



**University of
Zurich** ^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 337

Taxation and the Superrich

Florian Scheuer and Joel Slemrod

November 2019

Taxation and the Superrich*

Florian Scheuer Joel Slemrod
University of Zurich University of Michigan

November 2019

Abstract

This paper addresses the modern optimal tax progressivity literature, which clarifies the key role of the behavioral response to taxation and accounts for the incomes of the superrich being qualitatively different than others. Some may be “superstars,” for whom small differences in talent are magnified into much larger earnings differences, while others may work in winner-take-all markets, such that their effort to climb the ladder of success reduces the returns to others. We stress that pivotal tax-rate elasticities are not structural parameters, and will be smaller the broader and less plastic is the tax base and the more effective is the enforcement of tax evasion. For this reason, normative analysis of tax rates should be accompanied by attention to the tax base, with a special focus on capital gains, which comprise a large fraction of the taxable income of the superrich.

Keywords: Superrich, Tax Systems, Superstars, Winner-Take-All Markets, Plasticity of Taxable Income, Wealth Taxes

*We thank Alan Auerbach, Tim Besley, Casey Rothschild and Uwe Thümmel for helpful comments, and Emily Horton, Paul Kindsgrab and Simona Sartor for excellent research assistance.

1 Introduction

Proposals to increase taxes on the superrich have recently attracted much attention. In the United States, in 2019, newly-elected but prominent Congresswoman Alexandria Ocasio-Cortez proposed to add a new 70% income tax bracket on income in excess of \$10 million, and presidential candidate Senator Elizabeth Warren proposed a new annual wealth tax at a 2% rate on net worth in excess of \$50 million and a 3% rate, later amended to 6%, above \$1 billion.¹ In the United Kingdom, the 2017 Labour Party manifesto proposed re-introducing a 50% top income tax rate, 5% higher than the current rate. France levied a top marginal income “supertax” rate of 75% on earnings over €1 million from 2012 to 2014, but scrapped it and returned to a 45% top rate in 2015. In reaction to these proposals, some have suggested that there are better ways to raise taxes on the rich, such as targeting economic rents and windfalls, imposing higher property taxes, increasing the tax rate on capital gains, or levying a progressive consumption tax.²

The recent focus on how the very rich are taxed is tied to widespread concern about the extent and growth of inequality, which many view as inherently unacceptable or, in the words of [Simons \(1938\)](#), “distinctly evil or unlovely.” Others, such as [Persson and Tabellini \(1994\)](#), argue that inequality is harmful for growth. In addition to these economic concerns, some tie inequality to political instability ([Farhi et al., 2012](#); [Piketty, 2014](#); [Scheuer and Wolitzky, 2016](#)), and others fear political capture by the superrich ([Gilens, 2014](#); [Bartels, 2016](#)). Of course, inequality depends not only on the concentration of income and wealth at the top, but also on how well the poorest in a society are faring, and these two phenomena are not identical. This paper addresses the former only.

2 Who are the Superrich?

In this paper, we will use the term superrich to refer to the very rich. This is a term of art, not science, that has been around since the early 19th century, the threshold for which is essentially arbitrary.³ The data we present and discuss applies to different slices, often focusing on the top 400 in the U.S., which comprises the top 0.0003% of households, but occasionally addressing the top 0.01% or 0.1%. A conceptually more important definitional issue is whether to measure affluence by wealth or income.

¹Other Democratic presidential candidates have also proposed higher taxation of the rich (see [Matthews, 2019](#), for a summary).

²See e.g. [The Economist \(2019\)](#) and [Rattner \(2019\)](#).

³According to the OED online, the first use of the term superrich was in the early 19th century, from the *Daily National Intelligencer*, a Washington, DC newspaper that was published from about 1800 to 1870.

2.1 The Superrich by Wealth

The most detailed data about the superrich are the annual Forbes 400 list of the wealthiest Americans, going back to 1982, and the Forbes list of the world's billionaires, updated daily. These reports are based on public information supplemented by investigative reporting and "educated guesses" (Dolan, 2018).⁴ In recent years, Bloomberg has published its own Billionaires Index, which is a daily ranking of the world's 500 wealthiest individuals (Bloomberg, 2019).

Because billionaires have been singled out in some of the latest tax proposals in the U.S., it is worth studying them in more detail. The cutoff to make it into the Forbes 400 in 2018 was a net worth of \$2.1 billion, and the average wealth in this group was \$7.2 billion. The share of aggregate U.S. wealth owned by the Forbes 400 has increased from less than 1% in 1982 to more than 3% in 2018. The Forbes 400 are very well educated (86% graduated from college), up from 76% in 1982. 69% are "self-made," i.e., founders of their business (as opposed to inheriting it), nearly double the 44% in 1982. The industry composition of their wealth-generating businesses has changed as well between 1982 and 2018, with finance increasing in representation from 7% to 22% and technology from 4% to 17%, while energy fell from 22% to 6% and real estate from 17% to 9% (see Kaplan and Rauh, 2013, for a similar comparison using earlier data).

2.2 The Superrich by Income

For tax years 1992 to 2014, the U.S. Internal Revenue Service (IRS) released aggregated information about the 400 individual tax returns with the highest adjusted gross income (AGI). The cutoff for inclusion into the "Fortunate 400" rose from \$24 million in 1992 to \$127 million in 2014. The share of total AGI received by the top 400 more than doubled, from 0.5% in 1992 to 1.3% in 2014, while their share of total returns fell slightly (as it must, given the constant 400 number and the growing number of total returns filed). Their share of total income tax also increased, but by not quite as much, from 1% to 2%, reflecting in part that their average tax rate fell over this time from 26% to 23%. Notably, the average tax rate did not fall monotonically over time; its period low was 17% in 2007. Because capital gains comprise a majority of the income of the Fortunate 400, how they are taxed is crucial in affecting the average tax rate of their overall income. For example, the average tax rate jumped from 17% to 23% between 2012 and 2013, reflecting that the top rate on long-term gains rose from 15% to 20% between these two years.

⁴Other sources of data are surveys, administrative data from estate tax returns combined with mortality rates, capitalization from flows of capital income, and administrative data from wealth tax returns.

Bakija et al. (2012) study the occupational composition of top income earners in the U.S., defined as the top 0.01%, from 1979 to 2005 using tax administrative data. They find that executives, managers, supervisors and financial professionals accounted for approximately 60% of this group in 2005.⁵ Compared to 1979, the share of executives, managers, or supervisors decreased from roughly 48% to roughly 43%, while the share of financial professionals increased from 11% to 18%, and the increase in the share of income captured by these professions accounts for 70% of the growth in the income share captured by the top 0.1%; this is consistent with the trends in the characteristics of top wealth holders discussed above. Individuals in the top 0.1% saw their income grow at considerably larger rates than individuals in the same occupation within the 99th to 99.5th percentile range.

Guvenen and Kaplan (2017) use income data from the IRS and Social Security Administration, the latter of which includes only wage and self-employment income, and find that the 21st century rise in top income shares is due to the increasing importance of business income and is driven, to a large extent, by growth in the income share captured by the top 0.01%. **Smith et al. (2019)**, motivated by the observation that income from pass-through entities was a key driver of the growth in top income shares, address whether those individuals at the top of the income distribution receiving business income are passive rentiers or actively engaged in managing their firms, and interpret the evidence as favoring the latter.

The canonical definition of income is consumption plus the change in wealth. But the measure used in the aforementioned studies is based on a country's definition of taxable income. The principal omissions are inheritances and unrealized capital gains, but also often exclude unreported (i.e., evaded) income, legitimately tax-exempt income, the implicit flow of services from durable goods (mainly owner-occupied housing), and the accrual of rights via insurance and pension plans. **Steuerle (1985)** analyzed a collection of estate tax returns, income tax returns of decedents in years before death, and income tax returns of heirs in years both prior to and following the death of the persons granting the bequests for the U.S. between 1974 and 1980. He concluded that realized income is a poor measure of well-being, and that indicators of property or wealth may better capture real economic income. **Bourne et al. (2018)** obtain similar results using data from 2002 to 2006, and find that those at the top of the wealth distribution realize a lower taxable rate of return, suggesting that taxable income data understates economic inequality.

⁵**Denk (2015)** documents similar patterns in Europe.

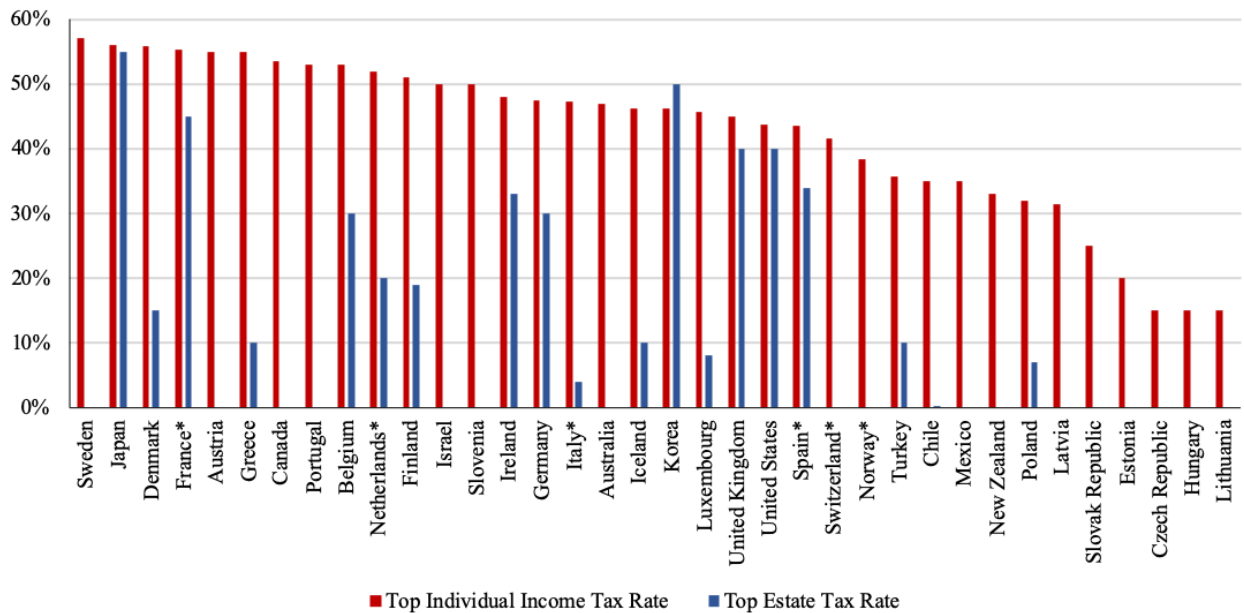


Figure 1: Maximum statutory tax rates. Source: OECD Tax Database, tables I.7 and II.4 and EY Global Tax Guide. Includes central and sub-central government taxes. Rates as of 2018. Countries with a wealth tax are indicated by a star (*).

