DISCUSSION PAPER SERIES

Are Two Tax Rates Better than One?

Felix FitzRoy
Jim Jin

May 2009

I Z A

Are Two Tax Rates Better than One?

Felix FitzRoy
*University of St Andrews
and IZA*

Jim Jin
University of St Andrews

ABSTRACT

Are Two Tax Rates Better than One?[*]

Should two-band income taxes be progressive given a general income distribution? We provide a negative answer under utilitarian and max-min welfare functions. While this result clarifies some ambiguities in the literature, it does not rule out progressive taxes in general. If we maximize total or weighted utility of the poor, as often intended by the society, progressive taxes can be justified, especially when the 'rich' are very rich. Under these objectives we obtain necessary conditions for progressive taxes, which only depend on aggregate features of income distributions. The validity of these conditions is examined using plausible income distributions.

Corresponding author:

Felix FitzRoy
School of Economics and Finance
University of St Andrews
St Andrews KY16 9AL
Scotland
United Kingdom
E-mail: frf@st-andrews.ac.uk

1. Introduction

Income taxes are usually progressive in developed countries. The justification is to benefit the poor by income redistribution from the rich, perhaps at the cost of reduced efficiency. The idea slipped down the political agenda in the U.S. during Bush's years, while it is gaining momentum in the new Obama administration. An alternative to the progressive tax system is a flat tax combined with a basic income, which reduces transaction costs and incentive complications. It has gained increasing support from various parts of the political spectrum. Several Eastern European countries have moved in this direction. A populist proposal has been made in the US to replace all welfare payments by a $10,000 p.a. basic income, combined with a flat tax (Murray 2006). Though generally considered sub-optimal, flat taxes have attracted serious interests in both politics and academia. Atkinson (1995) provides a comprehensive treatment of this issue.

Plausibly, administrative costs of multi-band tax systems increase with the number of bands, while the marginal benefit decreases. The advantage of progressive taxes over flat tax, if exists, can be best demonstrated by two-band taxes. In fact, many countries have slimmed down multiple tax bands to mainly two, such as in U.K. The literature has examined the two-band tax structure to justify progressive taxes or question their desirability (Sheshinski [1989], Slemrod et al [1994], and Kaplow [2008]). If we do not find the optimal tax rate for high income significantly higher than the rate for low income, the justification for progressive taxes becomes dubious. In this paper we investigate whether this is usually, or under which conditions, true.

Starting from Mirlees (1971), economists have made serious efforts to understand the optimal tax structure. Following Mirlees' model of general non-linear

taxes but finite income, Sadka (1976) and Seade (1977) show that the marginal tax rates for the highest and lowest income earners should be zero. But little can be said about the rest of the population without further specifying income distributions, utility functions and welfare objectives. These specifications make significant differences to the basic pattern of the optimal tax structure (see Tuomala [1984, 1990], Kanbur and Tuomala [1994], Dahan and Strawczynski (2000), Saez [2001], Tarkiainen and Tuomala [2007]). The differences are often too large for policy makers to follow one recommendation or the other. For instance, Diamond (1998) and Salanie (2003) obtain U–shaped tax curves, and Hindricks et al (2006) find increasing optimal average tax rates. But Boadway et al (2000), Hashimzade and Myles (2007) show that inverted U-shaped tax curves are optimal. These conflicting results do not tell us whether more realistic two-band taxes should be progressive or not.

Sheshinski (1989) is the first to examine two-band taxes and, as an exception in the literature, assumes general utility functions and income distributions. He finds that the optimal taxes cannot be regressive under utilitarian and max-min welfare functions. This is rather surprising given the conflicting results on non-linear taxes. Indeed, Slemrod et al (1994) point out that Sheshinski's proof for the utilitarian welfare case is flawed, because it ignores a possible discrete jump in the tax revenue from marginal households whose labour supplies are not unique. They further demonstrate that, to maximize weighted utility of two groups with different income levels, the tax rates can be either progressive or regressive. Moreover, using a CES utility function and a lognormal wage distribution, their numerical simulation shows that the optimal taxes are regressive under the utilitarian welfare and a rather egalitarian objective with heavy emphasis on the wellbeing of the poor. Salanie

(2003), Hindricks and Myles (2006) also obtain regressive taxes in a two-group economy, supporting the findings of Slemrod et al.

While these studies tell us that regressive taxes cannot be dismissed, certain questions remain unanswered. First, we do not know how likely the optimal two-band taxes are progressive or regressive. As Sheshinski (1989) assumes a general income distribution, the discrete jump in tax revenue might be insignificant for some or possibly most distributions. Should taxes be progressive in these cases? If so, the examples of Slemrod et al may be exceptional rather than general.

Secondly, the existing research has focused on utilitarian and max-min welfare functions. The former does not favor income redistribution and is unlikely to justify progressive taxes. The latter is often viewed as the most likely to yield progressive taxes. However, as demonstrated by Sheshinski's example, it needs not yield more progressive taxes than the utilitarian welfare. Then these two welfare functions may not suffice to dismiss progressive taxes, and the question should be examined under alternative welfare functions. In the real world the purpose of progressive taxes is not to benefit everyone equally or a few worst-off individuals exclusively, but to help a large number of poor people. The desired income redistribution is supposedly to benefit these low-income earners at the expense of the high-income earners. The literature has not explicitly investigated this type of welfare functions.

Thirdly, the literature has not clarified how the optimal tax structure depends on income distributions, which differ greatly across countries and may change dramatically as recently in the UK and US. Conclusions based on specific distributions are not reliable. We need simple criteria to evaluate the desirability of progressive taxes based on reasonable information about income distributions.

In the present paper we try to answer these questions. We first consider utilitarian and max-min welfare functions. In contrast to Sheshinski's claim, we show that the optimal taxes are generally not progressive. Then, we examine the optimal taxes under new welfare functions, the total and weighted utility of the poor. We find necessary conditions for progressive taxes, which only depend on aggregate and easily observable features of income distributions. We further discuss the validity of these conditions using plausible income distributions.

The plan of the paper is to develop the model in the next section 2. Sections 3 and 4 show optimal taxes are generally not progressive under utilitarian and max-min welfare functions. In sections 5 and 6 we derive necessary conditions for progressive taxes to maximize total and weighted utility of the poor. The last section concludes.

2. The 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 $a \geq 0$, b is either finite or infinite. The cumulative function is denoted by $G(w)$. A household's pre-tax earnings y are proportional to its wage w and labour supply x, i.e., $y = wx$.

The government observes earnings, and imposes two tax rates accordingly. We let t_1 be the marginal tax rate applicable to earnings up to a threshold $\bar{y}$. For earnings exceeding $\bar{y}$, the tax rate is t_2. If a household does not earn more than $\bar{y}$, its after-tax earnings are $wx(1 - t_1)$. Otherwise the after-tax earnings are $wx(1 - t_2) + \bar{y}(t_2 - t_1)$. We let $\bar{w}$ be the lowest wage of the households who pay t_2.

Besides wages, every household receives a basic income, denoted by B. Given our unit population B represents the basic income for each household as well as for

4

the total population. The utility function for every household is quasi-linear in income. Given t_1, t_2 and $\bar{y}$, the utility functions are assumed to be:

$$V_1 = wx(1 - t_1) - \frac{x^{1+1/\varepsilon}}{1+1/\varepsilon} + B \qquad\qquad \text{for } w \leq \bar{w} \qquad (1)$$

$$V_2 = wx(1 - t_2) + \bar{y}(t_2 - t_1) - \frac{x^{1+1/\varepsilon}}{1+1/\varepsilon} + B \qquad\qquad \text{for } w > \bar{w} \qquad (2)$$

Parameter ε represents the elasticity of labor supply. We assume $0 < \varepsilon < 1$, implying inelastic supply, which is consistent with empirical observations. This type of utility function has been used in the literature (e.g. Atkinson, 1995).

All tax revenue is spent on B after a fixed expenditure on public goods, P. Throughout the paper we assume that P is sufficiently low so that B can be positive if this is desirable. For simplicity we ignore P as it does not affect our solutions. Then we can write the basic income B as equal to the tax revenue:

$$B = t_1 \int_a^{\bar{w}} wx\, dG(w) + t_2 \int_{\bar{w}}^b wx\, dG(w) - (t_2 - t_1)\,\bar{y}\,[1 - G(\bar{w})] \qquad (3)$$

If there is no tax, every household's labor supply can be solved from the first-order condition $w - x^{1/\varepsilon} = 0$, as $x = w^\varepsilon$. This implies optimal-no-tax-earnings (ONTE) $w^{\varepsilon+1}$. We denote $w^{\varepsilon+1}$ by z. To simplify notations, we focus on the distribution of z, instead of w. The cumulative and density functions of z are $F(z)$ and $f(z)$, which can be derived from the wage distribution. The minimum and maximum ONTE, $a^{1+\varepsilon}$ and $b^{1+\varepsilon}$, are denoted by m and M respectively. For a household with wage $\bar{w}$, its ONTE is denoted by $\bar{z}$. Obviously $F(\bar{z}) = G(\bar{w})$.

