Higher Tax for Top Earners

Jim Jin and Felix FitzRoy

School of Economics and Finance Discussion Papers

School of Economics and Finance
Online Discussion Paper Series
issn 2055-303X
http://ideas.repec.org/s/san/wpecon.html
info: econ@st-andrews.ac.uk

School of Economics and Finance Discussion Paper No. 1702
16 Feb 2017 (revised 17 Feb 2017)
JEL Classification: D30, D60, H20
Keywords: flat tax, increasing marginal taxes, income redistribution

HIGHER TAX FOR TOP EARNERS

Felix FitzRoy*

Jim Jin**

School of Economics and Finance

University of St Andrews

JEL Classification Number: D30, D60, H20,

Key Words: flat tax, increasing marginal taxes, income redistribution,

February 2017

Abstract: The literature can justify increasing and decreasing marginal taxes (IMT & DMT) on top income under different social objectives and income distributions. Even if DMT are optimal, they are often politically infeasible. Then a flat tax seems to be a constrained optimal solution. We show however that, if we want to maximize the utility of a poor majority any flat tax can be inferior to some IMT. We provide a sufficient condition for (two-band) IMT to dominate any flat tax and further generalize this result to allow different welfare weights, declining elasticity of labor supply and more tax bands.

* School of Economics and Finance, University of St Andrews, St Andrews KY16 9AL, Scotland, UK, Tel. (01334)-462437, e-mail: frf@st-andrews.ac.uk.

** Corresponding author: School of Economics and Finance, University of St Andrews, St Andrews KY16 9AL, Scotland, UK, Tel. (01334)-462447, e-mail: jyj@st-andrews.ac.uk

1. Introduction

Throughout most developed as well as 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 are U-shaped, with increasing marginal tax (IMT) on high earnings (but not on capital gains). The justification of IMT 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 Salanie (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 earner. Following this line, the tax curve should be 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)). Then decreasing marginal taxes (DMT) on top income seem justifiable. 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 on DMT and seems to imply 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 of

IMT (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 the policy maker ignores the welfare of richest group (due to their low marginal utility of income) and focuses on the wellbeing of the poor majority, the marginal tax rate for top income should be 70-80%, substantially higher than its current level and the tax rates paid by the rest of the population. 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 the optimal DMT are not feasible, a flat tax may not be the next best alternative. If we want to maximize the utility of the poor, IMT on top income earners are superior to any flat tax under a simple condition, which means the optimal tax on top income derived by Saez (2001) is higher than the optimal flat tax. This condition generally holds if we want to maximize the utility of a large poor majority. It also depends on income distribution, but not necessarily on boundedness.

Following DSPS (though they consider more general cases), we ignore the interests of the rich group and only maximize the utility of 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, IMT are less likely to dominate a flat tax.

We first assume a constant elasticity of labour supply for the whole population. Later we make a more realistic assumption of declining elasticity with income and show that IMT are more likely to dominate any flat tax. 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.

Instead of a continuous tax structure, we first use a two-band tax model. The continuous tax curve has been criticized as “too far removed from the tax–benefit systems observed in practice to be a useful guide for policy” (Choné 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.” Diamond and Saez (2011) argue for practical and useful research on tax policy. The two-band tax model is the first step from a flat tax and can model IMT and DMT as well as a flat tax. Furthermore the two-band tax literature in particular supports DMT. Sheshinski (1989) first argued for rising two-band taxes under utilitarian and maximin objectives. However, Slemrod et al (1994) find errors in his proof and use numerical simulations to show that DMT are in fact optimal. Similarly Salanie (2003), Hindricks and Myles (2006) obtain optimal decreasing two-band taxes in a two-class economy. Hence it is interesting to see if two-band IMT can dominate any flat tax. We later allow more tax bands and generalize our result accordingly.

