Australian National University

The Australian National University

Centre for Economic Policy Research

DISCUSSION PAPER

Optimal Piecewise Linear Income Taxation

Patricia Apps
Faculty of Law, University of Sydney, and IZA

Ngo Van Long
Economics Dept., McGill University

Ray Rees
CES, University of Munich

DISCUSSION PAPER NO. 655

December 2011

ISSN: 1442-8636
ISBN: 978-1-921693-36-6

Optimal Piecewise Linear Income Taxation

Patricia Apps
Faculty of Law, University of Sydney, and IZA

Ngo Van Long
Economics Dept., McGill University

Ray Rees
CES, University of Munich

September 22, 2011

Abstract

Given its significance in practice, piecewise linear taxation has received relatively little attention in the literature. This paper offers a simple and transparent analysis of its main characteristics. We fully characterize optimal tax parameters for the cases in which budget sets are convex and nonconvex respectively. A numerical analysis of a discrete version of the model shows the circumstances under which each of these cases will hold as a global optimum. We find that, given plausible parameter values and wage distributions, the globally optimal tax system is convex, and marginal rate progressivity increases with rising inequality.

JEL classification H21, H31, J22.

Keywords: piecewise linear; income; taxation.

Acknowledgements

We would like to thank Michael Hoy, Efraim Sadka and two referees, as well as seminar participants at IZA and Vanderbilt University, for very helpful comments. We are also grateful to Yuri Andrienko and Piers Trepper for research assistance. The usual disclaimer applies. This research was supported under the Australian Research Council's DP Funding Scheme (DP 0881787).

1 Introduction

The foundations of the theory of optimal income taxation were provided by the theory of nonlinear taxation first developed by James Mirrlees (1971), and the theory of linear taxation formulated by Eytan Sheshinski (1972). In Mirrlees's analysis, the problem is seen as one of mechanism design. An optimally chosen menu of marginal tax rates and lump sum tax/subsidies is offered, and individuals select from this menu in a way that reveals their productivity type. As well as the government budget constraint therefore, a key role is played by incentive compatibility or self selection constraints.

In Sheshinski's linear tax analysis on the other hand, there is no attempt to solve the mechanism design problem. All individuals are pooled, and the problem is to find the optimal marginal tax rate and lump sum payment over the working population as a whole, subject only to the government budget constraint.

In each case, the theory provides an analysis of how concerns with the equity and efficiency of a tax system interact to determine the parameters of that system, and in particular its marginal rate structure and degree of progressivity.

As Boadway (1998) points out, the optimal nonlinear tax is Pareto superior to a linear tax for any given revenue requirement and set of consumers, implying a superior tradeoff between equity and efficiency. Nevertheless, tax policy makers or "central planners" do not seem to adopt the Mirrlees approach to the design of tax systems in practice.

In reality virtually all tax systems are neither linear in the sense of Sheshinski nor nonlinear in the sense of Mirrlees, but rather piecewise linear. Gross income is divided into (usually relatively few) brackets and marginal tax rates are constant within but vary across these brackets.[1] When we consider *formal* income tax systems, the marginal tax rates are typically strictly increasing with the income levels defining the brackets. We refer to this case of strict marginal rate progressivity as the *convex case*, since it defines for an income earner a convex budget set in the space of gross income-net income/consumption. However, when we widen the definition of the tax system to include cash transfers that are paid and withdrawn as a function of

[1] The German tax system is a rare exception to this. It has four brackets and in the second and third of these marginal tax rates increase linearly with income. However, this is likely to be replaced by a more conventional step function in future. For further discussion see Apps and Rees (2009).

gross income we see that typically this may lead marginal tax rates to fall over some range as gross income increases. Since this introduces nonconvexities into the budget set income earners actually face, we refer to this as the *nonconvex case*.[2]

The reason for planners' preference for piecewise linear as opposed to optimal nonlinear tax systems could be that the former overcome a large part of the inefficiency of a simple linear tax while remaining relatively simple to implement.[3] The present paper is concerned with the realistic case in which policy makers are not trying to solve a mechanism design problem. It can therefore be regarded as an extension of optimal linear taxation, rather than a restricted or approximative form of optimal nonlinear taxation. As we see below, interpretation of the results draws heavily on optimal linear taxation theory.

The problem of the empirical estimation of labour supply functions when a worker/consumer faces a piecewise linear budget constraint has been widely discussed in the econometrics literature.[4] Moreover, the literature[5] on the estimation of the marginal social cost of public funds has been concerned with the deadweight losses associated with raising a marginal unit of tax revenue in the context of some given piecewise linear tax system, which is assumed not to represent an optimal tax system. Yet there is surprisingly little analysis of the general problem of optimal piecewise linear income taxation.

There are two main papers in the theoretical literature on the continuum-of-types case, by Sheshinski (1989) and Slemrod et al. (1994).[6] We believe these papers leave the literature in a rather unfinished state, despite the fact that the paper by Slemrod et al gives a thorough and insightful discussion of the results of its simulation analysis of the nonconvex case, as well as of the problem of piecewise linear taxation in a model consisting of only two types.

The contribution by Sheshinski was the first to formulate and solve the problem of the optimal two-bracket piecewise linear tax system, including

[2]We generally try to avoid the terms "progressive" and "regressive" because of course the nonconvex case can be average rate progressive.

[3]It could also be mentioned that the mechanism design approach does not extend readily to deal with realistic aspects of tax systems, for example two-earner households and multiple time periods. For further discussion of this see Apps and Rees (2011).

[4]For a very extensive discussion see in particular Pudney (1989).

[5]See in particular Dahlby (1998), (2008).