In the following sections, we examine under various welfare functions if or under which conditions the optimal taxes are progressive, i.e., $t_2^* > t_1^*$.

3. Utilitarian Welfare

Following the literature, we choose tax rates t_1, t_2 and threshold earnings $\bar{y}$ to maximize utilitarian welfare. Sheshinski argues that if $t_1 > t_2$, there is no interior solution for $\bar{y}$, hence the optimal taxes cannot be regressive. His proof is flawed, as shown by Slemrod et al, as he ignores a possible jump in the tax revenue. However, as the magnitude of this jump depends on the income distribution, it is unclear whether his result still holds for some distributions. To solve this puzzle, we assume a general income distribution, which allows a minimum impact of the revenue jump.

We first consider households' labor supply given progressive taxes, i.e., $t_1 < t_2$. For any $\bar{y}$, we can divide households into three groups according to their ONTE z. The households with $z \leq \bar{y}/(1 - t_1)^\varepsilon$ choose $x = w^\varepsilon(1 - t_1)^\varepsilon$, and obtain pre-tax earnings $z(1 - t_1)^\varepsilon \leq \bar{y}$; those with $z > \bar{y}/(1 - t_2)^\varepsilon$ choose $x = w^\varepsilon(1 - t_2)^\varepsilon$, and earn $z(1 - t_2)^\varepsilon > \bar{y}$; those in the interval $\bar{y}/(1 - t_1)^\varepsilon < z \leq \bar{y}/(1 - t_2)^\varepsilon$ would choose $x = \bar{y}/w$ and just earn pre-tax earnings $\bar{y}$ (bunching).

To simplify our proof, we temporarily modify households' bunching behavior as follows. We allow those bunching households to pay the lower tax rate t_1 for earnings beyond $\bar{y}$. Then their work efforts will not stop at $\bar{y}/w$, but continue up to the first-order condition $w(1 - t_1) - x^{1/\varepsilon} = 0$, which implies $x = w^\varepsilon(1 - t_1)^\varepsilon$. This special treatment gives these households higher utility and adds extra tax revenue. Thus, it exaggerates the benefit of progressive taxes.

Now we have just two groups of households, separated by $\bar{z} = \bar{y}/(1 - t_2)^\varepsilon$. Substituting their efforts $w^\varepsilon(1 - t_1)^\varepsilon$ and $w^\varepsilon(1 - t_2)^\varepsilon$ into (1) and (2), and denoting $w^{1+\varepsilon}$ by z, we obtain the following maximized utility:

$$\hat{V}_1 = \frac{z(1-t_1)^{1+\varepsilon}}{1+\varepsilon} + B \qquad\qquad \text{for } z \leq \frac{\overline{y}}{(1-t_2)^{\varepsilon}}$$

$$\hat{V}_2 = \frac{z(1-t_2)^{1+\varepsilon}}{1+\varepsilon} + (t_2 - t_1)\,\overline{y} + B \qquad \text{for } z > \frac{\overline{y}}{(1-t_2)^{\varepsilon}}.$$

Integrating $\hat{V}_1$ over $[m, \overline{z}\,]$, $\hat{V}_2$ over $[\overline{z}\,, M]$, and adding them together, we get

the utilitarian welfare, over valuated due to our special treatment.

$$W = \frac{(1-t_1)^{1+\varepsilon}}{1+\varepsilon} \int_m^{\overline{z}} z\,dF(z) + \frac{(1-t_2)^{1+\varepsilon}}{1+\varepsilon} \int_{\overline{z}}^{M} z\,dF(z) + (t_2 - t_1)[1 - F(\overline{z}\,)]\,\overline{y} + B$$

$\int_m^{\overline{z}} z\,dF(z)$ and $\int_{\overline{z}}^{M} z\,dF(z)$ are the total ONTE of two groups of households with z

$\leq \overline{z}$ and $z > \overline{z}$. We denote them by Z_1 and Z_2 respectively. Using these notations and

substituting the optimal efforts into (3), we get the over-valued basic income:

$$B' = t_1(1 - t_1)^{\varepsilon}Z_1 + t_2(1 - t_2)^{\varepsilon}Z_2 - (t_2 - t_1)[1 - F(\overline{z}\,)]\,\overline{y} \qquad (4)$$

Plug (4) into function W above, we have our over-valued utilitarian welfare:

$$W = \frac{1+\varepsilon t_1}{1+\varepsilon}(1 - t_1)^{\varepsilon}Z_1 + \frac{1+\varepsilon t_2}{1+\varepsilon}(1 - t_2)^{\varepsilon}Z_2 \qquad (5)$$

We can show that, given $t_1 < t_2$, there is no interior solution for $\overline{y}$ to maximize

(5). Although (5) over valuates the welfare, it coincides with the true welfare when $\overline{y}$

takes an extreme value such that no bunching could occur. Hence, no interior solution

for $\overline{y}$ exists when we consider the true welfare function either (see Appendix A).

Proposition 1: The taxes maximizing utilitarian welfare are not progressive.

Hence the regressive taxes found by Slemrod et al (1994) represent a general

case. Sheshinski (1989) argues that the optimal taxes must be progressive if the poor

have sufficiently higher marginal utility of income than the rich[1]. This cannot happen here since our quasi-linear utility implies an identical marginal utility for all. We will show in section 5 that even a maximum gap between marginal utilities is not sufficient for progressive taxes. So to justify progressive taxes under utilitarian welfare requires complicated twists on utility functions.

4. Max-min objective

Sheshinski (1989) claims that the optimal taxes cannot be regressive under a max-min objective. He argues that, if taxes are regressive there is no interior solution for $\bar{y}$, because a higher $\bar{y}$ always raises tax revenue. Unfortunately, this argument is flawed again. When $t_1 > t_2$ and $\bar{y}$ increases, households who still earn more than $\bar{y}$ will pay more taxes. But some households will cut earnings below $\bar{y}$ after its increase, and pay less tax. If this effect dominates the positive impact on tax revenue, the total tax revenue may fall, and $\bar{y}$ may have an interior solution.

This error is similar to that pointed out by Slemrod et al (1994) for the utilitarian welfare case. However, they do not explicitly discuss the case of the max-min objective, nor provide an example of regressive taxes. It is worthwhile to give a counter example to Sheshinski's claim. We use an extreme case of our utility function with $\varepsilon = 1$, and assume three households with ONTE equal to 0, 1 and 2 respectively. As the worst-off household earns no income, the max-min objective is to maximize the basic income. We can show the maximum revenue of 9/8 is obtained by $t_1^* = $ ¾, $t_2^* = 0$, and $\bar{y}* = 5/4$[2]. So taxes can be regressive under the max-min objective.

---

[1] He does not provide a concrete example as his condition involves unsolved variables such as $\bar{y}$.

[2] If both households pay t_2^*, another solution is $t_1^* = 1$, $t_2^* = $ ¼, $\bar{y}* = 3/8$, and the same revenue 9/8.

Besides his (flawed) theoretical argument, Sheshinski (1989) gives an example where "*the optimum marginal tax increases moderately from 20 to 29 percent*". However, he soon recognizes that, "*Allowing for two different marginal rates shows that the linear tax schedule, with $1 - \widehat{\beta} = 0.5$, is optimum!*" (p. 212). This contradicts his claim that the progressive taxes of 20% and 29% are truly optimal. The next question is: Can optimal taxes be progressive under the max-min objective?

To answer this question, we first consider a special case. In our model, when $m = 0$, the worst-off households earn no income, so the max-min objective is to maximize the basic income. Our earlier example is such a case. To choose optimal taxes maximizing tax revenue, we now derive the true function of basic income, allowing bunching behavior. Recall that given $t_1 < t_2$, households are divided into three groups. Those with $z > \overline{z} = \overline{y}/(1 - t_2)^\varepsilon$ choose $x = w^\varepsilon(1 - t_2)^\varepsilon$; those with $z \leq \overline{e} \equiv \overline{y}/(1 - t_1)^\varepsilon$ choose $x = w^\varepsilon(1 - t_1)^\varepsilon$; the remaining ones with $\overline{e} < z \leq \overline{z}$ choose $x = \overline{y}/w$ and just earn $\overline{y}$. The total tax revenue from the three groups is:

$$B = t_1(1 - t_1)^\varepsilon \int_m^{\overline{e}} zdF(z) + t_1 \overline{y} \int_{\overline{e}}^{\overline{z}} dF(z) + t_2(1 - t_2)^\varepsilon \int_{\overline{z}}^{M} zdF(z) - (t_2 - t_1)[1 - F(\overline{z})] \overline{y} \quad (6)$$

We examine if the taxes maximizing (6) can be progressive. Instead of looking at the solution of $\overline{y}$, we allow it to be any value and obtain a stronger result.

Proposition 2: Given any $\overline{y}$, the optimal taxes to maximize the basic income B are either a single tax rate $t^* = 1/(1 + \varepsilon)$, or regressive.