We introduce our basic 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 IMT to dominate any flat tax and shows it is valid if we want to maximize the total utility of a large poor majority. Section 5 extends our model and generalizes the results allowing different welfare weights for the poor, declining elasticity of labor supply and multi-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 those with higher wages are 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 and will allow different weights given to them in section 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 section 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 section 5(iii).

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

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

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

Every household chooses labour supply x to maximize utility. We first consider IMT, i.e. $t_1 < t_2$ and assume $Y \ge \bar{w}^{1+\varepsilon}(1 - t_1)^{\varepsilon}$. 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 $Y \ge \bar{w}^{1+\varepsilon}(1 - t_1)^{\varepsilon}$ 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 $\hat{w} < w \le w_1$ 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:

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

3. DMT vs. flat tax

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 further show that a flat tax is always Pareto dominated by some DMT.

We start with a flat tax $t > 0$, and consider to lower tax rate t_2 for earnings beyond $Y = (1-t)^{\varepsilon}[1 - \varepsilon t/(1-t)]b^{1+\varepsilon}$. As $(1-t)^{\varepsilon}b^{1+\varepsilon}$ is the highest 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. The former assumes a constant labor supply and can be expressed as $[(1-t_2)^{\varepsilon}w^{\varepsilon+1} - Y]\Delta t_2$, and the latter reflects the response of labor supply and is 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 *B*. Moreover the poorer households are better off too due to higher basic income *B*. Therefore a lower t_2 benefits everyone, including the poor.

Proposition 1: 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 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 the constrained optimal solution if it is closer to the optimum than and thus dominates IMT by continuity. However, this monotonicity of tax policy may not be valid. Assuming the government is politically constrained to implement two-band IMT and maximize the total utility of the poor, we will show that under a simple condition the optimal flat tax is dominated by some IMT. Before proving this result, we first obtain the optimal flat tax *t**, which maximizes our objective function (3').

To simplify the notation, we denote the total earnings of the poor households under a flat tax by $E_1 \equiv \int_a^{\overline{w}} (1-t)^{\varepsilon} w^{1+\varepsilon} f(w)dw$ and denote the earnings of the rich by $E_2 \equiv \int_{\overline{w}}^{b} (1-t)^{\varepsilon} w^{1+\varepsilon} f(w)dw$. The total earnings of the whole population is $E = E_1 + E_2$. Since the population is normalized to 1, *E* is also the average earnings of the whole population.

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 $dW/dt = [1 - t(1 + \varepsilon)]F(\bar{w})E/(1 - t) - E_1$. It is positive if and only if $t < (1 - e_1/E)/(1 + \varepsilon - e_1/E]$. So we obtain the optimal flat tax t^*.

Proposition 2: The optimal flat tax to maximize (3') *is* $t^* = \dfrac{1 - e_1/E}{1 + \varepsilon - e_1/E}$.

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 is equal to the median income, our t^* becomes the median voter tax. Interestingly, the median income in the U.S. is roughly $26,000. Piketty and Saez estimate the average top 1% income as $1.2 million. Given the average earnings of $38,000, the average of the bottom 99%, e_1 is also about $26,000. Thus our flat tax for the 99% majority is equal to the median voter tax rate.

4. IMT vs. 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 find a sufficient condition for IMT to dominate t^*, this particular Y serves the purpose. $Y = \bar{w}^{1+\varepsilon}(1 - t_1)^{\varepsilon}$ implies $\hat{w} = \bar{w}$. The tax revenue (4) (hence B) simplifies to:

$$B = \int_a^{\bar{w}} t_1(1-t_1)^{\varepsilon} w^{1+\varepsilon} f(w)dw + \int_{w_1}^{b} t_2(1-t_2)^{\varepsilon} w^{1+\varepsilon} f(w)dw$$

$$+ \{t_1[1 - F(\bar{w})] - t_2[1 - F(w_1)]\}Y \qquad (4')$$

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').

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

$$\frac{e_1}{E} > \frac{Y}{e_2} \qquad (5)$$

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 $\partial W/\partial t_2 > 0$, it would be feasible 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').