3 The Tax Burden on the Superrich

3.1 Statutory Tax Features

For many reasons, it is not straightforward to measure the tax burden on the superrich. We begin by providing some relevant, although not dispositive, measures for developed (OECD) countries.

Income Tax. Figure 1 shows the top statutory personal income tax rate, defined according to the OECD as the combined central and sub-central government marginal personal income tax rate on wages at the earnings threshold where the top statutory personal income tax rate first applies, taking into account the effects of tax credits and the deductibility of sub-central taxes in central government taxes. For 2018, this varied from 15% (Lithuania) to a high of 57% (Sweden), with a median value of 46%. The threshold at which the top rate applies varies widely as a multiple of the average wage, from zero (in Hungary) where the rate schedule is flat, to 25 (in Mexico). For those with income far above the threshold, the average tax rate, tax liability divided by income, should—in principle—be well approximated by the top rate.

Annual Net Wealth Tax. While in 1990 twelve OECD countries had an annual tax on net wealth, by 2018 only four levied such a tax—France, Norway, Spain, and Switzerland—with Switzerland raising more than three times as much revenue as a fraction of total revenues (3.7%) as any of the other three countries (OECD, 2018). (In 2018, France replaced its annual wealth tax with a tax only on immovable property.) Annual taxes on (immovable) property are more widespread, and indeed are levied in all OECD countries. Italy levies an annual tax on financial assets. The Netherlands have a hybrid system that has similarities to an annual wealth tax, imputing an asset-type-specific rate of return to assets and assessing a 30% tax on those imputed returns.

Capital Gains Tax. Five OECD countries levy no tax on shareholders based on capital gains. Of those that do, all tax upon realization rather than accrual. Five more countries apply no tax after the end of a holding period test, while four others apply a more favorable rate afterwards. The tax rate varies widely, with the highest as of 2016 being Finland, at 34%. With a few exceptions, the accrued gains on assets in a decedent’s estate escape income taxation entirely, because the heir can treat the basis for tax purposes as the value upon inheritance.

Wealth Transfer Taxes: Estate, Inheritance, and Gift Taxes. As of 2017, 26 of the 35 OECD countries levied some kind of tax on wealth transfers; in a few countries inter-vivos gifts are included in the base, but in others a separate gift tax regime exists. With the prominent exception of the U.S., which levies an estate tax such that the heirs’ circumstances and relationship to the deceased (other than regarding spouses) do not affect tax liability, most countries levy an inheritance tax, where the liability lies with the recipient and the rate of tax depends on their relationship of the heir to the deceased. The exemption levels and rates vary starkly. Most countries have a relatively low exemption along with a graduated rate structure, the U.S. again being the exception, having an exemption of \$11.4 million per individual in 2019 and a flat rate of 40%.

3.2 The Effective Tax Burden on the Superrich

Assessing the burden of taxes on the superrich by examining statutory rates gives an incomplete picture. First is the issue of shifting, or incidence. Taxes that are “on” the superrich may be shifted via tax-induced changes in pre-tax prices. To what extent this happens depends on the nature of income, and we will discuss below how such general equilibrium effects impact the optimal taxation of the superrich; there is, however, little systematic evidence on the incidence of taxes on the superrich.

Second, the effective tax rate depends on how effective a country's enforcement system is in constraining tax evasion. The measurement of both income and wealth—and taxes on these bases—is distorted by evasion, especially when data is derived from tax returns. Evasion of the superrich is difficult to uncover through traditional means like random audits, as the auditor typically lacks the resources to trace the sophisticated means of evasion often involving layers of financial intermediaries. Using such audit data, [Johns and Slemrod \(2010\)](#) concluded that in the U.S. the ratio of aggregate misreported income to true income generally increases with income, although it peaks among taxpayers with AGI in the 99th to 99.5th percentile.

High-profile leaks from financial intermediaries, such as the 2007 leak from HSBC Bank in Switzerland and the 2015 “Panama Papers” from the firm Mossack Fonseca, have recently allowed researchers to gain insights into tax evasion by the richest. [Alstadsæter et al. \(2019\)](#) use data from these leaks along with administrative data from Norway, Sweden and Denmark to show that evasion rates rise across the income distribution, and conclude that the top 0.01% evade about 30% of the income and wealth taxes they owe. The authors link the account names from the HSBC leak with individual tax data and find that 95% of these foreign account holders did not report the existence of the account on their tax forms.

3.3 Capital Gains and the Plasticity of Taxable Income

As we explore in detail below, one key determinant of the optimal taxation of the superrich is the tax rate elasticity of the tax base. For the superrich, this elasticity depends crucially on what we call the “plasticity” of the tax base, i.e., its malleability. Plasticity is an issue when different kinds of income are subject to different effective tax rates, so that the ease with which higher-taxed income can be converted into lower-taxed income matters. By far the most important aspect of plasticity, with implications both for understanding the effective tax burden on the superrich and for measuring the extent of their income and therefore income inequality, concerns capital gains.

Realized capital gains represent a very high fraction of the reported income of the super-rich. For example, IRS data shows that in tax year 2014 realized capital gains represented 60% of total AGI for the 400 highest-AGI Americans. More generally, [Figure 2](#) plots the distribution of net capital gains as a share of AGI across income groups for tax year 2016, which reveals their concentration at the very top: those with AGI over \$10 million report net capital gains corresponding to 46% of their AGI, whereas it is a negligible fraction for those earning less than \$200,000.

This is not a new phenomenon. [Bailey \(1969\)](#) compared capital gain realizations reported on income tax returns with an estimate of accruing stock gains for individuals over the

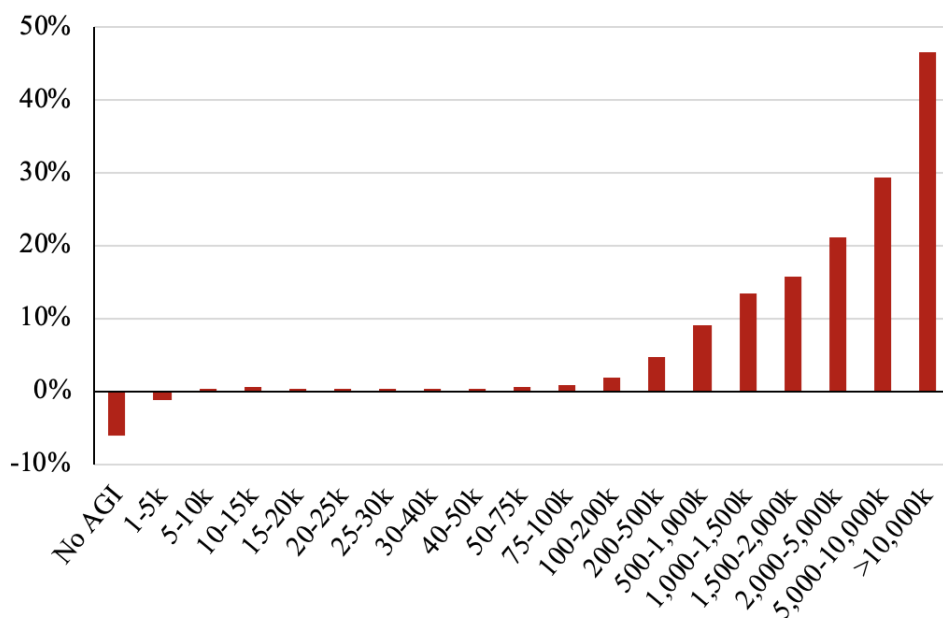


Figure 2: Net capital gains as a share of AGI across income groups (in U.S. \$). Source: IRS Table 1.4 for 2016.

1926–1961 period, and concluded that more than two-thirds of all individuals’ gains on corporate stock were never taxed because the gains were not realized during the holder’s lifetime and the stock holdings were passed along at death. Also telling is that, in tax year 1962, when the top marginal tax rate was 91% on ordinary taxable income over \$400,000, for the 1,376 taxpayers with AGI exceeding \$500,000, the net gain on capital assets—for which the maximum tax rate was 25%—comprised 50% of AGI, while fully-taxed wages and salaries comprised only 4%.

Arguably much of what shows up as capital gains of the superrich is in fact compensation to labor. Indeed, [Fleischer \(2015\)](#) argues that the rise of income inequality in the U.S. since 1990 is attributable mostly to such income, much of which was taxed as capital gains. He focuses on founders’ stock, as a large fraction of the gains of the superrich comes from the sale of stock in a company that was founded by the taxpayer or a relative of the taxpayer. Indeed, for a majority of those on the Forbes 400 list in 2018, founders’ stock was key to their wealth.⁶

If the plasticity of converting labor compensation into capital gains has changed over time, it has implications for interpreting tax-return-based measures of income inequality. Whereas in 1959, a successful inventor might have worked for a big company, earned a nice raise and increased income for his employer, in 2019 the income might barely show up at all

⁶Other examples of capital gains that are likely relabeled labor income are compensation in the form of stock options and the management of investment portfolios (the “carried interest loophole”). See [Goolsbee \(2000\)](#) on how taxes affect the exercise of stock options by corporate executives.

unless the founder sold her shares after an IPO, at which time taxable capital gains emerge. The implications for measured wealth would be different as well, as in 1959 the inventor would gradually accumulate wealth and the shareholders would become wealthier, while in 2019 the wealth would be much more concentrated in the founder.

