IFN Working Paper No. 1492, 2024

Top Income Taxation: Efficiency, Social Welfare and the Laffer Curve

Jacob Lundberg

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Top income taxation: Efficiency, social welfare and the Laffer curve

Jacob Lundberg*

18 June 2024

Abstract

This paper develops a comprehensive framework for analyzing the revenue, efficiency and social welfare implications of top income taxation. It generalizes the Saez (2001) formula for the optimal top tax rate by deriving analytical expressions for the Laffer curve and excess burden. Applied to the 2021 U.S. top federal tax bracket, assuming a taxable income elasticity of 0.25, the study finds an excess burden of $101 billion and a maximum potential revenue increase of $111 billion. In contrast, other English-speaking countries and Germany are positioned closer to their Laffer curve peaks, incurring greater efficiency losses, whereas the Nordic countries studied are on the downward-sloping part of the Laffer curve. Additionally, the paper endogenizes the marginal social welfare weight on high-income earners and, following an inverse optimal taxation approach, concludes that in none of the studied countries does the observed top marginal tax rate appear consistent with a conventional welfarist social welfare function.

Keywords: income taxation, optimal taxation, Laffer curve, excess burden

JEL codes: H21, H24

*Research Institute of Industrial Economics (IFN), Stockholm. Email: jacob.lundberg@ifn.se. An earlier version was published as Uppsala University Department of Economics Working Paper 2017:9 under the title "The Laffer curve for high incomes". This paper has benefited from excellent research assistance from Jonathan Olausson Toft. I am grateful for comments from Spencer Bastani, Emil Bustos, Per Engström, Magnus Henrekson, Johannes König, Gabriella Massenz, Katarina Nordblom, Håkan Selin, Tomas Sjögren, Helena Svaleryd, Hirofumi Takikawa, Daniel Waldenström and seminar participants at Uppsala University, Umeå University, ZEW and IFN. Financial support from the Jan Wallander and Tom Hedelius Foundation and the Taxation and Society programme at IFN is gratefully acknowledged.

1 Introduction

Determining the appropriate tax rate on high incomes is a major concern in both academia and politics. Modern top taxation theory began with Saez (2001), who, building on Diamond (1998), derived a much-used expression for the optimal top tax rate. Despite intense theoretical work since then, a coherent framework of top income taxation has remained elusive. This paper provides such a framework, focusing on three aspects: tax revenue, efficiency and social welfare. The main contributions are explicit expressions for the Laffer curve and the excess burden, which serve as generalizations of the Saez (2001) formula and can be applied using the same parameter values.

First, how tax revenue is affected by tax rate changes is obviously policy relevant, and considerable effort goes into scoring tax bills, typically using microsimulation models (e.g., Joint Committee on Taxation, 2015). My expression for the Laffer curve, presented in section 3, allows easy calculation of revenue changes. For instance, assuming a taxable income elasticity of 0.25, implementing the revenue-maximizing top marginal tax rate in the United States (72 percent) would raise revenue by $111 billion for 2021. The ability to draw the Laffer curve also has pedagogical value, as it is arguably the most widely known concept in public economics.

Second, I analyze the efficiency loss, or excess burden, of taxing top incomes in section 4. This is important because taxing incomes, particularly at the high rates in top tax brackets, inevitably implies sacrificing efficiency in order to raise revenue. The derived expression for the excess burden allows it to be calculated with minimal data requirements, aiding policymakers in this tradeoff. Applied to the United States, the excess burden is $101 billion, equivalent to 16 percent of revenue raised from the federal top tax bracket or 5 percent of top earners' total income. This is remarkably similar to Harberger's (1964) estimate, but lower than Feldstein's (1999).

Third, social welfare is discussed in section 5. The key concept is the marginal social welfare weight placed on high-income earners (Saez & Stantcheva, 2016). As emphasized by Chetty (2009a) and Hendren (2020), but overlooked by Diamond & Saez (2011), the social welfare weight is endogenous to the tax rate. Therefore, the current assessment of the social welfare weight cannot in general be used to directly calculate the optimal tax rate. I show that this deviation can be substantial.

Following the inverse optimal taxation literature (Bourguignon & Spadaro, 2012; Hendren, 2020), I infer the social welfare weight on high-income earners implicit in the United States income tax. With plausible magnitudes of behavioural responses and inequality aversion, the social welfare weight is unreasonably high for high tax rates, suggesting that the current top marginal tax rate is inconsistent with a standard welfarist social welfare function, where individuals' utilities are all that matter (cf. Berg & Piacquadio, 2023).

I demonstrate how to visualize the tradeoff between tax revenue and efficiency using a welfare possibilities frontier with tax revenue on one axis and taxpayer surplus (after-tax income minus the disutility of earning that income) on the other. Social optimum occurs at the tangency point with a social indifference curve, whose slope is the marginal social welfare weight. I also show how to draw indifference curves in Laffer space.

I explore top income taxation in a number of other advanced economies in section 6. Australia, New Zealand, Canada, the United Kingdom and Germany have revenue-maximizing rates 10–12 percentage points higher than today, where revenue from the top bracket would be 3–6 percent higher. The total excess burden amounts to roughly one quarter of revenue, which is more than in the United States. At the other end of the spectrum, Denmark, Finland and Sweden find themselves on the downward-sloping side of the Laffer curve and experience substantial efficiency losses. For no country can welfarism be regarded as a successful descriptive theory, in that observed tax rates are either too high or too low.

My analysis underscores that top income taxation poses a difficult tradeoff for policymakers. Most countries raise significant revenue from top incomes and have some potential to increase revenue further, but that would imply accepting even greater efficiency losses. Of course, if the current tax rate is higher than the revenue-maximizing rate there is a clear case for reducing it.

2 Theoretical framework

We begin by denoting the top marginal tax rate by τ, the top tax bracket threshold by b, taxable income by z and average income in the top tax bracket by $\bar{z}_b$. Revenues from the top bracket are then given by $R = \tau N(\bar{z}_b - b)$, where N is the number of taxpayers in that tax bracket.

The elasticity of taxable income with respect to the net-of-tax rate is defined as $\varepsilon = [dz/z] \,/\, [d(1-\tau)/(1-\tau)]$. It captures both labour supply and tax reporting responses to changing tax rates. Feldstein (1999) showed that the taxable income elasticity under some conditions is a sufficient statistic (Kleven, 2021) for evaluating the efficiency losses of income taxation; see Chetty (2009b) for when it may not be.

Next, the Pareto parameter, α, is the ratio of average income to the average tax base, i.e.

$$\alpha = \frac{\bar{z}_b}{\bar{z}_b - b} = 1 + \frac{\tau b N}{R}. \tag{1}$$

It is crucial that the Pareto parameter is consistent with actual revenues; it is preferably calculated directly from revenue statistics, as shown in the last step in equation 1. Either income or total tax revenues can be used, as long as the values for τ and R are internally consistent.

2.1 Small tax changes

It is useful to first consider marginal changes to the top marginal tax rate. We ask, if taxes on high-income earners are raised by a dollar, how much do revenues actually increase once behavioural responses as taken into account? I call this quantity q. As shown in appendix C, we can treat N as constant. Setting income effects aside, we have

$$q(\tau) = \frac{dR/d\tau}{dR/d\tau|_Z} = \frac{dR/R}{d\tau/\tau} = 1 - \frac{\alpha\varepsilon\tau}{1-\tau}, \tag{2}$$

where $|_Z$ denotes holding the tax base constant. We note that q can be interpreted as the elasticity of tax revenues with respect to the tax rate (because $dR/d\tau|_Z = Z = R/\tau$). The last expression for q is well known (e.g., Saez et al., 2012).[1] A rigorous derivation of q, also accounting for income effects, is given in appendix C, where it is shown that the expression is true regardless of utility function and income distribution.

[1]The quantity $1 - q(\tau) = \alpha\varepsilon\tau/(1-\tau)$ can also be termed the degree of self-financing of a tax cut (Sørensen, 2014) or the fiscal externality (Hendren, 2016).

2.2 Tax system and taxpayer behaviour

When considering non-marginal changes to τ and extrapolating from current parameter values of α and ε – for example setting $q = 0$ to recover the Saez (2001) revenue-maximizing tax rate $1/(1 + \alpha\varepsilon)$ – we need to make structural assumptions about taxpayer behaviour and the income distribution. I make three assumptions, which ensure that α and ε are invariant to changes in the tax rate. They are the simplest assumptions that lead to the Saez (2001) top tax rate; see appendix D.

Assumption 1. *The tax function is given by* $T(z) = \max\{0, \tau(z - b)\}$.

I set taxes below the top tax bracket to zero in order to ensure that the budget set is convex for all top marginal tax rates between 0 and 100 percent – the domain of the Laffer curve. When both the budget set and individuals' preferences are convex, jumping between the interiors of different segments of the tax schedule is ruled out and the top marginal tax rate does not affect revenues from lower tax brackets.

Next, I use the dominant functional form for the utility function.

Assumption 2. *The individual maximizes the isoelastic and quasilinear utility function*

$$u(z; z_0) = z - T(z) - \frac{z_0}{1 + \frac{1}{\varepsilon}} \left(\frac{z}{z_0}\right)^{1+\frac{1}{\varepsilon}}, \tag{3}$$

where z_0 is potential income (taxable income if the tax rate were zero) and the taxable income elasticity ε is the same for all individuals.

Other functional forms imply that the elasticity varies with the tax rate in opaque ways; see the discussion in appendix E. Since I assume that everyone has the same elasticity, individuals are heterogenous in potential income only. Like Saez (2001), I disregard extensive marginal responses, which are usually thought to be small for high-income earners, and income shifting, which is difficult to parameterize.

As the utility function is quasilinear, there are no income effects. The literature reviews by Saez et al. (2012) and Neisser (2021) both state that the empirical literature rarely reports the size of income effects, and therefore assume that compensated and uncompensated elasticities are equal. Income effects are instead discussed as an extension.

For potential top bracket earners, utility maximization yields the taxable income supply function

$$z(\tau; z_0) = \begin{cases} b & \text{if } b \leq z_0 \leq b/(1-\tau)^{\varepsilon} \\ z_0(1-\tau)^{\varepsilon} & \text{if } z_0 > b/(1-\tau)^{\varepsilon}. \end{cases} \tag{4}$$

2.3 Income distribution

I normalize the number of potential top bracket taxpayers to 1 and make the following assumption about the potential income distribution.

Assumption 3. *Potential incomes exceeding* b *follow a Pareto (type I) distribution with the density function*

$$f_0(z_0) = \frac{\alpha b^\alpha}{z_0^{\alpha+1}}. \tag{5}$$

From this we can derive the income distribution by inverting the taxable income supply function (equation 4) to obtain $z_0 = z/(1-\tau)^\varepsilon$. We then apply the formula for the transformation of random variables to equation 5:

$$f(z) = f_0(z_0)\frac{dz_0}{dz} = \frac{f_0(z/(1-\tau)^\varepsilon)}{(1-\tau)^\varepsilon} = \frac{\alpha b^\alpha (1-\tau)^{\alpha\varepsilon}}{z^{\alpha+1}} \quad \text{for } z > b. \tag{6}$$

Note that the resulting income distribution is the potential income distribution scaled down by the factor $(1-\tau)^{\alpha\varepsilon}$. Thus the density in the top bracket is $(1-\tau)^{\alpha\varepsilon}$. The rest of the potential top bracket taxpayers, $1-(1-\tau)^{\alpha\varepsilon}$, bunch at the kink point b, where the density is infinite. Bunching is a prediction of the standard model applied here, but is usually empirically small (Saez, 2010; Bastani & Selin, 2014; Kleven, 2016). Chetty (2012) shows that modest optimization frictions can reconcile this observation with the elasticities estimated in the quasiexperimental literature. Such optimization frictions – not considered in this paper – may, however, imply that the Pareto parameter will change with the tax rate.

Equation 6 illustrates that the Pareto distribution is scale invariant: A Pareto-distributed variable will still be Pareto distributed, with the same parameter α, if multiplied by a constant (Nair et al., 2022). Since the taxable income supply function (equation 4) implies a multiplicative relationship between potential and realized incomes in the top tax bracket, realized incomes are Pareto distributed if and only if potential incomes are.

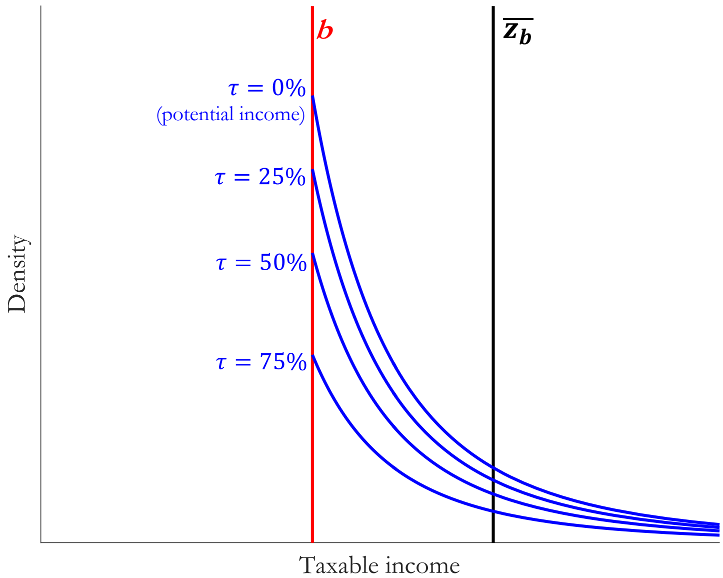
The original observation by Pareto (1896), that the right tail of an income distribution is well approximated by this distribution, has been confirmed repeatedly (Clementi & Gallegati, 2005; Atkinson et al., 2011; Bastani & Lundberg, 2017). Atkinson et al. (2011) describe some mechanisms that may explain this result theoretically.

We can now derive a crucial result that follows directly from assumptions 1–3.

Lemma 1. *Average income in the top tax bracket is independent of the tax rate.*

Proof. Since the Pareto distribution is scale invariant and since there is a multiplicative relationship between potential and realized incomes, the Pareto parameter of the realized incomes is the same as the Pareto parameter of potential incomes, and independent of the tax rate.[2] As the Pareto parameter (equation 1) is a function of the average income of top-bracket taxpayers ($\bar{z}_b$), it is also the same regardless of tax rate. This can be verified by calculating $\bar{z}_b$ directly from equation 6. □

Instead of average income, the *number* of people in the top tax bracket must change after a tax reform (as opposed to the marginal change in the tax rate considered in section 2). Figure 1 shows an example.



Note: Assumes a Pareto parameter of 2.5 and a taxable income elasticity of 0.25.

Figure 1: Illustration of the right tail of an income distribution

[2]See also the discussion in Saez (2001), p. 212.

3 Tax revenue

Using assumptions 1–3, we derive the following expression for the Laffer curve, which is the first main contribution of the paper.