Intuitively (5) can again be explained by Saez' (2001) concept of behavioural and mechanical responses as in Proposition 1. The difference is that here some households' tax payment increases with t_2, while others' may fall with t_2. Given $\Delta t_2 > 0$ at $t_1 = t_2 = t^*$, the mechanical effect on (4') is $\{\int_{\bar{w}}^{b}(1-t^*)^{\varepsilon} w^{1+\varepsilon} f(w)dw - [1-F(\bar{w})]Y\}\Delta t_2$, 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, i.e. $[1 - \varepsilon t^*/(1 - t^*)]\, e_2 > Y$, as shown by (A2) in Appendix A, the overall mechanical effect dominates the behavioral effect, and the total tax payment rises with t_2, i.e., $\partial B/\partial t_2 > 0$. Here the tax revenue rises, similar to the case of Proposition 1, due to a rise in t_2 instead of its fall. Since $1 - \varepsilon t^*/(1 - t^*)$ is equal to e_1/E, the condition reduces to (5), which guarantees the optimal flat tax t^* to be dominated by some IMT[1].

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 $(1 - g)/[1 - g + \varepsilon e_2/(e_2 - Y)]$, where g is the social welfare weight given to the rich (also see Piketty and Saez (2013)). In our model $g = 0$ since no welfare weight is given to the rich. Thus Saez' optimal top income tax rate

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

becomes $(e_2 - Y)/[(1+\varepsilon)e_2 - Y]$. 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 $[1 - \varepsilon t^a/(1-t^a)]e_2 = Y$. Hence $t^a > t^*$ if and only if (5) holds[2]. Otherwise Saez' marginal tax for top income will not be consistent with IMT.

Since (5) is evaluated under a flat tax, term $(1 - t^*)^\varepsilon$ drops out and the inequality only depends on the wage distribution, not on taxes. To evaluate (5), it is often convenient to consider the income distribution function $G(y)$, with $y = w^{1+\varepsilon}$, instead of wage distribution $F(w)$. We can calculate e_1, e_2 and E with zero taxes, and let Y be $\bar{y} \equiv \bar{w}^{1+\varepsilon}$. For some distributions the validity of (5) does not depends 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 $\bar{y}$[3]. This result is consistent with Diamond (1998). The reason may appear to be the Pareto distribution's thick-tail of top earnings. However, when α is large, the tail becomes very thin while (5) still holds. To illustrate 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 of rich households may not fall but even rise with income (if $\beta > 1$). But (5) never holds[4]. The validity of (5) may appear to require an unbounded income, but this is not necessarily true either. Let us 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 (5) is very sensitive to distributions.

---

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

[3] 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 $e_1 > 1$.

[4] 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 $e_2 > h$.

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 threshold earnings of the top 1% U.S. earners as \$0.4 million and their average earnings as \$1.2 million. This implies $Y/e_2 = 1/3$, 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 1% earners in the US, 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. $Y/e_2 = 0.5$). 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 Y 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. Y = the median earnings, (5) does not hold for any lognormal distribution. Therefore it seems undesirable to impose a higher tax rate on the rich half of the population on behalf of the poor half.

The question is to find a reasonable size of the rich group for justifiable IMT. 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 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* Y*.*

If the curve of e_1/e_2 is single peaked, after its maximum, the ratio will fall and IMT will dominate any flat tax. When earnings are unbounded and Y is sufficiently large, e_1 will approach to E but e_2 will go to infinity. So e_1/e_2 must fall and (5) holds eventually. In this case unbounded earnings are likely to justify IMT. The question is: how large Y is "sufficient". The answer may be difficult to obtain by theory alone and empirical data may reveal how large a rich group should be subject to a higher tax.

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 when we use the actual earnings ratios to justify IMT, it is an approximation, not accurate prediction. 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% to 90%, giving us 9 values of e_1/e_2. The results for G8 countries are given below.

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

Apparently, the ratio differs significantly between eight nations. However, e_1/e_2 exhibits a single peak in all G8 countries and surprisingly, it starts 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.