[6]Strawczynski (1998) also considers the optimal piecewise linear income tax, but gross income in his model is exogenous and attention is focussed, as in Varian (1980), on income uncertainty, so that taxation essentially becomes a form of social insurance.

the choice of the bracket threshold, for a continuum of worker/consumer-types. However, he claims to have shown that, under standard assumptions, marginal rate progressivity, the convex case, must *always* hold: in the social optimum the tax rate on the higher income bracket must always exceed that on the lower. Slemrod et al (1996) show that Sheshinski's proof does not hold in general because it ignores the existence of a discontinuity in the tax revenue function in the nonconvex case. They then carry out simulations which, using standard functional forms for the social welfare function, the individual utility function and the distribution of wage rates/productivities[7], in all cases produce the converse result - the upper-bracket marginal tax rate is optimally always lower.

The result that a nonconvex system *could* be optimal should not be surprising; for example it is foreshadowed by Sadka (1976), who established the "no distortion at the top" result for optimal nonlinear taxation and provided some intuition for why marginal tax rates could be lower at higher levels of income. The fact that the nonconvex case *always* turns out to be optimal is however also somewhat problematic, for two reasons.

First, in general non-parameterised models there is no reason to rule out the convex case, and there is the possibility that the specific functional forms and parameter values chosen by Slemrod et al for their simulations are biased toward producing the nonconvexity result. In particular their assumed wage distribution, taken from Stern (1976), is a lognormal distribution based on data from the late 1960's/early 1970's. Quite apart from the fact that wage inequality has increased considerably from that time, recent work[8] shows that the lognormal distribution is biased toward giving low and decreasing tax rates in the upper part of the income distribution, and argues strongly for using the Pareto distribution as a better representation of the data. In Section 4 below we present simulation results based upon Pareto wage rate distributions that are broadly consistent with current cross section data and with the evidence on increasing inequality over time. We find on reasonable elasticity assumptions that the globally optimal tax system is consistently convex and that, optimally, the degree of convexity should have increased in line with the rise in inequality over recent decades.[9]

[7]Essentially they take the model developed by Stern (1976) and extend the numerical analysis to the two-bracket case.

[8]See in particular Diamond and Saez (2011), and Atkinson, Piketty and Saez (2011).

[9]The "no distortion at the top" result, often cited as the basis of the intuition for the nonconvex case, does not imply that tax rates must fall over successive, relatively wide

Secondly, in practice, in virtually all countries, income tax systems do in fact exhibit a substantial degree of marginal rate progressivity. It is as if tax policy makers when designing the formal tax system aim for a basically convex system. However, cash transfer payments, most importantly payments to "in-work" households with dependent children, are typically withdrawn on income, and therefore have the effect of introducing nonconvexities.

In this paper, we find it useful first to separate the two types of system and examine the conditions that characterise a convex or nonconvex system when it is optimal. We provide a simple and transparent model which allows the characteristics of each type of tax system, and particularly the optimal bracket thresholds, to be easily seen and compared, and characterise the optimal tax parameters in the nonconvex case. We then go on to consider, in a numerical analysis, the determinants of whether one or the other system is in fact optimal.

2 Individual Choice Problems

We present first the analysis of the choice problems for the individual in the face of respectively convex and nonconvex tax systems. In the next section we discuss the optimal tax structures in each case.

Consumers have identical quasilinear utility functions[10]

$$u = x - D(l) \quad D' > 0,\ D'' > 0 \tag{1}$$

where x is consumption and l is labour supply. Gross income is $y = wl$, with the wage rate $w \in [w_0, w_1] \subset \mathbf{R}_{++}$. Given a two-bracket tax system with parameters $(a, t_1, t_2, \hat{y})$, with a the lump sum payment to all households, t_1 and t_2 the marginal tax rates in the first and second brackets respectively, and $\hat{y}$ the income level determining the upper limit of the first bracket, the consumer faces the budget constraint

$$x \leq a + (1 - t_1)y \quad y \leq \hat{y} \tag{2}$$

$$x \leq a + (t_2 - t_1)\hat{y} + (1 - t_2)y \quad y > \hat{y} \tag{3}$$

bands of income, and cannot provide any intuition here. See again Diamond and Saez (2011).

[10] Thus we are ruling out income effects. This considerably clarifies the results of the analysis.

We assume a differentiable wage distribution function, $F(w)$, with continuous density $f(w)$, strictly positive for all $w \in [w_0, w_1]$.

2.1 Convex case: $t_1 \leq t_2$

There are three solution possibilities:[11]

(i) Optimal income $y^* < \hat{y}$. In that case we have the first order condition

$$1 - t_1 - D'(\frac{y}{w})\frac{1}{w} = 0 \tag{4}$$

Defining $\psi(.)$ as the inverse function of $D'(.)$, this yields

$$y^* = w\psi((1-t_1)w) \equiv \phi(t_1, w) \tag{5}$$

giving in turn the indirect utility function

$$v(a, t_1, w) = a + (1-t_1)\phi(t_1, w) - D(\frac{\phi(t_1, w)}{w}) \tag{6}$$

Applying the Envelope Theorem to (6) yields the derivatives

$$\frac{\partial v}{\partial a} = 1; \quad \frac{\partial v}{\partial t_1} = -\phi(t_1, w); \quad \frac{dv}{dw} = D'(\frac{y}{w})\frac{\phi(t_1, w)}{w^2} > 0 \tag{7}$$

We define the unique value of the wage type $\tilde{w}$ by

$$\hat{y} = \phi(t_1, \tilde{w}) \tag{8}$$

Note that $w < \tilde{w} \Leftrightarrow y^* < \hat{y}$, and $\partial\tilde{w}/\partial\hat{y} > 0$.

