CrossMark

Higher Tax for Top Earners

Felix FitzRoy[1] · Jim Jin[1]

Abstract The literature can justify both increasing and decreasing marginal taxes (IMT & DMT) on top incomes under different welfare objectives and income distributions. Even when DMT are theoretically optimal, they are often politically infeasible. Then a flat tax seems to be a constrained optimal solution. We show however that, given any flat tax we can increase the total utility of a poor majority by raising the top income tax rate under a simple condition, which can be checked with empirical data. We further generalize our main results allowing different welfare weights, declining elasticity of labor supply and more tax bands.

1 Introduction

Throughout most developed and developing economies, income distributions have become increasingly skewed in recent decades (Stiglitz 2012; Piketty 2014). One reason has been declining marginal tax rates for top incomes. Another is the effective high marginal tax faced by low income earners due to withdrawal of benefits as earnings rise, leading to the poverty trap. The existing tax structures in most developed countries

✉ Jim Jin
  jyj@st-andrews.ac.uk

  Felix FitzRoy
  frf@st-andrews.ac.uk

[1] School of Economics and Finance, University of St Andrews, St Andrews KY16 9AL, Scotland, UK

 Springer

are U-shaped, with increasing marginal tax (IMT) on high income earnings (but not on capital gains). The justification of IMT on top incomes is to raise revenue from those most able to pay, and provide a social safety net for the poor. This view is theoretically justified by Diamond (1998) and Saez (2001) based on their assumption that top income follows Pareto distributions (see also Salanié 2003).

However, the shape of the optimal tax curve seems to be sensitive to income distributions. With a bounded distribution, Sadka (1976) and Seade (1977) find zero optimal marginal tax rate for the top earners. In general optimal tax curves are often inversely U-shaped or even declining (see Tuomala 1984; Kanbur and Tuomala 1994; Boadway et al. 2000; Tarkiainen and Tuomala 2007; Hashimzade and Myles 2007; Boadway and Jacquet 2008; Kaplow 2008), implying decreasing marginal taxes (DMT) on top incomes. But as Warren Buffet famously complained, the lower effective average tax rates paid by the rich, due to low capital gain taxes and various loopholes, are widely perceived to be unfair. This political problem often imposes a binding constraint and suggests a constrained optimal solution to be a flat tax, which by continuity should be closer to the optimal DMT and dominate IMT. Moreover, a flat tax will reduce administrative costs and avoid incentive distortions (see Atkinson 1995 for a good overview). Thus Mankiw et al. (2009) argue that "A flat tax, with a universal lump-sum transfer, could be close to optimal".

On the other hand, Diamond and Saez (2011), Piketty and Saez (2012, 2013) (DSPS) argue that if policy makers ignore the welfare of the richest group (due to their low marginal utility of income) and focus on the poor majority, the tax rate for top income should be 70–80%, supposedly higher than the tax rate for lower incomes. This policy has been successfully applied in the Scandinavian countries where high top tax rates co-exist with high labour force participation and the highest level of life satisfaction (Kleven 2014). The validity of different policy recommendations, IMT or flat tax, seems crucially dependent on social objectives as well as income distribution.

This paper shows that even when DMT are optimal but not feasible, a flat tax may not be the next best alternative. Given any flat tax, we can increase the total utility of a poor majority by raising the tax rate on top earners under a simple condition, which means the optimal tax on top earnings derived by Saez (2001) is higher than the optimal flat tax. This condition generally holds when the poor majority is sufficiently large.

Following DSPS (though they consider more general cases), we ignore the interests of the rich group and focus on the poor. Later we allow different weights given to different poor households, leading to a similar effect as decreasing marginal utility of income assumed by DSPS. Surprisingly, when we put more weight on the very poor households, a higher tax for top earners is less likely to benefit the poor.

We first assume a constant elasticity of labour supply for the whole population. Later we assume more realistically declining elasticity with income and show that a higher top tax is more likely to be justified. This is consistent with the optimal IMT obtained by Aaberge and Colombino (2013) and Andrienko et al. (2014), using data from Norway, US, UK and Australia, with declining elasticity of labour supply.

Continuous tax curves have been criticized as "too far removed from the tax–benefit systems observed in practice to be a useful guide for policy" (Chone and Laroque 2005, p. 396). Apps et al. (2009) remark that "Given its significance in practice, the

piecewise linear tax system seems to have received disproportionately little attention in the literature on optimal income taxation." Following Diamond and Saez (2011) who argue for practical and useful research on tax policy, we first consider two-band taxes. This system is the natural first step beyond a flat tax and can model IMT and DMT as well as a flat tax. Furthermore the two-band tax literature finds DMT optimal.

Sheshinski (1989) first argued for increasing two-band taxes. However, Slemrod et al. (1994) find errors in his proof and use numerical simulations to show that DMT maximize maximin and utilitarian objectives. Similarly Salanié (2003), Hindricks and Myles (2006) obtain optimal decreasing two-band taxes in a two-class economy. Hence it is interesting to see if both decreasing and increasing two-band taxes dominate any flat tax. We later allow more tax bands and generalize our result accordingly.

We introduce our two-band tax model in the next section. Section 3 shows that any flat tax is Pareto dominated by some DMT. Section 4 gives a sufficient condition for a higher top tax rate to benefit a poor majority and shows it is valid if the majority is sufficiently large. Section 5 extends our model and generalizes the results allowing different welfare weights for the poor, declining elasticity of labor supply and multiple-tax bands. Section 6 concludes the paper.

2 Basic Model

We assume that a population, normalized to unity, consists of a continuum of households, whose wage is denoted by w, and is distributed on $[a, b]$, where $a > 0$, b is bounded, but can be very large and approximately treated as infinite. The density and cumulative functions of w are denoted by $f(w)$ and $F(w)$. We define the poor population as those with wages below a fixed level $\bar{w}$, and denote those with higher wages as the rich. The government's objective is to maximize the total utility of the poor. This is similar to Diamond and Saez (2011) who give virtually zero weight to the rich in the social welfare function due to decreasing marginal utility of income. Our objective can be justified by the political goal of income redistributions. We first treat the poor equally but will give them different welfare weights in Sect. 5(i).

Every household has a quasi-linear utility, $m - x^{1+1/\varepsilon}/(1 + 1/\varepsilon)$, where m is net income, x is labour supply and ε is its elasticity. This simple utility function has been widely used in the literature (e.g. Atkinson 1995). We first assume an identical ε for the whole population and later allow declining elasticity in Sect. 5(ii).

Given wage w, a household's pre-tax earnings $y = wx$. The government imposes two tax rates, t_1 and t_2, for earnings below and above a threshold Y. The tax revenue, after a fixed expenditure is paid, is distributed to all households equally as a basic income, denoted by B. Given our unit population B is also equal to the total transfer received by the whole population. The two-band taxes reduce to a flat tax when $t_1 = t_2$. We will allow more tax bands in Sect. 5(iii).