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. Assigning decreasing weight to income has a similar effect as assuming decreasing marginal utility of income. Intuitively, one may expect that this should increase the chance of justifying IMT. This expectation is similar to the conventional belief that IMT are mostly justifiable under maximin, and is not correct.

Given $\overline{w}$ we assign welfare weight $s(w)$ to every poor household w ($\leq \overline{w}$) such that $\int_a^{\overline{w}} s(w) f(w) dw = F(\overline{w})$. 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 $[a, \overline{w}]$, to get a weighted utility of the poor as our new objective function:

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

(7) reduces to (3) when $s(w) = 1$. Since $s(w)$ generally falls with w, we have $\int_a^{\overline{w}} s(w) w^{1+\varepsilon} f(w) dw < \int_a^{\overline{w}} w^{1+\varepsilon} f(w) dw$. We use $\tilde{e}_1$ to denote the weighted average earnings of the poor, $(1 - t)^{\varepsilon} \int_a^{\overline{w}} s(w) w^{1+\varepsilon} f(w) dw / F(\overline{z})$. The more weight is given to the poorer households the lower $\tilde{e}_1$ is. Similar to 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 IMT dominate any flat tax if $\partial W/\partial t_1 < 0$ and $\partial W/\partial t_2 > 0$ when $t_1 = t_2 = \tilde{t}\,*$. Obtaining such a condition we can generalize the previous condition (5) for IMT to dominate any flat tax (see Appendix C).

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

$$\frac{\tilde{e}_1}{E} > \frac{Y}{e_2} \tag{8}$$

When $s(w) = 1$, $\tilde{e}_1 = e_1$ and (8) reduces to (5). So condition (8) is a generalization of (5), and 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 since it must maximize the tax revenue, but the optimal flat tax 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, thus IMT are 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 low income earners and more beneficial to them due to more tax revenue and money transfer from the rich. So a flat tax is more likely to dominate IMT if we give most weight to the poorest.

The validity of (5) only implies IMT can benefit the poor as a whole, not necessarily each of them. (8) can tell us if IMT benefit a particular household. Our objective to maximize (7) is identical to maximizing the utility of a household with earnings of $\tilde{e}_1$. This is consistent with the political agenda to help a representative family in the society. When (8) holds, IMT can benefit those with earnings higher than $\tilde{e}_1$. If it holds when $\tilde{e}_1$ is equal to the lowest earnings, IMT can benefit all poor. For instance, given a Pareto distribution with $\alpha = 1.5$ for the top 1% earners, $Y/e_2 = 1/3$, 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[5]. So a higher tax on top 1% can benefit all 99%. Similarly, with α = 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[6]. Thus virtually all poor can benefit from a higher tax on the rich.

(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 constant elasticity of labor supply is unrealistic. In fact this assumption is unfavorable for justifying 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 to 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 $t_1 = t_2 = \hat{t}*$, we know 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*

[5] See https://stats.oecd.org/Index.aspx?DataSetCode=MIN2AVE.
[6] See http://www.tradingeconomics.com/germany/wage.

$$1-\frac{e_1}{E} < (1-\frac{Y}{e_2})\frac{\hat{\varepsilon}}{\hat{\varepsilon}_2} \qquad (9)$$

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 IMT are more likely to dominate any flat tax, as expected. Given a Pareto distribution with $\alpha = 2$ for top income, $Y/e_2 = 0.5$, 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. Given $\hat{\varepsilon}_2$ we have Saez' revenue maximizing top income tax rate. But if the average elasticity $\hat{\varepsilon}$ is higher, the flat tax $\hat{t}^*$ must be lower. Hence the former is more likely to be higher than the latter, and (9) is more likely to hold than (5).

