x has a natural interpretation: it equals $\sigma_{\log X}/\sigma_{\log Y} = (1/\eta_x)/(1/\eta_y)$, confirming Corollary 1 (main text) for the Pareto case. Moreover, $\eta_y < \eta_x$ (progressivity) means the after-tax distribution has a heavier tail (smaller shape parameter) relative to its scale, i.e., after-tax inequality is lower.*

A4.2 Case 2: Lognormal Distribution

The lognormal distribution has been a standard model for the income distribution since Gibrat (1931)'s law of proportionate effect, which yields log-normality as the limiting distribution of a multiplicative growth process. Z is lognormal, $Z \sim \mathrm{LN}(\mu, \sigma^2)$, if $\log Z \sim \mathcal{N}(\mu, \sigma^2)$. The quantile function is $Q_Z(p) = e^{\mu + \sigma\, \Phi^{-1}(p)}$ where Φ is the standard normal CDF.

Proposition A3. *If $Y \sim \mathrm{LN}(\mu_y, \sigma_y^2)$ and $X \sim \mathrm{LN}(\mu_x, \sigma_x^2)$, the implied tax function is*

$$T(Y) \;=\; Y \;-\; \exp\!\left(\mu_x - \frac{\sigma_x}{\sigma_y}\mu_y\right) Y^{\,\sigma_x/\sigma_y},$$

a power tax function with residual elasticity $\epsilon = \sigma_x/\sigma_y$.

Proof. We have

$$F_y(Y) = \Phi\!\left(\frac{\log Y - \mu_y}{\sigma_y}\right).$$

Then

$$Q_x[F_y(Y)] = \exp\!\left\{\mu_x + \sigma_x\, \Phi^{-1}\!\left[\Phi\!\left(\frac{\log Y - \mu_y}{\sigma_y}\right)\right]\right\} = \exp\!\left\{\mu_x + \frac{\sigma_x}{\sigma_y}(\log Y - \mu_y)\right\} = e^{\mu_x - (\sigma_x/\sigma_y)\mu_y}\, Y^{\sigma_x/\sigma_y}.$$

Hence $T(Y) = Y - X(Y)$ takes the stated form, and $\epsilon = \sigma_x/\sigma_y$. ■

A4.3 Case 3: Log-Uniform Distribution

A random variable Z is log-uniformly distributed if $\log Z$ is uniformly distributed on $[\alpha, \alpha+k]$ for parameters $\alpha \in \mathbb{R}$ and $k > 0$. Its CDF is $F_Z(z) = (\log z - \alpha)/k$ for $z \in [e^\alpha, e^{\alpha+k}]$, and its quantile function is $Q_Z(p) = e^{\alpha+kp}$. Note that $\sigma_{\log Z} = k/\sqrt{12}$ (standard deviation of a uniform on $[0, k]$). The log-uniform distribution is the basis of the Chotikapanich Lorenz curve (Chotikapanich, 1993), as shown by Cortés and Gutiérrez Cubillos (2023).

Proposition A4. *If Y is log-uniform with parameters (α_y, k_y) and X is log-uniform with parameters (α_x, k_x), the implied tax function is*

$$T(Y) \;=\; Y \;-\; e^{\alpha_x - (k_x/k_y)\alpha_y}\, Y^{\,k_x/k_y},$$

a power tax function with residual elasticity $\epsilon = k_x/k_y$.

Proof. $F_y(Y) = (\log Y - \alpha_y)/k_y$, so

$$Q_x[F_y(Y)] = e^{\alpha_x + k_x\,(\log Y - \alpha_y)/k_y} = e^{\alpha_x - (k_x/k_y)\alpha_y}\, Y^{k_x/k_y}.$$

The result follows. ∎

Remark 2. *Since $\sigma_{\log Z} = k/\sqrt{12}$, the ratio $\epsilon = k_x/k_y = \sigma_{\log X}/\sigma_{\log Y}$, confirming Corollary 1 (main text).*

A4.4 Case 4: Weibull Distribution

The Weibull distribution has been applied to income data since D'Addario (1974) and has been revisited more recently in the context of global inequality (Mirzaei *et al.*, 2019). A random variable Z follows a Weibull distribution with scale $\lambda > 0$ and shape $\gamma > 0$, written $Z \sim \mathrm{W}(\lambda, \gamma)$, if its CDF is $F_Z(z) = 1 - e^{-(z/\lambda)^\gamma}$ for $z \geq 0$. The quantile function is $Q_Z(p) = \lambda\,[-\ln(1-p)]^{1/\gamma}$. Note that $\log Z = \log \lambda + (1/\gamma)\log[-\ln(1-U)]$ where $U \sim \mathrm{Uniform}(0, 1)$, so $\log Z$ follows a Gumbel (minimum extreme value) location-scale family, confirming that the Weibull is a log-location-scale distribution with shape parameter related to $1/\gamma$.