4 Optimal Income Taxes in the Classic View

4.1 A Simple Formula

We now turn to the normative question on the table—how *should* the superrich be taxed? The canonical model, going back to [Mirrlees \(1971\)](#), posits that top earners are rich simply because they have a greater income-earning ability than most everyone else. To the degree that such an unequal distribution of abilities is given exogenously, it naturally gives rise to a redistributive motive. For example, behind the veil of ignorance, we may agree to “insure” against the realization of ability draws, determined ultimately by the birth lottery. This justifies tax policies that redistribute some share of top incomes away from the superrich and towards less fortunate, lower-ability earners.

Of course, the inequality-reducing effects of redistributive taxation need to be balanced with the disincentive effects. Precisely how this tradeoff should be resolved depends crucially on the choice of social welfare function or, equivalently, one’s political attitudes towards inequality. Yet, it is hard to agree on this based on scientific principles (and, as a matter of fact, we observe a variety of redistributive preferences in the population). Therefore, we choose to ask a broader question: What is the range of Pareto efficient tax rates on top earners? In other words, rather than trying to determine *the* optimal tax policy, we accept that there will be different views on it, and instead focus on ruling out tax policies that we should all agree to be undesirable.

This approach, which follows [Werning \(2007\)](#) and [Scheuer and Werning \(2017\)](#), allows us to isolate the revenue effects of taxing the superrich. In fact, any tax rate to the right of the peak of the Laffer curve (which plots the relationship between a tax rate and revenue) is Pareto inefficient because taxing top earners at a lower rate both makes them better off and raises additional revenues, which can be used to benefit lower earners. The upper bound to the set of Pareto efficient tax rates is thus given precisely by the rate that achieves the peak of the curve, i.e., that maximizes tax revenue. Where that peak lies depends on the behavioral response of top earners to taxes. All tax rates to the left of the peak are Pareto efficient, and hence could be optimal under some social welfare function depending on the social welfare weights put on top earners.

This logic, familiar from optimal linear tax models (and reminiscent of the typical po-

litical rhetoric in favor of “self-financing” tax cuts), is straightforward to extend to nonlinear income taxes. Formally, whether a given tax schedule $T(y)$ (where y denotes income), together with an induced income distribution (described by the density function $h(y)$), is Pareto efficient or not depends on two parameters: Individuals’ behavioral responses to tax changes, as captured by the compensated earnings elasticity $\varepsilon(y)$ and income effects $\eta(y)$; and the shape of the income distribution $h(y)$, summarized by the local Pareto parameter

$$\rho(y) \equiv - \left(1 + \frac{yh'(y)}{h(y)} \right).$$

If the compensated earnings elasticity $\varepsilon(y)$ and marginal tax rate $T'(y)$ do not locally vary with y , [Scheuer and Werning \(2017\)](#) show that the following simple formula describes the upper bound to the range of Pareto optimal marginal tax rates $T'(y)$ at any income level y :

$$T'(y) \leq \frac{1}{1 + \rho(y)\varepsilon(y) - \eta(y)}. \quad (1)$$

Formula (1) has a straightforward intuition. A higher compensated elasticity ε means that income taxes induce greater distortions through behavioral responses, which leads to a tighter upper bound on marginal tax rates. By contrast, a stronger income effect η means that agents react to higher taxes by feeling poorer and thus working more. In this case the tax base shrinks less and hence marginal tax rates can be higher.

Finally, to grasp the intuition for the critical role of the local Pareto parameter ρ , imagine a reduction in the income tax burden $T(y)$ at one particular income level y and check whether the induced behavioral responses lead to an increase in revenue; if they do, there is a local Laffer effect, and the original tax rate cannot have been Pareto efficient. This is illustrated in [Figure 3](#). The first revenue effect is mechanical: we lose revenue from workers who have income of exactly y . In addition, there are two behavioral effects: Some workers who initially had income greater than y reduce their earnings and move to y . This also reduces tax revenues. However, neighboring workers initially with income below y increase their earnings (to benefit from the tax cut at y) and may thereby end up owing more taxes. The net effect on tax revenue is positive if the latter effect outweighs the former two. This in turn depends on the local shape of the income density $h(y)$: If it falls quickly at y , there are many more workers initially to the left of y than to the right, so the latter, positive revenue effect is more likely to dominate. This explains why a larger local Pareto parameter $\rho(y)$ implies that the peak of the local Laffer curve is reached at a smaller marginal tax rate.

We view the upper bound in formula (1) as a natural benchmark. First, as explained above, tax rates above it are guaranteed to be Pareto dominated. Second, in the context of the superrich, it is particularly relevant to know what rate to impose on, say, the top income

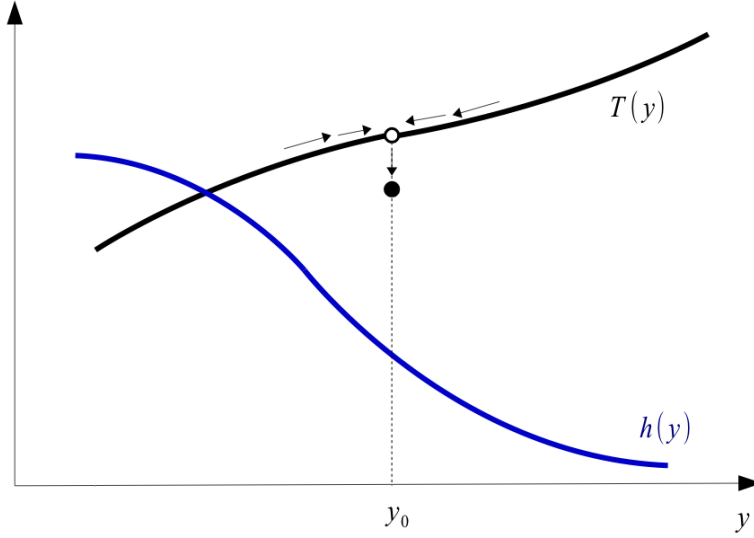


Figure 3: Revenue effects of a local tax cut

bracket or the top 1% if our goal is to raise as much revenue as possible from this group (Diamond and Saez, 2011). Third, it coincides with the optimal tax under a social welfare function that puts no weight on individuals who earn income y , such as (approximately) at the top of the distribution under a utilitarian criterion with diminishing marginal utility. Indeed, under the Rawlsian (maximin) criterion, it is optimal to set the marginal tax rate at the upper bound given by formula (1) at all incomes except the very bottom. Finally, if one puts positive welfare weight on the superrich, the optimal tax rate on their income is strictly lower than the upper bound (the more so the bigger the weight on top earners).

Formula (1) includes the classic zero top rate results due to Phelps (1973), Sadka (1976) and Seade (1977) as a special case. If the income distribution is bounded above, then the density $h(y)$ drops to zero at the top, which implies that the local Pareto parameter $\rho(y)$ diverges to infinity. Consequently, the only Pareto efficient marginal tax at the top is zero. However, this result is specific to the very top earner in the economy and therefore cannot be considered as policy relevant for a broader group of high-income earners. In fact, Pareto (1896) already observed that the top tail of the income distribution is well approximated by what has since then been called a Pareto distribution. This fat-tailed distribution has the convenient property that

$$\rho(y) = - \left(1 + \frac{h'(y)y}{h(y)} \right) = \frac{yh(y)}{1 - H(y)} = \frac{\mathbb{E}[y'|y' \geq y]}{\mathbb{E}[y'|y' \geq y] - y} \quad (2)$$

is independent of y , so the local Pareto parameter ρ is a constant. Hence, in this case, formula (1) also corresponds to the optimal asymptotic marginal tax rates characterized by Saez

(2001), or [Diamond \(1998\)](#) for the case without income effects.

This suggests a simple calibration of formula (1). In the U.S., $\rho(y)$ converges to about 1.4 at the top, and in fact is stable around that level for incomes above \$400,000.⁷ We will discuss evidence on behavioral responses to taxes in more detail below. Various studies find small and insignificant income effects, suggesting $\eta = 0$. Moreover, an elasticity of taxable income $\varepsilon = 0.3$ is approximately in the middle of the range of estimates. Taken together, we obtain an upper bound to the top marginal tax rate of roughly $1/(1 + 1.4 \times .3) = 70\%$. This calculation indicates that in the U.S. there is room to further increase top marginal tax rates, starting from current levels, without running into the downward-sloping part of the Laffer curve. One must bear in mind, though, that this calibration applies to the overall top marginal tax rate from a national perspective, taking into account both central and sub-central income taxes.⁸

Some recent proposals, however, such as Alexandria Ocasio-Cortez's envisioned 70% federal marginal tax bracket on incomes in excess of \$10 million, might exceed the upper bound according to this calibration.⁹ Of course, it is rather heroic to assume that the current shape of the income distribution would be preserved even after such a large tax reform, and the earnings elasticity at the top may also increase under a 70% top marginal tax rate, rather than staying at levels consistent with current estimates. Both ρ and ε cannot be viewed as primitive parameters, but are endogenous to the tax code, so holding them fixed is only valid for small tax reforms.

4.2 The Elasticity of Taxable Income

The foregoing discussion highlights the important role of the potential behavioral response of the superrich to changes in how they are taxed. Since [Feldstein \(1999\)](#), this has come to be referred to as the elasticity of taxable income (ETI) with respect to the net-of-tax rate. There is now a substantial literature that seeks to estimate the ETI, critically reviewed in [Saez et al. \(2012\)](#), but it rarely attempts to estimate the ETI that applies specifically to the rich, much less the superrich. An early study by [Gruber and Saez \(2002\)](#) estimated the ETI

⁷According to the most recent available IRS data from 2012, the average taxable income in the top 1% was roughly \$1.35 million, and the cutoff income for the top 1% was \$390,000, inflated to current U.S. dollars. Using equation (2), $\rho = 1.35/(1.35 - .39) = 1.41$.