(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 threshold Y_0, below which different tax rates may apply. So t_1 is only imposed on households with $w \geq w_0$, where $w_0^{\varepsilon+1}(1 - t_1)^{\varepsilon} = Y_0$. Let $u(w)$ be the utility of those households with $w \leq w_0$ ($< \overline{w}$), not subject to either t_1 or t_2. Then the utility of the poor, (3) can be rewritten as:

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

Let B_0 be the basic income transferred from earnings below Y_0, and $F(w_0)$ be the households with $w \leq w_0$. They are independent of t_1 and t_2. When $t_1 = t_2 = t$, we have:

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

The question is: whether a higher tax on income above Y ($t_2 > t_1$) can lead to a higher value of (10) than any partial flat tax t on income 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 income 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 E_0 denote the earnings of households with $w \geq w_0$ under a flat tax t, i.e., $E_0 = (1-t)^{\varepsilon}\int_{w_0}^{b} w^{1+\varepsilon} f(w)dw$. Their average earnings $e_0 = E_0/[1-F(w_0)]$. The optimal tax $\breve{t}$ * can be written as $(1-d)/(1+\varepsilon-d)$, where $d = Y_0/e_0 + (E_0 - E_2)/F(\overline{w})E_0$. 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 $t_1 = t_2 = \breve{t}$ *,

$$\frac{Y_0}{e_0} + \frac{E_0 - E_2}{F(\overline{w})E_0} > \frac{Y}{e_2} \tag{12}$$

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 $Y_0/e_0 \geq Y/e_2$. For instance, if earnings above Y_0 follow a Pareto distribution with $Y_0/e_0 = Y/e_2$, (12) must hold and a higher tax rate above 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 $Y/e_2 = 1/3$ 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 IMT are often better than any flat tax if we want to maximize the total utility of a large poor majority. We obtain a sufficient condition for IMT to dominate any flat tax, 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 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 multi-band taxes. 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 level of 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 $t_1 = t_2$, $w_1 = \overline{w}$, $t_1[1 - F(\overline{w})] - t_2[1 - F(w_1)] = 0$, so we can ignore the change of Y when we differentiate (4') with respect to t_1 and t_2.

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

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

Using E_1 and E_2, they reduce to $[1 - \varepsilon t/(1-t)]\, E_1 + [1 - F(\overline{w})]Y$ and $[1 - \varepsilon t/(1-t)]\, E_2 - [1 - F(\overline{w})]Y$. Substituting them into $\partial W/\partial t_1$ and $\partial W/\partial t_2$ given $t_1 = t_2$, we find

$$\frac{\partial W}{\partial t_1} = F(\overline{w})\{[(1-\frac{\varepsilon t}{1-t})E_1 + [1-F(\overline{w})]Y\} - E_1 \qquad \text{(A1)}$$

$$\frac{\partial W}{\partial t_2} = F(\overline{w})\{[(1-\frac{\varepsilon t}{1-t})E_2 - [1-F(\overline{w})]Y\} \qquad \text{(A2)}$$

As $t^* = (1 - e_1/E)/(1 + \varepsilon - e_1/E)$ and $E_1/F(\overline{w}) = e_1$, (A1) < 0 and (A2) > 0 if and only if $e_1 E_1/E + [1 - F(\overline{w})]Y - e_1 < 0$, and $e_1 E_2/E - [1 - F(\overline{w})]Y > 0$. Moreover as $e_1 E_1/E - e_1 = -e_1 E_2/E$, and $E_2/[1 - F(\overline{w})] = e_2$, both inequalities become $e_1 e_2 > EY$, i.e. (5).

When (5) holds, there exist a two-bracket tax schedule with $t_1 < t^* < t_2$, dominating t^*. Since t^* is the optimal linear tax rate, the two-bracket schedule dominates any flat tax.