Proposition 1. *Tax revenues from the top tax bracket are given by*

$$R(\tau) = \tau(1-\tau)^{\alpha\varepsilon} Z_0, \tag{7}$$

where $Z_0 = \bar{z}_b - b = b/(\alpha - 1)$ *is the potential tax base (total income in the top tax bracket in the absence of taxation).*

Proof. Follows immediately from equation 6:

$$R(\tau) = \tau \int_b^\infty (z-b)\, f(z) dz = \tau \int_b^\infty (z-b)\, \frac{\alpha b^\alpha (1-\tau)^{\alpha\varepsilon}}{z^{\alpha+1}} dz = \tau(1-\tau)^{\alpha\varepsilon} \frac{b}{\alpha-1}. \qquad \square$$

The maximum is given by the Saez (2001) formula $\tau = 1/(1+\alpha\varepsilon)$, and tax revenues are zero at tax rates of 0 and 100 percent, as expected.

The tax base is $Z(\tau) = (1-\tau)^{\alpha\varepsilon} Z_0$. Note that $\alpha\varepsilon$ can be interpreted as the elasticity of the tax base and that the relative magnitude of α and ε does not matter, only their product. Intuitively, because the top tax rate only applies to a part of income, the behavioural response is amplified by the Pareto parameter α when expressed in relation to tax revenue. Thus, top income taxation is especially distortionary.

Given that the potential tax base is not directly observable, it may be useful to express the Laffer curve in relation to current tax revenue:

$$R(\tau) = R_\mathrm{c} \frac{\tau}{\tau_\mathrm{c}} \left(\frac{1-\tau}{1-\tau_\mathrm{c}} \right)^{\alpha\varepsilon}, \tag{8}$$

where τ_c is the current tax rate. Applying this formula, one can show that implementing the revenue-maximizing rate would increase revenues by a factor of

$$\Delta\% R_{\max} = \frac{(\alpha\varepsilon)^{\alpha\varepsilon}}{\tau_\mathrm{c}(1-\tau_\mathrm{c})^{\alpha\varepsilon}(1+\alpha\varepsilon)^{1+\alpha\varepsilon}} - 1. \tag{9}$$

Table 1 shows this for some combinations of τ_c and $\alpha\varepsilon$.

The only other explicit expression for the high-income Laffer curve that I have found is derived by Badel (2013) and is given by $R = \tau(z_0(1-\tau)^\varepsilon - b)$.[3] This method only holds for a representative individual and thus fails to take into account the fact that the number of high-income taxpayers in general is a function of the tax rate. The peak of Badel's

[3] Giertz (2009) has a similar approach, but focuses on the change in revenue.

Table 1: Revenue increase potential

Current tax rate	$\alpha\varepsilon =$ 1.5×0.1 $= 3 \times 0.05$	1.5×0.2 $= 3 \times 0.1$	1.5×0.4 $= 3 \times 0.2$	1.5×0.6 $= 3 \times 0.3$	1.5×0.8 $= 3 \times 0.4$	1.5×1 $= 3 \times 0.5$
10%	551%	411%	270%	195%	149%	118%
20%	231%	165%	98%	64%	44%	30%
30%	125%	84%	43%	23%	12%	6%
40%	73%	44%	18%	6%	1%	0%
50%	42%	22%	5%	0%	**1%**	**5%**
60%	23%	9%	0%	**2%**	**10%**	**22%**
70%	10%	2%	**2%**	**13%**	**33%**	**62%**
80%	2%	**0%**	**14%**	**43%**	**89%**	**160%**
90%	**1%**	**10%**	**53%**	**137%**	**287%**	**553%**
$\tau_{\max}$	87%	77%	63%	53%	45%	40%

Notes: Shows the increase in top tax bracket revenue (equation 9) that would result from replacing the current top marginal tax rate with the revenue-maximizing rate $\tau_{\max} = 1/(1 + \alpha\varepsilon)$. Figures in **bold** indicate that the revenue-maximizing rate is lower than the current tax rate. As it is the product of the Pareto parameter (α) and the taxable income elasticity (ε) that is relevant for revenues, results are shown for both a fat-tailed country ($\alpha = 1.5$) like the United States and for a thin-tailed country ($\alpha = 3$) like Sweden.

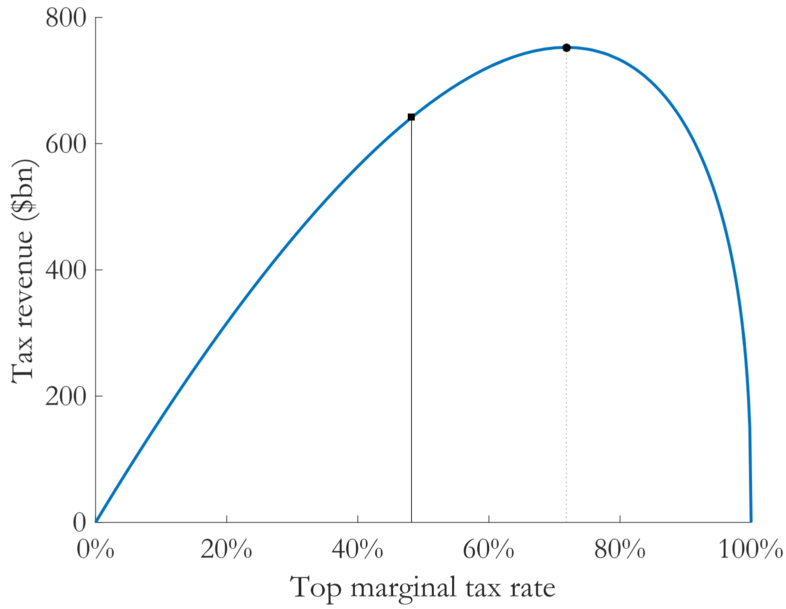
curve does not coincide with the Saez revenue-maximizing rate and the curve predicts negative tax revenue for high tax rates.

Equation 7 can be derived directly from the expression for q, since q is related to the slope of the Laffer curve: Equation 2 constitutes a differential equation in R and τ, which can be solved if $\alpha\varepsilon$ is treated as a fixed parameter. Intuitively, if we know the slope of the Laffer curve at each point, we can trace out the curve itself. The expression for q holds even if $b \to \infty$, given Pareto distribution of incomes, implying that the Laffer curve expression is true also asymptotically.

The Laffer curve for a proportional tax can easily be obtained by setting $\alpha = 1$ so that $R = \tau(1 - \tau)^{\varepsilon} Z_0$. This expression is known in the literature (e.g., Usher, 2014).

3.1 Empirical application

We can now apply the expression for the Laffer curve to the United States top federal income tax bracket (more countries are analyzed in section 6). The choice of taxable income elasticity is of course crucial. Chetty (2012) shows that an elasticity of 0.33 is consistent with 15 central papers. Saez et al. (2012) conclude that 0.25 is a reasonable midpoint of the literature. In a comprehensive meta-analysis, Neisser (2021) finds an average elasticity



Note: Assumes a taxable income elasticity of 0.25. Refers to the top federal income tax bracket, with state revenues included. The current tax rate is indicated.

Figure 2: The Laffer curve for top incomes in the United States, 2021

of 0.4 when looking at papers who consider taxable income after deductions.

It is often posited that high-income earners are more responsive than the population at large, for example because they may have more access to tax avoidance measures. E.g., Gruber & Saez (2002) found an elasticity of 0.57 for those earning more than $100,000. However, Neisser (2021) finds that analyses of the introduction of a top tax bracket are not significantly different from estimates in the literature at large.

Most studies in the new tax responsiveness literature have a follow-up period of a few years. It is likely that it takes longer than that for taxpayers to respond on some margins, so that the long-run elasticity is higher. For example, Kleven et al. (2023) find an elasticity of 0.4–0.5 in Denmark when looking only at income-earners who switch jobs. They argue that this is the relevant long-run elasticity.

In the main analysis, I will use an elasticity of 0.25, which has become the norm in the literature. Considering the above, this is probably a conservative estimate of the long-run elasticity for high-income earners. However, the literature is far from consensus. A sensitivity analysis is provided in supplementary figure 10. In addition, the mathematical expressions I provide enable the reader to easily plug in his or her preferred elasticity.

To calculate the Pareto parameter, we apply equation 1. Using the subscript n to clarify that we are looking only at the national income tax, for 2021 we have $\tau_n = 37\%$, $b = \$628{,}300$, $N = 1{,}189{,}000$ and $R_n = \$485$bn (IRS, 2024), so that $\alpha = 1.57$.[4] As shown

[4]This also implies $\bar{z}_b = \$1{,}731{,}000$. The bracket threshold refers to married joint filers, who constitute

in supplementary table 8, the US Pareto parameter has been remarkably stable over the last quarter-century, although exhibiting some countercyclicality, ranging between 1.54 and 1.69, with an average of 1.61.

When drawing the Laffer curve, all taxes that drive a wedge between before-tax and after-tax earnings need to be considered. Taking Medicare and state income and sales taxes into account, the effective marginal tax rate (τ) is 48 percent for the average top bracket taxpayer; see the calculation in supplementary table 5. Thus, total government revenue from the top tax bracket was $641 billion.[5]

Given these parameter values, figure 2 illustrates the Laffer curve specific to the top tax bracket. The revenue-maximizing rate is $1/(1 + 1.57 \times 0.25) = 72$ percent. This is close to Diamond & Saez (2011), who used the same elasticity but a slightly lower Pareto parameter to arrive at 73 percent. Raising the top rate to 72 percent would make the tax base contract by $283 billion but still raise an additional $111 billion in revenue. We can also calculate $q = 0.64$, meaning that if taxes are raised by a dollar, revenues increase by 64 cents once behavioural responses are taken into account.

3.2 Income effects

Applying the differential equation method described above, it is possible to account for income effects, albeit somewhat crudely.[6] Using the expression for q with income effects in appendix C (equation 26), we can write the Laffer curve as $R(\tau) = \tau(1-\tau)^{\alpha\varepsilon_c+\eta}Z_0$, where ε_c is the compensated taxable income elasticity and η is the income effect parameter, showing the impact on after-tax income of a one dollar increase in unearned income. This assumes that η is constant, which is a questionable assumption, so the expression for the Laffer curve is only approximately correct.[7]

---

the great majority in this bracket.

[5]In the theoretical derivation, I normalized the number of potential top bracket taxpayers to 1. Here, they number $1{,}189{,}000/(1-\tau)^{\alpha\varepsilon} = 1{,}189{,}000/(1-0.48)^{1.57\times0.25} \approx 1{,}539{,}000$. The employer FICA tax (1.45 percent of earnings) is not part of taxable income but is here added to the tax base. Thus the potential tax base (Z_0) is $1{,}539{,}000 \times 1.0145 \times (\$1{,}731{,}000 - \$628{,}300) = \$1.72$ trillion.

[6]Various extensions to the Saez (2001) formula for extensive margin responses, income shifting, bargaining, migration, human capital and innovation have been presented over the years (Saez et al., 2012, Piketty & Saez, 2013; Piketty et al., 2014; Badel & Huggett, 2017; Jones, 2022). While out of scope for the present paper, it should be possible to extend the Laffer curve to incorporate such effects using the differential equation method and possibly more theoretically grounded methods. Incorporating dynamic responses such as saving (Trabandt & Uhlig, 2011; Guner et al., 2016) probably requires numerical methods.

[7]Keane (2011) analyzes a utility function equivalent to equation 3 on page 5, but with income effects. His derivations imply that η will change after a tax reform since his variable S, which he shows determines the magnitude of income effects, is a function of virtual income in a piecewise linear tax system. Virtual income depends on the tax rate. See also the discussion in Golosov et al. (2021, p. 40). Saez (2001) considers the limiting case $b \to \infty$, in which η may be constant. Note that $\eta < 0$ and that by the Slutsky equation, the uncompensated elasticity is given by $\varepsilon_u = \varepsilon_c + \eta$.

As most of the empirical literature disregards income effects, it is not entirely clear whether empirically estimated elasticities should be interpreted as compensated or uncompensated, but it seems reasonable to assume the elasticity is uncompensated in most cases.[8] Thus, if the Saez et al. (2012) assessment of 0.25 is interpreted as uncompensated, and we calibrate income effects using the estimate $\eta \approx -0.1$ from the lottery study by Imbens et al. (2001), the compensated elasticity is 0.35.[9] Then, accounting for income effects lowers the revenue-maximizing rate from 72 to 69 percent and the maximum additional revenue potential is \$87 billion instead of \$111 billion. The resulting Laffer curve is illustrated in supplementary figure 11.

[8]Equation 2 in Neisser (2021), described as the most standard regression specification in the literature, does not feature any controls for changes in the absolute tax liability to capture income effects, implying that the coefficient on the log net-of-tax rate should be interpreted as uncompensated. Golosov et al. (2021, p. 40) also hold the uncompensated elasticity constant.

[9]This effect aligns with Cesarini et al. (2017), but is considerably smaller than estimated by Golosov et al. (2024).

4 Efficiency

From assumptions 1–3 it is possible to derive measures of efficiency, in particular the excess burden, or deadweight loss, of taxing top incomes. We begin by substituting the taxable income supply function (equation 4 on page 5) into the utility function (equation 3), yielding the indirect utility function

$$v(\tau; z_o) = \begin{cases} b - \dfrac{z_0}{1+\frac{1}{\varepsilon}} \left(\dfrac{b}{z_0}\right)^{1+\frac{1}{\varepsilon}} & \text{if } b \leq z_0 \leq \dfrac{b}{(1-\tau)^{\varepsilon}} \\ \tau b + \dfrac{z_0(1-\tau)^{\varepsilon+1}}{\varepsilon+1} & \text{if } z_0 > \dfrac{b}{(1-\tau)^{\varepsilon}}. \end{cases} \tag{10}$$

Given that the utility function is quasilinear with consumption as the numeraire, it is money-metric, allowing for direct comparison with tax revenues for welfare analysis. Moreover, equation 3 is the unique money-metric utility specification, so the efficiency measures derived in this section are not the result of some arbitrary utility normalization.

4.1 Excess burden

The excess burden at the individual level is the reduction in utility relative to a zero tax rate, over and above the amount of tax paid (Auerbach, 1985):

$$EB(\tau; z_0) = v(0; z_0) - v(\tau; z_0) - T(z(z_0)) = \begin{cases} \dfrac{1+\varepsilon\left(\dfrac{b}{z_0}\right)^{1+\frac{1}{\varepsilon}}}{\varepsilon+1} z_0 - b & \text{if } b \leq z_0 \leq b/(1-\tau)^{\varepsilon} \\ \dfrac{1-(1+\varepsilon\tau)(1-\tau)^{\varepsilon}}{\varepsilon+1} z_0 = & \\ = \dfrac{(1-\tau)^{-\varepsilon}-(1+\varepsilon\tau)}{\varepsilon+1} z & \text{if } z_0 > b/(1-\tau)^{\varepsilon}. \end{cases} \tag{11}$$

From this we can derive an expression for the aggregate excess burden. This is the second main contribution of the paper.