Given t_1, t_2, Y and B, households' utility functions can be written as:

$$u_1 = wx(1 - t_1) - \frac{x^{1+1/\varepsilon}}{1 + 1/\varepsilon} + B \qquad \text{for } wx \leq Y \qquad (1)$$

$$u_2 = wx(1 - t_2) + (t_2 - t_1)Y - \frac{x^{1+1/\varepsilon}}{1 + 1/\varepsilon} + B \qquad \text{for } wx > Y \qquad (2)$$

Every household chooses labour supply x to maximize utility. We first consider IMT, i.e. $t_1 < t_2$ and assume $Y \geq \bar{w}^{1+\varepsilon}(1 - t_1)^\varepsilon$. Thus every poor household faces the lower rate t_1, and chooses optimal labour supply $x = w^\varepsilon(1 - t_1)^\varepsilon$. This can be justified by the political agenda to help the poor by charging them a low tax rate t_1. Substituting it into (1), we obtain the maximized utility $w^{\varepsilon+1}(1 - t_1)^{\varepsilon+1}/(1+\varepsilon) + B$. Integrating it over $[a, \bar{w}]$, we get the total utility of the poor as our objective function:

$$W = \int_a^{\bar{w}} \frac{(1 - t_1)^{1+\varepsilon}}{1 + \varepsilon} w^{1+\varepsilon} f(w)dw + BF(\bar{w}) \qquad (3)$$

Given $Y \geq \bar{w}^{1+\varepsilon}(1 - t_1)^\varepsilon$ and $t_1 < t_2$, the population is divided into three groups. All poor households and some rich ones with $w < \hat{w} \equiv [Y/(1 - t_1)^\varepsilon]^{1/(1+\varepsilon)}$ choose labor supply $x = w^\varepsilon(1-t_1)^\varepsilon$ and pay tax of $t_1(1-t_1)^\varepsilon w^{\varepsilon+1}$. Very rich households with $w > w_1 \equiv [Y/(1-t_2)^\varepsilon]^{1/(1+\varepsilon)}$ choose $x = w^\varepsilon(1-t_2)^\varepsilon$ and pay tax $t_2(1-t_2)^\varepsilon w^{\varepsilon+1} + (t_1 - t_2)Y$. The remaining rich households with $\hat{w} < w \leq w_1$ choose $x = Y/w$, earning Y, i.e. bunching, and pay $t_1 Y$. As $t_1[F(w_1) - F(\hat{w})] + (t_1 - t_2)[1 - F(w_1)] = t_1[1 - F(\hat{w})] - t_2[1 - F(w_1)]$, the total tax revenue from these three groups is:

$$R = \int_a^{\hat{w}} t_1(1 - t_1)^\varepsilon w^{1+\varepsilon} f(w)dw + \int_{w_1}^b t_2(1 - t_2)^\varepsilon w^{1+\varepsilon} f(w)dw$$
$$+ \{t_1[1 - F(\hat{w})] - t_2[1 - F(w_1)]\}Y \qquad (4)$$

We assume the fixed expenditure is less than R, so B is positive and maximized whenever R is. So we can replace R by B. Under a flat tax, $t_1 = t_2 = t$, we have $\hat{w} = w_1$, and (4) reduces to $\int_a^b t(1 - t)^\varepsilon w^{1+\varepsilon} f(w)dw$. Then our objective function (3) reduces to:

$$W = \int_a^{\bar{w}} \frac{[(1 - t)w]^{1+\varepsilon}}{1 + \varepsilon} f(w)dw \qquad (3')$$
$$+ F(\bar{w}) \int_a^b t(1 - t)^\varepsilon w^{1+\varepsilon} f(w)dw$$

3 Flat Tax and DMT

The literature (e.g. Slemrod et al. 1994) has shown that DMT are generally optimal for two-band taxes under maximin or utilitarian objectives. In this section we show that a flat tax is always Pareto dominated by some DMT.

We first find the optimal flat tax which maximizes $(3')$. To simplify the notation we denote the total earnings of the poor under zero tax by $E_1 \equiv \int_a^{\bar{w}} w^{1+\varepsilon} f(w)dw$ and denote the corresponding earnings of the rich by $E_2 \equiv \int_{\bar{w}}^b w^{1+\varepsilon} f(w)dw$. The total zero-tax earnings of the whole population is $E = E_1 + E_2$. Since the population is

normalized to 1, E is also the average zero-tax earnings of the whole population. Then the average earnings of the poor and rich are $e_1 \equiv E_1/F(\bar{w})$ and $e_2 \equiv E_2/[1 - F(\bar{w})]$ respectively. By the definition we always have $e_1 \leq E \leq e_2$.

We differentiate (3′) and find $dW/dt = [1 - t(1+\varepsilon)](1-t)^{\varepsilon-1}F(\bar{w})E - E_1(1-t)^{\varepsilon}$. It is positive if and only if $t < (1 - e_1/E)/(1 + \varepsilon - e_1/E]$. Hence we get the following result.

Proposition 1 *The optimal flat tax to maximize (3′) is* $t^* = \frac{1 - e_1/E}{1 + \varepsilon - e_1/E}$.

This result is a special case of Piketty and Saez (2013), who derive an optimal linear tax of $(1 - \bar{g})/(1 + \varepsilon - \bar{g})$, where $\bar{g}$ is the average social welfare weight weighted by pre-tax incomes, which "*is also the ratio of the average income weighted by individual social welfare weights g_i to the actual average income*" (p. 21). Given our welfare function, which only values the utility of the poor, $\bar{g} = e_1/E$ and their formula reduces to our t^*. Piketty and Saez (2013) further discuss the median voter tax rate, which maximizes the utility of the median earner, and point out "a tight connection between optimal tax theory and political economy". If e_1 equals the median no-tax earnings, t^* is the median voter tax. Interestingly, the median income in the U.S. was roughly \$26,000, when the average top 1% income was estimated by Piketty and Saez as \$1.2 million. Given the average earnings of \$38,000, the average of the bottom 99%, e_1 was also about \$26,000. Thus our flat tax for the 99% majority is also the median voter tax.