⁸California, for instance, has a 13% top marginal income tax rate, which would translate a national revenue-maximizing tax rate of 70% into a top federal tax rate of 57% since the deduction for state and local taxes is capped. Also, this does not take into account the negative fiscal externalities on revenues from other (e.g. sales) taxes.

⁹In 2012, there were about 17,000 tax returns with taxable income above \$10 million, and their average income was \$30 million, all in current U.S. dollars. Hence, the Pareto parameter above this threshold was, by (2), $\rho = 30/(30 - 10) = 1.5$. This implies a national revenue-maximizing rate of $1/(1 + 1.5 \times .3) = 69\%$. The upper bound on the federal top marginal tax rate is lower than this unless the cap on the state and local tax deduction is removed (or state marginal tax rates reduced), and therefore likely considerably below 70%.

separately by income groups, finding 0.57 for the high-income group (with income in excess of \$100,000 in 1992 dollars). [Saez \(2017\)](#) analyzes the behavioral responses to taxing the rich, defined as the top 1%, to the 2013 U.S. income tax increase. He finds a short-term (2012 to 2013) elasticity larger than 1, due to year-to-year shifting of income, but this elasticity is not directly relevant to the optimal steady-state tax rate. The medium-term response, from 2011 to 2015, which would not include year-to-year shifting of taxable income, was noticeably lower, at about 0.25, and, strikingly, was lower for the top 0.01% than for the top 1%. Saez notes, though, that these estimates are very sensitive to the identification assumption about the counterfactual trends in taxable income.

Some proponents of much higher taxation of the superrich take solace from the fact that in the 1950s and 1960s, both the U.K. and the U.S. (until 1971) had top income tax rates that exceeded 90% and their economies did quite well. One factor was that the top tax rate on capital gains was much lower, 25% in the U.S. and 30% in the U.K., providing a safety valve for income that could be classified as capital gain, as discussed above. [Goolsbee \(1999\)](#) analyzed group-aggregated tax-return data for the U.S. for the period when the U.S. entered the 90% top tax rate era (1951) and when it left it (1964), and concluded that the ETI was about 0.15 in the first case and very close to zero in the latter case. However, there is no reason to expect that the ETI of the superrich has remained constant for the past 60 to 70 years.

As [Slemrod and Kopczuk \(2002\)](#) point out, the ETI is not a structural parameter (nor is the elasticity of any tax base); notably, the superrich will be less responsive to a tax rate increase the broader and less plastic is the tax base and the more effective is the enforcement of tax evasion. This suggests that proposals to raise rates should be accompanied by calls for a broader base and better enforcement. [Keen and Slemrod \(2017\)](#) show how the optimal tax rate structure should be determined simultaneously with the optimal setting of these other aspects of tax systems.

Finally, it is important to stress that the required elasticities in (1) should capture the long-run effects of taxes on revenue. The potentially long lags in the causal relationship are inherently difficult to estimate. Examples include career choices and the incentives to accumulate human capital early in life, which drive individuals' high-earnings potential much later in life ([Keane, 2011](#); [Badel et al., 2019](#)); the effects of taxes on entrepreneurial risk taking ([Cullen and Gordon, 2007](#)); and migration responses ([Kleven et al., 2019](#)).

5 Income Taxes in View of Complementarities

Recently, an active literature has started to incorporate richer labor markets into the canonical Mirrlees model. Rather than assuming that wage differences simply reflect exogenous

skill differences, labor economists have developed more realistic models of the labor market in order to explain recent inequality trends. Research in public economics has picked up these ideas from a normative perspective, characterizing optimal taxation including general equilibrium effects, superstar phenomena, and rent-seeking, to name just a few recent advances.

In the following, we focus on the implications of this work for the taxation of the super-rich, and begin with those views that regard complementarities as a driver of top incomes. We distinguish between complementarities at the worker or firm versus the aggregate level.

5.1 Superstar Effects

Many economists believe that “superstar effects,” to use the term coined by [Rosen \(1981\)](#), play an important role in the rise of top income inequality.¹⁰ The idea is that relatively small differences in ability or effort among workers are amplified by other factors, such as technology or globalization, leading to dramatic differences in productivity and, ultimately, pay. In a classic example, the advent of television enabled a small share of performers to capture a massive audience, leaving other artists in the dust.¹¹ A superstar story has also been used to explain the growing divergence in CEO compensation. In this narrative, more talented managers are snatched up by larger, more productive firms. Being surrounded by more efficient workers, having access to better resources and commanding larger markets gives these managers an extra kick to their productivity—a complementarity. Hence, the matching of slightly more talented managers with larger firms accounts for the large difference in income between superstar CEOs and the rest of the pack. Because the distribution of firm sizes is extremely fat-tailed (and has become increasingly so in the course of automation and globalization), top managers can make very large sums that eclipse any apparent differences in underlying skill or effort ([Gabaix and Landier, 2008](#); [Terviö, 2008](#)).

If superstars comprise a substantial portion of the superrich, how does this change how they should be taxed? With earnings disproportionate to inherent skills, it may appear intuitive that superstar effects tilt the calculus balancing efficiency and equality, exacerbating inequality, and leading to higher optimal tax rates on the rich. However, [Scheuer and Werning \(2017\)](#) uncover another, less obvious effect. A common feature of superstar models is the convex relationship between (skill-weighted) effort and earnings. This, in turn, increases the behavioral earnings response to any tax change. Intuitively, a worker induced to provide greater effort by way of lower taxes anticipates being matched with a better job, with better

¹⁰See [Garicano and Rossi-Hansberg \(2015\)](#) for a recent review.

¹¹[Koenig \(2019\)](#) provides evidence for superstar effects in this context, and [Krueger \(2019\)](#) studies this phenomenon in the music business.

pay, and this further amplifies the incentive for effort.¹² Thus, earnings elasticities that do not correctly capture this reassignment are too low.

Scheuer and Werning (2017) show that the revenue-maximizing marginal tax rate in a superstar economy is still given by the upper bound in condition (1), with the only difference that the compensated elasticity ε is now larger than in a standard labor market. Thus, superstar effects do not fundamentally alter the formula for optimal marginal tax rates, but the sufficient statistics featured in the formula—the elasticities and the local Pareto parameter of the income distribution—are endogenous to superstar effects. In particular, for a given observed income distribution (an empirical datum to policymakers at any moment), superstar effects increase the compensated earnings elasticity, which provides a force for lower taxes.

To quantify this effect, it would be important to have direct evidence on the earnings elasticity of superstar workers. Acquiring such evidence, however, encounters three challenges. For instance, one must capture sufficiently long-run responses to tax changes. Short-run estimates, or those based on temporary tax changes, would fail to properly capture the additional reassignment margin that drives the convexity of the earnings schedule in a superstar economy (i.e., in response to a tax cut, workers will not only increase their effort on the job, but also switch to a better job). Moreover, one cannot extrapolate elasticities from non-superstar workers. If what one has in hand are elasticities estimated based on the response of workers further down the income distribution, where superstar effects may be less prevalent, they would miss the fact that the earnings schedule becomes more convex as one moves up the scale.

In the absence of such measures, Scheuer and Werning (2017) show how more commonly estimated earnings elasticities would need to be adjusted in the presence of superstar phenomena. For instance, in the context of CEO compensation, Gabaix and Landier (2008) find empirical support that a CEO’s marginal product of effort increases one-for-one with the size of the firm he or she manages. Using the empirical firm size distribution (which follows Zipf’s law), Scheuer and Werning (2017) provide the following adjustment formula for the earnings elasticity:

$$\varepsilon^S = \frac{\varepsilon}{1 - \rho\varepsilon},$$

where ε^S is the superstar earnings elasticity and ε is the standard one that ignores the convexity of earnings. Using again an estimate for the standard earnings elasticity of $\varepsilon = 0.3$ and the Pareto coefficient of the top tail of the income distribution of $\rho = 1.4$, this would imply a superstar earnings elasticity of $\varepsilon^S = 0.52$, almost twice as high as the standard one. Substituting this in (1) delivers a revenue-maximizing top marginal tax rate of $1/(1 + 1.4 \times .52) =$

¹²We understand effort broadly, i.e., as a catch-all for whatever workers bring to the table, including the hours and intensity of their work effort while on the job, as well as what they do before they even reach the job market to make themselves more productive, such as their human capital investment.

57% instead of 70% under the standard parametrization.¹³

5.2 Aggregate Complementarities

A large literature in labor economics has emphasized the impact of technological change on wage inequality, with a focus on the assignment of skilled and unskilled workers to different tasks and occupations (Goldin and Katz, 1998; Autor et al., 2003). A key feature is imperfect substitutability across sectors in the economy (routine versus abstract tasks, manufacturing versus services, entrepreneurs versus workers), which implies that relative wages are endogenous to the aggregate allocation of employment. Stiglitz (1982) investigated the implications of such a richer labor market for tax policy in a simple two-type model. More recently, Rothschild and Scheuer (2013) have initiated the modern analysis of Mirrleesian taxation in general equilibrium, allowing for more realistic quantification, which has since developed into an active literature.

An insight from this line of work is that complementarities through aggregate technology imply a force for less progressive taxation relative to a world with exogenous wages. To understand the basic intuition, suppose there are two occupations, entrepreneurs and workers, and individuals are free to select into either depending on their skills, as in Rothschild and Scheuer (2013). There is an income tax (which does not condition on occupations) to redistribute across individuals. If there are disproportionately more entrepreneurs at higher incomes, the government can exploit tax-induced general equilibrium effects to enhance redistribution from high- to low-income individuals: Lowering taxes on high earners will disproportionately spur effort among entrepreneurs, which will raise the worker's wages (by stimulating labor demand due to complementarity across occupations) and lower entrepreneurs' wages (by diminishing marginal products within occupations).