Proposition 2. *The total excess burden of the top marginal tax rate is given by*

$$EB(\tau) = \frac{1-(1+\alpha\varepsilon\tau)(1-\tau)^{\alpha\varepsilon}}{1+\alpha\varepsilon} Z_0. \tag{12}$$

Proof. Found by integrating equation 11 over the potential income distribution: $EB(\tau) = \int_b^{\infty} EB(\tau; z_0) f_0(z_0) dz_0$. □

As expected, the excess burden equals zero when $\tau = 0$. Observe the similarity with the individual excess burden for the non-buncher (the second case in equation 11), but with

Table 2: Average excess burden

Top marginal tax rate (τ)	$\alpha\varepsilon =$ 1.5×0.1 $= 3 \times 0.05$	1.5×0.2 $= 3 \times 0.1$	1.5×0.4 $= 3 \times 0.2$	1.5×0.6 $= 3 \times 0.3$	1.5×0.8 $= 3 \times 0.4$	1.5×1 $= 3 \times 0.5$
10%	1%	2%	3%	5%	7%	8%
20%	2%	4%	7%	11%	15%	20%
30%	3%	6%	12%	19%	26%	34%
40%	4%	9%	19%	29%	42%	55%
50%	6%	12%	27%	44%	63%	86%
60%	8%	17%	39%	65%	97%	137%
70%	12%	25%	57%	100%	156%	231%
80%	17%	37%	90%	167%	281%	449%
90%	27%	62%	170%	359%	695%	1301%

Notes: Shows the excess burden of the top marginal tax rate as a percentage of top bracket tax revenues (equation 13). As the product of the Pareto parameter (α) and the taxable income elasticity (ε) is what is relevant, results are shown for both a fat-tailed country ($\alpha = 1.5$) like the United States and for a thin-tailed country ($\alpha = 3$) like Sweden.

$\alpha\varepsilon$ instead of ε.

We can express the excess burden in relation to revenue collected (the average excess burden):

$$\frac{EB(\tau)}{R(\tau)} = \frac{(1-\tau)^{-\alpha\varepsilon} - 1 - \alpha\varepsilon\tau}{\tau(1+\alpha\varepsilon)}. \tag{13}$$

Table 2 shows this for various parameter values.

How does this expression compare with the standard Harberger (1964) approach? At the individual level, the Harberger formula as corrected by Browning (1987) gives an excess burden of $z\varepsilon\tau^2/2(1-\tau)$. This assumes a linear taxable income supply function and hence calculates the excess burden as a triangle. For the average income earner, the excess burden as a share of tax revenue is $\bar{z}_b\varepsilon\tau^2/2(1-\tau)/\tau(\bar{z}_b - b) = \alpha\varepsilon\tau/2(1-\tau)$. This is equal to my expression if $\alpha\varepsilon = 1$. For $\alpha\varepsilon < 1$, my expression gives a lower number, and vice versa. See appendix E for a comparison between the linear and isoelastic taxable income supply functions, and an argument in favour of the latter.

4.2 Empirical application

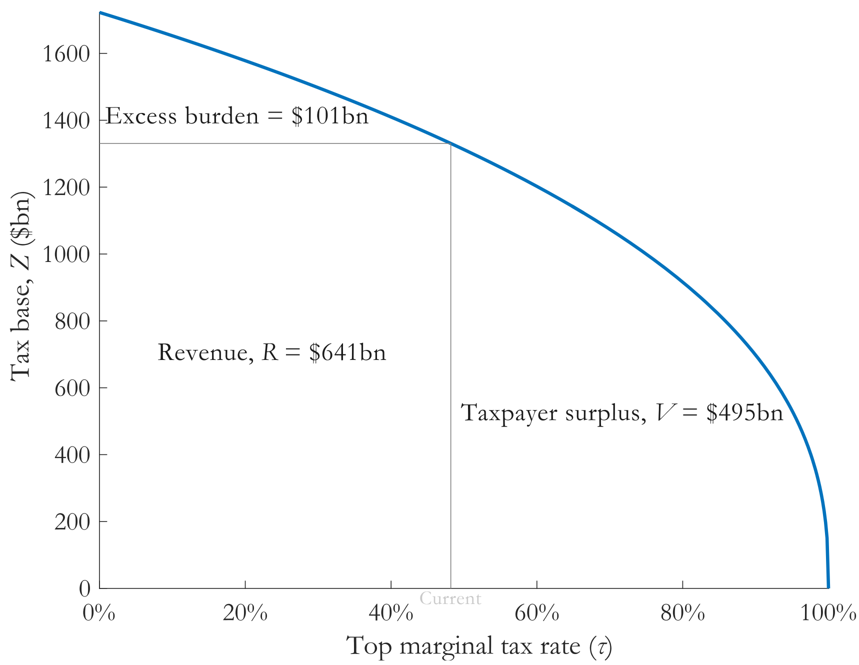
Applying equation 12 to the United States using the parameter values in the preceding section, the total excess burden caused by the taxation of top-bracket incomes is $101 billion – 16 percent of revenue raised or 5 percent of total income for affected taxpayers. This is strikingly similar to Harberger (1964), who found that the deadweight loss of the then 91 percent top federal income tax amounted to 5 percent of total income for those taxpayers, assuming a labour supply elasticity of 0.125. My estimate is lower than Feldstein (1999), who calculated the deadweight loss of the entire personal income tax (not just for high-income earners) to be 32 percent of revenue raised, using an elasticity of 1.04. Jorgenson & Yun (1991) came to a similar conclusion in a model that also incorporated saving, calculating an excess burden of 31 percent of total personal income tax revenue for 1986 tax law.

4.3 Taxpayer surplus and the policymaker's tradeoff

We can also find an expression for taxpayer surplus, which I denote V. Analogous to producer surplus, taxpayer surplus is after-tax income in the top bracket minus the disutility associated with earning that income. We calculate it by summing total utility while subtracting the utility level if the individual had earned b, $u(b; z_0)$:

$$V(\tau) = \int_b^\infty \left[v(\tau; z_o) - u(b; z_0)\right] f_0(z_0) dz_0 = \frac{(1-\tau)^{1+\alpha\varepsilon}}{1+\alpha\varepsilon} Z_0. \tag{14}$$

As expected, it decreases with the tax rate and reaches zero when the tax rate hits 100 percent and everyone is a buncher. The excess burden is how much lower the sum of tax revenue and taxpayer surplus is compared to a tax rate of zero ($EB(\tau) = V(0) - V(\tau) - R(\tau)$; see figure 4). As in a supply–demand diagram, taxpayer surplus and excess burden can be



Note: Plots the function $Z(\tau) = (1-\tau)^{\alpha\varepsilon} Z_0$, assuming $\varepsilon = 0.25$.

Figure 3: The tax base by tax rate in the United States, 2021

calculated as areas under the tax base curve (figure 3).[10]

The government's decision about the top marginal tax rate ultimately boils down to a choice between $V(\tau)$ and $R(\tau)$, which can be illustrated as a welfare possibilities frontier with the equation

$$\frac{R(V)}{Z_0} = \left[\frac{V}{Z_0}(1+\alpha\varepsilon)\right]^{\frac{\alpha\varepsilon}{1+\alpha\varepsilon}} - \frac{V}{Z_0}(1+\alpha\varepsilon). \tag{15}$$

It is illustrated for the United States in figure 6a.

The slope of the welfare possibilities frontier – with opposite sign – is the marginal revenue increase q (see section 2):

$$q(\tau) = -\frac{dR}{dV} = \frac{dR/d\tau}{dR/d\tau|_Z} = 1 - \frac{EB'(\tau)}{R'(\tau)|_Z} = 1 - \frac{\alpha\varepsilon\tau}{1-\tau}. \tag{16}$$

This follows from the the envelope theorem: Behavioural responses are of second-order importance for optimizing taxpayers, so their marginal welfare loss from a tax increase is given by the increase in tax revenues holding the tax base constant and therefore $dV = -dR|_Z$. The amount of tax revenue erased by behavioural responses is the increase in the excess burden, as famously illustrated by Okun's (1975) leaky bucket metaphor.[11]

[10] This is a partial equilibrium approach that does not take into account the demand side – the "demand" for taxable income is assumed perfectly elastic at one dollar per dollar of taxable income.

[11] The marginal excess burden per dollar of revenue actually raised is $EB'(\tau)/R'(\tau) = (1-q(\tau))/q(\tau) = \alpha\varepsilon\tau/(1-\tau-\alpha\varepsilon\tau) = 57\%$ for the United States. The marginal cost of public funds (Bastani, 2023) is $1 + EB'(\tau)/R'(\tau) = 1/q(\tau) = (1-\tau)/(1-\tau-\alpha\varepsilon\tau) = 1.57$.

5 Social welfare

In this section, I show how to calculate the optimal top marginal tax rate given normative parameters and quantify the social welfare impact of tax reforms. I also introduce new graphical tools to illustrate the tradeoff between efficiency and the need for government revenue. This section serves as a generalization and extension of the expression for the socially optimal top marginal tax rate derived by Saez (2001). The only additional assumption is that the social welfare function takes a standard isoelastic functional form (Atkinson, 1970; Moulin, 2004).

The structure of the tax system is taken as given: The government is constrained to impose a constant marginal tax rate τ above a given threshold b.[12] A key theoretical construct is the marginal social welfare weight on high-income individuals, g, which is the social value of an additional dollar in the pockets of those in the top tax bracket.[13]

5.1 The social welfare function

The government places the social welfare valuation $w(u)$, which is expressed in tax dollar equivalents, on individual utility u:

$$w(u) = \begin{cases} g(100\%)\dfrac{b^{\gamma}u^{1-\gamma} - b}{1-\gamma} & \text{if } \gamma \neq 1 \\ g(100\%)\, b \ln u & \text{if } \gamma = 1, \end{cases} \tag{17}$$

where $g(100\%)$ is the marginal social welfare weight at a tax rate of 100 percent and γ is the elasticity of marginal social welfare, an indicator of inequality aversion. These normative parameters determine the level and curvature of the social welfare function, respectively. Because everyone is bunching at the kink point b if the tax rate is 100 percent, $g(100\%)$ can be interpreted as the marginal social welfare of a dollar at the kink point. Since utility is money metric, the b^{γ} factor serves to express social welfare in money equivalents, as dimensional analysis confirms.[14]

[12] A fully nonlinear tax would of course allow the government to better achieve its objectives, but is not how actual tax systems look.

[13] $g(\tau)$ here is the same as $\bar{G}(b)$ in Saez & Stantcheva (2016), i.e., *average marginal* social welfare weight for those in the top tax bracket.

[14] The utility function (equation 3) is ordinal but expressed in money-metric terms. Social welfare, on the other hand, is cardinal and also money-metric. My approach is oblivious to the "true" cardinal utility function. For example, the same results could be obtained by treating the government as utilitarian, maximizing the sum of utilities, while individuals experience declining marginal utility as determined by γ (i.e., constant relative risk aversion); cf. Fleurbaey & Maniquet (2018).

In the literature, social welfare is often calculated as a function of income. My approach is similar, but I assume that the government also takes into account individuals' disutility of earning income.

We normally expect the government to be inequality averse (the Pigou–Dalton principle), implying $\gamma > 0$ and $w''(u) < 0$. A large literature has estimated γ, mostly by analyzing the lifetime allocation of individual consumption. Acland & Greenberg (2023) conduct a meta-analysis of 158 studies and conclude that 1.6 is a reasonable estimate, with 1.2 and 2 as lower and upper bounds. Kind et al. (2017) find that values used in cost–benefit analyses tend to range from 0.5 to 2, with a central value of 1.2. The graphs in this section are drawn assuming $\gamma = 1.5$. In table 3, results are also shown for $\gamma = 0.5$.

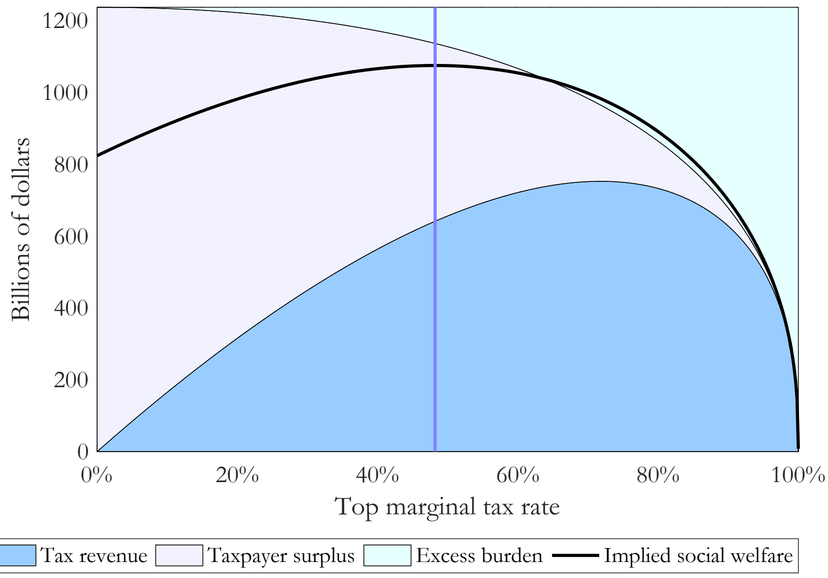
The aggregate social welfare of top bracket earnings is given by

$$G(\tau) = \int_{\frac{b}{(1-\tau)^{\varepsilon}}}^{\infty} \left[w(v(\tau; z_0)) - w(u(b; z_0))\right] f_0(z_0) dz_0, \tag{18}$$

where, as before, we disregard utility associated with earnings below the top tax bracket. Recalling that V is taxpayer surplus, social welfare is given by

$$W(\tau) = R(\tau) + G(\tau) = R(\tau) + \bar{g}(\tau)V(\tau) = \left(\tau + (1-\tau)\frac{\bar{g}(\tau)}{1+\alpha\varepsilon}\right)(1-\tau)^{\alpha\varepsilon} Z_0, \tag{19}$$

where $\bar{g}(\tau) = G(\tau)/V(\tau)$ is the *average* social welfare weight.[15] Social welfare is thus



Note: Assumes a taxable income elasticity of 0.25. The current effective marginal tax rate is indicated. Social welfare is drawn such that the current tax rate is optimal, given $\gamma = 1.5$. FICA and state taxes are included in revenue.

Figure 4: Revenue, excess burden and social welfare as a function of the top marginal tax rate, United States, 2021

[15] If $W(\tau)$ is maximized treating $\bar{g}$ as constant we recover the Saez (2001) optimal top tax rate (equation 21), underscoring that both quasilinearity and accounting for the disutility of supplying income – captured by the $(1 + \alpha\varepsilon)$ divisor (otherwise we are simply multiplying $\bar{g}$ with after-tax income) – are required to

quasilinear in tax revenue. This is reasonable if tax revenue from the top tax bracket constitutes a minor share of total tax revenue (in the United States, it is less than a tenth).