Proof: see Appendix B.

We then examine a general form of the max-min objective. In our model, if $m > 0$, the least productive households do work. The optimal labor supply yields utility

9

$m(1 - t_1)^{\varepsilon+1}/(1+\varepsilon)$. The max-min objective is no longer the basic income B alone, but has this extra term. Then we have a general max-min objective:

$$u_1 = \frac{m(1-t_1)^{1+\varepsilon}}{1+\varepsilon} + B \qquad (7)$$

This expression covers Sheshinski's earlier example. It is more likely to yield progressive taxes than B in (6) due to the extra term. A similar situation arises when the worst-off households are unemployed. Even though their utilities solely rely on the basic income, higher t_1 tends to reduce employment and tax revenue. This effect may increase the appeal of progressive taxes and deserves close examination. One way to model unemployment is to introduce a fixed cost of working. Households work only when they obtain higher utility than this cost. Let $\underline{z}$ be the lowest ONTE of working households. As households with $z < \underline{z}$ are unemployed, the tax revenue is reduced from (6) by $t_1(1 - t_1)^{\varepsilon} \int_{m}^{\underline{z}} z dF(z)$. So the utility of the unemployed becomes:

$$u_2 = B - t_1(1 - t_1)^{\varepsilon} \int_{m}^{\underline{z}} z dF(z) \qquad (8)$$

Moreover, unemployment can also arise due to a fixed unemployment benefit B_u, which will be lost if households work. A household works only when it obtains higher utility than this benefit. So B_u plays the same role as the fixed cost of working. The only difference is the additional income B_u, and a reduction of the basic income due to the unemployment payment, $B_u F(\underline{z})$, which adds further incentives to lower t_1 than (8), hence more chances for progressive taxes. The utility of the unemployed is:

$$u_3 = B + B_u - t_1(1 - t_1)^{\varepsilon} \int_{m}^{\underline{z}} z dF(z) - B_u F(\underline{z}) \qquad (9)$$

10

Welfare functions (7) – (9) provide additional incentives to lower t_1 than (6) due to extra terms, and hence more chances for progressive taxes. But these extra terms are not affected by $\bar{y}$. If we follow Sheshinski and focus on the interior solution for $\bar{y}$, they have no impact. We can still use B as the objective essentially, and show $\bar{y}$ has no interior solution if taxes are progressive. To simply our proof, we use the over-valued tax revenue B' in (4). It gives the same value as the true tax revenue B in (6) when $\bar{y}$ takes extreme values. If the maximization of B' does not yield an interior solution for $\bar{y}$, neither does B.

To obtain a clean result, however, we need to impose a mild restriction on income distribution functions.

Assumption: $\dfrac{zf(z)}{1-F(z)}$ *does not fall when z rises.*

This assumption requires that $f(z)$ does not fall with z too fast. It is not very restrictive as income distributions are usually continuous and change smoothly. In particular this assumption holds in all cases, which we will discuss later.

Proposition 3: Under the Assumption, the optimal taxes cannot be progressive under the max-min objectives (7) – (9).

Proof: see Appendix C.

In Sheshinski's example, the wage (earning) distribution satisfies our assumption. Proposition 3 confirms 20% and 29% progressive taxes are not optimal.

Given Propositions 1 – 3, neither utilitarian nor max-min welfare functions can justify progressive taxes. This should not be too surprising. As utilitarian welfare does not favor income redistributions from the rich to the poor, it should not be expected to justify progressive taxes. The max-min objective only concerns the worst-

off individuals, who are either unemployed or unproductive, and hence not much affected by the low-income tax. The objective is best achieved by maximizing the revenue, which favors regressive taxes due to different disincentive effects of t_1 and t_2. While t_2 discourages the work effort of high-income earners for every penny it generates, t_1 only has such an effect on low-income earners, and collects tax revenue from both. So it is optimal to set tax rate t_1 higher than t_2.

The failure to justify progressive taxes by utilitarian and max-min welfare functions does not mean that they should be dismissed generally. To evaluate the desirability of progressive taxes, we should look for alternative welfare functions. In the real world, the justification of progressive taxes is not to promote the interest of everyone or the worst-off individuals only, but to help the poor and working class. The low-income tax usually hurt them and this is why progressive taxes are preferred. Progressive taxes seem more justifiable if we maximize the utility of the poor.

5. Total Utility of the Poor

Sheshinski (1989) argues that optimal taxes maximizing utilitarian welfare should be progressive if the poor and rich have very different marginal utilities of income. Slemrod et al (1994) recognize that progressive taxes may maximize a weighted utility of two groups of households. They point to the same factor. When we put low weights on the rich, their loss is unimportant relative to the gain by the poor, and progressive taxes are likely. In the extreme case we can assign zero weight to the rich, and only maximize the utility of the poor. This is consistent with the political agenda of helping the poor and may give the best chance for progressive taxes.

We assume that the government maximizes the total utility of poor households whose ONTE are below an exogenously given level $\bar{e}$, which indicates the dividing

line between the poor and rich. When the government imposes progressive taxes, the poor should face the lower rate t_1, not the higher t_2. This can be achieved by setting the threshold income $\bar{y} = \bar{e}(1 - t_1)^\varepsilon$. If $\bar{y}$ is set lower, some poor households have to face t_2 reduce their work efforts. If $\bar{y}$ is set higher, it only affects poor households through the basic income. Proposition 2 implies that there is no interior solution for the optimal $\bar{y}$. Hence our choice of $\bar{y}$ is the most reasonable one. Then independent decision variables are t_1 and t_2 only. Given $t_1 < t_2$, poor households ($z \leq \bar{e}$) choose $x = w^\varepsilon(1 - t_1)^\varepsilon$, and obtain utility $z(1 - t_1)^{\varepsilon+1}/(1 + \varepsilon) + B$. Integrating this utility over [m, $\bar{e}$], we get the total utility of the poor, which the government maximizes:

$$W_1 = \frac{(1-t_1)^{1+\varepsilon}}{1+\varepsilon} \int_m^{\bar{e}} z dF(z) + F(\bar{e})B \qquad (10)$$

We investigate whether taxes maximizing (10) can be progressive. To simplify notations, we denote $\int_m^{\bar{e}} z dF(z)$ by E_1, $\int_{\bar{e}}^M z dF(z)$ by E_2, and $E_1 + E_2$ by E. Note that E is the average ONTE of the population. Further we let e_1 denote $E_1/F(\bar{e})$, and e_2 denote $E_2/[1 - F(\bar{e})]$. Note e_1 and e_2 are the average ONTE of the poor and rich respectively. From their definitions we always have $e_1 \leq \bar{e} \leq e_2$ and $e_1 \leq E \leq e_2$. All of them are determined by the income distribution and fixed $\bar{e}$, independent of taxes.

Proposition 4: Under the Assumption, the optimal taxes to maximize the total utility of the poor cannot be progressive if

$$e_1 e_2 \leq \bar{e} E \qquad (11)$$

Proof: see Appendix D.

We can write (11) as $e_1/E < \bar{e}/e_2$, where both sides are unit free ratios. If the rich are very rich, $\bar{e}/e_2$ is low, (11) is unlikely to hold, and progressive taxes are

possible. Conversely, if $e_2 \approx \bar{e}$, the rich are barely richer than the top poor, (11) holds as $e_1 < E$, and progressive taxes are impossible. Given $\bar{e}/e_2$, (11) is likely to hold when e_1/E is low, i.e., the poor are very poor relative to the population. If e_1 is close to zero, (11) must hold. By contrast, if the poor are quite "rich", such that e_1 is close to $\bar{e}$, (11) must be violated as $E < e_2$. It seems counter intuitive that optimal taxes tend to be regressive when the poor are very poor. The reason is the low earnings of the poor imply less loss when taxing them, and justify higher t_1 to raise tax revenue.

If we ignore the bunching behavior, it can be shown that (11) becomes a sufficient condition for progressive taxes as well. So it can be used approximately as a necessary and sufficient condition for progressive taxes if bunching is insignificant.

It is not always easy to evaluate (11). We present an equivalent condition, which depends on the change in the ratio of average ONTE of the poor and the rich.

Proposition 5: If f(z) > 0 around $\bar{e}$, (11) holds if and only if e_1/e_2 rises with $\bar{e}$.

Proof: see Appendix E.

Intuitively, a lower e_1/e_2 indicates a larger gap between the poor and the rich, and progressive taxes seem more plausible. However, it is the change, not the level of this ratio that matters. When e_1/e_2 is at its minimum, it will increase if $\bar{e}$ rises, taxes cannot be progressive until the ratio reaches its peak. Then (11) holds with equality. Further rise in $\bar{e}$ makes the ratio fall and the taxes may be progressive. In particular, if earnings are infinite, when $\bar{e}$ approaches infinity, e_1 is equal to E and e_2 is infinite, so e_1/e_2 is zero. That means the ratio must be falling when $\bar{e}$ is sufficiently large.