Next we show that any flat tax $t(0 < t < 1/(1 + \varepsilon))$, including t^*, is Pareto dominated by some DMT. $1/(1 + \varepsilon)$ is the revenue maximizing flat tax. We exclude the case of $t \geq 1/(1 + \varepsilon)$, which lies in the inefficient part of the Laffer curve. Now we lower tax rate t_2 for earnings beyond $Y = (1 - t)^{\varepsilon}[1 - \varepsilon t/(1 - t)]b^{1+\varepsilon}$, which is positive given $t < 1/(1 + \varepsilon)$. As $b^{1+\varepsilon}$ is the highest no-tax earnings, there is a positive mass earning more than Y, and we can show that each of them will pay more tax with a lower tax rate t_2. The tax payment from a household within this group is $t_2(1 - t_2)^{\varepsilon}w^{\varepsilon+1} + (t - t_2)Y$. According to Saez (2001) the impact of tax change can be decomposed into two effects, mechanical and behavioral[1]. The former can be obtained under a constant labor supply and expressed as $[(1 - t_2)^{\varepsilon}w^{\varepsilon+1} - Y]\Delta t_2$. The latter is due to the response of labor supply and indicated by $-\varepsilon t_2(1 - t_2)^{\varepsilon-1}w^{\varepsilon+1}\Delta t_2$. Adding them together the derivative of the tax payment respect to t_2 is negative at $t_2 = t$ if $(1 - t)^{\varepsilon}[1 - \varepsilon t/(1 - t)]w^{\varepsilon+1} < Y$, which is guaranteed for any $w < b$ given our definition of Y. Thus each household earning more than Y pays more tax when t_2 falls. These households must be better off due to a lower marginal tax rate and higher basic income B. Moreover the poorer households are better off too due to higher B. Therefore a lower t_2 benefits everyone.

Proposition 2 *Every flat tax is Pareto dominated by some DMT.*

The intuition follows from Saez' (2001) concept of behavioural and mechanical responses. A lower t_2 will motivate rich households to increase their labor supply. If the tax threshold Y is set sufficiently high, the tax revenue loss will be limited, and the

---

[1] We are grateful to an anonymous referee for pointing out this decomposition.

extra labor supply from each household can generate significant tax revenue due to its high productivity. So the behavioural effect dominates the mechanical effect, leading to a higher revenue. This lower tax applies to a positive mass, not just the highest earner, different from the zero top marginal tax obtained by Sadka (1976) and Seade (1977).

When DMT are optimal but politically infeasible, a flat tax seems to be better than IMT since it should be closer to the true optimum DMT by continuity, and thus dominates IMT. However, this monotonicity of tax policy may not be valid. Assuming the government is politically constrained to implement two-band IMT, we will show that the optimal flat tax is dominated by some IMT under a simple condition.

4 IMT Versus Flat Tax

Given the optimal flat tax t^*, the question now is whether some IMT ($t_1 < t_2$) can generate a higher value of (3) than t^* does. This must be true if we find $\partial W/\partial t_1 < 0$ and $\partial W/\partial t_2 > 0$ when $t_1 = t_2 = t^*$. In fact these two conditions are identical and we can focus on $\partial W/\partial t_2 > 0$. Notice that the first term in (3) does not depend on t_2. If t_2 maximizes (3), it must maximize B (i.e. R). This is essentially the approach taken by Saez (2001). To prove that IMT can dominate the optimal flat tax, we just need to show $\partial B/\partial t_2 > 0$ when $t_1 = t_2 = t^*$, instead of finding the optimal t_2.

For simple presentation we let $Y = \bar{w}^{1+\varepsilon}(1 - t_1)^\varepsilon$. This is not the only choice to obtain our results. For example, if we let $Y = \bar{w}^{1+\varepsilon}(1 - t^*)^\varepsilon$, the marginal poor ($w = \bar{w}$) will bunch when we lower t_1 and raise t_2, but this does not change the condition for $\partial W/\partial t_1 > 0$ and $\partial W/\partial t_2 < 0$ at $t_1 = t_2 = t^*$, and has no effect on our result. Since our goal is to show IMT can dominate t^*, this particular Y serves our purpose. $Y = \bar{w}^{1+\varepsilon}(1 - t_1)^\varepsilon$ implies $\hat{w} = \bar{w}$ and the tax revenue (4) (hence B) simplifies to:

$$
B = \int_a^{\bar{w}} t_1(1 - t_1)^\varepsilon w^{1+\varepsilon} f(w)dw \tag{4'}
$$
$$
+ \int_{w_1}^b t_2(1 - t_2)^\varepsilon w^{1+\varepsilon} f(w)dw
$$
$$
+ \{t_1[1 - F(\bar{w})] - t_2[1 - F(w_1)]\}Y
$$

Then we investigate whether two-band taxes with $t_1 < t_2$ can lead to a higher value of (3) than the optimal flat tax t^*, with B in (3) replaced by (4'). For simple expression we denote the marginal household's zero-tax earnings, $\bar{w}^{1+\varepsilon}$ by $\bar{y}$.

Proposition 3 *There exists a two-bracket tax schedule with $t_1 < t^* < t_2$ that dominates the optimal linear tax rate t^*, if at $t_1 = t_2 = t^*$, we have*

$$
\frac{e_1}{E} > \frac{\bar{y}}{e_2} \tag{5}
$$

Proof see "Appendix A". □

As we mentioned earlier, $\partial W/\partial t_1 < 0$ and $\partial W/\partial t_2 > 0$ depend on the same condition. This is not coincidental. If both $\partial W/\partial t_1 > 0$ and $\partial W/\partial t_2 > 0$ at $t_1 = t_2 = t^*$, it would be possible to increase (3) by raising t_1 and t_2 together. But this is impossible since t^* is the optimal flat tax to maximize (3′).

When (5) holds, any flat tax is dominated by both IMT and DMT. There is no monotonicity in the optimal taxes. However, the superior IMT and DMT require different thresholds for top tax rates. In proving the Pareto superiority of DMT, we assumed $Y = (1-t)^\varepsilon[1 - \frac{\varepsilon t}{1-t}]b^{1+\varepsilon}$. For (5) to hold, i.e. $\frac{\partial W}{\partial t_2} > 0$, we need $(1-\frac{\varepsilon t}{1-t})E_2 > [1-F(\bar{w})]\bar{y}$ (see "Appendix A"). Since $E_2/[1-F(\bar{w})] = e_2 < b^{1+\varepsilon}$, Y associated with DMT must be higher than $(1-t)^\varepsilon \bar{y}$ in the IMT case. Hence both a higher and a lower top tax rate would be desirable if implemented at different income thresholds.

Intuitively (5) can again be explained by Saez' (2001) concept of behavioural and mechanical responses as in Proposition 1. The difference is that here a higher t_2 may raise some households' tax payment and reduce others'. Given $\Delta t_2 > 0$, the mechanical effect on (4′) is equal to $\{\int_{\bar{w}}^b (1-t^*)^\varepsilon w^{1+\varepsilon} f(w)dw - [1 - F(\bar{w})]Y\}\Delta t_2$, and the behavioral effect is $-[\varepsilon \int_{\bar{w}}^b t^*(1-t^*)^{\varepsilon-1}w^{1+\varepsilon} f(w)dw]\Delta t_2$. If their net effect is positive, $[1 - \varepsilon t^*/(1-t^*)]e_2 > \bar{y}$, as shown by (A2) in "Appendix A", the total tax payment rises with t_2, i.e., $\partial B/\partial t_2 > 0$. Since $1 - \varepsilon t^*/(1-t^*)$ is equal to e_1/E, the condition reduces to (5)[2].