5.2 The marginal social welfare weight

The *marginal* social welfare weight is given by

$$g(\tau) = \frac{dG/d\tau}{dV/d\tau} = -\frac{1}{(1-\tau)^{\alpha\varepsilon} Z_0} \int_{\frac{b}{(1-\tau)^{\varepsilon}}}^{\infty} w'(u)v'(\tau)f_0(z_0)dz_0. \tag{20}$$

Unfortunately, no general analytical expressions for $g(\tau)$ or $G(\tau)$ can be found, but it is possible to numerically calculate $g(\tau)$ for various values of α, ε, γ and τ; see table 3.[16] As the table illustrates, $g(\tau)$ is increasing for reasonable parameter values: When taxes are higher, taxpayers are poorer and the government values a marginal increase in their consumption higher.[17]

The table shows that higher γ amplifies the variation in g (if $\gamma = 0$ then g is constant). We also see that g is higher for higher α, all else equal. This is intuitive since a high α means that top income earners have a lower average income and therefore a higher marginal

Table 3: Marginal social welfare weight (g) on top earners

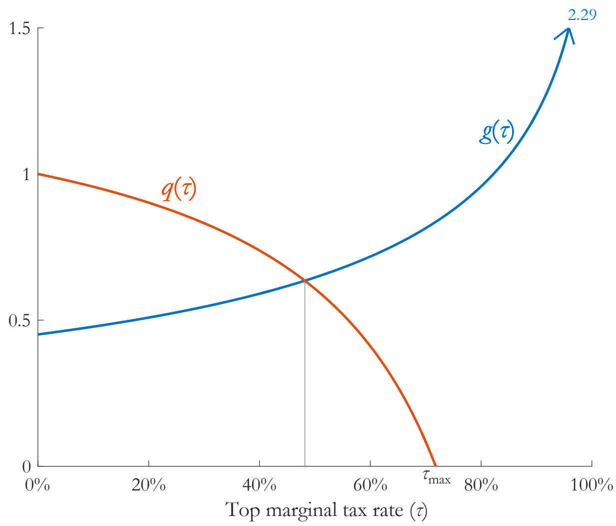
			$\tau =$										
			0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
$\varepsilon = 0.25$	$\alpha = 1.5$	$\gamma = 0.5$	0.42	0.43	0.45	0.46	0.48	0.51	0.53	0.57	0.62	0.70	1.00
		$\gamma = 1.5$	0.17	0.19	0.20	0.21	0.23	0.25	0.28	0.32	0.38	0.48	1.00
	$\alpha = 3$	$\gamma = 0.5$	0.77	0.78	0.79	0.81	0.82	0.84	0.86	0.89	0.91	0.95	1.00
		$\gamma = 1.5$	0.53	0.55	0.57	0.59	0.62	0.65	0.69	0.73	0.79	0.87	1.00
$\varepsilon = 0.5$	$\alpha = 1.5$	$\gamma = 0.5$	0.46	0.47	0.49	0.50	0.52	0.54	0.57	0.60	0.65	0.72	1.00
		$\gamma = 1.5$	0.23	0.24	0.25	0.26	0.28	0.30	0.33	0.37	0.42	0.52	1.00
	$\alpha = 3$	$\gamma = 0.5$	0.84	0.85	0.86	0.86	0.88	0.89	0.90	0.92	0.94	0.96	1.00
		$\gamma = 1.5$	0.70	0.71	0.71	0.73	0.74	0.76	0.78	0.81	0.85	0.91	1.00

Note: All values are expressed in relation to $g(100\%)$.

replicate this result.

[16]It can be shown that $g(0\%) = g(100\%)\frac{\alpha(\alpha-1)(\varepsilon+1)^{\gamma}}{(\alpha+\gamma)(\alpha+\gamma-1)}$.

[17]It is possible for $g(\tau)$ to be decreasing over some interval for high parameter values (e.g., $\alpha = 3$, $\varepsilon = 0.5$ and $\gamma = 3$). This occurs because a lower tax rate induces people with lower potential income, and therefore higher utility cost of supplying income, to enter the top tax bracket. If the elasticity – which determines the utility cost – is high enough, this effect implies that the marginal social welfare weight increases if the tax rate is reduced.



Note: q is the marginal increase in tax revenue, accounting for behavioural responses, per dollar of tax increase. g here is the path of the marginal social welfare weight that would rationalize the current tax rate, given a social welfare elasticity (γ) of 1.5. Assumes a taxable income elasticity of 0.25.

Figure 5: The marginal revenue increase (q) and implied marginal social welfare weight (g) by tax rate in the United States, 2021

social welfare. Higher ε also raises g compared to $g(100\%)$ since a higher elasticity means that the disutility of earning income is higher, and therefore utility is lower. Thus the marginal social welfare weight is an additional channel by which a higher Pareto parameter or elasticity lowers the optimal top tax rate.

The marginal social welfare weight shows how much revenue the government is willing to give up to raise taxpayer surplus by one dollar. Therefore, at optimum, it is equal to q, which shows how much tax revenue declines if taxes are cut by one dollar (which, by the envelope theorem, increases taxpayer surplus by one dollar):

$$g(\tau^*) = q(\tau^*) = 1 - \frac{\alpha\varepsilon\tau^*}{1-\tau^*} \Leftrightarrow \tau^* = \frac{1-g(\tau^*)}{1-g(\tau^*)+\alpha\varepsilon}. \tag{21}$$

As expected, we arrive at the Saez (2001) optimal top tax rate. The equation shows that $q(\tau)$ and $g(\tau)$ jointly determine the optimal tax rate τ^*.

5.3 Empirical application

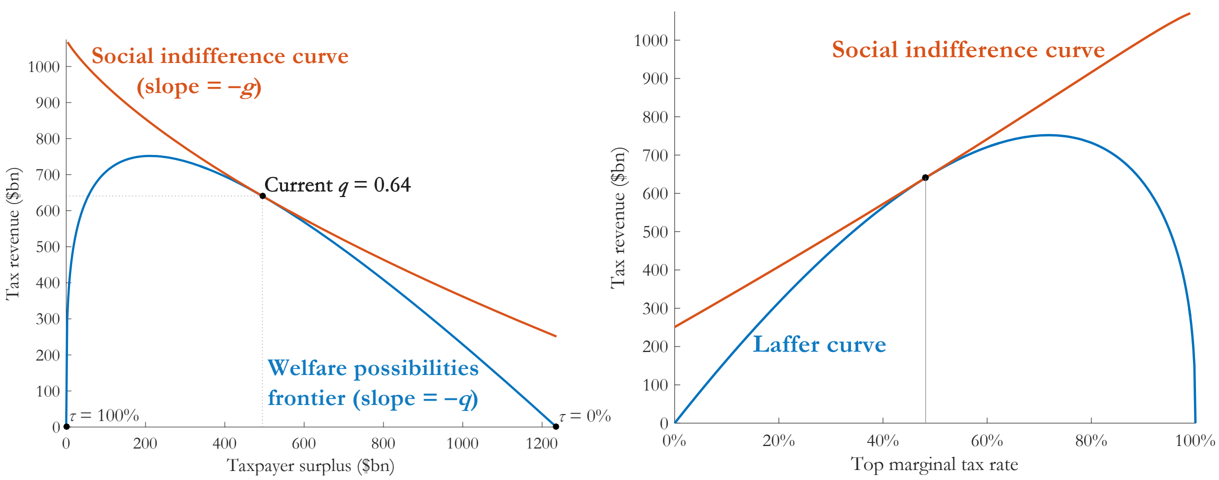
We saw in the preceding section that $q = 0.64$ in the United States if the elasticity is 0.25. Thus the current effective marginal tax rate of 48 percent is socially optimal if $g(48\%) = 0.64$. Assuming this and $\gamma = 1.5$, figure 5 shows how g varies with τ. Figure

4 depicts the implied level of social welfare by tax rate, $W(\tau)$.[18] The implied value of $g(100\%)$ is 2.29: An additional dollar in the pockets of those earning \$628,300 – the threshold for the top federal tax rate – is valued more than twice as much as a dollar of public spending. That would preclude imposing any taxes below the top tax bracket threshold, which clearly violates the government budget constraint, i.e., the current US income tax cannot be rationalized with a social welfare elasticity of 1.5, given a taxable income elasticity of 0.25.

For the reasonable case where $g(100\%)$ is lower than 1, γ must be lower than 0.33, which is a very low degree of inequality aversion. Thus we can conclude, like Berg & Piacquadio (2023), that the current US income tax is difficult to justify in the purely welfarist framework applied here. For example, policymakers may share the procedural justice sentiment that everyone *deserves* to keep a portion of their income (cf. Weinzierl, 2014). Alternatively, ε may be higher than 0.25.

5.4 The social indifference curve

Lastly, we can directly illustrate the tradeoff between tax revenue and taxpayer surplus; see figure 6a. The economically feasible combinations are given by the welfare possibilities frontier (equation 15), whose slope is $-q$. In the same diagram, we can draw draw social indifference curves which represent points with equal social welfare. The slope of these is



(a) With taxpayer surplus on the horizontal axis

(b) In Laffer space

Note: Assumes a taxable income elasticity (ε) of 0.25 and social welfare elasticity (γ) of 1.5.

Figure 6: The tradeoff between tax revenue and taxpayer surplus in the United States, 2021, with implied social indifference curves

[18]The implied current average social welfare weight ($\bar{g}$), the portion of taxpayer surplus included in social welfare, is 0.88.

$-g$ and social optimum naturally occurs at the tangency point.[19] The curvature of the indifference curves is determined by γ. If $\gamma = 0$, they are parallel straight lines.

It is also possible to draw indifference curves in Laffer space (figure 6b) – the government likes revenue but dislikes a high tax rate. The difference between the indifference curve through the optimal point and the Laffer curve is the social welfare loss of a suboptimal tax rate, i.e., how much the government would need to be compensated to deviate from social optimum.

[19] An indifference curve is given by the values of V and R for which $R + G(V) = \bar{W}$. The expression for $G(V)$ is obtained by inverting equation 14 to get $\tau(V)$, which is then plugged into equation 18. Zee (1995) draw a similar diagram for the two-person case.

6 International comparison

In this section, I apply the above expressions to eight other advanced economies. For comparability, I continue to assume a taxable income elasticity of 0.25, with a sensitivity analysis available in supplementary table 7 and figure 12. Pareto parameters are calculated using most recent revenue statistics provided by ministries of finance and tax agencies; see supplementary table 6. This is the limiting factor and explains why the analysis is confined to English-speaking countries, the Nordics and Germany. The effective marginal tax rate is computed for each country, taking into account payroll and consumption taxes; see supplementary table 5. The results are summarized in table 4 and illustrated in figures 8 and 9.

Contrary to the theoretical prediction, there is a positive correlation between the Pareto parameter and the tax rate in this sample of countries, as depicted in figure 7. At one extreme is the United States, and at the other are Denmark, Finland and Sweden, where Pareto parameters hover around 3 and effective marginal tax rates around 70 percent. In these Nordic countries, the elasticity would have to be approximately 0.15 for the current tax rate to be revenue-maximizing. Most recent studies on Nordic data find larger elasticities.[20] For the other countries the elasticity would need to be at least 0.4.

The efficiency losses of taxation are significant in all countries, ranging from 16 percent of tax revenue in the United States to about a quarter in the other Anglo-Saxon countries

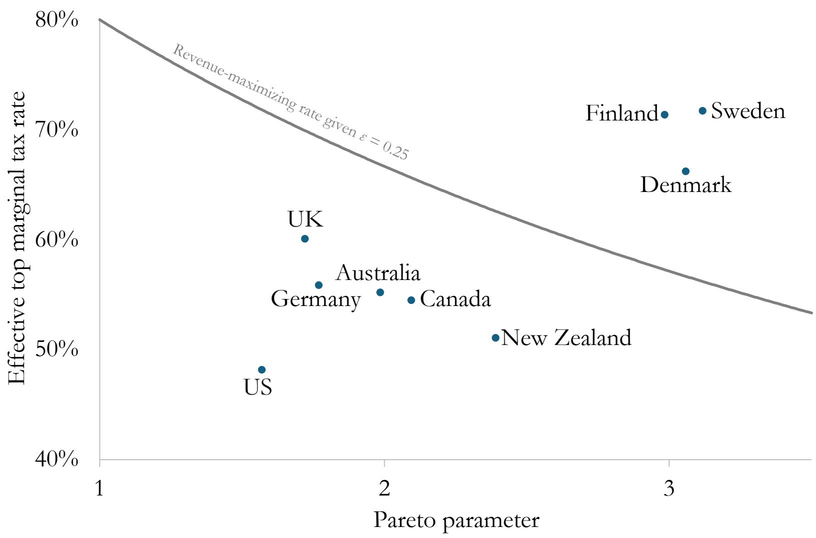


Figure 7: Country comparison of Pareto parameters and tax rates

[20]Gelber (2014), Matikka (2018), Miao et al. (2022), Kleven et al. (2023) and the difference-in-difference approach in Kleven & Schultz (2014). Smaller elasticities are estimated in the panel regression approach in Kleven & Schultz (2014), which the authors argue can be explained with optimization frictions, and by Thoresen & Vattø (2015), who look at earned, not taxable, income.

and Germany and 70–90 percent in the Nordic countries. This is even more true at the margin: Behavioural responses erase a third of additional revenue in the United States, all additional revenue and then some in the Nordic countries, and around 60 percent in the other countries.