(ii) Optimal income $y^* > \hat{y}$. In that case we have

$$1 - t_2 - D'(\frac{y}{w})\frac{1}{w} = 0 \tag{9}$$

implying

$$y^* = \phi(t_2, w) \tag{10}$$

[11] It is assumed throughout that all consumers have positive labour supply in equilibrium. It could of course be the case that for some lowest sub interval of wage rates consumers have zero labour supply. We do not explicitly consider this case but it is not difficult to extend the discussion to take it into account.

and the indirect utility

$$v(a, t_1, t_2, \hat{y}, w) = a + (t_2 - t_1)\hat{y} + (1 - t_2)\phi(t_2, w) - D(\frac{\phi(t_2, w)}{w}) \qquad (11)$$

and again the Envelope Theorem yields

$$\frac{\partial v}{\partial a} = 1; \quad \frac{\partial v}{\partial t_1} = -\hat{y}; \quad \frac{\partial v}{\partial t_2} = -(\phi(t_2, w) - \hat{y}); \quad \frac{\partial v}{\partial \hat{y}} = (t_2 - t_1) \qquad (12)$$

and $dv/dw > 0$ just as before. We define the unique wage type $\bar{w}$ by

$$\hat{y} = \phi(t_2, \bar{w}) \qquad (13)$$

and we have $w > \bar{w} \Leftrightarrow y^* > \hat{y}$, and $\partial \bar{w}/\partial \hat{y} > 0$.

(iii) Optimal income $y^* = \hat{y}$. In that case the consumer's indirect utility is

$$v(a, t_1, \hat{y}, w) = a + (1 - t_1)\hat{y} - D(\frac{\hat{y}}{w}) \qquad (14)$$

and the derivatives of the indirect utility function are

$$\frac{\partial v}{\partial a} = 1; \quad \frac{\partial v}{\partial t_1} = -\hat{y}; \quad \frac{\partial v}{\partial \hat{y}} = (1 - t_1) - D'(\frac{\hat{y}}{w})\frac{1}{w} \geq 0 \qquad (15)$$

The last inequality $\partial v/\partial \hat{y} \geq 0$ necessarily holds in this convex case[12] because these consumers, with the exception of type $\tilde{w}$, are effectively constrained at $\hat{y}$, in the sense that they would prefer to earn extra gross income if it could be taxed at the rate t_1, since $D'(\frac{\hat{y}}{w}) < (1 - t_1)w$, but since it would in fact be taxed at the higher rate t_2, they prefer to stay at $\hat{y}$. A small relaxation of this constraint increases net income by more than the money value of the marginal disutility of effort at this point.

To summarise these results: the consumers can be partitioned into three subsets according to their wage type, denoted $C_0 = [w_0, \tilde{w})$, $C_1 = [\tilde{w}, \bar{w}]$, $C_2 = (\bar{w}, w_1]$, with $C \equiv C_0 \cup C_1 \cup C_2 = [w_0, w_1]$. The subset C_0 consists of consumers choosing points at tangencies along the steeper part of the budget constraint, C_1 the consumers at the kink, and C_2 the consumers at tangencies on the flatter part of the budget constraint.[13]

[12] That is, one solves the problem for cases (i) and (iii) subject to the constraint $y \leq \hat{y}$, with (i) then the case with this constraint non-binding and (iii) that with it binding.

[13] We assume that the tax parameters are such that none of these subsets is empty.

Given the continuity of $F(w)$, consumers are continuously distributed around this budget constraint, with both maximised utility v and gross income y continuous functions of w. Utility v is strictly increasing in w for all w, and y^* is also strictly increasing in w except over the interval C_1. Finally, note that if $t_1 = t_2$, C_1 shrinks to a point.

2.2 Nonconvex case: $t_1 > t_2$

Here there are again three solution possibilities. Given $\hat{y}$, a, t_1 and t_2, with $t_1 > t_2$, there is a unique consumer type[14] denoted by $\hat{w}$, this being the solution to the equation

$$a+(1-t_1)\phi(t_1,w)-D(\frac{\phi(t_1,w)}{w}) = a+(1-t_2)\phi(t_2,w)-D(\frac{\phi(t_2,w)}{w})+(t_2-t_1)\hat{y} \quad (16)$$

where $\phi(.)$ has the same meaning as before. The left hand side of this equation is the consumer's utility at a tangency point with the first, flatter portion of the budget constraint, the right hand side her utility at a tangency point with the second, steeper portion of the budget constraint. Thus the condition specifies that this type is just indifferent between the two tax brackets. Note that

$$\phi(t_1,\hat{w}) < \hat{y} < \phi(t_2,\hat{w}) \quad (17)$$

and that $\partial\hat{w}/\partial\hat{y} > 0$. The income of consumers in $[w_0, \hat{w})$ is $\phi(t_1, w)$ and in $(\hat{w}, w_1]$ is $\phi(t_2, w)$. They pay taxes of $t_1\phi(t_1, w)$ and $t_2\phi(t_2, w) + (t_1 - t_2)\hat{y}$ respectively.

For individuals of type $\hat{w}$, the tax payments at the two local maxima are respectively $t_1\phi(t_1,\hat{w})$ and $[t_2(\phi(t_2,\hat{w}) - \hat{y}) + t_1\hat{y}] > t_1\phi(t_1,\hat{w})$. In this case, although maximised utility is a continuous function of w over $[w_0, w_1]$, optimal gross incomes and the resulting tax revenue are not. There is an upward jump in both at $\hat{w}$. Tax paid by a consumer of type $\hat{w}$ if she chooses to be in the higher tax bracket is always higher than that if she chooses the lower bracket, even though the tax rate in the latter is higher. Since however consumers of type $\hat{w}$ are a set of measure zero, their choice of gross income is of no consequence for social welfare or tax revenue. Nevertheless, this discontinuity will play an important role in the optimal tax analysis, as we see in the next section.

