Advanced Public Economics — Applied Tutorial

Session II: Estimating Optimal Top Tax Rates

Daniel Weishaar

Roadmap for Session II

Estimate the asymptotic **optimal top tax rate** under a **Pareto distribution**.

1. Using empirical CDF from micro data.
2. Estimate characteristics of Pareto distribution from tabulated income tax data.

Brief Recap of the Formula

Optimal Top Tax Rates

Following Diamond (1998), under the assumption of

- an unbounded distribution of skills,
- utility without income effects (quasi-linear), constant elasticity of labor supply ϵ,
- a constant social marginal welfare weight g at the top,
- assuming a Pareto distribution of skills at the top with a parameter a^{θ},

the asymptotic optimal top tax rate is given by $\frac{T'_{top}}{1-T'_{top}} = (1-g)\left(1+\frac{1}{\epsilon}\right)\left(\frac{1}{a^{\theta}}\right)$.

Optimal Top Tax Rates

Following Diamond (1998), under the assumption of

- an unbounded distribution of skills,
- utility without income effects (quasi-linear), constant elasticity of labor supply ϵ,
- a constant social marginal welfare weight g at the top,
- assuming a Pareto distribution of skills at the top with a parameter a^θ,

the asymptotic optimal top tax rate is given by $\frac{T'_{top}}{1-T'_{top}} = (1-g)\left(1+\frac{1}{\epsilon}\right)\left(\frac{1}{a^\theta}\right)$.

$\Rightarrow$ While a^θ is not observable, in the Diamond (1998) case, income is also Pareto distributed with $a = \frac{a^\theta}{1+\epsilon}$. Thus,

$$\frac{T'_{top}}{1-T'_{top}} = (1-g)\left(\frac{1}{\epsilon}\right)\left(\frac{1}{a}\right)$$

Optimal Top Tax Rates

Following Diamond (1998), under the assumption of

- an unbounded distribution of skills,
- utility without income effects (quasi-linear), constant elasticity of labor supply ϵ,
- a constant social marginal welfare weight g at the top,
- assuming a Pareto distribution of skills at the top with a parameter a^θ,

the asymptotic optimal top tax rate is given by $\frac{T'_{top}}{1-T'_{top}} = (1-g)\left(1+\frac{1}{\epsilon}\right)\left(\frac{1}{a^\theta}\right)$.

$\Rightarrow$ While a^θ is not observable, in the Diamond (1998) case, income is also Pareto distributed with $a = \frac{a^\theta}{1+\epsilon}$. Thus,

$$\frac{T'_{top}}{1-T'_{top}} = (1-g)\left(\frac{1}{\epsilon}\right)\left(\frac{1}{a}\right)$$

$\Rightarrow$ **Key Question: How do we estimate a?**

Empirical CDF from Micro Data

Empirical CDF from Micro Data — Pareto Distribution

The Pareto distribution has the following cumulative distribution function (**CDF**)

$$F(y) = 1 - \left(\frac{y_{min}}{y}\right)^{a}$$

with $y_{min} > 0$ representing a value at which the Pareto distribution begins.

Empirical CDF from Micro Data — Pareto Distribution

The Pareto distribution has the following cumulative distribution function (**CDF**)

$$F(y) = 1 - \left(\frac{y_{min}}{y}\right)^{a}$$

with $y_{min} > 0$ representing a value at which the Pareto distribution begins.

Can be reformulated as

$$1 - F(y) = \left(\frac{y_{min}}{y}\right)^{a}$$

$$\ln\left(1 - F(y)\right) = \ln\left(\left(\frac{y_{min}}{y}\right)^{a}\right)$$

Empirical CDF from Micro Data — Pareto Distribution

The Pareto distribution has the following cumulative distribution function (**CDF**)

$$F(y) = 1 - \left(\frac{y_{min}}{y}\right)^{a}$$

with $y_{min} > 0$ representing a value at which the Pareto distribution begins.