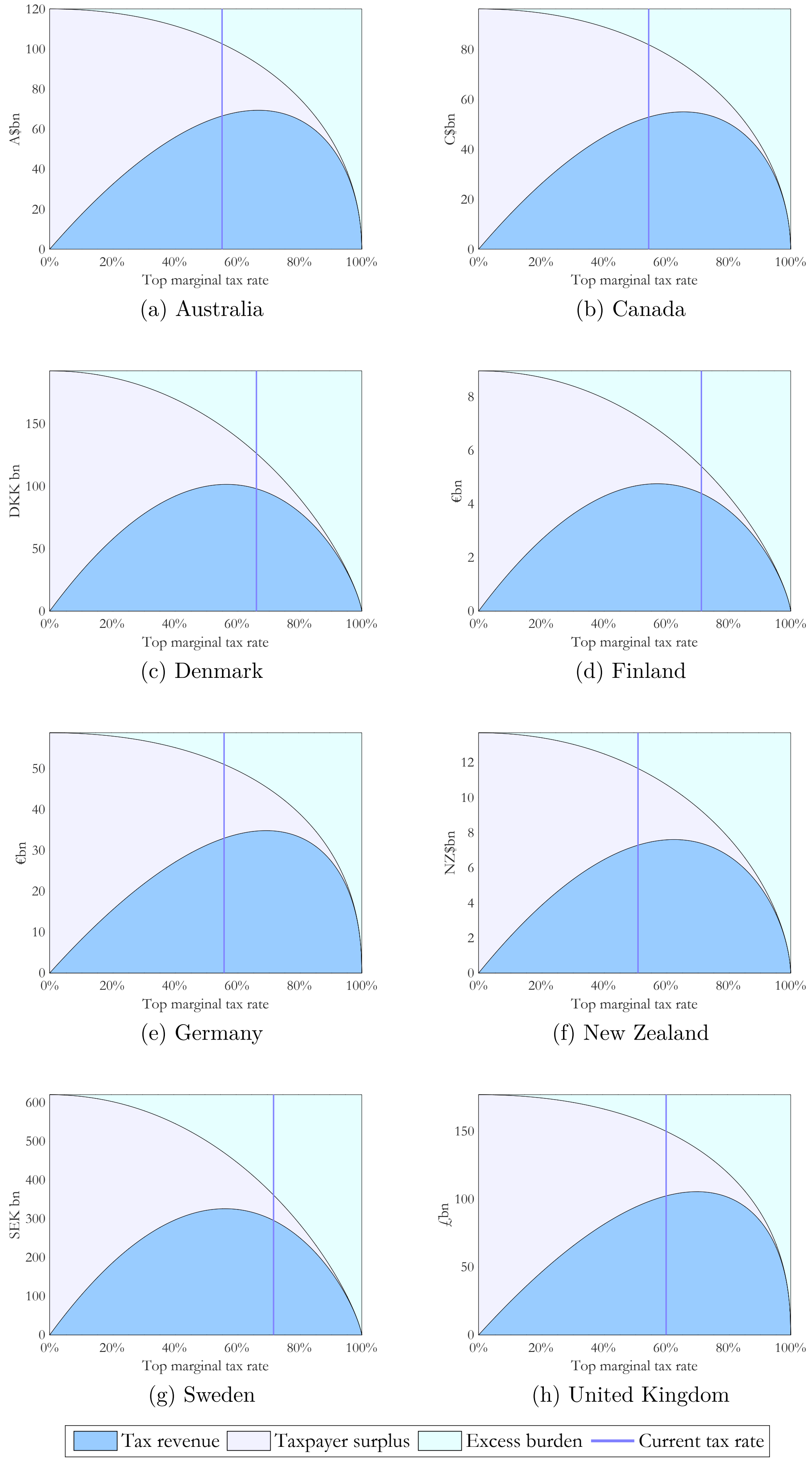
The welfarist approach to taxation employed in this paper does not fare very well as a positive theory. Under the assumptions of a taxable income elasticity of 0.25 and a social welfare elasticity of 1.5, the implied marginal social welfare weight at the threshold for the top tax bracket, $g(100\%)$, turns out to be either undefined because the country is on the wrong side of the Laffer curve (the three Nordic countries) or unreasonably high (all other countries).

Table 4: Top income taxation by country

	R	α	τ	$\tau_{\max}$	$\Delta\%R_{\max}$	AEB	q	$\varepsilon_{\max}$	$g(100\%)$
Australia	A$67bn	1.99	55%	67%	4%	26%	0.39	0.41	0.87
Canada	C$53bn	2.09	55%	66%	4%	27%	0.37	0.4	0.79
Denmark	DKK 98bn	3.06	66%	57%	**4%**	67%	–0.5	0.17	n/a
Finland	€4.4bn	2.98	71%	57%	**8%**	81%	–0.86	0.13	n/a
Germany	€33bn	1.77	56%	69%	6%	23%	0.44	0.45	1.17
New Zealand	NZ$7.3bn	2.39	51%	63%	4%	28%	0.38	0.4	0.7
Sweden	SEK 296bn	3.12	72%	56%	**10%**	88%	–0.98	0.13	n/a
UK	£102bn	1.72	60%	70%	3%	26%	0.35	0.39	0.94
US	$641bn	1.57	48%	72%	17%	16%	0.64	0.68	2.29

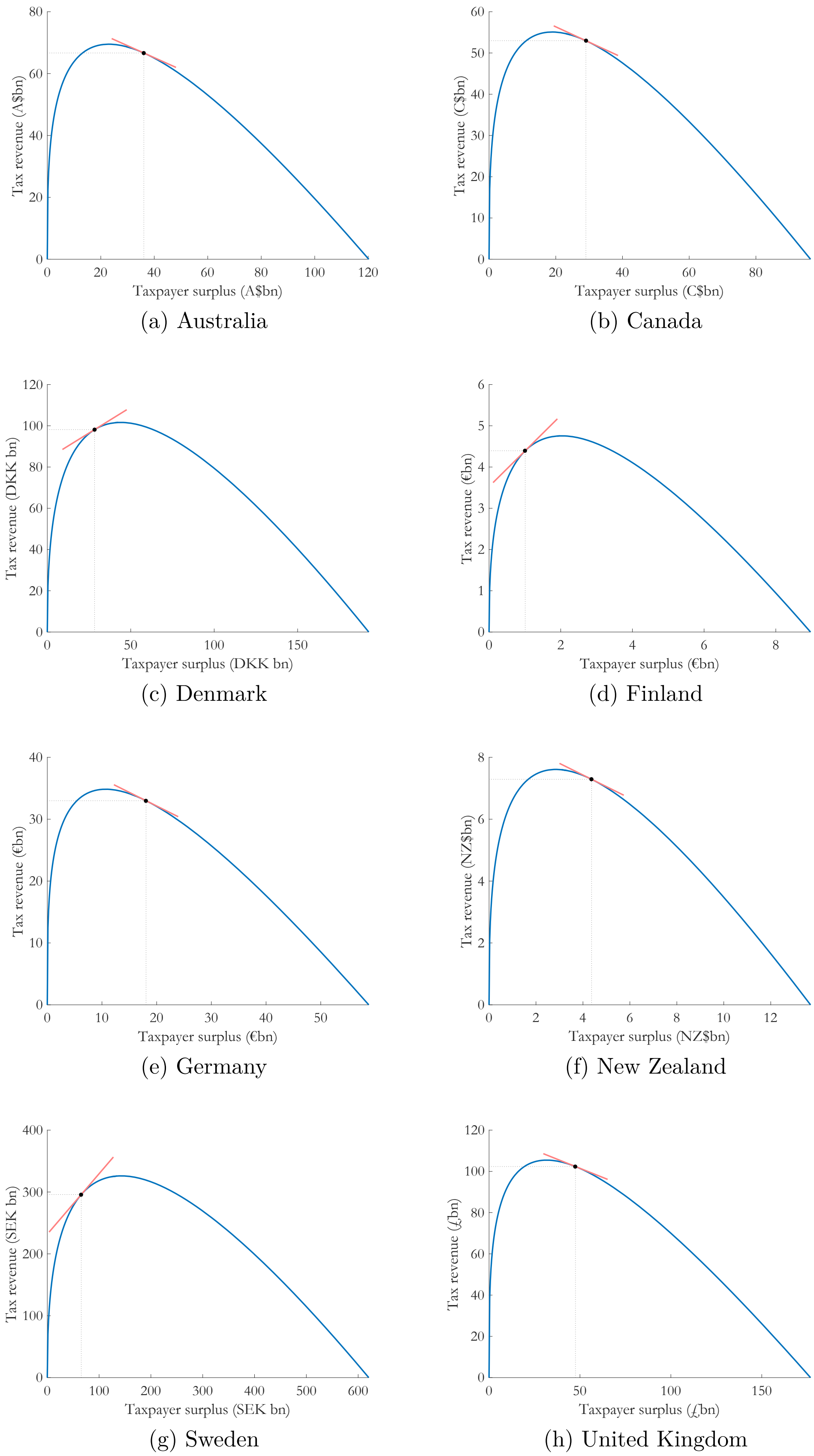
Note: Assumes a taxable income elasticity (ε) of 0.25.

Explanation of columns:
R is current total revenue (including subnational and indirect taxes) from the top tax bracket.
α is the Pareto parameter, as calculated from official revenue statistics for the top tax bracket; see supplementary table 6.
τ is the current effective top marginal tax rate; see supplementary table 5.
$\tau_{\max} = 1/(1 + \alpha\varepsilon)$ is the Saez (2001) revenue-maximizing tax rate.
$\Delta\%R_{\max}$ is how much top bracket revenues would increase if the revenue-maximizing rate were implemented (see equation 9). Countries in **bold** have revenue-maximizing rates lower than the current.
AEB is the current average excess burden of taxing top incomes, i.e., total excess burden as a percentage of total revenue (see equation 13).
$q = 1 - \alpha\varepsilon\tau/(1 - \tau)$ shows how much revenues actually go up relative to the mechanically calculated revenue gain of a marginal tax increase.
$\varepsilon_{\max} = (1 - \tau)/\alpha\tau$ is the elasticity for which the current tax rate is revenue-maximizing.
$g(100\%)$ is the marginal social welfare weight at the kink point b that would rationalize the current top tax rate, given a social welfare elasticity (γ) of 1.5. Countries with negative q cannot be analyzed in this framework and are indicated by "n/a".



Note: Assumes a taxable income elasticity of 0.25.

Figure 8: Tax revenue, taxpayer surplus and excess burden by country



(a) Australia (b) Canada (c) Denmark (d) Finland (e) Germany (f) New Zealand (g) Sweden (h) United Kingdom

Notes: Shows all combinations of tax revenue from the top bracket and taxpayer surplus in that bracket for all tax rates between 0 and 100 percent, assuming $\varepsilon = 0.25$, with the tax rate increasing from right to left. The current situation is indicated, with slope equal to $-q$.

Figure 9: The welfare possibilities frontier for top income taxation by country

7 Conclusion

This paper develops a framework for analyzing the taxation of top incomes. The empirical observation that high incomes are Pareto distributed, combined with the assumption of constant elasticity, enables the derivation of tractable mathematical expressions for not only the revenue-maximizing tax rate (Saez, 2001), but also tax revenues and the excess burden. These expressions, which I present in this paper, allow the calculation of fiscal and welfare impacts of tax reforms with minimal data requirements.

Utilizing the derived formula with an assumed taxable income elasticity of 0.25, I draw a Laffer curve for the United States. It peaks at 72 percent – 24 percentage points above the current level – potentially raising tax revenues by an additional $111 billion. In relation to top tax bracket revenue, I calculate an average excess burden of 16 percent and marginal excess burden of 57 percent.

In a comparison with eight other advanced economies, the United States is an outlier, having both the lowest marginal tax rate and lowest Pareto parameter. This implies that top income taxation is more distortionary in all other countries analyzed. The contrast with the Nordic countries is particularly stark.

From the findings in this paper it is clear that top income taxation is economically important, with implications for both government revenue and efficiency. My framework for analyzing social welfare can help illuminate the tradeoffs involved.

8 References

Acland, Daniel & Greenberg, David H. (2023), "The Elasticity of Marginal Utility of Income for Distributional Weighting and Social Discounting: A Meta-Analysis". *Journal of Benefit-Cost Analysis*, 14 (2).

Atkinson, Anthony B. (1970), "On the Measurement of Inequality". *Journal of Economic Theory*, 2 (3).

Atkinson, Anthony B., Piketty, Thomas & Saez, Emmanuel (2011), "Top Incomes in the Long Run of History". *Journal of Economic Literature*, 49 (1).

Auerbach, Alan J. (1985), "The theory of excess burden and optimal taxation". In Auerbach, Alan J. & Feldstein, Martin (eds.), *Handbook of Public Economics, Volume 1*. Amsterdam: North-Holland.

Badel, Alejandro (2013), "Higher taxes for top earners: Can they really increase revenue?". *The Regional Economist*, 21 (4).

Badel, Alejandro & Huggett, Mark (2017), "The sufficient statistic approach: Predicting the top of the Laffer curve". *Journal of Monetary Economics*, 87.

Bastani, Spencer & Selin, Håkan (2014), "Bunching and non-bunching at kink points of the Swedish tax schedule". *Journal of Public Economics*, 109.

Bastani, Spencer & Lundberg, Jacob (2017), "Political preferences for redistribution in Sweden". *Journal of Economic Inequality*, 15 (4).

Bastani, Spencer (2023), "The Marginal Cost of Public Funds: A Brief Guide". IFAU Working Paper 2023:14.

Berg, Kristoffer & Piacquadio, Paolo (2023), "Fairness and Paretian Social Welfare Functions". Unpublished manuscript.

Bergstrom, Katy & Dodds, William (2021), "Optimal Taxation with Multiple Dimensions of Heterogeneity". *Journal of Public Economics*, 200.

Bourguignon, François & Spadaro, Amedeo (2012), "Tax–benefit revealed social preferences". *Journal of Economic Inequality*, 10 (1).

Browning, Edgar K. (1987), "On the Marginal Welfare Cost of Taxation". *American Economic Review*, 77 (1).

Cesarini, David, Lindqvist, Erik, Notowidigdo, Matthew J. & Östling, Robert (2017), "The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries". *American Economic Review*, 107 (12).

Chetty, Raj (2009a), "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods". *Annual Review of Economics,* 1.

Chetty, Raj (2009b), "Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance". *American Economic Journal: Economic Policy*, 1 (2).

Chetty, Raj (2012), "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply". *Econometrica*, 80 (3).

Cirillo, Pasquale (2013), "Are your data really Pareto distributed?". *Physica A: Statistical Mechanics and its Applications*, 392 (23).

Clementi, Fabio & Gallegati, Mauro (2005), "Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States". In Chatterjee, A., Yarlagadda, S. & Chakrabarti, B.K. (eds.), *Econophysics of Wealth Distributions*. Milan: Springer.

Cowell, Frank (2011), *Measuring Inequality*. Third edition. Oxford: Oxford University Press.

Diamond, Peter (1998), "Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates". *American Economic Review*, 88 (1).

Diamond, Peter & Saez, Emmanuel (2011), "The case for a progressive tax: From basic research to policy recommendations". *Journal of Economic Perspectives*, 25 (4).

European Commission (2024), *Taxation Trends in the European Union – 2024 edition*. Directorate-General for Taxation and Customs Union.

Feldstein, Martin (1999), "Tax Avoidance and the Deadweight Loss of the Income Tax". *Review of Economics and Statistics*, 81 (4).

Fleurbaey, Marc & Maniquet, François (2018), "Optimal Income Taxation Theory and Principles of Fairness". *Journal of Economic Literature*, 56 (3).

Fritzon, Gustav & Lundberg, Jacob (2019), *Taxing high incomes: A comparison of 41 countries*. Timbro/Epicenter/Tax Foundation.

Gelber, Alexander M. (2014), "Taxation and the Earnings of Husbands and Wives: Evidence from Sweden". *Review of Economics and Statistics*, 96 (2).

Giertz, Seth H. (2009), "The elasticity of taxable income: Influences on economic efficiency and tax revenues, and implications for tax policy". In Viard, Alan D. (ed.), *Tax Policy Lessons from the 2000s*. Washington, D.C.: AEI Press.

Golosov, Mikhail, Graber, Michael, Mogstad, Magne & Novgorodsky, David (2021), "How Americans Respond to Idiosyncratic and Exogenous Changes in Household Wealth and Unearned Income". NBER Working Paper 29,000.