[14] A proof of this can be based on the fact that for given $a, t_1, t_2, \hat{y}$, each side of (16) is a continuous, strictly increasing function of w defined on the compact interval $[w_0, w_1]$.

Consumers with wages in $[w_0, \hat{w})$ have indirect utilities

$$v(a, t_1, w) = a + (1 - t_1)\phi(t_1, w) - D(\frac{\phi(t_1, w)}{w}) \tag{18}$$

with, again from the Envelope Theorem,

$$\frac{\partial v}{\partial a} = 1; \quad \frac{\partial v}{\partial t_1} = -\phi(t_1, w) \tag{19}$$

and for those in $(\hat{w}, w_1]$,

$$v(a, t_1, t_2, \hat{y}, w) = a + (t_2 - t_1)\hat{y} + (1 - t_2)\phi(t_2, w) - D(\frac{\phi(t_2, w)}{w}) \tag{20}$$

with

$$\frac{\partial v}{\partial a} = 1; \quad \frac{\partial v}{\partial t_1} = -\hat{y}; \quad \frac{\partial v}{\partial t_2} = -(\phi(t_2, w) - \hat{y}); \quad \frac{\partial v}{\partial \hat{y}} = (t_2 - t_1) < 0 \tag{21}$$

In contrast to the convex case, there is no bunching of consumers at the bracket limit $\hat{y}$. An interesting aspect of this nonconvex case is that consumers of types in a small neighbourhood below $\hat{w}$ have only small differences in productivity and achieved utilities but possibly large differences in labour supply, income and tax paid as compared to those in a small neighbourhood above $\hat{w}$.

We now turn to the optimal tax analysis.

3 Optimal Taxation

3.1 The optimal convex tax system

We assume that the optimal taxation system is convex. The planner chooses the parameters of the tax system to maximise a social welfare function defined as

$$\int_{C_0} S[v(a, t_1, w)]dF + \int_{C_1} S[v(a, t_1, \hat{y}, w)]dF + \int_{C_2} S[v(a, t_1, t_2, \hat{y}, w)]dF \tag{22}$$

where $S(.)$ is a continuously differentiable, strictly concave[15] and increasing function which expresses the planner's preferences over utility distributions.

[15] This therefore excludes the utilitarian case, which can however be arbitrarily closely approximated. As is well known, the strict utilitarian case, with $S' = 1$, presents technical problems when a quasilinear utility function with consumption as numeraire is also assumed.

The government budget constraint is

$$t_1[\int_{C_0} \phi(t_1, w)dF + \hat{y}\int_{C_1\cup C_2} dF] + t_2\int_{C_2}(\phi(t_2, w) - \hat{y})dF - a - G \geq 0 \quad (23)$$

where $G \geq 0$ is a per capita revenue requirement.

The first order conditions[16] for this problem can be written as:

Proposition 1: The values of the tax parameters $a^*, t_1^*, t_2^*, \hat{y}^*$ when an interior solution with $t_1^* < t_2^*$ is optimal satisfy the conditions

$$\int_C (\frac{S'(v(w))}{\lambda} - 1)dF = 0 \quad (24)$$

where λ is the shadow price of tax revenue;[17]

$$t_1^* = \frac{\int_{C_0}(\frac{S'}{\lambda} - 1)\phi(t_1^*, w)dF + \hat{y}^*\int_{C_1\cup C_2}(\frac{S'}{\lambda} - 1)dF}{\int_{C_0}\frac{\partial\phi(t_1^*, w)}{\partial t_1}dF} \quad (25)$$

$$t_2^* = \frac{\int_{C_2}(\frac{S'}{\lambda} - 1)[\phi(t_2^*, w) - \hat{y}^*]dF}{\int_{C_2}\frac{\partial\phi(t_2^*, w)}{\partial t_2}dF} \quad (26)$$

$$\int_{C_1}\{\frac{S'}{\lambda}v_{\hat{y}} + t_1^*\}dF = -(t_2^* - t_1^*)\int_{C_2}(\frac{S'}{\lambda} - 1)dF \quad (27)$$

The first of these conditions, that with respect to the uniform lump sum payment a, is essentially the same condition as for linear taxation. The marginal social utility of income averaged across the population is equated, by choice of a, to the marginal social cost of public expenditure, implying that the value of the average marginal social utility of income in terms of the numeraire is equated to the marginal cost of expenditure, which is 1.

[16] In deriving these conditions, it must of course be taken into account that the limits of integration $\tilde{w}$ and $\bar{w}$ are functions of the tax parameters. Because of the continuity of utility, optimal gross income and tax revenue in w, these effects all cancel and the first order conditions reduce to those shown here.

[17] Or the marginal social cost of public expenditure. Needless to say, if we assume that the planner has optimised the tax system, the problem of estimating this parameter becomes much simpler than it is taken to be in the literature on this problem. See for example Dahlby (1998), (2008).

The first term in the brackets is a measure of the marginal social utility of income to a consumer of type w. Since v is strictly increasing in w, strict concavity of S implies that the marginal social utility of income $S'(v(w))/\lambda$ falls monotonically with w. Thus income is redistributed from higher wage to lower wage types, the more so, the higher the value of a.

In the expression (25) for the first bracket's tax rate, the denominator, the sum of the (negative) compensated derivatives of earnings with respect to the tax rate, captures the deadweight loss or pure efficiency effect of the tax. The numerator is the equity effect. In Appendix A it is shown that this numerator is also negative and so the tax rate is positive.