Appendix B, Proof of Proposition 4:

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

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

$$\frac{\partial e_1}{\partial \overline{w}} = \frac{f(\overline{w})}{F(\overline{w})^2}[YF(\overline{w}) - E_1] = \frac{f(\overline{w})}{F(\overline{w})}(Y - e_1). \qquad \text{(B1)}$$

$$\frac{\partial e_2}{\partial \overline{w}} = \frac{f(\overline{w})}{[1-F(\overline{w})]^2}\{E_2 - Y[1-F(\overline{w})]\} = \frac{f(\overline{w})}{1-F(\overline{w})}(e_2 - Y) \qquad \text{(B2)}$$

Hence e_1/e_2 falls with $\overline{w}$ if and only if $e_2(Y - e_1)/F(\overline{w}) < e_1(e_2 - Y)/[1 - F(\overline{w})]$, i.e. $E_2(Y - e_1) < E_1(e_2 - Y)$, or $EY < E_1 e_2 + E_2 e_1 = e_1 e_2$, 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

$$[(1-\frac{\varepsilon t}{1-t})E_1 + [1-F(\overline{w})]Y - \tilde{e}_1 < 0 \qquad \text{(C)}$$

Substituting $\tilde{t}\,^* = (1 - \tilde{e}_1/E)/(1 + \varepsilon - \tilde{e}_1/E)$ into (C), we get $\tilde{e}_1 E_1/E + [1 - F(\overline{w})]Y < \tilde{e}_1$, or $[1 - F(\overline{w})]Y < \tilde{e}_1 E_2/E$, i.e. $EY < \tilde{e}_1 e_2$. 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 $t_1 = t_2$,

$$\frac{\partial B}{\partial t_2} = (1-\frac{\hat{\varepsilon}_2 t}{1-t})E_2 - [1-F(\overline{w})]Y \qquad \text{(D1)}$$

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

$$[1 - \frac{\hat{\varepsilon}_2}{\hat{\varepsilon}}(1-\frac{e_1}{E})]E_2 > [1-F(\overline{w})]Y \qquad \text{(D2)}$$

Dividing (D2) by E_2, we get $1 - (1 - e_1/E)\hat{\varepsilon}_2/\hat{\varepsilon} > Y/e_2$. 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 $w_0^{\varepsilon+1}(1-t)^{\varepsilon} = Y_0$. So we differentiate (11) given w_0 fixed, and find $\partial B/\partial t = [1 - \varepsilon t/(1-t)]E_0 - Y_0[1 - F(w_0)]$. From (10) we see $\partial W/\partial w_0 = 0$ since $u(w_0)$ must be equal to $[w_0(1-t)]^{\varepsilon+1}/(1+\varepsilon)$. 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

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

The optimal partial flat tax can be solved from (E) = 0, as $\breve{t}\,^* = (1-d)/(1+\varepsilon-d)$, where $d = \{Y_0[1-F(w_0)] + (E_0 - E_2)/F(\overline{w})\}/E_0 = Y_0/e_0 + (E_0 - E_2)/F(\overline{w})E_0$.

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

$$\frac{Y_0}{e_0} + \frac{E_0 - E_2}{F(\overline{w})E_0} > \frac{Y}{e_2}$$

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

References

Aaberge, R. and U. Colombino, 2013, Using a Microeconometric Model of Household Labour Supply to Design Optimal Income Taxes, *Scandinavian Journal of Economics*, vol. 115, pp. 449-475

Andrienko, Y., P. Apps and R. Rees, 2014, Optimal Taxation, Inequality and Top Incomes, IZA DP No. 8275

Apps, P., N. V. Long, and R. Rees, Optimal Piecewise Linear Taxation, CESifo *Working Paper No. 2565*, 2009

Atkinson, A. B., *Public Economics in Action: The Basic Income/Flat Tax Proposal*, Clarendon Press, Oxford, 1995

Bach, S., G. Corneo and V. Steiner, 2012, Optimal Top Marginal Tax Rates under Income Splitting for Couples, *European Economic Review* 56, pp. 1055–1069

Boadway, R, K. Cuff, and M. Marchand, 2000, Optimal Income Taxation with Quasi-Linear Preferences Revisited, *Journal of Public Economic Theory*, Vol. 2, pp. 435–60