Golosov, Mikhail, Graber, Michael, Mogstad, Magne & Novgorodsky, David (2024), "How Americans Respond to Idiosyncratic and Exogenous Changes in Household

Wealth and Unearned Income". *Quarterly Journal of Economics*, 139 (2).

Gruber, Jonathan & Saez, Emmanuel (2002), "The Elasticity of Taxable Income: Evidence and Implications". *Journal of Public Economics*, 84.

Guner, Nezih, Lopez-Daneri, Martin & Ventura, Gustavo (2016), "Heterogeneity and government revenues: Higher taxes at the top?". *Journal of Monetary Economics*, 80.

Harberger, Arnold (1964), "Taxation, Resource Allocation, and Welfare". In *The Role of Direct and Indirect Taxes in the Federal Reserve System*. Princeton, N.J.: Princeton University Press.

Heckman, James J. (1993), "What Has Been Learned About Labor Supply in the Past Twenty Years?". *American Economic Review*, 83 (2).

Hendren, Nathaniel (2016), "The Policy Elasticity". *Tax Policy and the Economy*, 30.

Hendren, Nathaniel (2020), "Measuring economic efficiency using inverse-optimum weights". *Journal of Public Economics*, 187.

Imbens, Guido W., Rubin, Donald B. & Sacerdote, Bruce I. (2001), "Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players". *American Economic Review*, 91 (4).

Internal Revenue Service (2024), "Statistics of Income". <https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-rates-and-tax-shares>

Jacquet, Laurence & Lehmann, Etienne (2020), "Optimal income taxation with composition effects". *Journal of the European Economic Association*, 19 (2).

Joint Committe on Taxation (2015), *Estimating changes in the federal individual income tax: Description of the Individual Tax Model*. JCX-75-15, United States Congress.

Jones, Charles I. (2022), "Taxing Top Incomes in a World of Ideas". *Journal of Political Economy*, 130 (9).

Jorgenson, Dale W. & Yun, Kun-Young (1991), "The Excess Burden of Taxation in the United States". *Journal of Accounting, Auditing & Finance*, 6 (4).

Keane, Michael P. (2011), "Labor supply and taxes: A survey". *Journal of Economic Literature*, 49 (4).

Kind, Jarl, Wouter Botzen, W.J. & Aerts, Jeroen C.J.H. (2017), "Accounting for risk aversion, income distribution and social welfare in cost-benefit analysis for flood risk management". *WIREs Climate Change*, 8 (2).

Kleven, Henrik (2016), "Bunching". *Annual Review of Economics*, 8.

Kleven, Henrik (2021), “Sufficient Statistics Revisited”. *Annual Review of Economics*, 13.

Kleven, Henrik, Kreiner, Claus, Larsen, Kristian & Søgaard, Jakob (2023), “Micro vs Macro Labor Supply Elasticities: The Role of Dynamic Returns to Effort”. NBER Working Paper 31,549.

Kleven, Henrik & Schultz, Esben Anton (2014), “Estimating Taxable Income Responses Using Danish Tax Reforms”. *American Economic Journal: Economic Policy*, 6 (4).

Matikka, Tuomas (2018), “Elasticity of Taxable Income: Evidence from Changes in Municipal Income Tax Rates in Finland”. *Scandinavian Journal of Economics*, 120 (3).

Miao, Dingquan, Selin, Håkan & Söderström, Martin (2022), “Earnings responses to even higher taxes”. IFAU Working Paper 2022:12.

Moulin, Hervé (2004), *Fair Division and Collective Welfare*. Cambridge, Mass.: MIT Press.

Nair, Jayakrishnan, Wierman, Adam & Zwart, Bert (2022), *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Cambridge: Cambridge University Press.

Neisser, Carina (2021), “The Elasticity of Taxable Income: A Meta-Regression Analysis”. *The Economic Journal*, 131 (640).

Okun, Arthur M. (1975), *Equality and Efficiency: The Big Tradeoff*. Washington, D.C.: The Brookings Institution.

Pareto, Vilfredo (1896), *Cours d’économie politique*. Lausanne: F. Rouge.

Piketty, Thomas & Saez, Emmanuel (2013), ”Optimal labor income taxation”. In Auerbach, Alan J. et al. (eds.), *Handbook of Public Economics, Volume 5*. Amsterdam: Elsevier.

Piketty, Thomas, Saez, Emmanuel & Stantcheva, Stefanie (2014), “Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities”. *American Economic Journal: Economic Policy*, 6 (1).

Saez, Emmanuel (2001), “Using elasticities to derive optimal income tax rates”. *Review of Economic Studies*, 68 (1).

Saez, Emmanuel (2010), “Do Taxpayers Bunch at Kink Points?”. *American Economic Journal: Economic Policy*, 2 (3).

Saez, Emmanuel, Slemrod, Joel & Giertz, Seth H. (2012), “The elasticity of taxable income with respect to marginal tax rates: A critical review”. *Journal of Economic Literature*, 50 (1).

Saez, Emmanuel & Stantcheva, Stefanie (2016), "Generalized Social Marginal Welfare Weights for Optimal Tax Theory". *American Economic Review*, 106 (1).

Simon, Hannah & Harding, Michelle (2020), "What drives consumption tax revenues? Disentangling policy and macroeconomic drivers". OECD Taxation Working Papers 47.

Sørensen, Peter Birch (2014), "Measuring the deadweight loss from taxation in a small open economy: A general method with an application to Sweden". *Journal of Public Economics*, 117.

Thoresen, Thor O. & Vattø, Trine E. (2015), "Validation of the discrete choice labor supply model by methods of the new tax responsiveness literature". *Labour Economics*, 37.

Trabandt, Mathias & Uhlig, Harald (2011), "The Laffer curve revisited". *Journal of Monetary Economics*, 58 (4).

Usher, Dan (2014), "How High Might the Revenue-maximizing Tax Rate Be?". Queen's Economics Department Working Paper No. 1334.

Weinzierl, Matthew (2014), "The promise of positive optimal taxation: normative diversity and a role for equal sacrifice". *Journal of Public Economics*, 118.

Zee, Howell H. (1995), "Theory of Optimal Income Taxation". In Shome, Parthasarathi (ed.), *Tax Policy Handbook*. Washington, D.C.: International Monetary Fund.

A List of variables

b	Top tax bracket threshold
f	Income distribution
g	Marginal social welfare weight
q	Marginal revenue increase
u	Utility
v	Indirect utility
w	Social valuation of individual utility
z	Taxable income
G	Social valuation of taxpayer surplus
N	Number of top bracket taxpayers
R	Tax revenue
T	Tax function
V	Taxpayer surplus
W	Social welfare
Z	Tax base
α	Pareto parameter
γ	Social welfare curvature
ε	Taxable income elasticity
τ	Top marginal tax rate

B Supplementary figures and tables

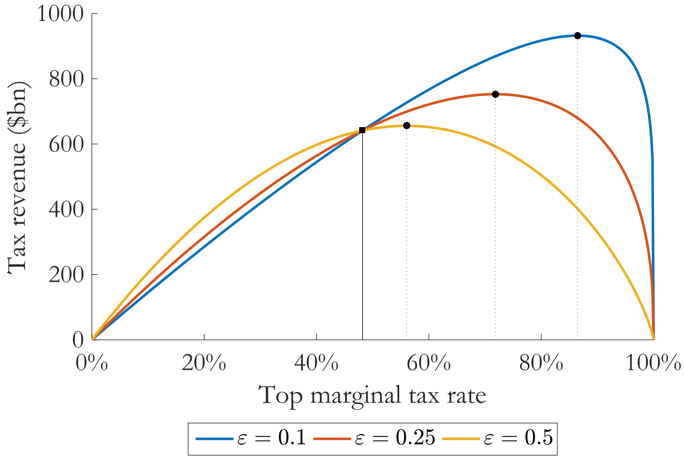
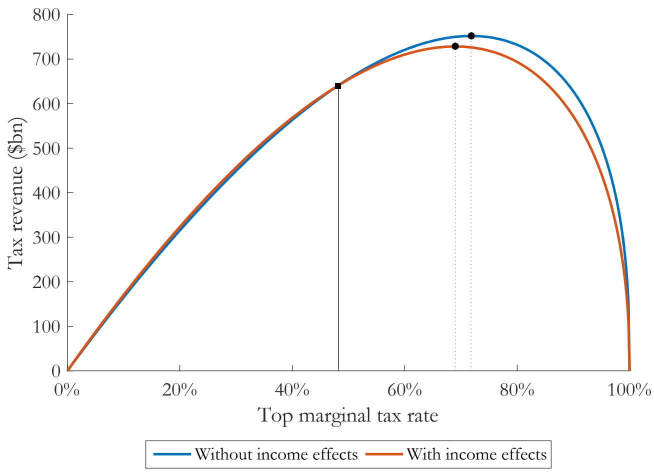


Figure 10: The Laffer curve for top incomes in the United States, 2021, depending on elasticity



Note: Refers to the top federal income tax bracket, with state revenues included. The uncompensated taxable income elasticity is held constant at 0.25, meaning that introducing income effects – here with an income effect parameter (η) of –0.1 – shifts the Laffer curve peak to the left; see the discussion in section 3.2.

Figure 11: The Laffer curve for top incomes in the United States, 2021, with and without income effects

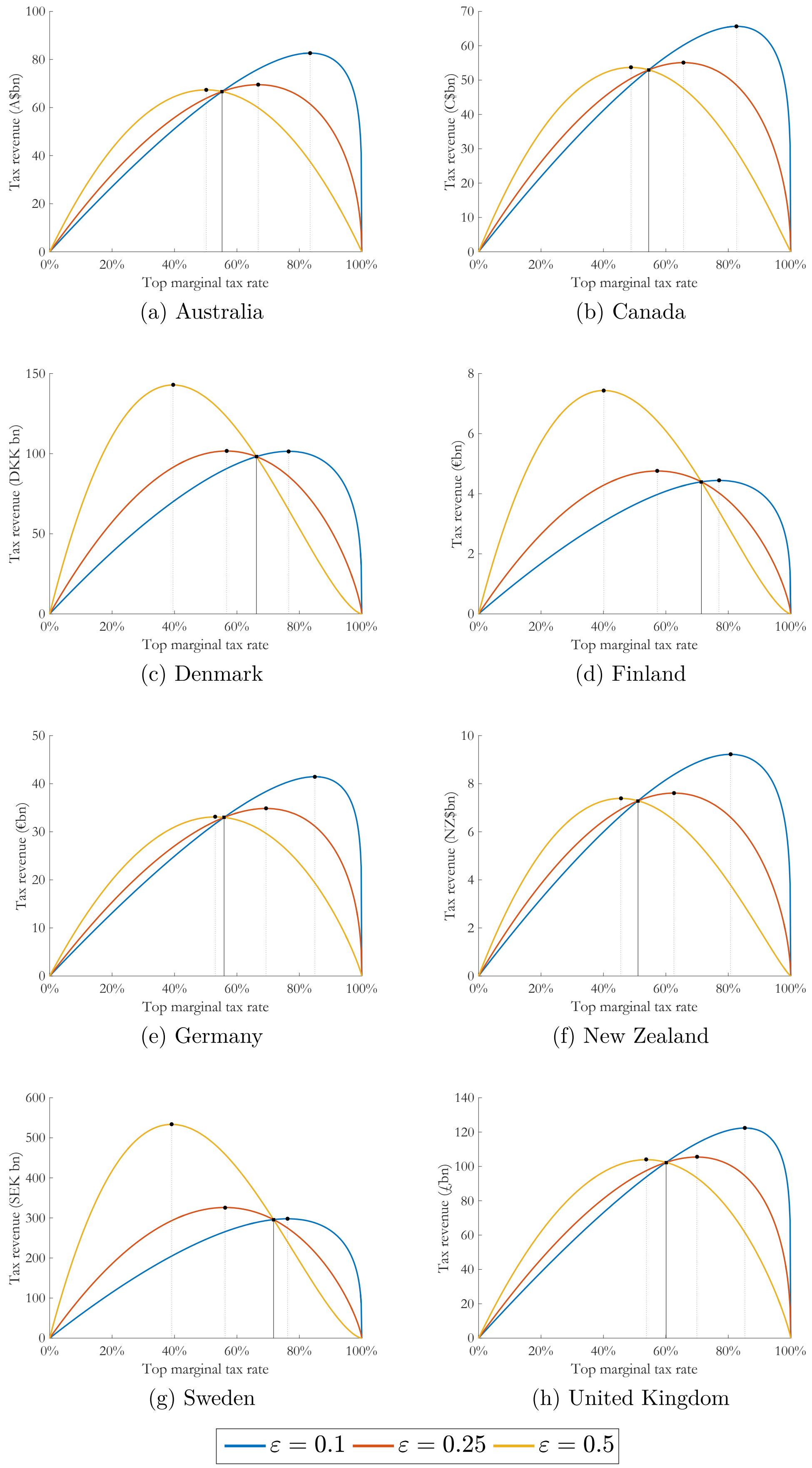


Figure 12: The Laffer curve for top incomes by country, depending on elasticity

Table 5: Effective top marginal tax rate by country

Country	**National income tax (τ_n)**	**Other direct taxes (τ_s)**	**Payroll tax (τ_p)**	**Consumption taxes (τ_c)***	**Effective marginal tax rate (τ)****
Australia (2020/21)	45%	2%	5.5%	11%	55%
Canada (2022)	33%	18.13%		7%	55%
Denmark (2022)	15%	40.93%		23%	66%
Finland (2022)	31.25%	25.05%	20.79%	21%	71%
Germany (2019)	45%	2.48%		16%	56%
New Zealand (2022)	39%			20%	51%
Sweden (2024)	20%	32.37%	31.42%	22%	72%
UK (2023/24)	45%	2%	13.8%	14%	60%
US (2021)	37%	7.44%	1.45%	5%	48%

* Implicit tax-inclusive tax rates on consumption calculated from aggregate data. For EU countries, the data is for 2022 and from the European Commission (2024). For other countries, data is for 2017 and from the OECD (Simon & Harding, 2020). Implicit consumption tax rates are generally stable over time.

** Calculated as $\tau = \dfrac{\tau_n + \tau_s + \tau_c(1 - \tau_n - \tau_s) + \tau_p}{1 + \tau_p}$.

Note: Shows the marginal tax rates that apply to the top tax bracket, using a similar methodology as Fritzon & Lundberg (2019).