Can be reformulated as

$$1 - F(y) = \left(\frac{y_{min}}{y}\right)^{a}$$

$$\ln\left(1 - F(y)\right) = \ln\left(\left(\frac{y_{min}}{y}\right)^{a}\right)$$

$$ln\big(1 - F(y)\big) = \underbrace{a\ ln(y_{min})}_{\text{Intercept}} - \underbrace{\color{red}{a}}_{\text{Slope}}\ \ln(y)$$

Empirical CDF from Micro Data — Pareto Distribution

The Pareto distribution has the following cumulative distribution function (**CDF**)

$$F(y) = 1 - \left(\frac{y_{min}}{y}\right)^a$$

with $y_{min} > 0$ representing a value at which the Pareto distribution begins.

Can be reformulated as

$$1 - F(y) = \left(\frac{y_{min}}{y}\right)^a$$

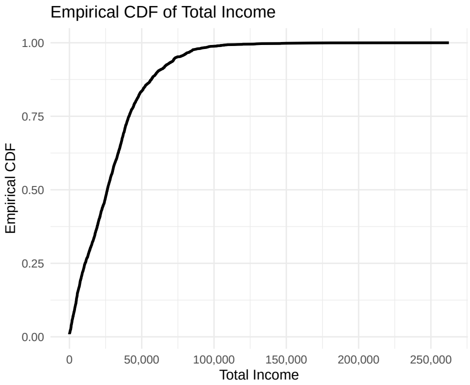
$$\ln\left(1 - F(y)\right) = \ln\left(\left(\frac{y_{min}}{y}\right)^a\right)$$

$$ln\big(1 - F(y)\big) = \underbrace{a\ ln(y_{min})}_{\text{Intercept}} - \underbrace{a}_{\text{Slope}}\ \ln(y)$$

- At the very top, slope should be approximately linear if the distribution is Pareto.
- Log-relationship can be estimated to obtain a → Hands on exercise

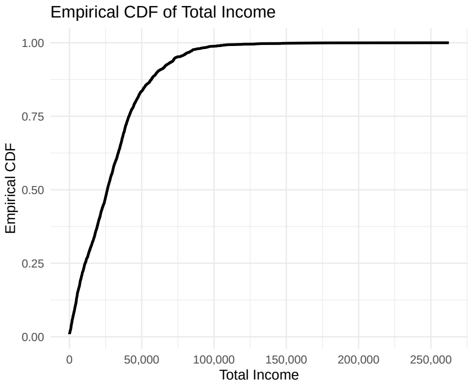
Empirical CDF from Micro Data — Estimate CDF

- When we have micro data on individual earnings, we can estimate the empirical CDF $F(y)$ by ranking incomes and estimate what share of incomes lies below a specific value.

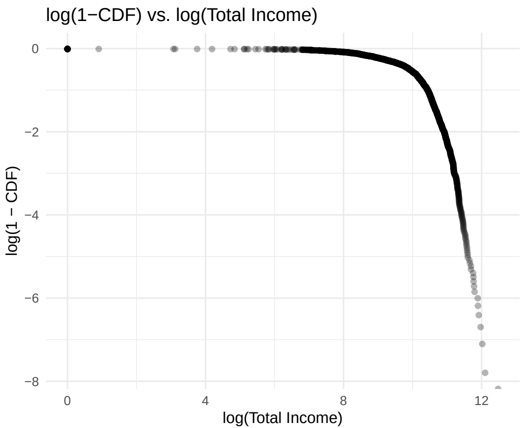


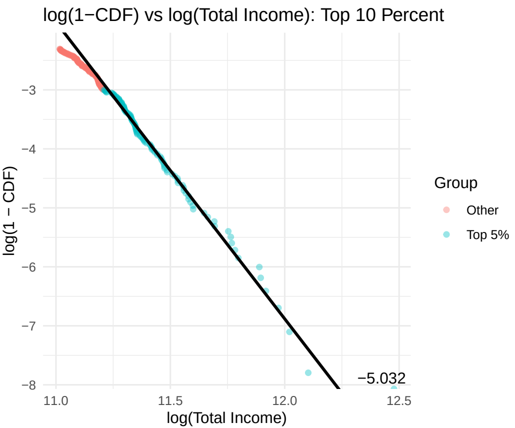
Empirical CDF from Micro Data — Estimate CDF

- When we have micro data on individual earnings, we can estimate the empirical CDF $F(y)$ by ranking incomes and estimate what share of incomes lies below a specific value.
- We can than take a look at how $\ln(1 - F(y))$ varies with income. Slope at the top of the distribution informs us about the Pareto coefficient a.