Corollary: If earnings are infinite, (11) must be violated when $\bar{e}$ is large.

Recall that if bunching is negligible, (11) becomes a sufficient condition for progressive taxes. Then the Corollary implies that, if the rich are very rich and our goal is to help a large poor majority, progressive taxes are justifiable.

To illustrate how the validity of (11) depends on income distributions, we present plausible examples and look at a few cases with analytic solutions.

Case 1: We first consider income distributions with finite earnings, where the lower and upper limits are normalized to 0 and 1. A simple density function is:

$$f(z) = \beta z^{\beta-1}, \qquad\qquad \text{where } \beta > 0, 0 \leq z \leq 1.$$

This function satisfies our Assumption but is restricted to be monotonic. If $\beta = 1$, it is a uniform distribution; if $\beta < (>) 1$, the density function decreases (increases) with z. (11) always holds[3], which means progressive taxes are impossible.

Case 2: Secondly, we consider a density function with infinite earnings. We assume a positive lower bound $m = 1$, and a decreasing density function.

$$f(z) = (\alpha - 1)z^{-\alpha}, \qquad\qquad \text{where } \alpha > 2, z \geq 1.$$

The Assumption holds. The shape of this distribution can be similar to Case 1. But the validity of (11) is very different. It never holds[4] for $\bar{e} > 1$.

---

[3] We have $F = \int_0^{\bar{e}} \beta z^{\beta-1} dz = \bar{e}^{\beta}$, $E_1 = \int_0^{\bar{e}} \beta z^{\beta} dz = \beta \bar{e}^{\beta+1}/(1 + \beta)$, and $E = \beta/(1 + \beta)$. Hence $e_1/E = \bar{e}$, and (11) becomes $e_2 \leq 1$, which is guaranteed.

[4] $F = (\alpha-1)\int_1^{\bar{e}} z^{-\alpha} dz = 1 - \bar{e}^{1-\alpha}$, $E_1 = (\alpha-1)\int_1^{\bar{e}} z^{1-\alpha} dz = (1-\bar{e}^{2-\alpha})(\alpha-1)/(\alpha-2)$, and $E = (\alpha-1)/(\alpha-2)$. So $e_1/E = (1-\bar{e}^{2-\alpha})/(1-\bar{e}^{1-\alpha})$, and $e_2 = \bar{e}(\alpha-1)/(\alpha-2)$. (11) becomes $(1-\bar{e}^{2-\alpha})/(1-\bar{e}^{1-\alpha}) < (\alpha-2)/(\alpha-1)$. It requires $\bar{e}^{\alpha-1} - \bar{e}(\alpha-1) + \alpha - 2 \leq 0$. The function rises with $\bar{e}$ from 0. So (11) never holds for $\bar{e} > 1$.

15

Case 3: In the real world, the monotonic distribution is uncommon, and single peaked distributions are usually observed. We consider these cases now. We assume finite earnings between 0 and 1. Two density functions serve as examples.

$$f(z) = \frac{(\alpha+1)(\beta+1)}{\beta-\alpha}(z^\alpha - z^\beta), \qquad \text{where } \beta > \alpha > 0,\, 0 \le z \le 1.$$

$$f(z) = \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha,\beta)}, \qquad \text{where } \beta,\, \alpha > 0,\, 0 \le z \le 1.$$

Both distributions satisfy the Assumption. The second is the Beta distribution, which allows monotonically increasing or decreasing density functions as well. Since $m = 0$, the ratio of e_1/e_2 must rise when $\bar{e}$ is very low. But it may not always be so. To show this, we look at a special case of both distributions, $f(z) = 6z(1 - z)$. It can be shown that (11) is violated if and only if $1 > \bar{e} > 5/6$ approximately [5].

Case 4: Next let us consider single peaked income distributions with infinite earnings. We first look at a case with a positive minimum ONTE $m = 1$.

$$f(z) = \frac{(\beta-1)(\alpha-1)}{\beta-\alpha}(z^{-\alpha} - z^{-\beta}), \qquad \text{where } \beta > \alpha > 2,\, z \ge 1.$$

The Assumption is again valid. It can be shown that (11) holds with equality when $\bar{e} = 1$, but is violated when $\bar{e}$ approaches to infinity. To illustrate how (11) depends on $\bar{e}$, we consider a special case with $\alpha = 3$ and $\beta = 4$, i.e., $f(z) = 6(z^{-3} - z^{-4})$. Then (11) holds if and only if $\bar{e} \le 4/3$ [6], which is the mode of the distribution.

---

[5] The mean $E = 0.5$ and $F = \bar{e}^2(3 - 2\bar{e})$. As $E_1 = \bar{e}^3(2 - 1.5\bar{e})$, we have $e_1 = \bar{e}(2 - 1.5\bar{e})/(3 - 2\bar{e})$. Also, $e_2 = 0.5(1 + 2\bar{e} + 3\bar{e}^2)/(1 + 2\bar{e})$. So (11) is $(2 - 1.5\bar{e})(1 + 2\bar{e} + 3\bar{e}^2) < (1 + 2\bar{e})(3 - 2\bar{e})$. It is $1 + 1.5\bar{e} - 7\bar{e}^2 + 4.5\bar{e}^3 = (1 - \bar{e})(1 + 2.5\bar{e} - 4.5\bar{e}^2) \ge 0$, violated if and only if $1 > \bar{e} > 5/6$ roughly.

[6] We first get $F = 1 - 3/\bar{e}^2 + 2/\bar{e}^3$, and $E_1 = 3 - 6/\bar{e} + 3/\bar{e}^2$. Hence $E = 3$, $e_1 = 3\bar{e}/(\bar{e} + 2)$, and $e_2 = E_2/(1 - F) = 3\bar{e}(2\bar{e} - 1)/(3\bar{e} - 2)$. Thus (11) becomes $3\bar{e}(2\bar{e} - 1) < (\bar{e} + 2)(3\bar{e} - 2)$, or $3\bar{e}^2 - 7\bar{e} + 4 = (\bar{e} - 1)(3\bar{e} - 4) < 0$. This holds when $\bar{e} \le 4/3$. 4/3 is the mode as $f'(z) \ge 0$ if and only if $z \le 4/3$.

Case 5: Finally, we consider single peaked income distributions with earnings covering all possible values. One such distribution often studied in the literature is the lognormal distribution. The Gamma distribution has similar properties.

$$f(z) = \frac{1}{z\sigma\sqrt{2\pi}} \exp[-\frac{(\ln z - \mu)^2}{2\sigma^2}] \qquad \text{where } z > 0, \, \sigma > 0, \, -\infty < \mu < \infty.$$

$$f(z) = \frac{z^{\alpha-1}}{\Gamma(\alpha)\theta^{\alpha}} \exp(-\frac{z}{\theta}) \qquad \text{where } z > 0, \, \alpha, \, \theta > 0.$$

Both density functions satisfy our Assumption. Since $\bar{e}$ covers the entire R_+, e_1/e_2 rises from zero ($0/E$) and falls to zero (E/∞) at the end. (11) must hold when $\bar{e}$ is small and be violated when $\bar{e}$ is large. Though (11) is difficult to evaluate in general, for a lognormal distribution it is valid for a particular value of $\bar{e}$, the median. This value is interesting because it divides the poor and rich equally. For any lognormal distribution (11) always holds when $\bar{e}$ is equal to the median[7].

A lognormal distribution has a median lower than the mean. If a distribution has a median larger than its mean, (11) always holds when $\bar{e}$ is equal to the median[8].

If (11) is violated, the optimal taxes may be progressive, but if the two tax rates are close, the advantage over a flat tax may not justify administrative costs. We can show that, if taxes are progressive, we have $t_2^* < (1 - \bar{e}/e_2)/(1 + \varepsilon - \bar{e}/e_2)$ (see Appendix D). The ratio decreases with $\bar{e}/e_2$. When $\bar{e}/e_2$ is close to 1, the ratio tends to be small. For instance, in the example of Case 3, the violation of (11) requires $\bar{e} > 5/6$. But when $\bar{e} = 5/6$, we have $\bar{e}/e_2 = 2\bar{e}(1 + 2\bar{e})/(1 + 2\bar{e} + 3\bar{e}^2) \approx 16/17$ (see

---

[7] The median is $\exp(\mu)$, the mean E is $\exp(\mu + 0.5\sigma^2)$, and $E_2 = \Phi(\sigma)\exp(\mu + 0.5\sigma^2)$, where $\Phi(\sigma)$ is the cumulative distribution of the standard normal. So $E_1 = E - E_2 = [1 - \Phi(\sigma)]\exp(\mu + 0.5\sigma^2)$. Substitute them and $F = 0.5$, (11) becomes $2[1 - \Phi(\sigma)] < 0.5\exp(-0.5\sigma^2)/\Phi(\sigma)$, or $4[1 - \Phi(\sigma)]\Phi(\sigma) - \exp(-0.5\sigma^2) < 0$. It falls from zero and rises to zero as σ changes from zero to infinity. So (11) holds.
[8] (11) becomes $E_1(E - E_1)/\bar{e}E < \frac{1}{4}$. As $\bar{e} \geq E$, this holds since $(1 - E_1/E)E_1/E \leq \frac{1}{4}$ for any E_1/E.

footnote 5). So t_2^* must be less than $1/(1 + 17\varepsilon)$. Unless ε is very small, t_2^* must be low, and significant tax progression is impossible. So a high value of $\overline{e}/e_2$ generally indicates the undesirability of progressive taxes.