Moreover, our result can be obtained by directly comparing the optimal flat tax t^* with the optimal top income tax rate obtained in Saez (2001). Without an income effect as assumed here, his tax rate becomes $(1-g)/[1 - g + \varepsilon e_2/(e_2 - \bar{y})]$, where g is the social welfare weight given to the rich (also see Piketty and Saez 2013). In our model $g = 0$ given zero welfare weight for the rich. Thus Saez' optimal top income tax rate becomes $(e_2 - \bar{y})/[(1+\varepsilon)e_2 - \bar{y}]$. If it is higher than t^*, IMT must dominate t^*. However no one has explicitly compared these two tax rates. In fact Saez' asymptotic marginal tax rate t^a can be obtained from $[1 - \varepsilon t^a/(1-t^a)]e_2 = \bar{y}$. So $t^a > t^*$ if and only if (5) holds[3]. Otherwise Saez' marginal tax for top income is inconsistent with IMT.

To evaluate (5), it is often convenient to consider the income distribution function $G(y)$, with $y = w^{1+\varepsilon}$, instead of the wage distribution $F(w)$. The validity of (5) may not depend on $\bar{y}$. For instance, when the income distribution is nearly unbounded, we may approximate it by a Pareto distribution, $G(y) = 1 - y^{-\alpha}$ for $y \geq 1, \alpha > 1$. Then condition (5) holds for any $\bar{y}$[4]. This result is consistent with Diamond (1998). One may attribute this result to the thick-tail of a Pareto distribution. However, if α is large, the tail becomes very thin while (5) still holds. To see this point further, we consider a thick-tailed distribution $G(y) = (y/h)^\beta$, with $0 \leq y \leq h$ and $\beta > 0$. The number

---

[2] We are very grateful to an anonymous referee for his suggestion on this interpretation.

[3] We thank an anonymous referee for pointing to this connection and implication.

[4] As $E_2 = \alpha \bar{y}^{1-\alpha}/(\alpha-1)$, $1 - G(\bar{y}) = \bar{y}^{-\alpha}$, $E = \alpha/(\alpha-1)$, $e_2 = \bar{y}E$, so (5) becomes $e_1 > 1$.

of rich households may even rise with income (if $\beta > 1$). But (5) never holds[5]. The validity of (5) does not require an unbounded income either. For instance, consider a bounded Pareto distribution with $G(y) = (1 - y^{-\alpha})/(1 - h^{-\alpha})$, with $1 \leq y \leq h$ and $\alpha > 1$. It can be shown that (5) holds for any h and $\bar{y}$, even when the maximum income h is very low and close to 1. These examples demonstrate the sensitivity of (5) to income distributions.

In spite of such complexity, the validity of (5) may be determined by simple data without knowing income distributions precisely. For instance, Diamond and Saez (2011) estimate the U.S. threshold of the top 1% as \$0.4 million and their average earnings as \$1.2 million. This implies $\bar{y}/e_2 = 1/3$, which is lower than e_1/E, given $e_1 = \$26,000$ and $E = \$38,000$ as we mentioned earlier. So condition (5) holds.

Moreover, if we know the income distribution above the threshold $\bar{y}$, (5) can be simplified. According to extreme value theory (Gnedenko 1943), for a wide range of random variables, the conditional probability approximately follows a Pareto distribution when they are sufficiently large. This theory and empirical evidence suggest a Pareto distribution as a good approximation for top earners. Let $G(y) = 1 - Ky^{-\alpha}$ for $y \geq \bar{y}, \alpha > 1$, we obtain $e_2/\bar{y} = \alpha/(\alpha - 1)$ and can simplify (5) to:

$$1 - \frac{e_1}{E} < \frac{1}{\alpha} \tag{6}$$

In this case a thick tail does have a crucial impact. Given e_1/E, a very thick tail (α close to 1) guarantees (6); while a thin tail (a large α) ensures its violation. For the top U.S. 1% earners Diamond and Saez (2011) estimate $\alpha = 1.5$, so (6) becomes $e_1/E > 1/3$. It holds as $e_1/E = 26/38$. For U.S. 1992 earnings above \$150,000, Saez (2001) shows $\alpha = 2$ (i.e. $\bar{y}/e_2 = 0.5$). Similarly Bach et al. (2012) find $\alpha = 2$ for German top earnings. Then (6) becomes $e_1 > 0.5E$. For any Pareto distribution with a finite α, when $\bar{w}^{1+\varepsilon}$ is sufficiently large, e_1 must be close to E and (6) will certainly hold.

Corollary *If high earnings follow a Pareto distribution, a higher tax on a small group of top earners always benefits the remaining population.*

This result supports DSPS's view about a higher tax on top earners. But this may only apply to a small rich group, e.g. 1%. The current top tax rate, however, usually applies to a much larger group. Bach et al. (2012) argue that their top tax rate of 2/3 in Germany should only apply to an income level much higher than the current threshold. Indeed when we consider a higher tax on a large group, (5) may not hold. For instance, if we consider a higher tax on the top 50%, i.e. $\bar{y}$ = the median earnings, (5) does not hold for any lognormal distribution. Therefore, starting from the optimal flat tax, raising the tax rate beyond the median earnings will not benefit the poor 50%.

The question is how large the rich group should face a higher tax. It is difficult to answer this question by (5) directly since it is very sensitive to income distributions which can hardly be identified precisely. It would be desirable to check its validity

---

[5] As $E_1 = \beta \bar{y}^{1+\beta}/h^{\beta}(\beta + 1)$, $E = \beta h/(\beta + 1)$, and $e_1 = \beta \bar{y}/(\beta + 1)$, (5) requires $e_2 > h$.

without assuming specific distributions. This is easier to do using another condition equivalent to (5). It depends on whether we have a decreasing e_1/e_2, the ratio of the average earnings of the poor and the rich (Proof: see "Appendix B").

Proposition 4 *(5) holds if and only if e_1/e_2 falls around $\bar{y}$.*

If e_1/e_2 is single peaked, it will fall after its maximum. If earnings are unbounded and $\bar{y}$ is sufficiently large, e_1 approaches to E but e_2 to infinity. So e_1/e_2 must fall and IMT must dominate any flat tax. The question is: how large $\bar{y}$ is "sufficient". The answer may not be obtained from the theory alone, but from empirical data.