These "trickle-down" effects (by which lower earners can benefit from tax cuts on higher earners) therefore push for less progressive taxes. Formally, Rothschild and Scheuer (2013) show that, for the case without income effects, the revenue-maximizing top marginal tax rate is given by

$$T'(y) = \frac{1 - \rho\varepsilon\tilde{\zeta}}{1 + \rho\varepsilon} \quad \text{with} \quad \tilde{\zeta} = \frac{C}{\sigma + S}.$$

The difference compared to the standard formula (1) is the adjustment factor $\tilde{\zeta} > 0$. Here, σ stands for the substitution elasticity between the two occupations. In the case of perfect substitutes, $\sigma = \infty$ and we are back to the canonical Mirrlees model because relative wages are exogenous. As σ becomes smaller, general equilibrium effects become stronger and hence

¹³See Ales and Sleet (2016) and Kapicka and Slavik (2019) for alternative, structural, calibrations, which also find lower taxes at the top than in a standard labor market.

the adjustment ζ larger. This illustrates that, in this framework, the revenue-maximizing top marginal tax rate is always lower than in the standard model. For example, if the income distribution were bounded above, as in [Stiglitz \(1982\)](#), general equilibrium effects would imply that the top marginal tax rate is negative rather than zero: $T'(y) = -\zeta$.

The term $C > 0$ captures the wage compression effects discussed above. However, [Rothschild and Scheuer \(2013\)](#) show that there is an additional effect from endogenous occupational choice, captured by the shift term $S > 0$: Lowering taxes at the top, in order to increase the effort of the top earners and hence raise the workers' wages through general equilibrium effects, will induce some individuals to exit entrepreneurship and become workers, undoing some of the original increase in aggregate effort among entrepreneurs. In other words, occupational shift effects dampen the degree to which general equilibrium effects can be exploited for redistribution, which reduces the adjustment factor ζ relative to a model with fixed assignment.

A challenge in this literature is to quantify the importance of “trickle down” as well as the countervailing shift effects. [Rothschild and Scheuer \(2013\)](#) provide a stylized calibration of their two-sector model assuming Cobb-Douglas technology ($\sigma = 1$) and find small adjustments to be optimal relative to a model with fixed wages. [Rothschild and Scheuer \(2014\)](#) and [Ales et al. \(2015\)](#) develop extensions to more than two sectors, and [Tsyvinski and Werquin \(2017\)](#) and [Sachs et al. \(2019\)](#) to a continuum of sectors; the latter papers also pursue richer calibrations and find quantitatively more important general equilibrium effects.¹⁴

6 Income Taxes in View of Spillovers

In the Mirrlees model and its descendants discussed so far, the return to an individual equals his or her contribution to society. In this view, top incomes, no matter how high, reflect a correspondingly high social marginal product. This means that the economy without taxes is Pareto efficient, and the only reason for taxation is to redistribute.

Some recent policy discussion about rising inequality, however, has questioned whether top incomes result from extraordinary economic productivity. The financial crisis, for instance, exposed numerous examples of highly compensated individuals whose apparent contributions to aggregate output proved illusory. The “Occupy” movement lamented that some of the income growth for the top 1% may have been at the expense of the bottom 99%. Accordingly, the view that some top incomes reflect rent-seeking—i.e., the pursuit of

¹⁴The results also depend on whether sector-specific tax instruments are available, such as the corporate income tax in the context of entrepreneurship ([Scheuer 2013; 2014](#)), and on the specification of technology ([Scheuer and Werning, 2016](#)). Subsequent studies include [Slavik and Yazici \(2014\)](#), [Ales et al. \(2017\)](#), [Costinot and Werning \(2018\)](#), [Thümmel \(2018\)](#), and [Lawson \(2019\)](#).

personal enrichment by enlarging one's slice of the existing economic pie rather than by increasing the size of that pie—has inspired calls for a more steeply progressive tax code.¹⁵

Rent-seeking is an example of a negative externality. Intuitively, the optimal policy is to levy a “corrective” Pigouvian tax equal to the marginal social damage. For example, if an activity provides a private return of one dollar but reduces others' income by 50 cents, the Pigouvian correction is a tax of 50%. If some sectors or professions are more prone to rent-seeking than others (Lockwood et al., 2017), sector-specific corrective surcharges could be used to discourage such activities while redirecting individuals to more productive behavior. Besley and Ghatak (2013), for instance, have proposed higher taxes on financial-sector bonus payments based on the idea that they result from contest-like tournaments or races with “winner-takes-all” compensation, such as high-speed trading. Brock and Magee (1984) argue that many lawyers engage in rent-seeking activities akin to zero-sum games. Bertrand and Mullainathan (2001) and Piketty et al. (2014) suggest that CEOs can raise their pay artificially, for instance due to luck or by stacking a board of directors in their favor.

The first challenge in mapping this intuition into practice is to measure the extent of rent-seeking. Estimating how much individuals are overpaid relative to their social contribution, even at the aggregate or sectoral level, is difficult, although it is certainly an important area for future research. A second challenge is that it may be hard to perfectly target rent-seeking incomes through the tax code. Differential taxation across professions might be constrained if individuals can easily relabel their occupations, if it encourages special interest lobbying for preferential tax treatment, or due to concerns about empowering the government to determine how socially productive different professions really are. Moreover, even sector- or profession-specific taxes are likely imperfectly targeted, as they apply to multiple different activities within such sectors. For example, while some of what lawyers do might amount to rent-seeking, many lawyers produce socially valuable services such as upholding property rights and providing incentives to abide by useful rules.

Rothschild and Scheuer (2016) have incorporated rent-seeking into the Mirrlees model and shown that, under these constraints, it is not generally enough to know how much rent-seeking there is at any given income level in order to determine the optimal “corrective” adjustment to the income tax. It is also critical *at whose expense* overpaid individuals are benefiting. If CEO pay hikes are at the expense of productive workers further down the distribution, then raising top tax rates leads to an increase in more fruitful activities, and hence the correction should be even higher than the simple Pigouvian intuition above would suggest. But if top earners are making outsize incomes by winning against others in the same line of work, raising taxes could backfire. One example is high-speed trading. If

¹⁵Of course, regulation that reduces the amount of rents in the economy might also be called for, but in general there remains a case for tax policy even if these other instruments are set optimally.

the most profitable traders faced higher taxes, that would discourage their activity—the intended effect. But this in turn would also make it easier for other, less efficient, traders who compete against them, with the unintended effect of potentially drawing even more traders into the fray. As a result, the optimal income tax surcharge lies strictly below the Pigouvian benchmark in this case. In fact, [Rothschild and Scheuer \(2016\)](#) provide examples where the optimal top tax rate is unchanged relative to the standard Mirrlees model even though all top incomes result from completely unproductive rent-seeking.¹⁶

Other forms of spillovers could be relevant as well. Some top earners might be paid below their marginal product, such as innovators who only appropriate a fraction of the value of their innovations ([Jones, 2019](#); [Jaravel and Olivi, 2019](#)). Yet another kind of externality, perhaps particularly relevant at the top, arises from “positional” concerns ([Oswald, 1983](#); [Frank, 1999](#)) where individuals fail to internalize that increasing their income makes some others envious. [Rothschild and Scheuer \(2014\)](#) analyze a general model where individuals may pursue activities with positive or negative externalities and argue that operationalizing the “corrective” approach to income taxation requires better measures of both the externalities in the economy—and their distribution across income levels—and of the relative impact of these spillovers on the returns to all activities. These are not easy-to-obtain quantities, suggesting the importance of research similar, for instance, to the work by economists and scientists on the magnitude and distributional impacts of environmental externalities.¹⁷

7 Economic Mobility

7.1 Lifetime versus Snapshot Inequality

The work reviewed so far is based on a one-shot static view of income inequality. In reality, individuals face a productivity profile over their lifetime, as well as random shocks, which produce “churn” in the cross-sectional income distribution from year to year ([Slemrod, 1992](#)). As to the taxation of top incomes, it is then crucial to determine how the composition of the superrich varies over time.

To see this, suppose the government sets an income tax based on annual income, as is the case in practice (i.e., without age- or history-dependence). Moreover, assume individuals have access to complete financial markets (likely a reasonable approximation for many

¹⁶Only in the special case where rent-seeking incomes are equally at the expense of productive workers and other rent-seekers does the Pigouvian correction apply. This is the assumption underlying [Piketty et al. \(2014\)](#) and [Lockwood et al. \(2017\)](#).

¹⁷In the case of behavioral distortions or “internalities,” a paternalistic government could use “corrective” income taxes on similar grounds (see e.g. [Farhi and Gabaix, 2019](#)). For instance, top earners may be disproportionately plagued by workaholicism ([Hamermesh and Slemrod, 2008](#)), although this may be hard to distinguish from the “rat race” effects of positional externalities, which also lead to excessive labor supply.

top earners), abstracting from insurance motives for taxation (Varian, 1980). Scheuer and Werning (2016) show that, without income effects, the optimal marginal tax rate at income y is then given by

$$T'(y) = \frac{1 - \bar{\beta}(y)}{1 - \bar{\beta}(y) + \varepsilon(y)\rho(y)}, \quad (3)$$

where $\bar{\beta}(y)$ is the average welfare weight on those who earn income y . The difference with the static model is that welfare weights are now driven by lifetime inequality rather than cross-sectional inequality at any given moment. Because there may be substantially less of the former than the latter, the average welfare weights naturally vary less across incomes than in the static framework. An extreme case occurs when there is no lifetime inequality, so all inequality in annual incomes is driven by temporary productivity shocks or lifecycle effects (such as retirement). When viewed over their entire lifetimes, all individuals face the same distribution of productivities. In this case, $\bar{\beta}(y)$ is independent of y (and equal to one), so optimal annual income taxes are zero, even though the observed cross-sectional income inequality at any point in time could be arbitrarily large.¹⁸

More generally, this illustrates that the extent of income mobility is an important ingredient to calibrate formula (3) in a dynamic setting. Take, for instance, the top marginal tax rate. Do top earners have consistently high incomes, or are many of them only temporarily at the top of the distribution? In the former case, the average welfare weight on high incomes is low and we effectively collapse back to the static formula (1). But, in the latter case, setting the tax at the revenue-maximizing level becomes a less natural benchmark.