The first term in the numerator of (25) is the sum of deviations of households' marginal social utilities of income from the population mean, weighted by their gross incomes. Thus the higher the marginal social utility of income of low-wage individuals relative to the average, the smaller will be the absolute value of the numerator and therefore the tax rate in the first bracket, other things equal. This reflects the fact that the social planner seeks to redistribute income towards the lower wage types. This is done by a combination of paying the lump sum transfer to all households and then "withdrawing" it, i.e. funding it, through the tax rate structure. The lower the tax rate on the first bracket, other things equal, the smaller the contribution made by low wage households to this funding and the larger their net transfer - lump sum *minus* tax payment. It is possible for this term to be positive,[18] implying that in the absence of the second term the lower bracket tax rate would be negative. Lower income wage types would benefit from a wage subsidy as well as the lump sum. Appendix A however shows that this is never optimal, because of the presence of the second term in the numerator, which must dominate the first in the case where this is positive.

The second term in the numerator of (25) works in opposition to the first in tending to raise the tax rate in the first bracket, and indeed, given the choice of optimal bracket limit, must be larger in absolute value than the first term if this is positive. It expresses the fact that the lower bracket tax rate is a lump sum tax on the incomes of the upper bracket wage-earners, as well as those at the kink, and therefore in respect of these has no deadweight loss associated with it. Its marginal contribution to tax revenue is measured by the bracket limit $\hat{y}^*$. The overall effect on welfare of this lump sum tax is positive, because the integral term is negative - the marginal social utility of

[18] See Appendix A.

income to wage types in the upper bracket is on average below the population average.

The upper bracket tax rate characterised in (26) reflects similarly a trade off between equity and efficiency. Here $\phi(t_1^*, w) - \hat{y}^* > 0$ and is increasing in w, and so the numerator here can also be shown to be negative, along the same lines as for t_1^*. The optimal tax rate for the subgroup C_2 is determined (given $\hat{y}^*$) entirely by the characteristics of the wage types in this group, since there is no higher group for which this tax rate is a lump sum tax.

Thus, given the optimal choice of tax brackets and of the lump sum a, the tax rates are set optimally over the sub-populations within each bracket. The advantage over a strictly linear tax is therefore that the tax rates can more closely take account of differences in the relationships between income and the marginal social valuation of income, and in the average deadweight losses, across the subsets of the population. This suggests the intuition that there would be little to gain from deviating from a linear ("flat") tax when the ratio of the equity effect to the efficiency effect remains constant as we move through the wage type distribution. As we show in the numerical analysis in Section 4 below, however, given realistic wage distributions the two bracket progressive tax does deliver higher social welfare than the linear tax, even when the elasticity on gross income with respect to the tax rate is constant throughout the population.

Defining

$$\varepsilon(t_i, w) \equiv -\frac{\partial \phi(t_i^*, w)}{\partial(1-t_i)} \frac{(1-t_i)}{y^*} \qquad i = 1, 2 \tag{28}$$

as the compensated elasiticity of earned income with respect to (1 *minus*) the tax rate, and denoting $[\int_{C_0} dF]^{-1} F'$ by $f_0(w)$ and $[\int_{C_2} dF]^{-1} F'$ by $f_2(w)$, we can write (25) and (26), using (24)[19] as

$$\frac{t_1^*}{1-t_1^*} = \frac{\int_{C_0} (\frac{S'}{\lambda} - 1)[\phi(t_1^*, w) - \hat{y}^*] f_0(w) dw}{\int_{C_0} \varepsilon(t_1^*, w) y^* f_0(w) dw} \tag{29}$$

$$\frac{t_2^*}{1-t_2^*} = \frac{\int_{C_2} (\frac{S'}{\lambda} - 1)[\phi(t_2^*, w) - \hat{y}^*] f_0(w) dw}{\int_{C_2} \varepsilon(t_2^*, w) y^* f_2(w) dw} \tag{30}$$

From there it is just a short step to argue that if the compensated elasticities are constant with respect both to wage type and tax rate, then we have the

[19] See the Appendix.

result that, other things equal, the tax rate will be lower, the higher the mean income in a tax bracket. This means in turn that for a convex tax system to be optimal, the equity (numerator) term for the lower bracket would have to be correspondingly lower in absolute value, since average income in the second tax bracket will obviously be higher than in the first. Thus an apparently innocuous constant elasticity assumption makes more stringent the condition for finding that a convex tax system is optimal. The constant elasticity assumption places an important restriction on the compensated income derivatives - marginal deadweight losses of a tax - namely that they must increase proportionately with income.

The left hand side of (27), the condition with respect to $\hat{y}$, gives the marginal social benefit of a relaxation of the constraint on the consumer types in C_1 who are effectively constrained by $\hat{y}$. First, for $w \in (\tilde{w}, \bar{w}]$ the marginal utility with respect to a relaxation of the gross income constraint is $v_{\hat{y}} = (1 - t_1) - D'(\frac{\hat{y}}{w})\frac{1}{w} > 0$, as shown earlier. This is weighted by the marginal social utility of income to these consumer types. Moreover, since they increase their gross income, this increases tax revenue at the rate t_1^*. The right hand side is also positive and gives the marginal social cost of increasing $\hat{y}$, because, since $t_2^* > t_1^*$, this reduces the tax burden on the higher income group. An increase in $\hat{y}$ can be thought of as equivalent to giving a lump sum payment to higher rate taxpayers proportionate to the difference in marginal tax rates, and this is weighted by a term reflecting the net marginal social utilities of income to consumers in this group. An implication of this solution is that $[\int_{C_2} dF]^{-1} \int_{C_2} \frac{S'}{\lambda} dF < 1$, so that the average of the marginal social utilities of income of the upper bracket consumers is below the population average. The planner then suffers a distributional loss from giving this group a lump sum income increase.