Our data are obtained from the United Nation's "World Income Inequality Database" (May (2008)), and provide each decile's earnings as percentages of aggregate earnings. The data set does not contain the relevant information for all years. To avoid subjective bias we use the most recent data for each country. Unfortunately, our ratio of e_1/e_2 does not take into account complex tax systems which generate real data. So we use the actual earnings ratios as approximation for zero-tax earnings ratios. On the other hand, despite complex tax systems in G8 countries, we find their e_1/e_2 curves are all single peaked and fall from similar thresholds of income deciles.

We use a decile's earnings as a percentage of the aggregate earnings to calculate e_1/e_2. The ratio of this group's earnings to that of the whole population is given as $r \equiv E_1/E$. So $e_1 = E_1/G(y) = rE/G(y)$. $e_2 = (E - E_1)/[1 - G(y)]$, i.e. $(1-r)E/[1 - G(y)]$. Hence $e_1/e_2 = r[1 - G(y)]/(1-r)G(y)$. The data provide us the values of r for $G(y) = 10$–90%, giving us 9 values of e_1/e_2. The results for G8 countries are given in Table 1.

Apparently, the e_1/e_2 ratios differ significantly between eight countries. As we mentioned earlier, these values are only approximations since we do not take into account complex non-linear tax systems different from flat tax assumed here. Nonetheless these e_1/e_2 ratios all exhibit a single peak in G8 countries and surprisingly, they start to decline around 80% of income levels. Hence a higher tax can be justified when it is imposed on less than 20% of top earners on the behalf of more than 80% poor majority.

Table 1 Ratio of e_1/e_2 for G8 countries

Country	Year	10%	20%	30%	40%	50%	60%	70%	**80%**	90%
Canada	2000	0.25	0.32	0.35	0.38	0.39	0.39	0.39	**0.38**	0.34
France	2000	0.38	0.40	0.44	0.45	0.45	0.46	0.46	**0.44**	0.42
Germany	2000	0.31	0.37	0.40	0.42	0.43	0.43	0.43	**0.41**	0.38
Italy	2002	0.21	0.27	0.31	0.32	0.34	0.35	0.35	**0.34**	0.31
Russia	2000	0.13	0.20	0.24	0.27	0.29	0.29	0.29	**0.27**	0.23
UK	1999	0.23	0.28	0.31	0.32	0.34	0.34	0.34	**0.32**	0.28
USA	2000	0.17	0.23	0.26	0.29	0.31	0.32	0.32	**0.31**	0.27
Japan	1971	0.05	0.16	0.22	0.26	0.29	0.30	0.30	**0.29**	0.25

5 Extensions

(i) *Welfare weight* So far we have treated all the poor equally. Ideally we may give them different welfare weights, and allow a continuous treatment across the rich and the poor. This is similar to the approach taken by DSPS based on decreasing marginal utility of income. Decreasing welfare weights have a similar effect as assuming decreasing marginal utility of income. Intuitively, one may expect that this should increase the chance of justifying a higher top tax rate. However, like the conventional belief that the maximin is most likely to justify IMT, this conjecture is not correct.

Given $\bar{w}$ we assign welfare weight $s(w)$ to every poor household w ($\leq \bar{w}$) such that $\int_a^{\bar{w}} s(w) f(w) dw = F(\bar{w})$. We then multiply $s(w)$ with each poor household's net utility $[(1 - t_1)w]^{1+\varepsilon}/(1 + \varepsilon) + B$, and integrate the product over $[a, \bar{w}]$, to get a weighted total utility of the poor as our new objective function:

$$W = \frac{(1 - t_1)^{1+\varepsilon}}{1 + \varepsilon} \int_a^{\bar{w}} s(w) w^{1+\varepsilon} f(w) dw + B F(\bar{w}) \qquad (7)$$

Objective (7) reduces to (3) when $s(w) = 1$ for any $w \leq \bar{w}$. Since $s(w)$ falls with w, $\int_a^{\bar{w}} s(w) w^{1+\varepsilon} f(w) dw < \int_a^{\bar{w}} w^{1+\varepsilon} f(w) dw$. We use $\tilde{e}_1$ to denote the weighted average no-tax earnings of the poor, $\int_a^{\bar{w}} s(w) w^{1+\varepsilon} f(w) dw / F(\bar{w})$. The more weight is given to the poorer households the lower $\tilde{e}_1$ is. As in the previous case, we first obtain the optimal flat tax $\tilde{t}^*$ which maximizes (7). It is similar to t^*, except for e_1 being replaced by $\tilde{e}_1$, i.e. $\tilde{t}^* = (1 - \tilde{e}_1/E)/(1 + \varepsilon - \tilde{e}_1/E)$. Then a higher top tax rate raises (7) if $\partial W/\partial t_1 < 0$ and $\partial W/\partial t_2 > 0$ when $t_1 = t_2 = \tilde{t}^*$, which holds under a new condition (see "Appendix C").

Proposition 5 *IMT give a higher value of (7) than any flat tax if at $t_1 = t_2 = \tilde{t}^*$*

$$\frac{\tilde{e}_1}{E} > \frac{\bar{y}}{e_2} \qquad (8)$$

When $s(w) = 1$, $\tilde{e}_1 = e_1$ and (8) reduces to (5). Condition (8) can also be linked to Saez' asymptotic marginal tax rate. Given any $\tilde{e}_1 < e_1$, the optimal tax rate for top income remains the same as before, but the optimal flat tax $\tilde{t}^*$ is higher given higher welfare weights on the very poor. So the former is less likely to be higher than the latter, and (8) is less likely to hold than (5) is, and a higher top tax is less likely to be justifiable, unexpectedly. The intuition is that the poorer households are less productive, and rely more on income transfer. A higher tax on low earnings is less damaging to them and more beneficial due to more money transfer from the rich. So a flat tax is less likely to be dominated if we give most weight to the poorest.

The validity of (5) only implies a higher top tax rate can benefit the poor as a whole, not necessarily each of them. (8) can tell us if it benefits a particular household. Our objective (7) is identical to maximizing the utility of a household with earnings of $\tilde{e}_1$, as a representative family. When (8) holds, a higher top tax rate benefit those with earnings equal or higher than $\tilde{e}_1$. If $\tilde{e}_1$ is the lowest earnings, all poor will be better off. For instance, given a Pareto distribution with $\alpha = 1.5$ for the top 1% earners,

$\bar{y}/e_2 = 1/3$, and (8) becomes $\tilde{e}_1/E > 1/3$. In most OECD countries (except for US), the ratio of the minimum wage to average wage is more than $1/3$[6]. So a higher tax on top 1% can benefit all 99%. Similarly, with $\alpha = 2$ for the top German earnings, (8) becomes $\tilde{e}_1/E > 0.5$. The lowest and average monthly German salaries are €1,832 and €3449[7]. Thus virtually all poor can benefit from a higher top tax.