The “Fortunate 400” data collected by the IRS provides some information on the persistence of top incomes. Over the 23 tax years the data covers (1992 to 2014), 4,584 unique taxpayers made the group, compared to 9,200 if there was absolutely no overlap each year. 138 taxpayers made the top 400 in 10 or more years, while the vast majority of 3,262 showed up in just one of the years. In interpreting this apparently low persistence, however, one must keep in mind the importance of realized capital gains in the AGI of the Fortunate 400: Realized gains show particularly high volatility that does not usually reflect variation in true annual economic income. Thus, the numbers in large part tell us that there is a lot of churn year to year among those who realize large gains.

Considering larger groups of top earners, Auten and Gee (2009) show that, among those in the top 0.01% of the income distribution in 1996 (roughly 12,000 households), only 23%

¹⁸If the productivity shocks needed to earn top incomes are mostly due to luck, this justifies higher taxes (Kindermann and Krueger, 2014; Frank, 2016). Indeed, this would be reflected in formula (3) through a low elasticity $\varepsilon(y)$ at high incomes because individuals respond little to a tax they only face with small probability. Nonetheless, the ETI remains the sufficient statistic to determine the optimal marginal tax, whether luck plays an important role or not. On the other hand, productivity shocks might not be exogenous but affected by human capital accumulation.

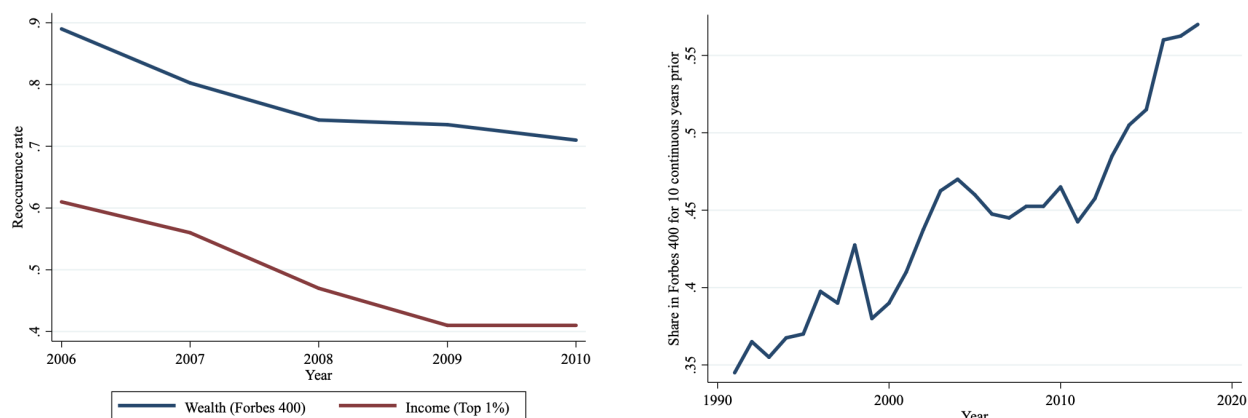


Figure 4: Left panel: Survival rates of the top 1% income earners in 2005 (from [Auten et al., 2013](#)) and those with wealth in the Forbes 400 (own calculations). Right panel: Share of those in the Forbes 400 for 10 continuous years prior, from 1992-2018.

still showed up in this group 10 years later. However, over 80% of them remained within the top 1%, and only 6% dropped out of the top quintile in 2005. The left panel of Figure 4 plots the one- to five-year persistence rates of those in the top 1% of the income distribution in 2005. For instance, 39% of individuals exit the top income percentile already the following year.

Given the churn in annual incomes at the top, one might consider consumption or wealth as better indicators of lifetime inequality. Using the Forbes 400 data on the wealthiest Americans, we compute the one- to 5-year survival rates of those on this list in 2005, and compare them to the survival rates for income in the left panel of Figure 4. This indeed reveals the much higher persistence of top wealth. For instance, of those who were among the Forbes 400 in 2005, 89% percent were still listed in 2006, and 71% still in 2010. In fact, the wealth persistence of the superrich has risen considerably over the last few decades. In the right panel of Figure 4, we plot the share of those in the Forbes 400 in each given year who were on the list also in all 10 prior years. This share has risen from less than 35% in 1992 to almost 60% in 2018, even though the share of “self-made” business founders has increased during the same period. Indeed, while the persistence of inherited wealth has remained relatively stable, the overall trend is driven by the fact that even self-made wealth has become more and more persistent over time.

7.2 Intergenerational Mobility

The transfer of wealth from one generation to the next is another important determinant of economic mobility. Optimal estate taxes must reflect the perspectives of both the (inheriting) children as well as the (bequeathing) parents. From the children’s viewpoint, inheritances

are largely luck, so the principles of meritocracy and equality of opportunity provide widely shared support for high bequest taxes. On the other hand, from the parents' viewpoint, consider two individuals, with identical lifetime earnings, but one of them consumes all of the accumulated resources whereas the other is altruistic and prefers to leave most of them to his or her children. Bequest taxes imply that the latter pays higher lifetime taxes than the former; this pushes for low or even negative bequest taxes. Hence, the level of estate taxes depends on how this tradeoff is resolved, but there is a general case for making these taxes progressive (Farhi and Werning, 2010; 2013; Piketty and Saez, 2013).

Many countries have highly progressive bequest taxes, with considerable exemption amounts but sizable tax rates above them. In principle, this should mean that the superrich are exposed to high estate taxes, yet, in practice, many additional exceptions allow them to significantly reduce their exposure. Notably, capital gains unrealized at death (typically in the form of closely held stocks in private businesses) are excluded from taxation due to step-up provisions. In the U.S., the effective tax rate is reduced by extensive undervaluation of wealth transfers via, for example, family limited partnerships and irrevocable life insurance trusts. Of course, such provisions also affect the behavioral response of bequests to taxation, which is an important determinant of the level of optimal estate taxes. Kopczuk (2013) surveys the evidence and concludes that, despite limitations in empirical strategies, the literature suggests an elasticity of reported estates with respect to the net-of-tax rate between 0.1 and 0.2.

8 Optimal Wealth Taxes

8.1 Arguments for Wealth Taxation

Recently, the discussion about wealth taxation has gained steam in the U.S., partly fueled by evidence of rising wealth inequality (Saez and Zucman, 2016). There is relatively little rigorous analysis of optimal wealth taxation per se, especially when progressive income and bequest taxes are already in place. In part this may be due to the close connection between wealth taxes and taxes on capital income. To see this, assume that all wealth earns an annual return of r ; then levying a tax rate t on capital income is equivalent to imposing a tax rate of rt directly on the wealth each year. For example, a 2% wealth tax is equivalent to a 100% capital income tax if $r = 2\%$ and to a 50% capital income tax if $r = 4\%$.

Capital income taxes have been amply studied, and we do not have the space here to even summarize this literature. The classic benchmark, going back to Judd (1985), used to be that zero capital taxation is optimal in the long run, even in the face of extreme wealth inequality and redistributive preferences. Modern arguments, however, have challenged

this view and argue that there is a case for steady-state capital taxes in this setting (Straub and Werning, 2019). Moreover, capital taxes can be optimal in other models, such as when agents face uninsurable productivity shocks or when wealth is a source of utility (see e.g. Golosov et al., 2007, and Banks and Diamond, 2010, for overviews).

Given that most countries already have capital income taxes, what might justify levying wealth taxes in addition? Guvenen et al. (2019) provide an argument when there is heterogeneity in the return to capital, such as when r is interpreted as the return to a real investment. In this case, the capital-income-tax equivalent of a given wealth tax rate is also heterogeneous: the higher is the return, the lower is the equivalent capital income tax, so that more productive entrepreneurs face a lower capital-income-tax equivalent. As a result, a wealth tax encourages the reallocation of capital from unproductive to productive entrepreneurs, and the authors find efficiency gains of 8% compared to a uniform capital income tax.

On the other hand, if heterogeneous returns reflect heterogeneous windfall gains, rents or excess profits (e.g. due to market power), rather than actual productivity differences, then taxing those away has well-known efficiency benefits. But a wealth tax gets this exactly reversed—it taxes the normal rate of return rather than the excess return. For example, if all investors have a real rate of return of 2%, but some earn additional excess profits on their investments, then a 2% wealth tax would not target any of those rents, whereas a capital income tax would. This is because a wealth tax is equivalent to a unit tax on the rate of return rather than an ad valorem tax. To balance this tradeoff, it would be necessary to decompose the heterogeneity in returns observed in the data (see e.g. Fagereng et al., 2016) into actual productivity differences versus differential rents.

Other arguments for wealth taxes include concerns about wealth inequality per se, for instance when wealth is related to political influence. If this is an urgent problem, annual wealth taxes are also able to compress the wealth distribution relatively quickly because they can correspond to capital income taxes of a rate higher than 100% and have an immediate effect compared to, say, bequest taxes. Relatedly, their introduction may be a means to exploit the lump-sum redistribution of existing capital if it can be implemented unexpectedly, although that of course does not apply in the long-run. Finally, one may view them as yet another backstop against the evasion of other taxes (on income, bequests, etc.), even if the enforcement of wealth taxes is likely also imperfect and costly. Saez and Zucman (2019) and Scheuer and Slemrod (2019) provide recent accounts of these ideas and the optimal design of wealth taxes.¹⁹

¹⁹Another question relevant to interpreting the wealth of the superrich is to what degree they give it away to charities, rather than consuming or bequeathing it. Philanthropist initiatives, such as Bill Gates' and Warren Buffet's "Giving Pledge," receive much attention, but also raise further concerns about the political power of the superrich. Moreover, charitable donations are affected by the tax system, for instance through deductions

8.2 Behavioral Responses to Wealth Taxes

Once one accepts the case for wealth taxes, the question remains how much redistribution they can achieve. Empirical studies of the behavioral response to wealth taxes are much sparser than for income taxes, largely because the taxes themselves are much rarer. They are also harder to generalize from, as the tax bases and relevant enforcement details vary widely.