6. Weighted Utility of the Poor

When we maximize the total utility of the poor, we treat the poor equally. It seems more reasonable to give a continuous discrimination to the poor, favouring the poorest over the less poor. Given $\overline{e}$ we can construct a weighted utility of the poor. We assign weight $s(z)$ to poor households with $z \leq \overline{e}$, subject to $\int_m^{\overline{e}} s(z)dF(z) = F(\overline{e})$.

Household z's optimal work leads to utility $z(1-t_1)^{1+\varepsilon}/(1+\varepsilon) + B$. Multiply it with $s(z)$ and integrate the product over $[m, \overline{e}]$. We get the weighted utility of the poor:

$$W_2 = \frac{(1-t_1)^{1+\varepsilon}}{1+\varepsilon} \int_m^{\overline{e}} s(z)zdF(z) + F(\overline{e})B \tag{12}$$

This function covers (10) as a special case with $s(z) = 1$. Since we want to help the poor, we should not assign a lower weight to poorer households than less poor ones. So $s(z)$ should not increase with z. Then the value of $\int_m^{\overline{e}} s(z)zdF(z)$ cannot exceed E_1. If we divide it by $F(\overline{e})$, the ratio must lie between m and e_1. The more weight on the poorer households, the closer the ratio is to m, while the less weight, the closer to e_1. Dividing (12) by $F(\overline{e})$ and denoting $\int_m^{\overline{e}} s(z)zdF(z)/F(\overline{e})$ by $\tilde{e}$, we have:

$$W_3 = \frac{\tilde{e}(1-t_1)^{1+\varepsilon}}{1+\varepsilon} + B \tag{13}$$

This function is identical to the utility of a poor household with ONTE $\tilde{e}$, which can be interpreted a representative household. To maximize (13) is equivalent

18

to maximizing the utility of this representative household. This is also consistent with the political agenda to help a chosen type of family in the society.

Proposition 6: The optimal taxes to maximize the weighted utility of the poor cannot be progressive if

$$\tilde{e}\, e_2 \le \bar{e}\, E \tag{14}$$

Proof: see Appendix F.

The right hand side of (14) is the same as that in (11), but the left hand side is smaller as $\tilde{e} < e_1$. So (14) is more likely to hold. If (11) holds, no weighted utility of the poor can justify progressive taxes either. Thus (11) can be taken as a necessary condition for progressive taxes under most reasonable welfare functions. Moreover, although our model assumes constant marginal utility of income, our general utility weighting plays a similar role as assigning high marginal utility to the poor. Since more weights on the very poor cannot justify progressive taxes, our result will not be totally invalidated by decreasing marginal utility of income.

Given $\bar{e}$, the poorer household we choose (lower $\tilde{e}$), the less likely are taxes to be progressive. For instance, in the example of Case 3, the violation of (11) requires $\bar{e}/e_2 \ge 16/17$. As $E = 0.5$, (14) holds if $\tilde{e} \le 8/17$. Progressive taxes are possible only if the income of the target household is close to the average. Moreover, as shown in Appendix F, the lower limit of t_1^* falls with $\tilde{e}$. The lower $\tilde{e}$ we choose, the higher t_1^*. If we target a very poor representative, even if the taxes are progressive, the progression tends to be small. This confirms our earlier finding that progressive taxes are unlikely to be justified by an objective to help the very poor.

6. Concluding remarks

In this paper we try to answer three questions. First we show that, progressive taxes are generally not justified by utilitarian and max-min welfare functions. This result clarifies some ambiguities in the existing literature and questions the general desirability of progressive taxes. However, it does not reject progressive taxes completely, because these two welfare functions are unlikely to justify progressive taxes and are not consistent with the usual political grounds for progressive taxes.

Our second goal is to evaluate the desirability of progressive taxes when we aim to help the poor exclusively. Under our new welfare functions, progressive taxes are possible but still far from guaranteed. We provide necessary conditions for progressive taxes. The conditions only depend on aggregate features of income distributions, which indicate how poor the poor are relative to the population and how rich the rich are relative to the dividing line between the poor and rich.

This result allows us to fulfil our third goal, to evaluate the desirability of progressive taxes with feasible information. For instance, unbounded earnings, meaning that the rich are very rich, and a high dividing line between the poor and rich are likely to justify progressive taxes. The opposite is true when the rich are not so rich (high $\overline{e}/e_2$) or the poor are very poor (low e_1/E). In particular if the former ratio is close to one, even if taxes are progressive, the difference between two tax rates will be small, and hence may not justify the extra administrative costs.

When two-band taxes are not better than a flat tax, a more complicated tax structure is unlikely to be so either, as administrative costs increase with the number of tax bands faster than the benefit of income redistribution. Then our result raises a general question about the desirability of progressive taxes under certain identifiable conditions. Our main contribution may be summarised as showing that, progressive taxes are optimal only under surprisingly restrictive conditions.

To ensure the tractability of the model, we assume a particular utility function with constant elasticity of labour supply and no income effects. It allows us to analyze the impact of income tax with general income distributions. Though it is restrictive, the compromise seems necessary to obtain clear results, and similar assumptions are often made in the existing literature. Though highly stylized, we hope that this paper contributes to the debate on a flat tax system combined with a basic income.

Appendix A, Proof of Proposition 1:

We show that there does not exist an interior solution for $\overline{y}$ if $t_2 > t_1$. When we differentiate (5) with respect to $\overline{y}$, we can write $\partial W/\partial \overline{y}$ as $\partial W/\partial \overline{z} \times \partial \overline{z}/\partial \overline{y}$. When $t_2 > t_1$, we have $\partial \overline{z}/\partial \overline{y} = 1/(1 - t_2)^{\varepsilon} > 0$. Hence $\partial W/\partial \overline{y} = 0$ only if $\partial W/\partial \overline{z} = 0$.

We notice that $\partial Z_1/\partial \overline{z} = \overline{z} f(\overline{z}) = -\partial Z_2/\partial \overline{z}$. So the first-order condition for $\overline{y}$ is:

$$\frac{\partial W}{\partial \overline{z}} = [(1 + \varepsilon t_1)(1 - t_1)^{\varepsilon} - (1 + \varepsilon t_2)(1 - t_2)^{\varepsilon}] \frac{\overline{z} f(\overline{z})}{1 + \varepsilon} = 0 \qquad (A)$$

As $f(z) \geq 0$, (A) is non-negative if $(1 + \varepsilon t_1)(1 - t_1)^{\varepsilon} > (1 + \varepsilon t_2)(1 - t_2)^{\varepsilon}$. Differentiate function $(1 + \varepsilon t)(1 - t)^{\varepsilon}$, we get $-\varepsilon(1 + \varepsilon)t(1 - t)^{\varepsilon - 1} < 0$. So $(1 + \varepsilon t)(1 - t)^{\varepsilon}$ falls with t. If $t_2 > t_1$, we have $\partial W/\partial \overline{z} \geq 0$. W always increases with $\overline{y}$ and no interior solution for $\overline{y}$ exists. Thus the optimal $\overline{y}$ must take an extreme value, which means the maximum level of (5) is obtained by a single tax rate. In this case, bunching cannot occur, the maximum (5) represents a feasible welfare level. Recall that the true welfare is lower than (5), except at the extreme values of $\overline{y}$. It cannot exceed the maximum (5) under a single tax rate. So the optimal taxes cannot be progressive.