Empirical CDF from Micro Data — Log(1-CDF) vs. Log(Income)

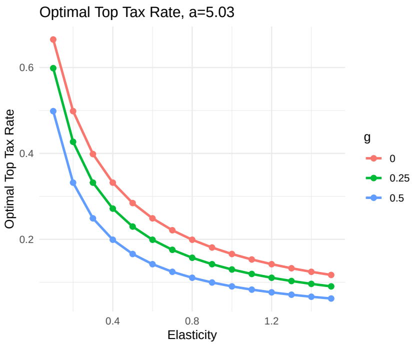




For robustness, check how this varies at the top (e.g. P90-P95 vs. P95-P99).

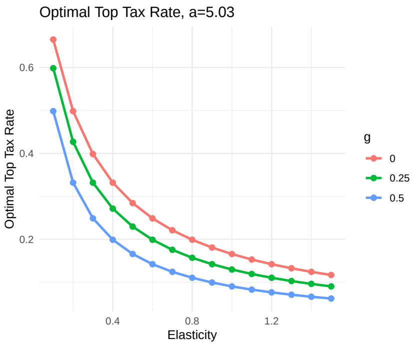
Empirical CDF from Micro Data — Optimal Top Tax Rates

- We can use our estimate for $a = 5.03$ to calculate the optimal top tax rate for different elasticity values and different values of g.



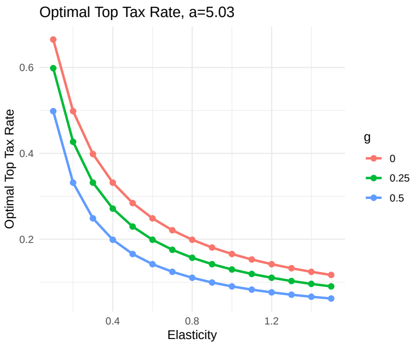
Empirical CDF from Micro Data — Optimal Top Tax Rates

- We can use our estimate for $a = 5.03$ to calculate the optimal top tax rate for different elasticity values and different values of g.
- g represents how much we value 1\$ for the rich relative to 1\$ of government revenue.



Empirical CDF from Micro Data — Optimal Top Tax Rates

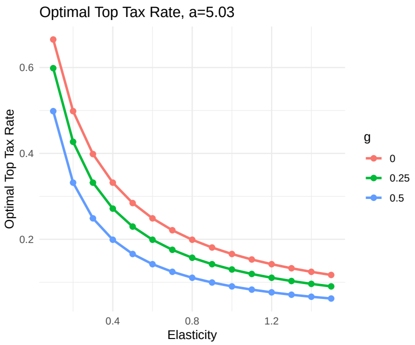
- We can use our estimate for $a = 5.03$ to calculate the optimal top tax rate for different elasticity values and different values of g.
- g represents how much we value 1\$ for the rich relative to 1\$ of government revenue.
- $g = 0$ follows from concave social welfare function (SWF) / utilitarian SWF with decreasing marginal utility of consumption.



Empirical CDF from Micro Data — Optimal Top Tax Rates

- We can use our estimate for $a = 5.03$ to calculate the optimal top tax rate for different elasticity values and different values of g.
- g represents how much we value 1\$ for the rich relative to 1\$ of government revenue.
- $g = 0$ follows from concave social welfare function (SWF) / utilitarian SWF with decreasing marginal utility of consumption.
- Setting $g = 0$ gives us the upper Pareto bounds for top tax rates (rev.-maximizing tax rates).