(ii) *Declining elasticity* Empirical data show that full-time and high income earners are less responsive to tax changes than part-time and low income earners (see Aaberge and Colombino 2013 and Andrienko et al. 2014). So our assumption of constant elasticity of labor supply is unrealistic. In fact this assumption is unfavorable for IMT. Now we allow the elasticity to be declining with income. Our objectives (3) and (3'), and the tax revenue (4') remain valid, except that ε cannot be taken out of the integrals. We follow the same approach as before, i.e. first obtain the optimal flat tax $\hat{t}^*$, which maximizes (3'), then evaluate $\partial W/\partial t_1$ and $\partial W/\partial t_2$ when $t_1 = t_2 = \hat{t}^*$.

Following DSPS we define the average elasticity of labor supply, weighted by earnings, as $\hat{\varepsilon} \equiv \int_a^b \varepsilon w^{1+\varepsilon} f(w) dw / \int_a^b w^{1+\varepsilon} f(w) dw$, and define the average elasticity of the rich as $\hat{\varepsilon}_2 \equiv \int_{\bar{w}}^b \varepsilon w^{1+\varepsilon} f(w) dw / \int_{\bar{w}}^b w^{1+\varepsilon} f(w) dw$. Declining elasticity implies $\hat{\varepsilon}_2 < \hat{\varepsilon}$. Then we differentiate (3') to get the optimal flat tax $\hat{t}^* = (1 - e_1/E)/(1 + \hat{\varepsilon} - e_1/E)$. If $\partial W/\partial t_1 < 0$ and $\partial W/\partial t_2 > 0$ when $t_1 = t_2 = \hat{t}^*$, some IMT dominate any flat tax.

Proposition 6 *With declining elasticity of labor supply, some IMT dominate any flat tax if at $t_1 = t_2 = \hat{t}^*$, we have*

$$1 - \frac{e_1}{E} < \left(1 - \frac{\bar{y}}{e_2}\right) \frac{\hat{\varepsilon}}{\hat{\varepsilon}_2} \tag{9}$$

Proof see "Appendix D". □

Inequality (9) reduces to (5) if $\hat{\varepsilon}_2 = \hat{\varepsilon}$. Given $\hat{\varepsilon}_2 < \hat{\varepsilon}$, (9) is more likely to hold than (5) is, and a higher top tax rate is more likely to benefit the poor, as expected. Given a Pareto distribution with $\alpha = 2$ for top incomes, we have $\bar{y}/e_2 = 0.5$, and (9) becomes $1 - e_1/E < 0.5\hat{\varepsilon}/\hat{\varepsilon}_2$. If $\hat{\varepsilon}/\hat{\varepsilon}_2 = 2$ (e.g. $\hat{\varepsilon} = 0.4$, $\hat{\varepsilon}_2 = 0.2$), (9) is guaranteed. Once again an intuitive explanation emerges from the comparison of the optimal flat tax and Saez' asymptotic marginal tax rate. We know $\hat{t}^*$ depends on $\hat{\varepsilon}$. But Saez' revenue maximizing top tax rate will be higher given a lower $\hat{\varepsilon}_2$. Hence (9) is more likely to hold than (5) due to declining elasticity of labor supply.

(iii) *More tax bands* Finally, we consider the case of more than two tax bands. We assume t_1 only applies to incomes between Y and another lower threshold Y_0, below which different tax rates may apply. So t_1 is imposed on households with $w \geq w_0$ where $w_0^{1+\varepsilon}(1 - t_1)^\varepsilon = Y_0(w_0 < \bar{w}$ as $Y_0 < Y)$. Let $u(w)$ be the utility of households with $w \leq w_0$, not subject to either t_1 or t_2. Then the utility of the poor, (3) can be

---

[6] See https://stats.oecd.org/Index.aspx?DataSetCode=MIN2AVE.

[7] See http://www.tradingeconomics.com/germany/wage.

rewritten as:

$$W = \int_a^{w_0} u(w)f(w)dw + \int_{w_0}^{\bar{w}} \frac{(1-t_1)^{1+\varepsilon}}{1+\varepsilon} w^{1+\varepsilon} f(w)dw + BF(\bar{w}) \qquad (10)$$

Let $F(w_0)$ be the proportion of the households with $w \le w_0$. Note that the tax revenue from earnings below Y_0 is independent of t_1 and t_2. When $t_1 = t_2 = t$, with B_0 representing the part independent of t, we can write the basic income as:

$$B = B_0 + t(1-t)^\varepsilon \int_{w_0}^b w^{1+\varepsilon} f(w)dw - tY_0[1 - F(w_0)] \qquad (11)$$

The question is: whether a higher tax on income above $Y(t_2 > t)$ can lead to a higher value of (10) than any partial flat tax t on incomes above Y_0, given other tax rates below Y_0 fixed. To answer this question, we follow the same approach again as before. We first obtain the optimal partial flat tax $\breve{t}^*$ on incomes above Y_0. Then we find the condition for $\partial W/\partial t_1 < 0$ and $\partial W/\partial t_2 > 0$ when $t_1 = t_2 = \breve{t}^*$.

We let $y_0 \equiv w_0^{1+\varepsilon}$, and E_0 be the zero-tax earnings of households with $w \ge w_0$, i.e. $E_0 = \int_{w_0}^b w^{1+\varepsilon} f(w)dw$. Their average earnings $e_0 = E_0/[1 - F(w_0)]$. The flat tax $\breve{t}^*$ can be written as $(1-d)/(1+\varepsilon-d)$, where $d = y_0/e_0 + (E_0 - E_2)/F(\bar{w})E_0$. Thus we can generalize (5) to the case with more than two tax bands (see "Appendix E").

Proposition 7 *IMT can do better than any partial flat tax if at $t_1 = t_2 = \breve{t}^*$,*

$$\frac{y_0}{e_0} + \frac{E_0 - E_2}{F(\bar{w})E_0} > \frac{\bar{y}}{e_2} \qquad (12)$$

In our previous two-band tax case, $y_0 = 0$, $w_0 = a$, $E_0 = E$, (12) reduces to (5). Although (12) is more complex than (5), its validity may be determined with simple data. In particular (12) must hold when $y_0/e_0 \ge \bar{y}/e_2$. For instance, if earnings above y_0 follow a Pareto distribution with $y_0/e_0 = \bar{y}/e_2$, (12) must hold and a higher tax rate above $\bar{y}$ is desirable. Moreover, let $Y_0 = \$0.15$ million and $Y = \$0.4$ million, we have $y_0/e_0 = 0.5$ according to Saez (2001), and $\bar{y}/e_2 = 1/3$ according to Diamond and Saez (2011). Again (12) holds and the tax rate above \$0.4 million should be higher. These results again support DSPS' higher taxes for top earners.