Appendix B, Proof of Proposition 2:

Given $t_2 > t_1$ and a fixed $\overline{y}$, we differentiate B in (6) with respect to t_1 and t_2. Note that $\overline{z}$ does not depend on t_1 and $\overline{e}$ does not depend on t_2. Hence we get

$$\frac{\partial B}{\partial t_1} = (1 - t_1)^{\varepsilon}[1 - (1+\varepsilon)t_1] \int_m^{\overline{e}} z dF(z) + t_1(1 - t_1)^{\varepsilon} \overline{e} f(\overline{e}) \frac{\partial \overline{e}}{\partial t_1}$$

$$+ \overline{y} \int_{\overline{e}}^{\overline{z}} dF(z) - t_1 \overline{y} f(\overline{e}) \frac{\partial \overline{e}}{\partial t_1} + [1 - F(\overline{z})] \overline{y} = 0 \qquad (B1)$$

$$\frac{\partial B}{\partial t_2} = (1 - t_2)^{\varepsilon}[1 - (1+\varepsilon)t_2] \int_{\overline{z}}^{M} z dF(z) - t_2(1 - t_2)^{\varepsilon} \overline{z} f(\overline{z}) \frac{\partial \overline{z}}{\partial t_2}$$

$$+ t_1 \overline{y} f(\overline{z}) \frac{\partial \overline{z}}{\partial t_2} - [1 - F(\overline{z})] \overline{y} + (t_2 - t_1) \overline{y} f(\overline{z}) \frac{\partial \overline{z}}{\partial t_2} = 0 \qquad (B2)$$

As $(1 - t_1)^{\varepsilon} \overline{e} = (1 - t_2)^{\varepsilon} \overline{z} = \overline{y}$, (B1) and (B2) simplify to:

$$\frac{\partial B}{\partial t_1} = (1-t_1)^\varepsilon[1-(1+\varepsilon)t_1]\int_m^{\bar{e}} z\,dF(z) + [1-F(\bar{e})]\,\bar{y} = 0 \tag{B3}$$

$$\frac{\partial B}{\partial t_2} = (1-t_2)^\varepsilon[1-(1+\varepsilon)t_2]\int_{\bar{z}}^M z\,dF(z) - [1-F(\bar{z})]\,\bar{y} = 0 \tag{B4}$$

Given $1 - F(\bar{e}) > 0$, (B3) implies $t_1 > 1/(1+\varepsilon)$. Given $1 - F(\bar{z}) > 0$, (B4) implies $t_2 < 1/(1+\varepsilon)$. It is impossible to have $t_2 > t_1$. If $t_2 = t_1$, (6) simplifies to $t(1-t)^\varepsilon\int_m^M z\,dF(z)$.

Hence the optimal tax rate is $1/(1+\varepsilon)$.

Appendix C, Proof of Proposition 3:

We first show that if $t_2^* > t_1^*$, there is no interior solution for $\bar{y}$ to maximize B' in (4). If such a solution exists, we must have $\partial B'/\partial\bar{y} = 0$, or $\partial B'/\partial\bar{z}\times\partial\bar{z}/\partial\bar{y} = 0$. As $\partial\bar{z}/\partial\bar{y} > 0$, we need $\partial B'/\partial\bar{z} = 0$. Substitute $\bar{y} = (1-t_2)^\varepsilon\bar{z}$ in (4), and differentiate it with respect to $\bar{z}$, as $\partial Z_1/\partial\bar{z} = \bar{z}f(\bar{z}) = -\partial Z_2/\partial\bar{z}$, we obtain

$$\frac{\partial B'}{\partial\bar{z}} = t_2[(1-t_1)^\varepsilon - (1-t_2)^\varepsilon]\,\bar{z}f(\bar{z}) - (t_2-t_1)(1-t_1)^\varepsilon[1-F(\bar{z})] \tag{C1}$$

Since $\partial\bar{z}/\partial\bar{y}$ does not depend on $\bar{y}$, $\partial^2 B'/\partial\bar{y}^2 = \partial^2 B'/\partial\bar{z}^2\times(\partial\bar{z}/\partial\bar{y})^2$. So the sign of $\partial^2 B'/\partial\bar{y}^2$ is the same as that of $\partial^2 B'/\partial\bar{z}^2$. We differentiate (C1) again and get:

$$\frac{\partial^2 B'}{\partial\bar{z}^2} = t_2[(1-t_1)^\varepsilon - (1-t_2)^\varepsilon][\bar{z}f'(\bar{z}) + f(\bar{z})] + (t_2-t_1)(1-t_1)^\varepsilon f(\bar{z}) \tag{C2}$$

Substitute (C1) into (C2), assuming $f(\bar{z}) > 0$, we find

$$\frac{\partial^2 B'}{\partial\bar{z}^2} = (t_2-t_1)(1-t_1)^\varepsilon\{[1-F(\bar{z})]\frac{f(\bar{z})+\bar{z}f'(\bar{z})}{\bar{z}f(\bar{z})} + f(\bar{z})\} \tag{C3}$$

Note that $[1 - F(\bar{z})][\bar{z}f'(\bar{z}) + f(\bar{z})] + \bar{z}[f(\bar{z})]^2 \geq 0$ as $zf(z)/[1 - F(z)]$ does not fall with z. Hence if $t_2 > t_1$, we have $\partial^2 B'/\partial\bar{z}^2 \geq 0$ when $\partial B'/\partial\bar{z} = 0$. So B' does not reach its maximum for any interior $\bar{y}$. The optimal $\bar{y}$ has to be an extreme value, which means two taxes effectively become a flat tax. Recall that the true tax revenue B in (6) is lower than B', but the difference disappears with a single tax. Hence the maximum B' represents feasible tax revenue, and B cannot have an interior solution for $\bar{y}$.

Appendix D, Proof of Proposition 4:

(i) We first show that if $t_2^* > t_1^*$, they must be interior solutions. In this proof, since $F(\overline{e})$ is constant, we just write it as F. Assuming $t_2^* > t_1^*$, we substitute $(1 - t_1)^{\varepsilon}\overline{e} = \overline{y}$ into (6), and write the basic income as:

$$B = t_1(1 - t_1)^{\varepsilon}[\int_{m}^{\overline{e}} zdF(z) + \overline{e}(1-F)] + t_2(1 - t_2)^{\varepsilon}\int_{\overline{z}}^{M} zdF(z) - t_2(1 - t_1)^{\varepsilon}\overline{e}[1 - F(\overline{z})] \quad (D1)$$

Note $\overline{e}$ is constant, while $\overline{z} = \overline{e}(1 - t_1)^{\varepsilon}/(1 - t_2)^{\varepsilon}$. We substitute (D1) for B in (10) and differentiate it with respect to t_1 and t_2. Writing $\int_{m}^{\overline{e}} zdF(z)/F$ as e_1, we get:

$$\frac{(1-t_1)^{1-\varepsilon}}{F}\frac{\partial W_1}{\partial t_1} = -(1 - t_1)e_1 + [1 - (1+\varepsilon)t_1][E_1 + \overline{e}(1 - F)] + \varepsilon\overline{e}t_2[1 - F(\overline{z})] \quad (D2)$$

$$\frac{1}{F}\frac{\partial W_1}{\partial t_2} = (1 - t_2)^{\varepsilon-1}[1 - (1+\varepsilon)t_2]\int_{\overline{z}}^{M} zdF(z) - (1 - t_1)^{\varepsilon}\overline{e}[1 - F(\overline{z})] \quad (D3)$$

When $t_1 = 0$, we have $\partial W_1/\partial t_1 > 0$ if $-e_1 + E_1 + \overline{e}(1 - F) > 0$. As $E_1 = Fe_1$ and $\overline{e} > e_1$, the inequality holds. Hence, if the optimal taxes are progressive, t_1^* must satisfy the first-order condition $\partial W_1/\partial t_1 = 0$. Similarly, we see $\partial W_1/\partial t_2 < 0$ if $t_2 > 1/(1 + \varepsilon)$. So t_2^* must satisfy the first-order condition $\partial W_1/\partial t_2 = 0$.

(ii) We now show that if $t_2^* > t_1^*$, we must have $t_2^*[1 - F(\overline{z})] > t_1^*(1 - F)$. If W_1 is maximized by t_2^* and t_1^*, B must be higher when $t_2 = t_2^*$ than $t_2 = t_1^*$. Note when $t_2 = t_1^*$, we have $\overline{z} = \overline{e}$. Then a higher B with $t_2 = t_2^*$ implies

$$t_2^*\{(1-t_2^*)^{\varepsilon}\int_{\overline{z}}^{M} zdF(z) - \overline{y}[1 - F(\overline{z})]\} > t_1^*[(1-t_1^*)^{\varepsilon}\int_{\overline{e}}^{M} zdF(z) - \overline{y}(1 - F)]$$

(D4)

Note both sides of (D4) are positive. Then we have $t_2^*[1 - F(\overline{z})] > t_1^*(1 - F)$ if

$$\frac{(1-t_2^*)^{\varepsilon}\int_{\overline{z}}^{M} zdF(z)}{1 - F(\overline{z})} \leq \frac{(1-t_1^*)^{\varepsilon}\int_{\overline{e}}^{M} zdF(z)}{1 - F} \quad (D5)$$

Since $(1 - t_2^*)^{\varepsilon}\overline{z} = (1 - t_1^*)^{\varepsilon}\overline{e}$, we can write (D5) as

$$\frac{\int_{\bar{z}}^{M} zdF(z)}{\bar{z}[1-F(\bar{z})]} \leq \frac{\int_{\bar{e}}^{M} zdF(z)}{\bar{e}(1-F)} \equiv \frac{e_2}{\bar{e}} \tag{D6}$$

Given $t_2^* > t_1^*$, we have $\bar{z} > \bar{e}$. Then (D6) holds if $e_2/\bar{e}$ does not rise or equivalently, $\bar{e}/e_2$ does not fall with $\bar{e}$, which will be shown given our Assumption.