  $\Rightarrow$ Taxes above Pareto inefficient!



Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

- Even though survey is representative, coverage of (very) high incomes not good (e.g., due to non-response bias).

Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

- Even though survey is representative, coverage of (very) high incomes not good (e.g., due to non-response bias).
- Income values in survey data oftentimes top coded for privacy or data protection concerns, e.g. all values above one million are coded as one million.

Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

- Even though survey is representative, coverage of (very) high incomes not good (e.g., due to non-response bias).
- Income values in survey data oftentimes top coded for privacy or data protection concerns, e.g. all values above one million are coded as one million.

Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

- Even though survey is representative, coverage of (very) high incomes not good (e.g., due to non-response bias).
- Income values in survey data oftentimes top coded for privacy or data protection concerns, e.g. all values above one million are coded as one million.

$\Rightarrow$ Survey-based a tends to be too large, implying too low optimal tax rates.

Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

- Even though survey is representative, coverage of (very) high incomes not good (e.g., due to non-response bias).
- Income values in survey data oftentimes top coded for privacy or data protection concerns, e.g. all values above one million are coded as one million.

$\Rightarrow$ Survey-based a tends to be too large, implying too low optimal tax rates.

$\rightarrow$ **First-best solution:** use administrative tax return micro data, e.g. tax payer panel in Germany or IRS tax return micro data in the United States.

Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

- Even though survey is representative, coverage of (very) high incomes not good (e.g., due to non-response bias).
- Income values in survey data oftentimes top coded for privacy or data protection concerns, e.g. all values above one million are coded as one million.

$\Rightarrow$ Survey-based a tends to be too large, implying too low optimal tax rates.

$\rightarrow$ **First-best solution:** use administrative tax return micro data, e.g. tax payer panel in Germany or IRS tax return micro data in the United States.

But! Not always available or (very) restricted access.

Empirical CDF from Micro Data — Problems

When we use survey data (like GSOEP in Germany or CPS in the United States), direct estimation of a via empirical CDF might be problematic:

- Even though survey is representative, coverage of (very) high incomes not good (e.g., due to non-response bias).
- Income values in survey data oftentimes top coded for privacy or data protection concerns, e.g. all values above one million are coded as one million.

$\Rightarrow$ Survey-based a tends to be too large, implying too low optimal tax rates.

$\rightarrow$ **First-best solution:** use administrative tax return micro data, e.g. tax payer panel in Germany or IRS tax return micro data in the United States.

But! Not always available or (very) restricted access.

$\rightarrow$ **Second-best solution:** use tabulated tax return data.

Tabulated Income Tax Data

Tabulated Income Tax Data — Overview

Objective: Use grouped income tax statistics (e.g., from Destatis in Germany or from IRS in the United States) to estimate the Pareto parameter a without microdata.

Example for Germany (Source):

Adjusted gross income from ... to under ... Euro	Adjusted gross income			
	taxpayers[1]	%	1,000 EUR	%
0 to 5,000	4,820,115	11.2	7,480,434	0.4
5,000 to 10,000	2,353,352	5.5	17,851,525	0.9
10,000 to 15,000	3,155,869	7.3	39,902,164	2.0
15,000 to 20,000	3,594,246	8.3	62,692,463	3.2
20,000 to 25,000	3,333,587	7.7	75,016,710	3.8
...				
50,000 to 60,000	3,044,466	7.1	166,523,902	8.4
60,000 to 70,000	2,147,178	5.0	139,003,998	7.0
70,000 to 125,000	4,951,807	11.5	448,444,143	22.7
125,000 to 250,000	1,607,958	3.7	262,642,612	13.3
250,000 to 500,000	321,834	0.7	107,367,770	5.4
500,000 to 1,000,000	83,268	0.2	55,537,482	2.8
1,000,000 or more	34,509	0.1	98,259,228	5.0

Table shows number of tax payers and total income in a bracket.