[Brülhart et al. \(2016\)](#) take advantage of variations in the Swiss wealth tax rate across cantons and over time and find that a 0.1 percentage-point increase in wealth taxes leads to a 3.5% decline in taxable wealth, of which about one-tenth is due to the mechanical effect that tax remittance lowers wealth even if behavior were fixed. [Jakobsen et al. \(2019\)](#) study the cuts in the Danish wealth tax implemented in 1989, and find that reducing the wealth tax rate for the top 1% of the wealth distribution by 1.56 percentage points would raise wealth after 30 years by 70%, corresponding to an elasticity with respect to the net-of-tax rate of return of 2; about half of the response comes from a mechanical effect. They argue that, because the estimated effect grows over time, it could not be all a one-time avoidance effect.

[Seim \(2017\)](#) exploits bunching around a kink in the Swedish wealth tax rate schedule where the rate changes from 0% to 1.5% and estimates an elasticity of taxable wealth with respect to the net-of-tax-rate of between 0.1 and 0.3. He concludes that the elasticity mainly represents reporting responses and finds no evidence of households changing their saving or portfolio composition. Finally, [Durán-Cabé et al. \(2017\)](#) study the surprise re-introduction of a wealth tax in Catalonia in 2011. They find no evidence of it reducing wealth accumulation, and that it triggered substantial tax avoidance via taxpayers changing their asset composition toward exempt assets (mainly company shares) so as to reduce wealth tax liability.

In sum, the small set of empirical studies of wealth taxes in developed countries have reached no consensus on its effects, with the results ranging from a large negative effect on wealth accumulation but minimal avoidance (Denmark and Switzerland) to no noticeable effects on wealth accumulation but non-trivial avoidance (Spain and Sweden).

9 Conclusion

The recent explosion of income and wealth inequality, led by an extraordinary increase in concentration among the very richest swath of households, has focused policy attention on the superrich. Prominent politicians in the U.S. and elsewhere have called for increasing the tax burden on these households, in the form both of higher top rates for existing income taxes as well as new tax levies, notably on wealth, targeting the superrich.

from income and estate tax liability ([Saez, 2017](#)), which make them a vehicle for tax avoidance.

Whether these proposals constitute good policy depends on a number of factors, such as the effects of income and wealth concentration on the desired functioning of the political system, and the appropriate relative weight to put on the well-being of the superrich versus other citizens. In this paper, we put aside these political issues and concentrate on the critical factors in designing the taxation of the superrich that are right in the wheelhouse of modern economics.

Those readers hoping that we will end by laying out exactly how any country should tax the superrich will be disappointed. Modern research allows us to provide upper bounds to the optimal rate of tax, but they depend on a number of empirical magnitudes and modeling assumptions. For instance, we do not know with certainty the crucial elasticity of behavioral responses—and they are unlikely the same everywhere and forever. And while there has been recent progress in understanding how complementarities and spillovers—positive or negative—in the superrich’s activities should shape tax policy, we cannot, at this point, precisely pin down their empirical importance.

The absence of definitive policy conclusions comes with the silver lining of a clear agenda for future research. We need to understand better the elasticity of the tax bases that might be used to tax the superrich, and especially how these elasticities depend on the non-rate aspects of a tax system. This will inform the optimal rate structure and the optimal setting of non-rate tax instruments. And we need to better understand the sources of the riches of the superrich, their effects on the rest of the economy, as well as their dynamics. Are they superstars, rent seekers, or job creators, and how does their composition change over time? New evidence on these questions, tied in with the optimal tax implications laid out here, will inject much-needed scientific objectivity into the ongoing public debate about taxation and the superrich.

References

- ALES, L., A. BELLOFATTO, AND J. WANG (2017): “Taxing Atlas: Executive compensation, firm size, and their impact on optimal top income tax rates,” *Review of Economic Dynamics*, 26, 62–90. 14
- ALES, L., M. KURNAZ, AND C. SLEET (2015): “Technical Change, Wage Inequality and Taxation,” *American Economic Review*, 105, 3061–3101. 5.2
- ALES, L. AND C. SLEET (2016): “Taxing Top CEO Incomes,” *American Economic Review*, 106, 3331–3366. 13

- ALSTADSÆTER, A., N. JOHANNESSEN, AND G. ZUCMAN (2019): "Tax evasion and inequality," *American Economic Review*, 109, 2073–2103. 3.2
- AUTEN, G. AND G. GEE (2009): "Income Mobility in the United States: New Evidence from Income Tax Data," *National Tax Journal*, 62, 301–328. 7.1
- AUTEN, G., G. GEE, AND N. TURNER (2013): "New Perspectives on Income Mobility and Inequality," *National Tax Journal*, 62, 893–912. 4
- AUTOR, D., F. LEVY, AND R. MURNANE (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, 118, 1279–1333. 5.2
- BADEL, A., M. HUGGETT, AND W. LUO (2019): "Taxing Top Earners: A Human Capital Perspective," *Working Paper, Georgetown University*. 4.2
- BAILEY, M. (1969): "Capital gains and income taxation," in *The Taxation of Income from Capital*, ed. by A. Harberger and M. Bailey, Brookings Institution Press, 11–49. 3.3
- BAKIJA, J., A. COLE, AND B. T. HEIM (2012): "Jobs and income growth of top earners and the causes of changing income inequality: Evidence from US tax return data," *Working Paper, Williams College*. 2.2
- BANKS, S. AND P. DIAMOND (2010): "The Base for Direct Taxation," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba, Oxford University Press, chap. 6, 548–648. 8.1
- BARTELS, L. (2016): *Unequal Democracy, Political Economy of the New Gilded Age*, Princeton University Press. 1
- BERTRAND, M. AND S. MULLAINATHAN (2001): "Are CEOs Rewarded for Luck?" *Quarterly Journal of Economics*, 116, 901–932. 6
- BESLEY, T. AND M. GHATAK (2013): "Bailouts and the Optimal Taxation of Bonus Pay," *American Economic Review Papers and Proceedings*, 103, 163–167. 6
- BLOOMBERG (2019): "Methodology," <https://www.bloomberg.com/billionaires/methodology/>. 2.1
- BOURNE, J., E. STEUERLE, B. RAUB, J. NEWCOMB, AND E. STEELE (2018): "More than they realize: The income of the wealthy," *National Tax Journal*, 71, 335–356. 2.2

- BROCK, W. AND S. MAGEE (1984): “The invisible foot and the waste of nations,” in *Neoclassical political economy. The analysis of rent-seeking and DUP activities*, ed. by D. Colander, Ballinger, 177–186. 6
- BRÜLHART, M., J. GRUBER, M. KRAPP, AND K. SCHMIDHEINY (2016): “The elasticity of taxable wealth: Evidence from Switzerland,” *NBER Working Paper 22376*. 8.2
- COSTINOT, A. AND I. WERNING (2018): “Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation,” *NBER Working Paper 25103*. 14
- CULLEN, J. AND R. GORDON (2007): “Taxes and Entrepreneurial Risk-Taking: Theory and Evidence for the U.S.” *Journal of Public Economics*, 9, 1479–1505. 4.2
- DENK, O. (2015): “Who are the top 1% earners in Europe?” *OECD Working Paper*. 5
- DIAMOND, P. (1998): “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates,” *American Economic Review*, 88, 83–95. 4.1
- DIAMOND, P. AND E. SAEZ (2011): “The case for a progressive tax: from basic research to policy recommendations,” *Journal of Economic Perspectives*, 25, 165–190. 4.1
- DOLAN, K. (2018): “Forbes 400 2018: A New Number One And A Record-Breaking Year For America’s Richest People,” <https://www.forbes.com/sites/kerryadolan/2018/10/03/forbes-400-2018-a-new-number-one-and-a-record-breaking-year-for-americas-richest-people/#7e332f4560b7>. 2.1
- DURÁN-CABÉ, J., A. ESTELLER-MORÉ, AND M. MAS-MONTSERRAT (2017): “Behavioral responses to the (re)introduction of wealth taxes: Evidence from Spain,” *Working Paper, Universitat de Barcelona*. 8.2
- FAGERENG, A., L. GUISO, D. MALACRINO, AND L. PISTAFERRI (2016): “Heterogeneity and Persistence in Returns to Wealth,” *NBER Working Paper 22822*. 8.1
- FARHI, E. AND X. GABAIX (2019): “Optimal Taxation with Behavioral Agents,” *American Economic Review*, forthcoming. 17
- FARHI, E., C. SLEET, I. WERNING, AND S. YELTEKIN (2012): “Non-linear Capital Taxation Without Commitment,” *Review of Economic Studies*, 79, 1469–1493. 1
- FARHI, E. AND I. WERNING (2010): “Progressive Estate Taxation,” *Quarterly Journal of Economics*, 125, 635–673. 7.2

- (2013): “Estate Taxation with Altruism Heterogeneity,” *American Economic Review Papers and Proceedings*, 103, 489–495. 7.2
- FELDSTEIN, M. (1999): “Tax avoidance and the deadweight loss of the income tax,” *Review of Economics and Statistics*, 81, 674–680. 4.2
- FLEISCHER, V. (2015): “Alpha: Labor is the new capital,” https://ostromworkshop.indiana.edu/pdf/seriespapers/2015f_c/fleischerpaper.pdf. 3.3
- FRANK, R. (1999): *Luxury Fever: Money and Happiness in an Era of Excess*, Free Press. 6
- (2016): *Success and Luck: Good Fortune and the Myth of Meritocracy*, Princeton University Press. 18
- GABAIX, X. AND A. LANDIER (2008): “Why Has CEO Pay Increased So Much?” *Quarterly Journal of Economics*, 123, 49–100. 5.1
- GARICANO, L. AND E. ROSSI-HANSBERG (2015): “Knowledge-Based Hierarchies: Using Organizations to Understand the Economy,” *Annual Review of Economics*, 7, 1–30. 10
- GILENS, M. (2014): *Affluence and Influence, Economic Inequality and Political Power in America*, Princeton University Press. 1
- GOLDIN, C. AND L. KATZ (1998): “The Origins of Technology-Skill Complementarity,” *Quarterly Journal of Economics*, 113, 693–732. 5.2
- GOLOSOV, M., A. TSYVINSKI, AND I. WERNING (2007): “New Dynamic Public Finance: A User’s Guide,” *NBER Macroeconomics Annual 2006, Volume 21*, 317–388. 8.1
- GOOLSBEE, A. (1999): “Evidence on the High-Income Laffer Curve from Six Decades of Tax Reform,” *Brookings Papers on Economic Activity*, 2, 1–64. 4.2
- (2000): “What Happens When You Tax the Rich? Evidence from Executive Compensation,” *Journal of Political Economy*, 108, 352–378. 6
- GRUBER, J. AND E. SAEZ (2002): “The Elasticity Of Taxable Income: Evidence And Implications,” *Journal of Public Economics*, 84, 1–32. 4.2
- GUVENEN, F., G. KAMBOUROV, B. KURUSCU, S. OCAMPO, AND D. CHEN (2019): “Use it or lose it: Efficiency gains from wealth taxation,” *Working Paper, University of Minnesota*. 8.1
- GUVENEN, F. AND G. KAPLAN (2017): “Top income inequality in the 21st century: Some cautionary notes,” *NBER Working Paper 23321*. 2.2