Source: Government and other country-specific sources (see country notes)

Country notes:
Australia: The Medicare levy is 2%. Payroll taxes are levied at the state level; shown here is the simple average for large employers.
Canada: τ_s is the simple average of top provincial income tax rates, taking into account surtaxes if applicable.
Denmark: The national income tax refers to the national "top tax". The national "bottom tax" was 12.1% in 2022 and the average municipal tax rate was 25% according to Statistics Denmark. Together with the deductible 8% labour market contribution this implies $\tau_s = 12.1\% + 25\% + 8\% \times (1 - 15\% - 12.1\% - 25\%) = 40.93\%$.
Finland: The average municipal tax rate was 20% and the sum of deductible employee contributions for healthcare, sickleave, unemployment insurance and pensions (aged 53 or younger) was $0.53\% + 1.18\% + 1.5\% + 7.15\% = 10.36\%$ in 2022, implying $\tau_s = 20\% + 10.36\% \times (1 - 20\% - 31.25\%) = 25.05\%$. For employers, the contribution rates were 1.34% for healthcare, 2.05% for unemployment insurance and 17.4% for pensions, for a total of 20.79% (tax rates from the Finnish Ministry of Social Affairs and Health).
Germany: τ_s refers to the solidarity tax, which is 5.5% of the tax liability.
New Zealand: No social security contributions or local income taxes.
Sweden: τ_s is the average municipal tax rate according to Statistics Sweden. The EITC phaseout for higher incomes is not accounted for.
United Kingdom: The additional rate is slightly lower on dividends and slightly higher in Scotland, which is ignored for simplicity.
United States: The Social Security tax is capped and does not affect the top marginal tax rate. The Medicare tax is 1.45 percent on the employer and 2.35 percent on the employee. The simple average top state income tax rate, adjusting for deductibility of federal taxes when applicable, is 5.09 percent, using 2021 data from the NBER TAXSIM model. Thus $\tau_s = 5.09\% + 2.35\% = 7.44\%$.

Table 6: Pareto parameter by country

Country	**Tax rate** (τ_n)	**Top bracket threshold** (b)	**Number of returns** (N)	**Revenue at rate** (R_n)	**Pareto parameter*** (α)
Australia (2020/21)	45%	A$180,000	626,392	A$51bn	1.99
Canada (2022)	33%	C$221,708	479,750	C$32bn	2.09
Denmark (2022)	15%	DKK 552,500	552,000	DKK 22bn	3.06
Finland (2022)	31.25%	€82,900	122,028	€1.6bn	2.98
Germany (2019)	45%	€265,327	171,251	€27bn	1.77
New Zealand (2022)	39%	NZ$180,000	110,240	NZ$5.6bn	2.39
Sweden (2024)	20%	SEK 615,300	1,080,000	SEK 63bn	3.12
UK (2023/24)	45%	£125,140	862,000	£67bn	1.72
US (2021)	37%	$628,301	1,189,107	$485bn	1.57

* Calculated as $\alpha = 1 + \tau_n bN/R_n$.

Note: Considers revenue from the national income tax only. Some countries report total income tax paid by top bracket taxpayers, in which case the tax liability at the top bracket threshold, $T(b)$, multiplied by N, is deducted in order to calculate R_n.

Sources:
Australia: Australian Taxation Office, *Taxation Statistics 2020–21*, table 5. Calculated using $T(b)$ = A$51,667.
Canada: Canada Revenue Agency, "Individual Tax Statistics by Tax Bracket", tables 1 and 3. Calculated using $T(b)$ = C$51,343.
Denmark: Statistics Denmark. Revenues have been adjusted for the marginal tax rate ceiling (*skatteloft*). I am grateful to Cepos for assistance with this.
Finland: Finnish Tax Administration. Calculated using $T(b)$ = €11,350.
Germany: Federal Ministry of Finance, *Datensammlung zur Steuerpolitik 2024*, table 2.7.8. Married joint filers are counted separately in N, so b refers to the threshold for single filers. Calculated using $T(b)$ = 45% × €265,327 – €16,741 = €102,657.
New Zealand: Inland Revenue, "Taxable income distribution of individuals". Calculated using $T(b)$ = NZ$50,320.
Sweden: Estimates by the Swedish Ministry of Finance.
United Kingdom: HMRC, "Income Tax statistics", tables 2.5 and 2.6.
United States: IRS, *Individual Income Tax Returns Complete Report 2021,* table 3.6. The threshold is for married joint filers, who constititute the vast majority in this bracket.

Table 7: Top income taxation by country, sensitivity analysis

		$\varepsilon = 0.1$				$\varepsilon = 0.5$			
Country	τ	$\tau_{\max}$	$\Delta\% R_{\max}$	AEB	q	$\tau_{\max}$	$\Delta\% R_{\max}$	AEB	q
Australia	55%	83%	24%	10%	0.75	50%	**1%**	61%	–0.23
Canada	55%	83%	24%	10%	0.75	49%	**1%**	64%	–0.25
Denmark	66%	77%	3%	22%	0.40	40%	**46%**	194%	–2.00
Finland	71%	77%	1%	26%	0.26	40%	**69%**	248%	–2.72
Germany	56%	85%	26%	9%	0.78	53%	**0%**	54%	–0.12
New Zealand	51%	81%	26%	10%	0.75	46%	**1%**	66%	–0.25
Sweden	72%	76%	1%	28%	0.21	39%	**80%**	275%	–2.96
UK	60%	85%	20%	10%	0.74	54%	**2%**	61%	–0.30
US	48%	86%	45%	6%	0.85	56%	2%	35%	0.27

Note: Refer to table 4 for an explanation of columns.

Table 8: Historical Pareto parameter for the US top federal tax bracket

Year	Tax rate (τ_n)	Top bracket threshold* (b)	Number of returns (N)	Revenue at rate (R_n)	Pareto parameter ($\alpha = 1 + \tau_n bN/R_n$)
1996	39.6%	$263,750	624,601	$95bn	1.69
1997	39.6%	$271,050	691,359	$115bn	1.64
1998	39.6%	$278,450	753,426	$133bn	1.63
1999	39.6%	$283,150	864,306	$164bn	1.59
2000	39.6%	$288,350	921,396	$194bn	1.54
2001	39.1%	$297,350	846,345	$162bn	1.61
2002	38.6%	$307,050	766,125	$141bn	1.64
2003	35%	$311,950	752,028	$129bn	1.64
2004	35%	$319,100	842,749	$158bn	1.59
2005	35%	$326,450	953,005	$198bn	1.55
2006	35%	$336,550	1,002,051	$218bn	1.54
2007	35%	$349,700	1,060,714	$240bn	1.54
2008	35%	$357,700	971,591	$218bn	1.56
2009	35%	$372,950	790,063	$170bn	1.61
2010	35%	$373,650	854,212	$189bn	1.59
2011	35%	$379,150	922,346	$194bn	1.63
2012	35%	$388,350	1,034,695	$243bn	1.58
2013	39.6%	$450,000	892,420	$243bn	1.66
2014	39.6%	$457,601	978,900	$279bn	1.64
2015	39.6%	$464,851	1,026,445	$296bn	1.64
2016	39.6%	$466,950	1,017,009	$278bn	1.68
2017	39.6%	$470,700	1,109,602	$323bn	1.64
2018	37%	$600,000	841,679	$288bn	1.65
2019	37%	$612,350	871,748	$285bn	1.69
2020	37%	$622,050	922,362	$341bn	1.62
2021	37%	$628,301	1,189,107	$485bn	1.57

Source: IRS, *Individual Income Tax Returns Complete Report*, table 3.6, various years

Note: Refers to the federal income tax only. The calculations in the main text also consider Medicare and state taxes.

* Married filing jointly. The threshold for single filers and heads of households has been slightly lower since 2013. The threshold for married persons filing separately is half as high as for joint filers, but this group accounted for only 4 percent of top bracket income in 2021.

C Marginal tax changes

Here I derive an expression for the for the ratio between the revenue increase of a small tax hike with and without behavioural responses, which I term q. The purpose is to show that this expression can be derived without imposing structural assumptions on individual utility and the income distribution. In addition, I use my expression for q to derive an expression for the Laffer curve with income effects (see section 3).

We assume that individuals are heterogenous in earnings capacity, z_0, only. Income increases monotonically with earnings capacity and all individuals with the same earnings capacity have the same income. The tax function is the same as in the main text. Individuals locate in the first segment of the tax schedule, on the kink (bunching) or in the second segment. For those in the second segment – those whose incomes strictly exceed b – virtual income is given by $y(\tau) = \tau z - T(z) = \tau b$ so that the budget constraint is $c = (1-\tau)z + y$, where c is consumption. Introducing virtual income is a method of linearizing a piecewise linear budget constraint. The taxable income supply function for this group is denoted $z(1-\tau, y(\tau); z_0)$.

We are interested in how the individual's optimal taxable income will change when the tax rate is increased, considering both income and substitution effects. When the tax rate changes, virtual income will also change. Because virtual income is the intercept of the linearized budget constraint, changing the tax rate will shift the budget constraint and this will induce income effects. For a taxpayer in the top tax bracket, $\partial y/\partial\tau = b$. Applying the definitions of the uncompensated taxable income elasticity ($\varepsilon_u = \partial z/\partial(1-\tau) \times (1-\tau)/z \mid_y$) and the income effect parameter ($\eta = (1-\tau)\partial z/\partial y$), the change in taxable income brought about by a tax reform can be expressed

$$\frac{dz(1-\tau, y(\tau))}{d\tau} = -\frac{\partial z}{\partial(1-\tau)} \mid_y + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \tau} = -\frac{\varepsilon_u z - \eta b}{1-\tau} = -\frac{\varepsilon_c z + \eta(z-b)}{1-\tau}, \qquad (22)$$

where the elasticity version of the Slutsky equation ($\varepsilon_u = \varepsilon_c + \eta$) is used in the last step.

Turning to considering aggregate effects, the density function of the earnings capacity distribution is denoted $f_0(z_0)$. The proportion of taxpayers in the top tax bracket (those who earn more than b) is given by

$$N(\tau) = \int_{b_0(\tau)}^{\infty} f_0(z_0)dz_0, \qquad (23)$$

where $b_0(\tau)$ is such that $z(1-\tau, y(\tau); b_0(\tau)) = b$, i.e., earnings capacity for the taxpayer who is at the margin of entering the top tax bracket. The density function is normalized so that $N(0) = \int_{b_0(0)}^{\infty} f_0(z_0)dz_0 = 1$, i.e., the population of potential (if the tax rate were zero) top-bracket taxpayers is one.

The average income of top-bracket taxpayers is

$$\bar{z_b}(\tau) = \int_{b_0(\tau)}^{\infty} z(1-\tau, y(\tau); z_0) f_0(z_0)dz_0/N(\tau). \tag{24}$$

Integrating over taxpayers, tax revenues from the top tax bracket can be expressed

$$R(\tau) = \tau \int_{b_0(\tau)}^{\infty} \left[z(1-\tau, y(\tau); z_0) - b\right] f_0(z_0)dz_0 = \tau \left[\bar{z_b}(\tau) - b\right] N(\tau). \tag{25}$$

Using the Leibniz rule for the derivative of integrals, the revenue impact of a small tax increase is

$$\frac{dR}{d\tau} = \underbrace{\left[\bar{z_b} - b\right] N}_{\text{Mechanical effect}} + \tau \left[\underbrace{\int_{b_0(\tau)}^{\infty} \frac{dz(1-\tau, y(\tau); z_0)}{d\tau} f_0(z_0)dz_0}_{\text{Effect of average income}} - \underbrace{(z(1-\tau, y(\tau); b_0) - b)\frac{db_0(\tau)}{d\tau}}_{\text{Taxpayer number effect } (=0)} \right].$$

The third term is equal to zero because $z(1-\tau, y; b_0) = b$ by definition. Thus N is of second-order importance for revenues and can be regarded as constant for small changes in τ (see also Saez et al., 2012, footnote 7).

Next, we define the income-weighted average compensated taxable income elasticity to be

$$\bar{\varepsilon_c}(\tau) = \frac{\int_{b_0}^{\infty} \varepsilon_c(\tau; z_0) z(1-\tau, y; z_0) f_0(z_0)dz_0}{\int_{b_0}^{\infty} z(1-\tau, y; z_0) f_0(z_0)dz_0} = \frac{\int_{b_0}^{\infty} \varepsilon_c(\tau; z_0) z(1-\tau, y; z_0) f_0(z_0)dz_0}{\bar{z_b}N}.$$

The tax-base-weighted average income effect parameter is

$$\tilde{\eta}(\tau) = \frac{\int_{b_0}^{\infty} \eta(\tau; z_0)[z(1-\tau, y; z_0) - b] f_0(z_0)dz_0}{\int_{b_0}^{\infty}[z(1-\tau, y; z_0) - b] f_0(z_0)dz_0} = \frac{\int_{b_0}^{\infty} \eta(\tau; z_0)[z(1-\tau, y; z_0) - b] f_0(z_0)dz_0}{(\bar{z_b} - b)N}.$$

The existence of these averages is guaranteed by the second mean value theorem of integrals. To explain the different weightings, we note that the substitution effect is proportional to all of taxable income (because the elasticity is defined in terms of the proportional change in taxable income) while the income effect depends only on the part that is in the top tax bracket (the tax base), because this determines how disposable income will change.[21]

Applying these definitions and the Slutsky equation (equation 22), we are now ready to derive our parameter of interest, q, the increase in revenue divided by the mechanically

[21] In contrast, Saez (2001, p. 210) considers an income-weighted *uncompensated* elasticity and an *unweighted* income effect parameter.

calculated revenue increase:

$$q(\tau) = \frac{\dfrac{dR}{d\tau}}{\left.\dfrac{dR}{d\tau}\right|_Z} = 1 + \frac{\tau \int_{b_0}^{\infty} \dfrac{dz(1-\tau, y(\tau); z_0)}{d\tau} f_0(z_0) dz_0}{(\bar{z}_b - b)N}$$

$$= 1 - \frac{\tau \int_{b_0}^{\infty} \dfrac{\varepsilon_c(\tau; z_0) z(1-\tau, y(\tau); z_0) + \eta(\tau; z_0)[z(1-\tau, y(\tau); z_0) - b]}{1-\tau} f_0(z_0) dz_0}{(\bar{z}_b - b)N} =$$

$$= 1 - \frac{\tau \left[\bar{\varepsilon_c} \bar{z}_b N + \tilde{\eta}(\bar{z}_b - b)N\right]}{(1-\tau)(\bar{z}_b - b)N} = 1 - \frac{\tau \left[\alpha \bar{\varepsilon_c} + \tilde{\eta}\right]}{1-\tau}. \quad (26)$$

This expression is well known in the literature (e.g., Saez et al., 2012).

Intuitively, q is decreasing in the Pareto parameter because that means a smaller fraction of average income is in the top tax bracket. It is decreasing in the taxable income elasticity because this implies larger behavioural responses. It is decreasing in the current tax rate partly because the revenue impact is larger if the tax rate is larger (τ in the numerator) and partly because a higher tax rate means the net-of-tax rate will be affected more proportionately by a given tax change (τ in the denominator). Income effects are less important for high-income taxation, as the compensated response is amplified by the Pareto parameter $\alpha > 1$ while the income effect is not. The intuition is that a tax cut for the top tax bracket may increase the incentive to earn taxable income at the margin considerably while net income does not increase much, implying that demand for leisure will not increase much either.