(iii) As $\bar{e}/e_2 = \bar{e}[1 - F(\bar{e})]/E_2$, it does not fall with $\bar{e}$ if its derivative with respect to $\bar{e}$ is non-negative, i.e., $E_2[1 - F(\bar{e}) - \bar{e}f(\bar{e})] + \bar{e}^2 f(\bar{e})[1 - F(\bar{e})] \geq 0$. Divide this inequality by $\bar{e}f(\bar{e})E_2$, we get

$$\frac{1-F(\bar{e})}{\bar{e}f(\bar{e})} + \frac{\bar{e}}{e_2} \geq 1 \tag{D7}$$

If (D7) is violated for any $\bar{e}$, $\bar{e}/e_2$ will fall as $\bar{e}$ increases. Given the Assumption, we know $[1 - F(\bar{e})]/[\bar{e}f(\bar{e})]$ does not rise with $\bar{e}$. So (D7) will never hold as $\bar{e}$ continues to rise, and must be violated when $\bar{e} = M$.

However, if M is finite, when $\bar{e} = M$, $\bar{e}/e_2$ must be equal to 1 and (D7) is valid. If M is infinite, $\bar{e}/e_2$ must converge to a limit, so its derivative must be zero. This implies the equality of (D7). As (D7) holds when $\bar{e} = M$, it cannot be violated for any $\bar{e}$. Thus $\bar{e}/e_2$ never falls with $\bar{e}$. So (D6) holds, and we have $t_2^*[1 - F(\bar{z})] > t_1^*(1 - F)$.

(iv) We now find the lower bound of t_1^*. If we substitute $t_1^*(1 - F)$ for $t_2^*[1 - F(\bar{z})]$ in (D2), it falls, and $-(1 - t_1^*)e_1 + [1 - (1+\varepsilon)t_1^*][E_1 + \bar{e}(1 - F)] + \varepsilon\bar{e}t_1^*(1 - F) < 0$, i.e.,

$$t_1^* > \frac{\bar{e} - e_1}{\bar{e} - e_1 + \varepsilon E_1/(1-F)} \tag{D8}$$

(v) Find the upper bound of t_2^*. As $(1 - t_2^*)^\varepsilon \bar{z} = (1 - t_1^*)^\varepsilon \bar{e}$, we rewrite (D3) as

$$\frac{1}{F}\frac{\partial W_1}{\partial t_2} = \bar{z}(1-t_2)^\varepsilon [1 - F(\bar{z})]\{\frac{1-(1+\varepsilon)t_2}{1-t_2}\frac{\int_{\bar{z}}^{M} zdF(z)}{\bar{z}[1-F(\bar{z})]} - 1\} \tag{D9}$$

Given (D6), if we replace $\int_{\bar{z}}^{M} zdF(z)/\{\bar{z}[1 - F(\bar{z})]\}$ by $e_2/\bar{e}$, the right hand side of (D9) must increase. Hence we have $[1 - (1+\varepsilon)t_2^*]e_2 - (1 - t_2^*)\bar{e} > 0$. It implies

$$t_2^* < \frac{e_2 - \bar{e}}{(1+\varepsilon)e_2 - \bar{e}} \tag{D10}$$

Combining (D8) and (D10), we find impossible to have $t_2^* > t_1^*$ if

$$\frac{\bar{e} - e_1}{\bar{e} - e_1 + \varepsilon E_1 / (1 - F)} \geq \frac{e_2 - \bar{e}}{(1 + \varepsilon)e_2 - \bar{e}} \tag{D11}$$

(D11) reduces to $\bar{e}\,[(1 - F)e_2 + E_1] \geq e_2[(1 - F)e_1 + E_1]$. As $(1 - F)e_2 = E_2$ and $E_1 = Fe_1$, we have $\bar{e}\,E \geq e_2e_1$.

Appendix E, Proof of Proposition 5:

Differentiating e_1/e_2 with respect to $\bar{e}$, it is positive if and only if $e_2\dfrac{\partial e_1}{\partial \bar{e}} \geq e_1\dfrac{\partial e_2}{\partial \bar{e}}$.

Recall that $e_1 = E_1/F(\bar{e})$, $e_2 = E_2/[1 - F(\bar{e})]$, $\partial F/\partial \bar{e} = f(\bar{e})$ and $\partial E_1/\partial \bar{e} = \bar{e}f(\bar{e}) = -\partial E_2/\partial \bar{e}$. Hence $\partial e_1/\partial \bar{e} = [\bar{e}f(\bar{e})F(\bar{e}) - E_1f(\bar{e})]/F(\bar{e})^2 = f(\bar{e})[\bar{e} - e_1]/F(\bar{e})$. Also, $\partial e_2/\partial \bar{e} = f(\bar{e})\{E_2 - [1 - F(\bar{e})]\bar{e}\}/[1 - F(\bar{e})]^2 = f(\bar{e})[e_2 - \bar{e}]/[1 - F(\bar{e})]$. We substitute theses into the inequality. Given $f(\bar{e}) > 0$, the inequality holds if $E_2(\bar{e} - e_1) \geq E_1(e_2 - \bar{e})$. We can write it as $E\bar{e} \geq E_1e_2 + E_2e_1 = F(\bar{e})e_1e_2 + [1 - F(\bar{e})]e_2e_1 = e_1e_2$.

Appendix F, Proof of Proposition 6:

As this proof is very similar to Appendix D, we only present it briefly. The argument regarding t_2^* is identical to the earlier case. We focus on t_1^*. Substitute (D1) into (13), and differentiate, we get the first-order condition:

$$(1 - t_1)^{1-\varepsilon}\frac{\partial W_3}{\partial t_1} = -(1 - t_1)\tilde{e} + [1 - (1+\varepsilon)t_1][E_1 + \bar{e}(1 - F)] + t_2\varepsilon\bar{e}[1 - F(\bar{z})] = 0 \tag{F1}$$

If $t_2^* > t_1^*$, we have $t_2^*[1 - F(\bar{z})] > t_1^*(1 - F)$ (Appendix D). Replacing $t_2^*[1 - F(\bar{z})]$ by $t_1^*(1 - F)$ must lowers (F1). So $-(1 - t_1^*)\tilde{e} + [1 - (1 + \varepsilon)t_1^*][E_1 + \bar{e}(1 - F)] + \varepsilon\bar{e}\,t_1^*(1 - F) < 0$. Solving this inequality we get:

$$t_1^* > \frac{E_1 + \bar{e}(1 - F) - \tilde{e}}{(1 + \varepsilon)E_1 + \bar{e}(1 - F) - \tilde{e}} \tag{F2}$$

As shown in Appendix D, $t_2^* < (e_2 - \bar{e})/[(1+\varepsilon)e_2 - \bar{e}]$, so $t_2^* > t_1^*$ is impossible if

$$\frac{E_1 + \bar{e}(1 - F) - \tilde{e}}{(1 + \varepsilon)E_1 + \bar{e}(1 - F) - \tilde{e}} \geq \frac{e_2 - \bar{e}}{(1 + \varepsilon)e_2 - \bar{e}} \tag{F3}$$

One can see that (F3) holds if and only if $\bar{e}\,E \geq e_2\tilde{e}$.

References

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

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

Dahan, M. and M. Strawczynski, 2000, Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates: Comment, *American Economic Review*, vol. 90, pp. 681-6.

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

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., E. Lehmann and A. Parmentier, 2006, Optimal Income Taxation and the Shape of Average Tax Rates, *Economic Bulletin*, vol. 8, pp. 1−6

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

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, Princeton and Osford 2008

Mirrlees, J. A., 1971, An Exploration in the Theory of Optimum Income Taxation, *Review of Economic Journal*, vol. 88, pp. 175–208

Murray, C.A., *In our hands: a plan to replace the welfare state*, Washington D. C., A. E. I. Press, 2006

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

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., *Optimal Income Tax and Distribution*, Oxford University Press, 1990

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

Appendix G (not for publication), Proof that the Assumption holds in the five cases:

Case 1: $f(z) = \beta z^{\beta-1}$, where $\beta > 0$, $0 \le z \le 1$. Since $F(z) = z^{\beta}$, we have $zf(z)/[1 - F(z)] = \beta z^{\beta}/(1 - z^{\beta})$. It always rises with z, so the Assumption holds.

Case 2: $f(z) = (\alpha - 1)z^{-\alpha}$, where $\alpha > 2$, $z \ge 1$. Since $F(z) = 1 - z^{1-\alpha}$, we get a constant $zf(z)/[1 - F(z)] = \alpha - 1$. So the Assumption holds.