6 Concluding remarks

In this paper we argue that a large poor majority are often better off under IMT than any flat tax. We obtain a sufficient condition, which only depends on aggregate features of the income distribution and the tax threshold. Using empirical data from G8 countries we find supporting evidence that a higher tax rate is justifiable when it is imposed on a small group (less than 20%). However, IMT become less likely to dominate any flat tax if we give more welfare weights to the very poor households. Similar to our

original condition (5), more general results are obtained with declining elasticity of labor supply and multiple tax bands. These findings support the argument of DSPS for higher taxes on top earners. It also has interesting political economy implications, and might perhaps be interpreted as an explanation for—or at least consistent with—IMT on high income earners in most democracies, in contrast to much optimal tax theory.

In this paper we do not consider categorical benefits associated with unemployment or low income. Those benefits create high marginal tax rates for participation in the labour market—the 'poverty trap'. This phenomenon, however, does not affect the larger part of the working population. We focus on the tax rates relevant to the working population and do not consider more complex structures. We do not focus on the optimal difference in tax rates and the magnitude of social gains. Both tend to be small in our model, but would be more significant given low marginal utility of income and low elasticity of labour supply for the rich. Though highly stylized, we hope that this paper contributes to the debate on tax policies.

Appendix A: Proof of Proposition 3

We show that $\partial W/\partial t_1 < 0$ and $\partial W/\partial t_2 > 0$ when $t_1 = t_2 = t^*$ if and only if (5) holds. From (4') we see $\partial B/\partial w_1 = t_2 f(w_1)Y - t_2(1-t_2)^\varepsilon f(w_1)w_1^{\varepsilon+1} = 0$ as $(1-t_2)^\varepsilon w_1^{1+\varepsilon} \equiv Y$. So we can differentiate B given w_1 fixed. As $t_1 = t_2, w_1 = \bar{w}, t_1[1 - F(\bar{w})] - t_2[1 - F(w_1)] = 0$, so we can ignore the change of Y when we differentiate (4') with respect to t_1 and t_2.

$$\frac{\partial B}{\partial t_1} = (1-t_1)^{\varepsilon-1}[1-(1+\varepsilon)t_1]\int_a^{\bar{w}} w^{1+\varepsilon}f(w)dw + [1-F(\bar{w})]Y$$

$$\frac{\partial B}{\partial t_2} = (1-t_2)^{\varepsilon-1}[1-(1+\varepsilon)t_2]\int_{w_1}^{b} w^{1+\varepsilon}f(w)dw - [1-F(w_1)]Y$$

Using our notations of E_1, E_2 and $\bar{y}$, they reduce to $[1 - \varepsilon t/(1-t)]E_1 + [1 - F(\bar{w})]\bar{y}$ and $[1 - \varepsilon t/(1-t)]E_2 - [1 - F(\bar{w})]\bar{y}$. Substituting them and $t_1 = t_2$, we find

$$\frac{\partial W}{\partial t_1} = (1-t)^\varepsilon\{F(\bar{w})\left[\left(1 - \frac{\varepsilon t}{1-t}\right)E_1 + (1 - F(\bar{w}))\bar{y}\right] - E_1\} \quad \text{(A1)}$$

$$\frac{\partial W}{\partial t_2} = (1-t)^\varepsilon F(\bar{w})\left\{\left(1 - \frac{\varepsilon t}{1-t}\right)E_2 - [1 - F(\bar{w})]\bar{y}\right\} \quad \text{(A2)}$$

As $t^* = (1 - e_1/E)/(1 + \varepsilon - e_1/E)$ and $E_1/F(\bar{w}) = e_1$, (A1) < 0 and (A2) > 0 if and only if $e_1 E_1/E + [1 - F(\bar{w})]\bar{y} - e_1 < 0$, and $e_1 E_2/E - [1 - F(\bar{w})]\bar{y} > 0$ respectively. Moreover as $e_1 E_1/E - e_1 = -e_1 E_2/E$, and $E_2/[1 - F(\bar{w})] = e_2$, both

inequalities hold if and only if $e_1 e_2 > E\bar{y}$, i.e. (5). If this holds, a two-bracket tax schedule with $t_1 < t^* < t_2$ dominates t^*. Since t^* is the optimal flat tax, this schedule must dominate any flat tax.

Appendix B: Proof of Proposition 4

The derivative of e_1/e_2 with respect to $\bar{w}$ is negative if $e_2 \frac{\partial e_1}{\partial \bar{w}} < e_1 \frac{\partial e_2}{\partial \bar{w}}$.

Note $e_1 = \frac{E_1}{F(\bar{w})}$, $e_2 = \frac{E_2}{1-F(\bar{w})}$, $\frac{\partial E_1}{\partial \bar{w}} = \bar{y} f(\bar{w}) = -\frac{\partial E_2}{\partial \bar{w}}$. So we obtain

$$\frac{\partial e_1}{\partial \bar{w}} = \frac{f(\bar{w})}{F(\bar{w})^2}[\bar{y} F(\bar{w}) - E_1]$$
$$= \frac{f(\bar{w})}{F(\bar{w})}(\bar{y} - e_1). \tag{B1}$$

$$\frac{\partial e_2}{\partial \bar{w}} = \frac{f(\bar{w})}{[1 - F(\bar{w})]^2}\{E_2 - \bar{y}[1 - F(\bar{w})]\}$$
$$= \frac{f(\bar{w})}{1 - F(\bar{w})}(e_2 - \bar{y}) \tag{B2}$$

So e_1/e_2 falls with $\bar{w}$ (or $\bar{y}$) if and only if $e_2(\bar{y} - e_1)/F(\bar{w}) < e_1(e_2 - \bar{y})/[1 - F(\bar{w})]$, i.e. $E_2(\bar{y} - e_1) < E_1(e_2 - \bar{y})$, or $E\bar{y} < E_1 e_2 + e_1 = e_1 e_2$, which is (5).

Appendix C: Proof of Proposition 5

Since (7) is similar to (3), $\partial W/\partial t_1$ is similar to (A1) and < 0 if and only if

$$\left[\left(1 - \frac{\varepsilon t}{1 - t}\right) E_1 + [1 - F(\bar{w})]\right] \bar{y} - \tilde{e}_1 < 0 \tag{C}$$

Substituting $\tilde{t}^* = (1 - \tilde{e}_1/E)/(1 + \varepsilon - \tilde{e}_1/E)$ into (C), we get $\tilde{e}_1 E_1/E + [1 - F(\bar{w})]\bar{y} < \tilde{e}_1$, or $[1 - F(\bar{w})]\bar{y} < \tilde{e}_1 E_2/E$, i.e. $E\bar{y} < \tilde{e}_1 e_2$. This also applies to $\partial W/\partial t_2 > 0$.