- HAMERMESH, D. AND J. SLEMROD (2008): “The economics of workaholism: we should not have worked on this paper,” *The BE Journal of Economic Analysis Policy*, 8. 17
- JAKOBSEN, K., K. JAKOBSEN, H. KLEVEN, AND G. ZUCMAN (2019): “Wealth taxation and wealth accumulation: Theory and evidence from Denmark,” *Quarterly Journal of Economics*, forthcoming. 8.2
- JARAVEL, X. AND A. OLIVI (2019): “Optimal Taxation with Market Size Effects,” *Working Paper, London School of Economics*. 6
- JOHNS, A. AND J. SLEMROD (2010): “The distribution of income tax noncompliance,” *National Tax Journal*, 63, 397–418. 3.2
- JONES, C. (2019): “Taxing Top Incomes in a World of Ideas,” *NBER Working Paper 25725*. 6
- JUDD, K. (1985): “Redistributive taxation in a simple perfect foresight model,” *Journal of Public Economics*, 28, 59–83. 8.1
- KAPICKA, M. AND C. SLAVIK (2019): “Organization of Knowledge and Taxation,” *Working Paper, CERGE-EI*. 13
- KAPLAN, S. AND J. RAUH (2013): “It’s the Market: The Broad-Based Rise in the Return to Top Talent,” *Journal of Economic Perspectives*, 27, 35–56. 2.1
- KEANE, M. (2011): “Labor Supply and Taxes: A Survey,” *Journal of Economic Literature*, 49, 961–1075. 4.2
- KEEN, M. AND J. SLEMROD (2017): “Optimal tax administration,” *Journal of Public Economics*, 152, 133–142. 4.2
- KINDERMANN, F. AND D. KRUEGER (2014): “High marginal tax rates on the top 1%? Lessons from a lifecycle model with idiosyncratic income risk,” *NBER Working Paper 20601*. 18
- KLEVEN, H., C. LANDAIS, M. MUNOZ, AND S. STANTCHEVA (2019): “Taxation and Migration: Evidence and Policy Implications,” *NBER Working Paper 25740*. 4.2
- KOENIG, F. (2019): “Superstar Effects and Market Size: Evidence from the Roll-Out of TV,” *Working Paper, London School of Economics*. 11
- KOPCZUK, W. (2013): “Taxation of intergenerational transfers and wealth,” in *Handbook of Public Economics*, ed. by A. Auerbach, R. Chetty, M. Feldstein, and E. Saez, Elsevier, vol. 5. 7.2

- KRUEGER, A. (2019): *Rockonomics: A Behind-The-Scenes Tour of What the Music Industry Can Teach Us about Economics and Life, Currency*. 11
- LAWSON, N. (2019): "Taxing the job creators: Efficient taxation with bargaining in hierarchical firms," *Labour Economics*, 56, 1–25. 14
- LOCKWOOD, B., C. NATHANSON, AND G. WEYL (2017): "Taxation and the Allocation of Talent," *Journal of Political Economy*, 125, 1635–1682. 6, 16
- MATTHEWS, D. (2019): "How to tax the rich, explained," <https://www.vox.com/2019/3/19/18240377/estate-tax-wealth-tax-70-percent-warren-sanders-aoc>. 1
- MIRRLLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208. 4.1
- OECD (2018): *The Role and Design of Net Wealth Taxes in the OECD*. 3.1
- OSWALD, A. (1983): "Altruism, jealousy and the theory of optimal non-linear taxation," *Journal of Public Economics*, 20, 77–87. 6
- PARETO, V. (1896): *La courbe de la répartition de la richesse*. 4.1
- PERSSON, T. AND G. TABELLINI (1994): "Is inequality harmful for growth?" *American Economic Review*, 84, 600–621. 1
- PHELPS, E. (1973): "Taxation of Wage Income for Economic Justice," *Quarterly Journal of Economics*, 87, 331–354. 4.1
- PIKETTY, T. (2014): *Capital in the Twenty-first Century*, Harvard University Press. 1
- PIKETTY, T. AND E. SAEZ (2013): "A Theory of Optimal Inheritance Taxation," *Econometrica*, 81, 1851–1886. 7.2
- 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, 230–271. 6, 16
- RATTNER, S. (2019): "A Better Way to Tax the Rich," <https://www.nytimes.com/2019/01/28/opinion/aoc-wealth-tax.html>. 2
- ROSEN, S. (1981): "The Economics of Superstars," *American Economic Review*, 71, 845–858. 5.1

- ROTHSCHILD, C. AND F. SCHEUER (2013): “Redistributive Taxation in the Roy Model,” *Quarterly Journal of Economics*, 128, 623–668. 5.2
- (2014): “A Theory of Income Taxation under Multidimensional Skill Heterogeneity,” *NBER Working Paper 19822*. 5.2, 6
- (2016): “Optimal Taxation with Rent-Seeking,” *Review of Economic Studies*, 83, 1225–1262. 6
- SACHS, D., A. TSYVINSKI, AND N. WERQUIN (2019): “Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium,” *Econometrica*, forthcoming. 5.2
- SADKA, E. (1976): “Income distribution, incentive effects and optimal income taxation,” *Review of Economic Studies*, 43, 261–267. 4.1
- SAEZ, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68, 205–29. 4.1
- (2017): “Taxing the rich more: Preliminary evidence from the 2013 tax increase,” *Tax Policy and the Economy*, 31, 71–120. 4.2, 19
- SAEZ, E., J. SLEMROD, AND S. GIERTZ (2012): “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Perspectives*, 50, 3–50. 4.2
- SAEZ, E. AND G. ZUCMAN (2016): “Wealth inequality in the United States since 1913: Evidence from capitalized income tax data,” *Quarterly Journal of Economics*, 131, 519–578. 8.1
- (2019): “Progressive Wealth Taxation,” *Brookings Papers on Economic Activity*, forthcoming. 8.1
- SCHEUER, F. (2013): “Adverse Selection In Credit Markets and Regressive Profit Taxation,” *Journal of Economic Theory*, 148, 1333–1360. 14
- (2014): “Entrepreneurial Taxation with Endogenous Entry,” *American Economic Journal: Economic Policy*, 6, 126–163. 14
- SCHEUER, F. AND J. SLEMROD (2019): “Taxing Our Wealth,” *Working Paper, University of Michigan*. 8.1
- SCHEUER, F. AND I. WERNING (2016): “Mirrlees Meets Diamond-Mirrlees,” *NBER Working Paper 22076*. 14, 7.1

- (2017): “The Taxation of Superstars,” *Quarterly Journal of Economics*, 132, 211–270. 4.1, 5.1
- SCHEUER, F. AND A. WOLITZKY (2016): “Capital Taxation under Political Constraints,” *American Economic Review*, 106, 2304–2328. 1
- SEADE, J. (1977): “On the shape of optimal tax schedules,” *Journal of Public Economics*, 7, 203–235. 4.1
- SEIM, D. (2017): “Behavioral Responses to Wealth Taxes: Evidence from Sweden,” *American Economic Journal: Economic Policy*, 9, 395–421. 8.2
- SIMONS, H. (1938): *Personal Income Taxation*, University of Chicago Press. 1
- SLAVIK, C. AND H. YAZICI (2014): “Machines, buildings, and optimal dynamic taxes,” *Journal of Monetary Economics*, 66, 47–61. 14
- SLEMROD, J. (1992): “Taxation and inequality: a time-exposure perspective,” *Tax Policy and the Economy*, 6, 105–127. 7.1
- SLEMROD, J. AND W. KOPCZUK (2002): “The optimal elasticity of taxable income,” *Journal of Public Economics*, 84, 91–112. 4.2
- SMITH, M., D. YAGAN, O. ZIDAR, AND E. ZWICK (2019): “Capitalists in the twenty-first century,” *Quarterly Journal of Economics*, 134, 1675–1745. 2.2
- STEUERLE, E. (1985): “Wealth, realized income, and the measure of well-being,” in *Horizontal Equity, Uncertainty, and Economic Well-being*, ed. by M. David and T. Smeeding, University of Chicago Press, 91–124. 2.2
- STIGLITZ, J. (1982): “Self-Selection and Pareto Efficient Taxation,” *Journal of Public Economics*, 17, 213–240. 5.2
- STRAUB, L. AND I. WERNING (2019): “Positive Long Run Capital Taxation: Chamley-Judd Revisited,” *American Economic Review*, forthcoming. 8.1
- TERVIÖ, M. (2008): “The difference that CEOs make: An assignment model approach,” *The American Economic Review*, 642–668. 5.1
- THE ECONOMIST (2019): “How to tax the rich,” <https://www.economist.com/leaders/2019/02/02/how-to-tax-the-rich>. 2
- THÜMMEL, U. (2018): “Optimal Taxation of Robots,” *CESifo Working Paper 7317*. 14

TSYVINSKI, A. AND N. WERQUIN (2017): "Generalized Compensation Principle," *NBER Working Paper 23509*. 5.2

VARIAN, H. (1980): "Redistributive taxation as social insurance," *Journal of Public Economics*, 14, 49–68. 7.1

WERNING, I. (2007): "Pareto Efficient Income Taxation," *Working Paper, MIT*. 4.1