Case 3: (i) $f(z) = \dfrac{(\alpha+1)(\beta+1)}{\beta-\alpha}(z^{\alpha} - z^{\beta})$, where $\beta > \alpha > 0$, $0 \le z \le 1$. We first find $F(z)$

$= \dfrac{(\alpha+1)(\beta+1)}{\beta-\alpha}(\dfrac{z^{\alpha+1}}{\alpha+1} - \dfrac{z^{\beta+1}}{\beta+1})$. It suffices to show the following ratio rises in z.

$$\frac{zf(z)}{1-F(z)} = (z^{\alpha+1} - z^{\beta+1})[\frac{1-z^{\alpha+1}}{\alpha+1} - \frac{1-z^{\beta+1}}{\beta+1}]^{-1}$$

$$= [1 - z^{\beta+1} - (1 - z^{\alpha+1})][\frac{1-z^{\alpha+1}}{\alpha+1} - \frac{1-z^{\beta+1}}{\beta+1}]^{-1} \tag{G1}$$

Let $x \equiv (1 - z^{\alpha+1})/(1 - z^{\beta+1})$, we can write (G1) as $(1 - x)(\dfrac{x}{\alpha+1} - \dfrac{1}{\beta+1})^{-1}$. Obviously,

(G1) falls with x. Hence it rises with z if x decreases with z, which we will show next. Differentiating x, we find $dx/dz < 0$ if $-(1 + \alpha)(1 - z^{\beta+1})z^{\alpha} + (1 + \beta)(1 - z^{\alpha+1})z^{\beta} < 0$, or $-(1 + \alpha)(1 - z^{\beta+1}) + (1 + \beta)(z^{\beta-\alpha} - z^{\beta+1}) = (1 + \beta)z^{\beta-\alpha} - (\beta - \alpha)z^{\beta+1} - 1 - \alpha < 0$. This function monotonically increases with z, and is equal to zero when $z = 1$. So the inequality holds for any $z < 1$. Thus (G1) rises with z, and the Assumption holds.

(ii) $f(z) = \dfrac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha,\beta)}$, where β, $\alpha > 0$, $0 \le z \le 1$. $F(z) = \dfrac{1}{B(\alpha,\beta)}\int_0^z x^{\alpha-1}(1-x)^{\beta-1}dx$.

$$\frac{zf(z)}{1-F(z)} = z^{\alpha}(1-z)^{\beta-1}[B(\alpha,\beta) - \int_0^z x^{\alpha-1}(1-x)^{\beta-1}dx]^{-1} \tag{G2}$$

If $\alpha(1 - z) \ge z(\beta - 1)$, (G2) must increases since $zf(z)$ is non-decreasing with z. So we only need to consider the case of $\alpha(1 - z) < z(\beta - 1)$. Differentiating (G2), we see that it increases with z if and only if

$$[\mathrm{B}(\alpha,\beta) - \int_0^z x^{\alpha-1}(1-x)^{\beta-1}dx\,][\alpha(1-z) - z(\beta-1)] + z^\alpha(1-z)^\beta > 0 \quad \text{(G3)}$$

Since $\alpha(1-z) < z(\beta-1)$, (G3) holds if and only if

$$L = \mathrm{B}(\alpha,\beta) - \int_0^z x^{\alpha-1}(1-x)^{\beta-1}dx + \frac{z^\alpha(1-z)^\beta}{\alpha(1-z) - z(\beta-1)} < 0 \qquad \text{(G4)}$$

Differentiating (G4) with respect to z, we get

$$\frac{dL}{dz} = \{\,-1 + \frac{[\alpha(1-z) - z(\beta-1)][\alpha(1-z) - \beta z] + (\alpha+\beta-1)z(1-z)}{[\alpha(1-z) - z(\beta-1)]^2}\,\}z^{\alpha-1}(1-z)^{\beta-1}$$

$$= \frac{(\beta-1)}{[\alpha(1-z) - z(\beta-1)]^2}z^\alpha(1-z)^{\beta-1}$$

The condition $\alpha(1-z) < z(\beta-1)$ implies $\beta > 1$. So L increases with z. But when z = 1, L is zero. Hence it must be negative given $1 > z > \alpha/(\alpha+\beta-1)$, and (G3) holds.

Case 4: $f(z) = \dfrac{(\beta-1)(\alpha-1)}{\beta-\alpha}(z^{-\alpha} - z^{-\beta})$, where $\beta > \alpha > 2$, $z \geq 1$. We obtain $F(z) = 1 +$

$\dfrac{(\beta-1)(\alpha-1)}{\beta-\alpha}(\dfrac{z^{1-\beta}}{\beta-1} - \dfrac{z^{1-\alpha}}{\alpha-1})$. So

$$\frac{zf(z)}{1-F(z)} = (z^{1-\alpha} - z^{1-\beta})[\frac{z^{1-\alpha}}{\alpha-1} - \frac{z^{1-\beta}}{\beta-1}]^{-1}$$

$$= (z^{\beta-\alpha} - 1)[\frac{z^{\beta-\alpha}}{\alpha-1} - \frac{1}{\beta-1}]^{-1} \qquad \text{(G5)}$$

Given $\beta > \alpha$, (G5) rises with z, hence the Assumption holds.

Case 5: (i) $f(z) = \dfrac{1}{z\sigma\sqrt{2\pi}}\exp[-\dfrac{(\ln z - \mu)^2}{2\sigma^2}\,]$ where z and $\sigma > 0$, $-\infty < \mu < \infty$.

We obtain $F(z) = 0.5[1 + \mathrm{erf}(\dfrac{\ln z - \mu}{\sqrt{2}\sigma})]$, where erf($w$) is defined as $\dfrac{2}{\sqrt{\pi}}\int_0^w e^{-t^2}dt$. So

$$\frac{zf(z)}{1-F(z)} = \frac{\sqrt{2}}{\sigma\sqrt{\pi}}\exp[-\frac{(\ln z - \mu)^2}{2\sigma^2}][1 - \mathrm{erf}(\frac{\ln z - \mu}{\sqrt{2}\sigma})]^{-1} \qquad \text{(G6)}$$

When $\ln z < \mu$, $f(z)$ increases with z, the Assumption always holds. Hence, we let $w = (\ln z - \mu)/(\sqrt{2}\,\sigma)$. To show (G6) rises with z, we just need to show that function $L = \exp(w^2)[1 - \mathrm{erf}(w)]$ falls with w for $w > 0$. Differentiating L, we need to show that

$$\frac{dL}{dw} = 2w \times exp(w^2)[1 - \text{erf}(w)] - \frac{2}{\sqrt{\pi}} \le 0 \text{ or}$$

$$S = w[1 - \text{erf}(w)] - \frac{1}{\sqrt{\pi}} exp(-w^2) \le 0 \tag{G7}$$

Further differentiating function S with respect to w, we get

$$\frac{dS}{dw} = 1 - \text{erf}(w) - w \times exp(-w^2) + \frac{2w}{\sqrt{\pi}} exp(-w^2) = 1 - \text{erf}(w) + (\frac{2}{\sqrt{\pi}} - 1)w \times exp(-w^2).$$

Since $\text{erf}(w) \le 1$ and $2 > \sqrt{\pi}$, we know $dS/dw > 0$. Using l'Hopital's rule, we see that $w[1 - \text{erf}(w)]$ goes to zero when w approaches infinity. Since S is zero when w is infinite, and $dS/dw > 0$, we know that (G7) holds for any w.

(ii) $f(z) = \dfrac{z^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} exp(-\dfrac{z}{\theta})$, where z, α and $\theta > 0$. $F(z) = \dfrac{1}{\Gamma(\alpha)\theta^\alpha} \displaystyle\int_0^z x^{\alpha-1} exp(\dfrac{x}{\theta})dx$.

$$\frac{zf(z)}{1-F(z)} = z^\alpha exp(-\frac{z}{\theta})[\Gamma(\alpha)\theta^\alpha - \int_0^z x^{\alpha-1} exp(\frac{x}{\theta})dx]^{-1} \tag{G8}$$

If $z \le \alpha\theta$, (G8) must rises with z as $zf(z)$ is non-decreasing. We only need to consider the case of $z > \alpha\theta$. Differentiating (G8), we see it increases with z if and only if

$$[\Gamma(\alpha)\theta^\alpha - \int_0^z x^{\alpha-1} exp(\frac{x}{\theta})dx]\frac{\alpha\theta - z}{\theta} + z^\alpha exp(-\frac{z}{\theta}) > 0 \tag{G9}$$

Since $z > \alpha\theta$, (G9) holds if and only if

$$L = \Gamma(\alpha)\theta^\alpha - \int_0^z x^{\alpha-1} exp(\frac{x}{\theta})dx + \frac{\theta}{\alpha\theta - z} z^\alpha exp(-\frac{z}{\theta}) < 0 \tag{G10}$$

Differentiating (G10) with respect to z, we get

$$\frac{dL}{dz} = -z^{\alpha-1} exp(-\frac{z}{\theta}) + \frac{\theta z^{\alpha-1}}{(\alpha\theta - z)^2} exp(-\frac{z}{\theta})[(\alpha\theta - z)(\alpha - \frac{z}{\theta}) + z]$$

$$= \frac{\theta z^\alpha}{(\alpha\theta - z)^2} exp(-\frac{z}{\theta}) > 0.$$

Hence L is increasing for all $z > \alpha\theta$. But when z approaches infinity, L is zero. So it must be negative for all finite $z > \alpha\theta$, and (G8) cannot decrease with z.