Appendix D: Proof of Proposition 6

Similar to Appendix A, except for varying ε, we find when $t_1 = t_2$,

$$\frac{\partial B}{\partial t_2} = \left(1 - \frac{\hat{\varepsilon}_2 t}{1 - t}\right)(1 - t)^\varepsilon E_2 - [1 - F(\bar{w})]Y \tag{D1}$$

When $t_1 = t_2 = \hat{t}^*$, (D1) is positive if and only if

$$\left[1 - \frac{\hat{\varepsilon}_2}{\hat{\varepsilon}}\left(1 - \frac{e_1}{E}\right)\right] E_2 > [1 - F(\bar{w})]\bar{y} \tag{D2}$$

Dividing (D2) by E_2, we get $1 - (1 - e_1/E)\hat{\varepsilon}_2/\hat{\varepsilon} > \bar{y}/e_2$. One can check that the same condition holds for $\partial W/\partial t_1 < 0$.

Appendix E: Proof of Proposition 7

Given (11) and $t_1 = t_2 = t$, we have $\partial B/\partial w_0 = 0$ as $w_0^{1+\varepsilon}(1-t)^\varepsilon = Y_0$. So we differentiate (11) given w_0 fixed, and find $\partial B/\partial t = [1 - \varepsilon t/(1-t)](1-t)^\varepsilon E_0 - Y_0[1 - F(w_0)]$. From (10) we see $\partial W/\partial w_0 = 0$ since $u(w_0)$ must be equal to $w_0^{1+\varepsilon}(1-t)^{\varepsilon+1}/(1+\varepsilon)$. So we differentiate (10) given $t_1 = t_2$ and w_0 fixed. Substitute $\partial B/\partial t$ into $\partial W/\partial t$, we get

$$\frac{\partial W}{\partial t} = F(\bar{w}) \left\{ \left(1 - \frac{\varepsilon t}{1-t} \right) (1-t)^\varepsilon E_0 - Y_0[1 - F(w_0)] \right\} \qquad \text{(E)}$$
$$- (E_0 - E_2)(1-t)^\varepsilon$$

The optimal partial flat tax can be solved from (E) $= 0$, as $\breve{t}^* = (1-d)/(1+\varepsilon-d)$, where $d = \{y_0[1 - F(w_0)] + (E_0 - E_2)/F(\bar{w})\}/E_0 = y_0/e_0 + (E_0 - E_2)/F(\bar{w})E_0$.

Then we check if $\partial W/\partial t_2 > 0$ by substituting $\breve{t}^*$ into (A2). This is equivalent to check if $\breve{t}^* < (1 - \bar{y}/e_2)/(1 + \varepsilon - \bar{y}/e_2)$, i.e., $d > \bar{y}/e_2$, which holds if and only if

$$\frac{y_0}{e_0} + \frac{E_0 - E_2}{F(\bar{w})E_0} > \frac{\bar{y}}{e_2}$$

One can check that the same condition holds for $\partial W/\partial t_1 < 0$.

References

Aaberge R, Colombino U (2013) Using a microeconometric model of household labour supply to design optimal income taxes. Scand J Econ 115:449–475

Andrienko Y, Apps P, Rees R (2014) Optimal taxation, inequality and top incomes. IZA DP No, p 8275

Apps P, Long NV, Rees R (2009) Optimal piecewise linear taxation, CESifo Working Paper No. 2565

Atkinson AB (1995) Public economics in action: the basic income/flat tax proposal. Clarendon Press, Oxford

Bach S, Corneo G, Steiner V (2012) Optimal top marginal tax rates under income splitting for couples. Eur Econ Rev 56:1055–1069

Boadway R, Cuff K, Marchand M (2000) Optimal Income Taxation with Quasi-Linear Preferences Revisited. J Public Econ Theory 2:435–460

Boadway R, Jacquet L (2008) Optimal marginal and average income taxation under maximin. J Econ Theory 143:425–441

Chone P, Laroque G (2005) Optimal incentives for labour force participation. J Public Econ 89(2–3):395–425

Diamond PA (1998) Optimal income taxation: an example with a U-shaped pattern of optimal marginal rates. Am Econ Rev 88(1):83–95

Diamond PA, Saez E (2011) The case for a progressive tax: from basic research to policy recommendations. J Econ Perspect 25(4):165–190

Gnedenko DV (1943) Sur la distribution limite du terme maximum d'une serie aleatoire. Ann Math 44:423–453

Hashimzade N, Myles G (2007) Structure of the optimal income taxation in the quasi-linear model. Int J Econ Theory 3:5–33

Hindricks J, Myles G (2006) Intermed Public Econ. MIT Press, London

Kanbur R, Tuomala M (1994) Inherent inequality and the optimal graduation of marginal tax rates. Scand J Econ 96:275–282

Kaplow L (2008) The theory of taxation and public economics. Princeton University Press, Princeton

Kleven HL (2014) How can Scandinavians tax so much? J Econ Perspect 28(4):77–98

Mankiw NG, Weinzierl M, Yergan D (2009) Optimal taxation in theory and practice. NBER Working Paper 15071

Piketty T, Saez E (2012) Optimal labor income taxation, NBER working paper 18521

Piketty T, Saez E (2013) Optimal labor income taxation. In: Auerbach AJ, Chetty R, Feldstein M and Saez E (eds) Handbook of public economics, vol 5. Elsevier, pp 391–474

Piketty T (2014) Capital in the twenty-first century. Harvard University Press, Cambridge

Sadka E (1976) On income distribution, incentive effects and optimal income taxation. Rev Econ Stud 43:261–268

Saez E (2001) Using elasticities to derive optimal income tax rates. Rev Econ Stud 68:205–229

Salanié B (2003) The economics of taxation. MIT Press, Cambridge

Seade J (1977) On the shape of optimal income schedules. J Public Econ 7:203–236

Seade J (1977) On the shape of optimal income schedules. J Public Econ 7:203–236

Slemrod J, Yitzhaki S, Mayshar J, Lundholm M (1994) The optimal two-bracket linear income tax. J Public Econ 53:269–290

Stiglitz J (2012) The price of inequality. W. W. Norton & Company, New York

Tarkiainen R, Tuomala M (2007) On optimal income taxation with heterogeneous work preference. Int J Econ Theory 3:35–46

Tuomala M (1984) On the optimal income taxation: some further numerical results. J Public Econ 23:351–366

World income inequality database V2.0c May (2008) United Nations University, UNU-WIDER at http://www.wider.unu.edu/research/Database/en_GB/database/