Tabulated Income Tax Data — Use Property of Pareto Distribution

Pareto Tail Property (van der Wijk's law) For a Pareto distribution with parameter a and threshold z:

$$E[Y \mid Y \geq z] = \underbrace{\frac{a}{a-1}}_{b} \, z.$$

Tabulated Income Tax Data — Use Property of Pareto Distribution

Pareto Tail Property (van der Wijk's law) For a Pareto distribution with parameter a and threshold z:

$$E[Y \mid Y \geq z] = \underbrace{\frac{a}{a-1}}_{b} z.$$

Inverted Pareto coefficient: The inverted Pareto coefficient b can be estimated from tabulated data as

$$b(z) = \frac{E[Y \mid Y \geq z]}{z}.$$

- z: bracket threshold, e.g. 250,000.
- $E[Y \mid Y \geq z]$: average income for incomes greater or equal than threshold.

Tabulated Income Tax Data — Use Property of Pareto Distribution

Pareto Tail Property (van der Wijk's law) For a Pareto distribution with parameter a and threshold z:

$$E[Y \mid Y \geq z] = \underbrace{\frac{a}{a-1}}_{b} z.$$

Inverted Pareto coefficient: The inverted Pareto coefficient b can be estimated from tabulated data as

$$b(z) = \frac{E[Y \mid Y \geq z]}{z}.$$

- z: bracket threshold, e.g. 250,000.
- $E[Y \mid Y \geq z]$: average income for incomes greater or equal than threshold.

$\Rightarrow$ See also Saez (2001), Piketty and Saez (2003) or for more advanced techniques when Pareto parameter varies, see Blanchet et al. (2022).

Tabulated Income Tax Data — Derivation of Property (1/2)

We are interested in

$$E[Y \mid Y \geq z] = \frac{\int_z f(y)\, y \, dy}{P(Y \geq z)} \tag{1}$$

Differentiating the cdf (see Slide 6), the probability density function reads

$$f(y) = a \left(\frac{y_{min}}{y} \right)^a \frac{1}{y} \tag{2}$$

From the cdf (see Slide 6) we also know that

$$P(Y \geq z) = 1 - F(z) = \left(\frac{y_{min}}{z} \right)^a \tag{3}$$

Tabulated Income Tax Data — Derivation of Property (2/2)

Plugging (2) and (3) into (1) we obtain

$$
\begin{aligned}
E[Y \mid Y \geq z] &= \frac{\int_z a \left(\frac{y_{min}}{y}\right)^a \frac{1}{y}\ y\ dy}{\left(\frac{y_{min}}{z}\right)^a} \\
&= a\ y_{min}^a\ \frac{\int_z y^{-a}\ dy}{\left(\frac{y_{min}}{z}\right)^a} \\
&= a\ \frac{\left[\frac{1}{1-a} y^{1-a}\right]_z^\infty}{z^{-a}} \\
&= a\ \frac{\frac{-1}{1-a} z^{1-a}}{z^{-a}} \\
&= \frac{a}{a-1} z
\end{aligned}
$$

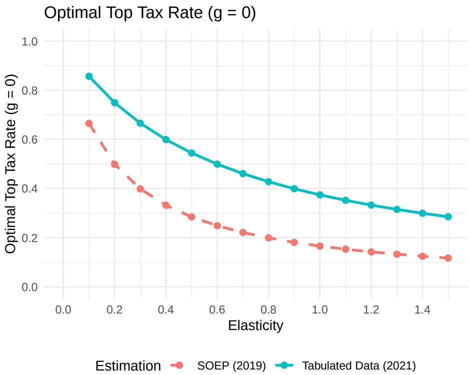
Tabulated Income Tax Data — Hands-on Exercise

In the hands-on exercise, we use tabulated data from income tax records for Germany (2021) to estimate Pareto coefficients and optimal top tax rates.

Tabulated Income Tax Data — Hands-on Exercise

In the hands-on exercise, we use tabulated data from income tax records for Germany (2021) to estimate Pareto coefficients and optimal top tax rates.