Boadway, R, and L. Jacquet, 2008, Optimal Marginal and Average Income Taxation under Maximin, *Journal of Economic Theory*, vol. 143, pp. 425–441

Chone, P., and G. Laroque, 2005, Optimal Incentives for Labour Force Participation, *Journal of Public Economics*, vol. 89(2-3), pp. 395–425

Diamond, P. A., 1998, Optimal Income Taxation: An Example with a U – shaped Pattern of Optimal Marginal Rates, *American Economic Review*, vol. 88(1), pp. 83–95

Diamond, P. A., and E. Saez, 2011, the Case for a Progressive Tax: From Basic Research to Policy Recommendations, *The Journal of Economic Perspectives*, Vol. 25, No. 4 (Fall 2011), pp. 165-190

Gnedenko, D. V., 1943, Sur la distribution limite du terme maximum d'une serie aleatoire", *Annals of Mathematics* vol. 44: pp. 423-453

Hashimzade, N. and G. Myles, 2007, Structure of the Optimal Income Taxation in the Quasi-Linear Model, *International Journal of Economic Theory*, vol. 3, pp. 5–33

Hindricks, J. and G. Myles, 2006, *Intermediate Public Economics*, MIT Press, London,

Kanbur, R. and M. Tuomala, 1994, Inherent Inequality and the Optimal Graduation of Marginal Tax Rates, *Scandinavian Journal of Economics*, vol. 96, pp. 275–82

Kaplow, L., *The Theory of Taxation and Public Economics*, Princeton University Press, 2008

Kleven, H. L., 2014, "How can Scandinavians tax so much?", Journal *of Economic Perspectives*, 28, 4, pp. 77 - 98

Mankiw, N. G., M. Weinzierl and D. Yergan, 2009, "Optimal Taxation in Theory and Practice", NBER Working Paper 15071

Piketty, T. and E. Saez, 2012, Optimal Labor Income Taxation, NBER working paper 18521.

Piketty, T. and E. Saez, 2013, Optimal Labor Income Taxation, *Handbook of Public Economics*, Chapter 7, vol. 5, pp. 391-474.

Piketty, T., *Capital in the Twenty-First Century*, Harvard University Press 2014

Sadka, E., 1976, On Income Distribution, Incentive Effects and Optimal Income Taxation, *Review of Economic Studies*, vol. 43, pp. 261–8

Saez, E., 2001, Using Elasticities to Derive Optimal Income Tax Rates, *Review of Economic Studies*, vol. 68, pp.205–29

Salanié, B., *the Economics of Taxation*, MIT Press, 2003

Seade, J., 1977, On the Shape of Optimal Income Schedules, *Journal of Public Economics*, vol. 7, pp. 203–36

Sheshinski, E., 1989, Note on the Shape of the Optimum Income Tax Schedule, *Journal of Public Economics*, vol. 40, pp. 201–15

Slemrod, J., S.Yitzhaki, J. Mayshar, and M. Lundholm, 1994, The Optimal Two–Bracket Linear Income Tax, *Journal of Public Economics*, vol. 53, pp. 269–90

Stiglitz, J., *The Price of Inequality*, W. W. Norton & Company, 2012

Tarkiainen, R. and M. Tuomala, 2007, On Optimal Income Taxation with Heterogeneous Work Preference, *International Journal of Economic Theory*, vol. 3, pp. 35–46

Tuomala, M., 1984, On the Optimal Income Taxation: Some Further Numerical Results, *Journal of Public Economics*, vol. 23, pp. 351–66

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

***Appendix F* (*not for publication*), (5) *holds for any bounded Pareto distribution*:**

From $G(y)$, we obtain $1 - G(y) = (y^{-\alpha} - h^{-\alpha})/(1 - h^{-\alpha})$, and $g(y) = \alpha y^{-\alpha-1}/(1 - h^{-\alpha})$.