Sheshinski argued that if $t_2^* < t_1^*$ the term on the right hand side of (27) must be negative, thus yielding a contradiction, and therefore ruling out the possibility of nonconvex taxation. However, because of the discontinuity in the tax revenue function in the nonconvex case, this is not the appropriate necessary condition, as pointed out by Slemrod et al., who did not however provide a complete characterisation of the optimum for this case. We now go on to provide the appropriate necessary conditions, which have as yet not been given in the literature.

3.2 The optimal nonconvex tax system

We can state the optimal tax problem in this case as

$$\max_{a,t_1,t_2,\hat{y}} \int_{w_0}^{\hat{w}} S[v(a,t_1,w)]dF + \int_{\hat{w}}^{w_1} S[v(a,t_1,t_2,\hat{y},w)]dF \tag{31}$$

$$s.t. \quad \int_{w_0}^{\hat{w}} t_1\phi(t_1,w)dF + \int_{\hat{w}}^{w_1} [t_2\phi(t_2,w) + (t_1 - t_2)\hat{y}]dF - a - G \geq 0 \tag{32}$$

where it has to be remembered that indirect utility is continuous in w, but that there is a discontinuity in tax revenue at $\hat{w}$.

From (16) it is easy to see that a change in a does not affect the value[20] of $\hat{w}$, and so the first order condition with respect to a is just as before, and can again be written as

$$\int_{w_0}^{w_1} (\frac{S'}{\lambda} - 1)dF = 0 \tag{33}$$

However, for each of the remaining tax parameters the discontinuity in gross income will be relevant, because a change will cause a change in $\hat{w}$, the type that is just indifferent to being in either of the tax brackets. A small discrete increase (decrease) in t_1 relative to t_2 will induce a subset of consumers in a neighbourhood below (above) $\hat{w}$ to choose to be in the higher (lower) tax bracket, thus reducing (increasing) the value of $\hat{w}$.

Now define

$$\Delta R = [t_2\phi(\hat{w},t_2) - (t_2 - t_1)\hat{y}] - t_1\phi(\hat{w},t_1) > 0 \tag{34}$$

This is the value of the jump in tax revenue at $\hat{w}$.

The remaining first order conditions for the above problem are then given by

Proposition 2:

$$t_1^* = \frac{\int_{w_0}^{\hat{w}} (\frac{S'}{\lambda} - 1)[\phi(t_1^*,w) - \hat{y}^*]dF + \frac{\partial\hat{w}}{\partial t_1}\Delta R f(\hat{w})}{\int_{w_0}^{\hat{w}} \frac{\partial\phi(t_1^*,w)}{\partial t_1}dF} \tag{35}$$

[20] Note the usefulness of the quasilinearity assumption in this respect.

$$t_2^* = \frac{\int_{\hat{w}}^{w_1} (\frac{S'}{\lambda} - 1)[\phi(t_2^*, w) - \hat{y}^*]dF + \frac{\partial \hat{w}}{\partial t_2}\Delta R f(\hat{w})}{\int_{\hat{w}}^{w_1} \frac{\partial \phi(t_2^*, w)}{\partial t_2} dF} \tag{36}$$

and the condition with respect to the optimal bracket value $\hat{y}^*$ is

$$\frac{\partial \hat{w}}{\partial \hat{y}} \Delta R f(\hat{w}) = (t_2 - t_1) \int_{\hat{w}}^{w_1} \left\{ \frac{S'}{\lambda} - 1 \right\} dF \tag{37}$$

The new element in the condition for t_1^*, as compared to the convex case, is the second term in the numerator, which, since $\partial \hat{w}/\partial t_1 < 0$, is also negative. Thus this term acts so as to increase the absolute value of the numerator, and therefore the value of t_1. The intuition for this term is that an increase in t_1 expands the subset of consumers who prefer to be in the upper tax bracket (with the lower tax rate) and so causes an upward jump in tax revenue, equal in the limit, as the change in t_1 goes to zero, to $\Delta R f(\hat{w})$.

In the condition for t_2, again the new element is the second term in the numerator, which, since $\partial \hat{w}/\partial t_2 > 0$, is positive. Thus this tends to reduce the tax rate in the upper bracket. The intuition for this term is that an increase in t_2 widens the subset of consumers who prefer to be in the lower bracket and so causes a downward jump in tax revenue. This then makes for a lower tax rate in this bracket.

In the final condition it can be shown that $\partial \hat{w}/\partial \hat{y} > 0$, and, on the same arguments as used before, but with $(t_2 - t_1) < 0$, the right hand side is also positive. Thus, there is nothing *a priori* to rule this case out, contrary to Sheshinski's assertion. The intuition is straightforward. The right hand side now gives the marginal benefit of an increase in $\hat{y}$ to the planner, namely a lump sum reduction in the net income of higher bracket consumers with, on average, below-average marginal social utility of income. The marginal cost of this is a jump downward in tax revenue from consumers who now find the first tax bracket better than the second. More precisely, a discrete increase $\Delta \hat{y}$ would cause a discrete interval of consumers to jump down into the lower bracket, and, in the limit, as $\Delta \hat{y} \to 0$, the resulting revenue loss is given by $\Delta R f(\hat{w})$. Both marginal benefit and marginal cost are positive

The above discussion proceeded by assuming either that the convex case or the nonconvex case was optimal, and then examining the necessary conditions for the optimal tax parameters in each case. This does not however help us determine which of the two is in fact the optimal tax system. As is well-known, even in a model as simple as that analysed here, it cannot be assumed that a local optimum is unique, or that any local optimum is global.

In the space of tax parameters $(a, t_1, t_2, \hat{y})$, we cannot make the convexity assumptions that would guarantee that any local optimum is global or unique. Nevertheless, as we now go on to show, the conditions can be used to help us understand the circumstances under which one or the other system will in fact be optimal.