Proposition A5. *If $Y \sim \mathrm{W}(\lambda_y, \gamma_y)$ and $X \sim \mathrm{W}(\lambda_x, \gamma_x)$, the implied tax function is*

$$T(Y) \;=\; Y \;-\; \lambda_x \left(\frac{Y}{\lambda_y}\right)^{\gamma_y/\gamma_x},$$

a power tax function with residual elasticity $\epsilon = \gamma_y/\gamma_x$.

Proof. We have $F_y(Y) = 1 - e^{-(Y/\lambda_y)^{\gamma_y}}$, so $1 - F_y(Y) = e^{-(Y/\lambda_y)^{\gamma_y}}$. Then

$$Q_x[F_y(Y)] = \lambda_x \left[-\ln(1 - F_y(Y))\right]^{1/\gamma_x} = \lambda_x \left[(Y/\lambda_y)^{\gamma_y}\right]^{1/\gamma_x} = \lambda_x \,(Y/\lambda_y)^{\gamma_y/\gamma_x}.$$

Hence $X(Y) = \lambda_x(\lambda_y)^{-\gamma_y/\gamma_x}\, Y^{\gamma_y/\gamma_x}$, which is a power function with exponent γ_y/γ_x. ■

Remark 3. *The condition $\gamma_y/\gamma_x < 1$ (progressivity) requires $\gamma_y < \gamma_x$, i.e., the after-tax Weibull distribution must have a larger shape parameter than the pretax distribution. Since larger γ corresponds to a more concentrated (less dispersed) Weibull distribution, this is intuitive: a progressive tax compresses the distribution.*

A4.5 Case 5: Normal Distribution

The normal distribution has occasionally been used to model the income distribution (Lebergott, 1959) and is included here as the canonical example of a location-scale family.

Proposition A6. *If $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, the implied tax function is*

$$T(Y) \;=\; \Big(1 - \tfrac{\sigma_x}{\sigma_y}\Big)Y \;+\; \Big(\tfrac{\sigma_x}{\sigma_y}\mu_y - \mu_x\Big),$$

a linear (flat-rate plus lump-sum) tax schedule.

Proof. This follows directly from Proposition 1 since the normal family is a location-scale family. Explicitly: $F_y(Y) = \Phi((Y - \mu_y)/\sigma_y)$, so

$$Q_x[F_y(Y)] = \mu_x + \sigma_x\, \Phi^{-1}\Big[\Phi\Big(\tfrac{Y-\mu_y}{\sigma_y}\Big)\Big] = \mu_x + \sigma_x\, \tfrac{Y-\mu_y}{\sigma_y} = \tfrac{\sigma_x}{\sigma_y}Y + \mu_x - \tfrac{\sigma_x}{\sigma_y}\mu_y. \qquad \square$$

■

Remark 4. *The flat tax rate is $1 - \sigma_x/\sigma_y$, which is positive (progressive lump-sum effect) whenever $\sigma_x < \sigma_y$. The lump-sum component $\sigma_x\mu_y/\sigma_y - \mu_x$ may be positive or negative depending on parameters. Notably, the normal-implied tax function is* not *a power function: the residual elasticity is not constant but instead varies with income as $\epsilon(Y) = \sigma_x/\sigma_y \cdot Y/(Y - (\sigma_x/\sigma_y)\mu_y + \mu_x)$, converging to σ_x/σ_y as $Y \to \infty$.*

A4.6 Case 6: Shifted Exponential Distribution

The shifted (biparametric) exponential distribution is $Z \sim \mathrm{EXP}(\theta, \gamma)$ with CDF $F_Z(z) = 1 - e^{-(z-\gamma)/\theta}$ for $z \geq \gamma$, where $\theta > 0$ is the scale and $\gamma \geq 0$ is the location (shift). Its quantile function is $Q_Z(p) = \gamma - \theta \ln(1-p)$. Note that $Z - \gamma \sim \mathrm{Exp}(\theta)$, and $\log(Z-\gamma)$ is not normally distributed, so the shifted exponential is a *location-scale family in Z* (not log-Z).