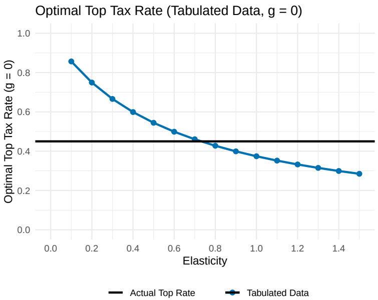
- Much lower coefficient $a = 1.67$ than the one obtained from survey micro data (5.03).
- Lower a (higher b) $\rightarrow$ fatter tail $\rightarrow$ many more extreme values $\rightarrow$ higher optimal top tax rates.
- **Data quality is important!**



Additional Remarks

Additional Remarks — Comparison with Actual Top Tax Rate

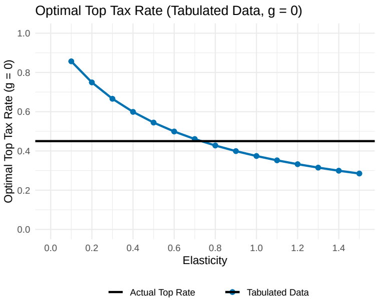
- Top marginal income tax rate in Germany (*Reichensteuersatz*) lies at 45% (from taxable income of 277,826 EUR in 2025).



Additional Remarks — Comparison with Actual Top Tax Rate

- Top marginal income tax rate in Germany (*Reichensteuersatz*) lies at 45% (from taxable income of 277,826 EUR in 2025).

→ For plausible elasticity values, still room for higher top tax rates.

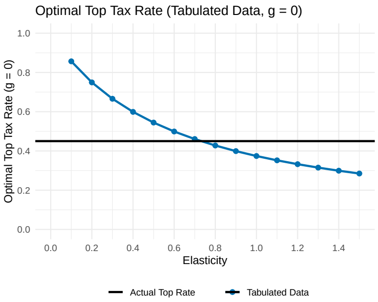


Additional Remarks — Comparison with Actual Top Tax Rate

- Top marginal income tax rate in Germany (*Reichensteuersatz*) lies at 45% (from taxable income of 277,826 EUR in 2025).

→ For plausible elasticity values, still room for higher top tax rates.

- **But!** Need to take into account that there are also other taxes distorting labor-leisure trade-off, e.g., consumption taxes like VAT (19% in Germany on most goods and services).



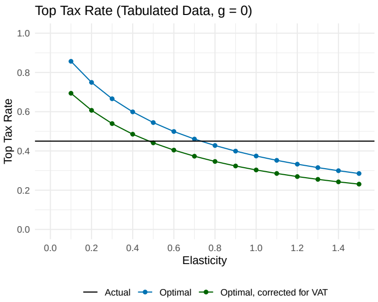
Additional Remarks — Comparison with Actual Top Tax Rate

- Top marginal income tax rate in Germany (*Reichensteuersatz*) lies at 45% (from taxable income of 277,826 EUR in 2025).

$\rightarrow$ For plausible elasticity values, still room for higher top tax rates.

- **But!** Need to take into account that there are also other taxes distorting labor-leisure trade-off, e.g., consumption taxes like VAT (19% in Germany on most goods and services).

$\rightarrow$ Scale down optimal tax rates by $(1 - t^c)$.



Additional Remarks — Pareto Distribution and Top Income Shares

Properties of Pareto distribution also used to estimate inequality at the top:

- See for instance the methodology of the World Inequality Database (WID).
- Under a Pareto distribution, the top income shares can be estimated as

$$\text{Top } p \text{ share} = (p)^{\left(\frac{1}{b}\right)}$$

- With our parameter of $a = 1.67 (b = 2.49)$, the 2021 top income shares read

Top 10%	39%
Top 5%	30%
Top 1%	16%

$\Rightarrow$ Close to numbers from WID.

Additional Remarks — Optimal Top Tax Rate Formula (1/2)

By estimating

$$\frac{T'_{top}}{1 - T'_{top}} = (1 - g)\left(\frac{1}{\epsilon}\right)\left(\frac{1}{a}\right)$$

we assumed that a is invariant to the tax rate. Is this correct?