4 Comparing tax systems

In this section, we first give a graphical comparison of the convex and nonconvex systems with an initially optimal linear tax, to give some intuition on the distributional outcomes of each system. We then set out the general discrete model with n household types and go on to show, using numerical solutions to the optimal tax problem, how the optimality of each type of system depends critically on the assumed wage distribution, and also on how it is affected by changes in the other parameters of the model - the distributional preferences embodied in the social welfare function (SWF) and the compensated labour supply elasticities.

To gain some insight into the distributional implications of switching between convex and nonconvex tax systems, we compare each in general terms with an initially optimal linear tax. Figure 1 illustrates the comparison between the optimal linear tax and the optimal convex two-bracket tax. The line $a_L L$ represents the budget constraint facing all consumers under the optimal linear tax, $a_C CD$ that under the optimal convex piecewise linear tax. Given that each tax system satisfies the government revenue requirement, one budget constraint cannot lie entirely above the other over the whole domain of y-values. Thus there must be at least one intersection point within this domain. Cases however can, by suitable choices of the wage distribution, parameters of the SWF and compensated labour supply elasticities, also be constructed in which $a_C \geq a_L$, the lump sum transfer is at least as high in the convex piecewise linear case.

Figure 1 about here

The essential feature of the illustration is that the convex piecewise linear tax system redistributes welfare towards the middle and away from the ends, as compared to the linear tax, since over the range AB the budget constraint lies above that in the linear case, outside that range it is everywhere below it. All consumers in the lower tax bracket under the piecewise linear tax will expand their (compensated) labour supplies, all those in the higher bracket will contract theirs, as compared to the linear tax. However, in the case in which optimally $a_C > a_L$ (not shown), only consumers in the upper part of the higher tax bracket would be worse off. In this case a higher tax rate in the upper bracket funds a larger lump sum transfer as well as a lower tax rate in the lower bracket.

Figure 2 compares the optimal linear and nonconvex piecewise linear tax systems. The budget constraint corresponding to the linear tax is again $a_L L$, that of the piecewise linear tax is $a_N EF$. Thus we see that, as compared to the linear tax, the nonconvex piecewise linear tax redistributes welfare from the middle towards the bottom and top. Lower bracket consumers, who now pay a higher marginal rate, reduce their labour supplies and gross incomes, higher bracket consumers increase theirs. Cases are also possible in which $a_N \leq a_L$, and so only the upper segment of the higher bracket would be made better off. In this case, a constant or reduced lump sum transfer and a higher tax rate in the lower bracket funds a lower tax rate in the upper bracket. There is a good deal of evidence to suggest that tax reforms over the last couple of decades in a number of OECD countries, notably the US, UK and Australia, have had this outcome. Tax cuts at the top have, in effect, been funded by higher taxes on the middle, often made less than transparent by expressing the changes in rate structure in terms of an income supplement to the lowest wage types with a high withdrawal rate as a function of income over the lower and middle income ranges.[21]

Figure 2 about here

4.1 A general discrete model

The simplicity of the model presented here means that the effects of changes in the parameters, especially the wage distribution, are very transparent. This contrasts with the study by Slemrod et al., which, following Stern (1976), assumed CES utility functions, a utilitarian SWF[22] and a lognormal wage distribution, the parameters of which were taken from Stern, and which relate to an estimated wage distribution dating from the late 1960's/early 1970's. This model was solved for varying values of the elasticity of substitution in the utility function which, as Saez (2001) points out, bear no

[21] See Apps and Rees (2010), where this is discussed at some length for the case of Australia. Similar points apply to the UK and US.

[22] A SWF with a positive degree of inequality aversion was also considered, with no significant change in results.

simple relationship to the compensated labour supply elasticities on which our intuition is more usually based. No changes in the wage distribution were considered, yet, as our results show, this is in fact the most important driver of the results, as far as the general shape of the tax structure is concerned.

Consistent with the theoretical discussion given earlier, we find that for selected values of ρ, the parameter expressing the planner's inequality aversion, and for plausible elasticities:

- The optimal rate structure is always convex when the wage distribution is at first relatively flat and then rises steeply in the higher deciles, implying greater wage and income inequality among these deciles and in the distribution overall, as compared to the distribution used by Slemrod et al. This pattern in fact characterises the existing wage distributions of fully employed individuals of prime age in many OECD countries, in particular the US, UK, Germany and Australia.[23]

- The progressivity of the optimal convex rate scale increases as the inequality in wage rates among the top deciles increases, as we would expect from the theoretical conditions presented earlier in (25) and (26).

- The nonconvex case is optimal if inequality is concentrated in the bottom percentiles and the remainder of the distribution is relatively flat, again as suggested by the theoretical results presented earlier.

- The nonconvex case can also be obtained with the more realistic distribution if we assume an implausibly large gap between elasticities for the lower and upper parts of the distribution, with a very high elasticity at the top.

In the general discrete model the n household types have gross incomes y_i, each corresponding to a wage type w_i, $i = 1, .., n$, We assume two tax brackets, and the bracket limit is again denoted by $\hat{y}$. The SWF is given by $[\sum_{i=1}^{n} v_i^{1-\rho}]^{1/(1-\rho)}$, with $\rho \neq 1$ a measure of inequality aversion. The indirect utility functions v_i are derived just as in Section 2, with the quasilinear utility function $u = x - kl^{\alpha}$, $\alpha > 1$. The parameters k and α are calibrated so as to yield empirically reasonable values of labour supplies and compensated

[23] See, for example, the distributions of the earnings and hours of primary earners reported in Apps and Rees, 2009, Ch 1.

labour supply elasticities respectively, given the distributions of wage types. In the solution to the optimal tax problem, each wage type i will have a corresponding labour supply and gross income y_i^*, increasing in the wage, and we let j denote the type such that $y_1^* < ... \leq y_j^* \leq \hat{y} < y_{j+1}^* < ... < y_n^*$, that is, the highest wage type in the lower income bracket.