Then we have $E_1 = \frac{\alpha(1-y^{1-\alpha})}{(\alpha-1)(1-h^{-\alpha})}$, $E = \frac{\alpha(1-h^{1-\alpha})}{(\alpha-1)(1-h^{-\alpha})}$, and $E_2 = \frac{\alpha(y^{1-\alpha}-h^{1-\alpha})}{(\alpha-1)(1-h^{-\alpha})}$.

So $e_1 = \frac{\alpha(1-y^{1-\alpha})}{(\alpha-1)(1-y^{-\alpha})}$, $e_2 = \frac{\alpha(y^{1-\alpha}-h^{1-\alpha})}{(\alpha-1)(y^{-\alpha}-h^{-\alpha})}$, and (5) becomes:

$$\frac{(1-y^{1-\alpha})}{(1-y^{-\alpha})}\frac{(1-h^{-\alpha})}{(1-h^{1-\alpha})} \geq \frac{(\alpha-1)y(y^{-\alpha}-h^{-\alpha})}{\alpha(y^{1-\alpha}-h^{1-\alpha})} \qquad \text{(F1)}$$

When h = its minimum y, the L'Hôpital's rule implies the equality of (F1). So (5) always holds if $L = \frac{(1-h^{-\alpha})}{(1-h^{1-\alpha})}\frac{(y^{1-\alpha}-h^{1-\alpha})}{(y^{-\alpha}-h^{-\alpha})}$ rises with h, which must be true if:

$$\frac{\partial \ln(L)}{\partial h} = \frac{\alpha h^{-\alpha-1}}{1-h^{-\alpha}} - \frac{(\alpha-1)h^{-\alpha}}{1-h^{1-\alpha}} + \frac{(\alpha-1)h^{-\alpha}}{y^{1-\alpha}-h^{1-\alpha}} - \frac{\alpha h^{-\alpha-1}}{y^{-\alpha}-h^{-\alpha}} \geq 0 \qquad \text{(F2)}$$

When y = its minimum 1, (F2) = 0 for any h. So we just need to show (F2) increasing with y. This is true if $\frac{(\alpha-1)h^{-\alpha}}{y^{1-\alpha}-h^{1-\alpha}} - \frac{\alpha h^{-\alpha-1}}{y^{-\alpha}-h^{-\alpha}}$ increases with y. Let $s \equiv h/y \geq 1$, this holds if $L_1 \equiv \frac{\alpha-1}{s^{\alpha-1}-1} - \frac{\alpha}{s^{\alpha}-1}$ falls with s, i.e. $\frac{\partial L_1}{\partial s} \leq 0$, or $\frac{(\alpha-1)^2 s^{\alpha-2}}{(s^{\alpha-1}-1)^2} > \frac{\alpha^2 s^{\alpha-1}}{(s^{\alpha}-1)^2}$, i.e.

$$L_2 = (\alpha-1)(s^{\alpha}-1) - \alpha s^{1/2}(s^{\alpha-1}-1) \geq 0 \qquad \text{(F3)}$$

When $s = 1$, $L_2 = 0$. So it suffices to show L_2 increases with s, i.e.

$$\frac{\partial L_2}{\partial s} = \alpha(\alpha-1)s^{\alpha-1} - \alpha(\alpha-1)s^{\alpha-3/2} - 0.5\alpha s^{-1/2}(s^{\alpha-1}-1) \geq 0 \qquad \text{(F4)}$$

This is true if $L_3 = (\alpha-1)(s^{\alpha-0.5} - s^{\alpha-1}) - 0.5(s^{\alpha-1} - 1) = (\alpha-1)s^{\alpha-0.5} - (\alpha-0.5)s^{\alpha-1} + 0.5 \geq 0$. When $s = 1$, $L_3 = 0$. So $L_3 \geq 0$ if $\partial L_3/\partial s = (\alpha-1)(\alpha-0.5)s^{\alpha-2}(\sqrt{s} - 1) \geq 0$, which is guaranteed. Hence (5) must hold for any $\alpha > 1$, $h \geq y \geq 1$.