Note that α, $\bar{\varepsilon_c}$ and $\tilde{\eta}$ in general are functions of τ. Equation 26 is thus valid for the evaluation of small tax changes, where no functional form assumptions are needed for the income distribution or utility function. When evaluating non-marginal tax changes, stronger assumptions are needed.

D Connection with Saez (2001)

In this appendix I show how my derivations in sections 2–4 above, and the assumptions behind them, are connected to Saez (2001).

Proposition 3. *The expressions for tax revenue and total excess burden in the present paper serve as a generalization of the Saez (2001) optimal top marginal tax formula without additional assumptions.*

Proof. Saez (2001) derives the top marginal tax rate by noting that raising taxes by one dollar increases tax revenues by $q(\tau) = 1 - \alpha\varepsilon\tau/(1-\tau)$ dollars after accounting for behavioural responses (see also appendix C). This holds regardless of utility function and income distribution. At social optimum, q is equal to the the marginal social welfare weight on top earners, g. The optimal tax rate is then given by the Saez (2001) formula $\tau^* = (1-g)/(1-g+\alpha(\tau^*)\varepsilon(\tau^*))$. Note that any tax rate between 0 and 100 percent can be rationalized as optimal by setting $g \in (-\infty, 1]$. As the term $\alpha(\tau^*)\varepsilon(\tau^*)$ in general is endogenous, there could be several critical points. In order to rule this out and to be able to extrapolate from current parameter values, we require that it is independent of τ. In section 3 I use a differential equation approach to show that this assumption allows the derivation of my expression for the Laffer curve (equation 7) and the tax base. From the expression for the tax base it is possible to derive my expression for the excess burden (equation 12), since it is the sum of the differences between the current tax base $Z(\tau)$ and the tax base at each tax rate lower than τ: $EB(\tau) = \int_0^\tau (Z(\hat{\tau}) - Z(\tau))\, d\hat{\tau} = \int_0^\tau ((1-\hat{\tau})^{\alpha\varepsilon} - (1-\tau)^{\alpha\varepsilon})\, d\hat{\tau} Z_0$, as illustrated in figure 3. □

Under what conditions is $\alpha\varepsilon$ independent of τ? While it is theoretically possible for them to vary with τ in some complicated way that makes their product constant, Occam's razor dictates that we primarily consider the case where α and ε (both in aggregate and for any given individual) are constant across tax rates. I will now show that this entails that all individuals must have the same elasticity, which implies my assumption 2, and that incomes are Pareto distributed, which implies my assumption 3. Finally, both my derivation (assumption 1) and the Saez top tax formula require that the budget set is convex in order to rule out discontinuous jumping (Saez, 2001, footnote 7; Bergstrom & Dodds, 2021).

Lemma 2. *A constant elasticity for a given individual implies an isoelastic utility function.*

Proof. The taxable income elasticity is given by $\varepsilon = [dz/z] \,/\, [d(1-\tau)/(1-\tau)] = d\ln z / d\ln(1-\tau)$. Holding the elasticity constant, this is a differential equation with the solution

$z = z_0(1-\tau)^\varepsilon$. (If $z_0(1-\tau)^\varepsilon \leq b$, the individual will not earn income in the top tax bracket.) By the utility maximization theorem, the only utility function (up to a positive monotonic transformation) that represents such preferences is equation 3 on page 5, i.e., an isoelastic utility function. □

To say that individuals must have the same elasticity is the same as saying that individuals must be heterogenous in one dimension only (potential income). In his derivation of the optimal tax schedule, Saez (2001) only considered unidimensional heterogeneity. When adding additional dimensions of heterogeneity, the optimal marginal tax rate at particular income level z depends on the average elasticity of individuals at that income level (Jacquet & Lehman, 2020; Bergstrom & Dodds, 2021). Intuitively, this creates a circularity that implies that the average elasticity in general will depend on the tax rate. I show this formally below.

Proposition 4. *For the aggregate elasticity to be independent of the tax rate, all potential top tax bracket individuals must have the same elasticity, given that the elasticity is constant across tax rates for any given individual.*

Proof. Assume that two individuals with potential incomes z_0^1 and z_0^2 have different elasticities, where $\varepsilon^1 > \varepsilon^2 > 0$. For each individual, the elasticity is independent of the tax rate, so by lemma 2, the utility function is isoelastic. Both are potential top bracket taxpayers, i.e., $z_0 > b$. They will bunch at the kink if $z_0 < b/(1-\tau)^\varepsilon \Leftrightarrow b > z_0(1-\tau)^\varepsilon$ and earn income in the top bracket otherwise. Hence if $\tau = 0$ both are top bracket taxpayers. As $\tau \to 1$, $b/(1-\tau)^\varepsilon \to \infty$ and neither is a top bracket taxpayer.

We consider the three possible cases $z_0^1 = z_0^2$, $z_0^1 > z_0^2$ and $z_0^1 < z_0^2$. In the first case, there exists a τ such that $b/(1-\tau)^{\varepsilon^2} < z_0^2 = z_0^1 < b/(1-\tau)^{\varepsilon^1}$.

For the second case, note that $(1-\tau)^\varepsilon$ is continuous and decreasing in ε. By the intermediate value theorem and because $\varepsilon^1 > \varepsilon^2$, there exists a τ such that $\hat{z} := z_0^2(1-\tau)^{\varepsilon^2} = z_0^1(1-\tau)^{\varepsilon^1}$. First assume $\hat{z} < b$ (both are bunching). Then, there exists a $\delta > 0$ such that $z_0^2(1-\tau+\delta)^{\varepsilon^2} < b < z_0^1(1-\tau+\delta)^{\varepsilon^1}$. I.e., a small tax cut will induce the more elastic individual to enter the top tax bracket, but not the less elastic individual. If, on the other hand, $\hat{z} > b$, there exists a $\delta > 0$ such that $z_0^1(1-\tau-\delta)^{\varepsilon^1} < b < z_0^2(1-\tau-\delta)^{\varepsilon^2}$.

For the third case, as $z_0^1 < z_0^2$ and $(1-\tau)^{\varepsilon^1} < (1-\tau)^{\varepsilon^2}$, we have $z_0^1(1-\tau)^{\varepsilon^1} < z_0^2(1-\tau)^{\varepsilon^2}$ for all τ. Trivially there exists a τ such that $z_0^1(1-\tau)^{\varepsilon^1} < b < z_0^2(1-\tau)^{\varepsilon^2}$.

We have shown for all three possible cases that there exists a tax rate where one individual bunches and the other does not. Since no one bunches if the tax rate is 100 percent and both bunch if the tax rate is zero, which individuals are in the top tax bracket is determined by the tax rate. Further, the average elasticity is the effect of a small tax change on average taxable income in the top tax bracket. Only those in the top bracket, not the bunchers, determine this elasticity. Thus the average elasticity depends on the tax rate if any two individuals have different elasticities. □

Lastly, I show that the assumption of constant Pareto parameter implies a Pareto distribution, a fact also noted by Saez (2001, p. 212).

Proposition 5. *For the Pareto parameter to be independent of the tax rate, high potential incomes must be Pareto distributed, given that the elasticity is constant.*

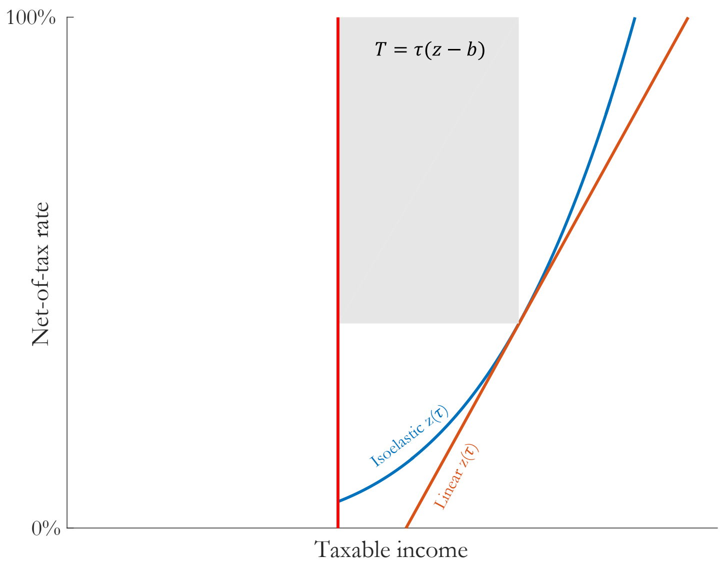
Proof. The Pareto parameter for realized incomes is is defined as $\alpha = \bar{z}_b/(\bar{z}_b - b)$, where $\bar{z}_b$ is average income in the top tax bracket. By lemma 2, individuals earn income in the top tax bracket if their potential income exceeds $b/(1-\tau)^\varepsilon$. Expressing the Pareto parameter in terms of potential income, we have $\alpha = \bar{z}_b/(1-\tau)^\varepsilon/(\bar{z}_b/(1-\tau)^\varepsilon - b/(1-\tau)^\varepsilon)$, which must be constant for any $\tau \in [0, 1)$. That is, the ratio between any arbitrary potential income threshold (between b and infinity) to average potential income over that threshold must be constant. By van der Wijk's law, the Pareto distribution is the only probability distribution that has this property (Cowell, 2011; Cirillo, 2013). □

E Comparison with Harberger (1964) and linear labour supply

In this paper, I consistently use an isoelastic utility function, implying that the taxable income supply function is linear in logarithms. A common assumption in the older literature is a linear labour supply function, which in the modern framework corresponds to a linear taxable income supply function. It is therefore interesting to compare the expressions above to the corresponding expressions under a linearity assumption – in particular, the standard Harberger (1964) excess burden formula.

Let us consider the average top earner and linearize the taxable income supply function from the current income $\bar{z}_b$, current tax rate τ_c and current elasticity ε, so that $\varepsilon = -z'(\tau)(1-\tau_c)/\bar{z}_b \Rightarrow z'(\tau) = -\varepsilon\bar{z}_b/(1-\tau) \Rightarrow z(\tau) = \bar{z}_b\,[1-(\tau-\tau_c)\varepsilon/(1-\tau_c)]$. This is illustrated in figure 13, along with the isoelastic case $z(\tau) = [(1-\tau)/(1-\tau_c)]^{\varepsilon}\,\bar{z}_b$.

For the linear case, income in the case of no taxation is $z_0 = \bar{z}_b\,[1+\varepsilon\tau_c/(1-\tau_c)]$. The excess burden is the area of a triangle with the sides τ_c and $\bar{z}_b\varepsilon\tau_c/(1-\tau_c)$, i.e., $\bar{z}_b\varepsilon\tau_c^2/2(1-\tau_c)$. This is the Harberger formula as corrected by Browning (1987); see also Feldstein (1999). When comparing with the isoelastic case for the non-buncher (see equation 11), we note that they give the same result for $\varepsilon = 1$. For the realistic case $\varepsilon < 1$, my expression gives a lower number; if $\varepsilon < 1$ the elasticity is monotonically declining with the tax rate under linearity, implying that it gets higher for lower tax rates. Recall that calculating the excess



Note: The excess burden is the area bordered by the tax revenue rectangle (to the left) and the taxable income supply function (to the right).

Figure 13: Tax revenue and excess burden for a representative individual

Table 9: Comparison between the linear and the isoelastic taxable income supply function

	Linear–representative	Isoelastic–Pareto
$z(\tau) =$	$\left(1 - \varepsilon \frac{\tau - \tau_c}{1 - \tau_c}\right) \bar{z}_b$	$\left(\frac{1 - \tau}{1 - \tau_c}\right)^{\varepsilon} \bar{z}_b$
$AEB =$	$\frac{\alpha\varepsilon\tau}{2(1 - \tau)}$	$\frac{(1 - \tau)^{-\alpha\varepsilon} - 1 - \alpha\varepsilon\tau}{\tau(1 + \alpha\varepsilon)}$
$R(\tau) =$	$\tau \left(1 - \alpha\varepsilon \frac{\tau - \tau_c}{1 - \tau_c}\right) (\bar{z}_b - b)$	$\tau(1 - \tau)^{\alpha\varepsilon}(\bar{z}_b - b)$
$\tau_{\max} =$	$\frac{1 + (\alpha\varepsilon - 1)\tau_c}{2\alpha\varepsilon}$	$\frac{1}{1 + \alpha\varepsilon}$

Note: The isoelastic case normalizes the number of potential top bracket taxpayers to 1. τ_c denotes the current top marginal tax rate.

burden entails estimating how high taxable income would be if the tax rate were cut to zero. Figure 13 illustrates: With $\varepsilon < 1$ the isoelastic taxable income supply function is concave in $1 - \tau$, which means that the area bordered by the curve is smaller. If $\varepsilon > 1$, the isoelastic excess burden is greater than the linear excess burden.

Let us now consider the aggregate excess burden as a share of tax revenue (the average excess burden). The Harberger approach usually treats the average income earner as a representative taxpayer, so the average excess burden is $\bar{z}_b\varepsilon\tau^2/2(1 - \tau)/\tau(\bar{z}_b - b) = \alpha\varepsilon\tau/2(1 - \tau)$. My expression (equation 13) accounts for the income distribution (assumed Pareto) and for bunching. Similar to the individual case, we see that the isoelastic excess burden is lower if and only if $\alpha\varepsilon < 1$ (this can be seen intuitively in figure 3).

How does the Laffer curve look under a linear taxable income supply function? Again, we consider a representative taxpayer and write $R(\tau) = \tau(z(\tau) - b) = \tau(\bar{z}_b [1 - (\tau - \tau_c)\varepsilon/(1 - \tau_c)] - b) = \tau [1 - \alpha\varepsilon(\tau - \tau_c)/(1 - \tau_c)] (\bar{z}_b - b)$. The maximum occurs at $\tau = (1 + (\alpha\varepsilon - 1)\tau_c)/2\alpha\varepsilon$. This Laffer curve exhibits some odd behaviour at high tax rates. Tax revenues can become negative, or the revenue-maximizing rate – which is different depending on the current tax rate – can be higher than 100 percent.

Table 9 summarizes the various quantities of interest under a linear and an isoelastic taxable income supply function. First, we can say that the researcher should be consistent with which expressions are used. It is not uncommon in the literature to use the linear Harberger (1964) formula to calculate the excess burden and the isoelastic Saez (2001) formula for the revenue-maximizing rate. Given that the current tax rate is lower than the Laffer curve peak and $\alpha\varepsilon < 1$, this approach entails assuming that the elasticity would be constant if the tax rate were increased, and at the same time that it would go up if the tax rate were cut.

The historical research endeavour of estimating labour supply functions (Heckman, 1993) has largely been abandoned in favour of a sufficient statistics approach (Kleven, 2021). Until that endeavour is resumed, we need to make an assumption about how the elasticity will change if the tax rate changes. This assumption naturally becomes more heroic the larger the tax reform under consideration is.[22] For the reasons outlined above, I advocate the isoelastic taxable income supply function.

[22] As already mentioned, no assumptions at all need to be made for marginal tax changes: $q(\tau) = 1 - \alpha\varepsilon\tau/(1-\tau)$ regardless.