Proposition A7. *If $Y \sim \mathrm{EXP}(\theta_y, \gamma_y)$ and $X \sim \mathrm{EXP}(\theta_x, \gamma_x)$, the implied tax function is*

$$T(Y) \;=\; \Big(1 - \tfrac{\theta_x}{\theta_y}\Big)Y \;+\; \Big(\tfrac{\theta_x}{\theta_y}\gamma_y - \gamma_x\Big),$$

a linear tax function with flat-tax rate $1 - \theta_x/\theta_y$ and lump-sum $\theta_x\gamma_y/\theta_y - \gamma_x$.

Proof. $F_y(Y) = 1 - e^{-(Y-\gamma_y)/\theta_y}$, so $1 - F_y(Y) = e^{-(Y-\gamma_y)/\theta_y}$. Then

$$Q_x[F_y(Y)] = \gamma_x - \theta_x \ln(1 - F_y(Y)) = \gamma_x + \tfrac{\theta_x}{\theta_y}(Y - \gamma_y) = \tfrac{\theta_x}{\theta_y}Y + \gamma_x - \tfrac{\theta_x}{\theta_y}\gamma_y.$$

Hence $T(Y) = Y - X(Y) = (1 - \theta_x/\theta_y)Y + (\theta_x\gamma_y/\theta_y - \gamma_x)$. ■

A4.7 Case 7: Generalized Pareto Distribution

The generalized Pareto distribution (GPD) has been proposed as a more flexible alternative to the standard Pareto for modeling the upper tail (Charpentier and Flachaire, 2022; Jenkins, 2017). It nests the standard Pareto as a special case while allowing richer tail behavior. $Z \sim \mathrm{GPD}(\mu, \sigma, \alpha)$ has CDF

$$F_Z(z) \;=\; 1 - \left(1 + \frac{z-\mu}{\sigma}\right)^{-\alpha}, \quad z \geq \mu,$$

for $\sigma > 0$ and $\alpha > 0$. Setting $\mu = \sigma$ reduces to the standard Pareto with scale μ and shape α. The quantile function is $Q_Z(p) = \mu + \sigma[(1-p)^{-1/\alpha} - 1]$.

Proposition A8. *If $Y \sim \text{GPD}(\mu_y, \sigma_y, \alpha_y)$ and $X \sim \text{GPD}(\mu_x, \sigma_x, \alpha_x)$, the implied tax function is*

$$T(Y) \;=\; Y \;-\; \mu_x \;-\; \sigma_x\left[\left(1 + \frac{Y-\mu_y}{\sigma_y}\right)^{\alpha_y/\alpha_x} - 1\right].$$

Proof. $1 - F_y(Y) = (1 + (Y-\mu_y)/\sigma_y)^{-\alpha_y}$. Then

$$\begin{aligned} Q_x[F_y(Y)] &= \mu_x + \sigma_x\big[(1 - F_y(Y))^{-1/\alpha_x} - 1\big] \\ &= \mu_x + \sigma_x\Big[\Big(1 + \tfrac{Y-\mu_y}{\sigma_y}\Big)^{\alpha_y/\alpha_x} - 1\Big]. \end{aligned}$$

■

Remark 5. *The GPD-implied tax function is* not *a power function in general, since the exponent on $(1 + (Y-\mu_y)/\sigma_y)$ does not simplify to an exponent on Y alone unless $\mu_y = 0$. However, the asymptotic residual elasticity satisfies*

$$\lim_{Y\to\infty} \epsilon(Y) \;=\; \frac{\alpha_y}{\alpha_x},$$

which recovers the Pareto ratio. This reflects the GPD's asymptotic power-law behavior: for very high incomes, the GPD converges to the Pareto distribution, and the implied tax function converges to a power function. In finite samples, the GPD provides a more flexible approximation to the true distribution and hence potentially a more accurate estimate of the (non-constant) residual elasticity for moderately high incomes.

Remark 6 (Setting $\mu = \sigma$)**.** *When $\mu_y = \sigma_y$ and $\mu_x = \sigma_x$, the GPD reduces to the standard Pareto, and the implied tax function reduces to*

$$T(Y) = Y - \sigma_x\left[\left(\frac{Y}{\sigma_y}\right)^{\alpha_y/\alpha_x} - 1 + 1\right] - \sigma_x = Y - \left(\frac{\sigma_x^{\alpha_x}}{\sigma_y^{\alpha_y}}\right)^{1/\alpha_x} Y^{\alpha_y/\alpha_x},$$

which matches Proposition A2 with $b_z = \sigma_z$ and $\eta_z = \alpha_z$.