- Crucial assumption is the assumption that elasticity is constant at the top.
- In this case, and under $T'_{top} < 1$, behavioral responses triggered by changes in tax rates scale all incomes up or down by the same amount, i.e., the ratio b and thus a is unchanged, see Saez (2001).

Additional Remarks — Optimal Top Tax Rate Formula (2/2)

We ignored income effects, i.e., $\epsilon = \epsilon^u = \epsilon^c$.

Without Income Effects:

$$\frac{T'_{top}}{1 - T'_{top}} = (1 - g)\left(\frac{1}{\epsilon}\right)\left(\frac{1}{a}\right)$$

With Income Effects, and using $\epsilon^u = \epsilon^c + \eta$, see Saez (2001):

$$\frac{T'_{top}}{1 - T'_{top}} = (1 - g)\left(\frac{1}{\epsilon^u + \epsilon^c(a - 1)}\right) = (1 - g)\left(\frac{1}{\epsilon^c \; a + \eta}\right)$$

- Presence of income effects ($\eta \leq 0$ if leisure is not inferior good) implies that optimal top tax rates can be higher.

Additional Remarks — Optimal Top Tax Rate Formula (2/2)

We ignored income effects, i.e., $\epsilon = \epsilon^u = \epsilon^c$.

Without Income Effects:

$$\frac{T'_{top}}{1 - T'_{top}} = (1 - g) \left(\frac{1}{\epsilon}\right) \left(\frac{1}{a}\right)$$

With Income Effects, and using $\epsilon^u = \epsilon^c + \eta$, see Saez (2001):

$$\frac{T'_{top}}{1 - T'_{top}} = (1 - g) \left(\frac{1}{\epsilon^u + \epsilon^c (a - 1)}\right) = (1 - g) \left(\frac{1}{\epsilon^c \; a + \eta}\right)$$

- Presence of income effects ($\eta \leq 0$ if leisure is not inferior good) implies that optimal top tax rates can be higher.
- Also, it is important what ϵ captures, real responses vs. tax evasion / bargaining behavior (see Piketty et al., 2014).

Additional Remarks — Why Pareto?

The Pareto distribution is frequently used in empirical research to describe the top of the income distribution — Why?

- **Empirical Observation:** Pareto distribution provides a good empirical fit, but controversies remain, see, e.g. Mankiw et al. (2009).
- **Theoretical Explanations:** Properties in growth models (mostly related to exponential growth) can explain why incomes are Pareto distributed, see Jones (2015) for an overview.

Important! "Pareto is a good approximation only for the upper tail; the bulk of the income distribution requires other distributions. $\rightarrow$ Relevant for next session when we discuss the shape of optimal marginal tax rates.

That's it! See you next time!

References

References I

- BLANCHET, T., J. FOURNIER, and T. PIKETTY (2022). "Generalized Pareto Curves: Theory and Applications." *Review of Income and Wealth* 68 (1), pp. 263–300.
- DIAMOND, P. A. (1998). "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88 (1), pp. 83–95.
- JONES, C. I. (2015). "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality." *Journal of Economic Perspectives* 29 (1), pp. 29–46.
- MANKIW, N. G., M. WEINZIERL, and D. YAGAN (2009). "Optimal Taxation in Theory and Practice." *Journal of Economic Perspectives* 23 (4), pp. 147–174.
- PIKETTY, T. and E. SAEZ (2003). "Income Inequality in the United States, 1913–1998." *The Quarterly Journal of Economics* 118 (1), pp. 1–39.
- PIKETTY, T., E. SAEZ, and S. STANTCHEVA (2014). "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities." *American Economic Journal: Economic Policy* 6 (1), pp. 230–271.

References II

SAEZ, E. (2001). “Using Elasticities to Derive Optimal Income Tax Rates.” *Review of Economic Studies* 68 (1), pp. 205–229.