We write the optimal tax problem as:

$$\max_{a,t_1,t_2,\hat{y}} \{\sum_{i=1}^{n} [v_i(a, t_1, t_2, \hat{y})]^{1-\rho}\}^{1/(1-\rho)} \tag{38}$$

$$\text{s.t.} \quad t_1 \sum_{i=1}^{j} y_i(t_1) + t_2 \sum_{i=j+1}^{n} (y_i(t_2) - \hat{y}) + (n-j)t_1\hat{y} - na \geq 0 \tag{39}$$

We then solve this problem numerically for the optimal lump sum transfer a, tax rates t_1, t_2 and bracket limit $\hat{y}$, given assumed parameter values for ρ and the compensated labour supply elasticity ε (which implies a unique value for α), and given the wage distributions that were discussed in general terms above, and are described in more detail in the next subsection. The numerical analysis presented below is based on $n = 1,000$. The procedure is to assume successive values of the tax bracket $\hat{y}$ at \$100 intervals throughout the income distribution and solve numerically for the optimal tax rates and lump sum transfer at each bracket value. We then take the bracket value which yields the global maximum of the SWF.[24]

4.2 Numerical Results

We solve for globally optimal tax structures for two sets of wage distributions.[25] In the first set we begin with a Pareto wage distribution defined to approximate that of primary earners, aged from 25 to 59 years and earning above the minimum wage, in a sample of couples selected from a recent house-

[24] Two methods were used: general grid search and global optimisation software. They gave virtually identical results.

[25] These distributions were constructed by first taking 1 million random draws from a Pareto distribution with the given parameter, truncated in the way described in the text and arranged into 1000 equally-sized blocks in ascending order of size. The mean of each block was then calculated, to give the discrete distribution for 1000 wage types used in the numerical analysis.

hold survey.[26] Limiting the sample to primary earners on above minimum wages excludes those on very low earnings and recipients of unemployment and disability benefits, who make up around 20 per cent of the full sample. Wage rates that closely match those of the selected sample are generated by a Pareto distribution[27] with a beta parameter of 3.5, a lower bound of $20 per hour and an upper bound at the 98th percentile. These parameters set wage rates in the upper percentiles at higher rates than in the data to adjust for top-end coding. To illustrate the implications of rising top incomes for the structure of optimal tax rates, we then vary the beta parameter to construct two further distributions with lower degrees of inequality in the top percentiles. Thus we have the following three distributions, which we label Distribution Set 1:

Distribution 1a: $\beta = 3.5$. Average wage = \$48.10

Distribution 1b: $\beta = 2.0$. Average wage = \$35.03

Distribution 1c: $\beta = 1.5$. Average wage = \$28.34

We would argue that the economic circumstances of households in this type of sample are, at least to some extent, consistent with two assumptions of optimal tax theory - that productivities are innate and cannot be observed and, therefore, that wage rates representing productivities can be treated as exogenous and unobservable. These assumptions cannot plausibly be considered to hold in the case of recipients of disability pensions or long term unemployment benefits. Many types of disabilities are observable and disability pensions are individual-specific and not part of the general tax system. In the case of the long term unemployed, the available empirical evidence suggests that their earnings possibilities reflect the need for further education, training and work experience, implying that a broader set of policy instruments than income taxation are relevant, and indeed are in use. We therefore regard the results for the above set of distributions as being the most relevant for the general analysis of tax systems in present-day economies.

To show the extent to which a nonconvex income tax structure can result from including these categories of welfare recipients, we construct a second

[26]We draw on data for primary earnings and hours reported for couple income unit records in the Australian Bureau of Statistics 2008 Income and Housing Survey.

[27]The *cdf* of the Pareto distribution for a variate x is given by $F(x) = 1 - (\frac{A}{x})^a$ if $x \geq A$ and $F(x) = 0$ if $x < A$, with Pareto index $a > 1$ and parameter $\beta = a/(a-1)$. Increasing β is associated with increasing inequality in the distribution.

set of three wage distributions, which we label Distribution Set 2. In each distribution we allow wage rates to rise steeply in the first two deciles and then become relatively flat, by taking the same Pareto distributions as above but combining each with a uniform distribution up to the 20th percentile, a lower bound of \$20 at this point and an upper bound at the 90th percentile. The new distributions show considerable inequality at the bottom but relatively little at the top, in contrast to the actual wage distribution, and represent rising inequality over time. Average wage rates are as follows:

Distribution 2a: Average wage = \$31.98

Distribution 2b: Average wage = \$26.29

Distribution 2c: Average wage = \$22.92

Percentile wage distributions for each set are shown in Figures 3a and 3b.

Figures 3a and 3b about here

Table 1 reports the optimal tax parameters and bracket points for the first set of wage distributions. The top panel gives the results for Distribution 1a. When elasticities are constant throughout the wage distribution, at values of 0.01, 0.1, 0.2, and 0.3 respectively, the tax system is convex, with, as we would expect, values of the marginal tax rate (mtr) in both brackets falling as the elasticity increases. Likewise the value of the lump sum transfer falls as the elasticity increases, indicating that the extent of redistribution falls with increasing elasticity. We find that the nonconvex case can be obtained if the elasticity rises sufficiently across the wage distribution. We test this by setting the elasticity in the first eight deciles below that of the top two deciles. We find that we need to make the upper deciles' elasticity many times larger than that of the deciles below, as for example 0.3 to 0.01. The assumed elasticities tend to limit redistribution to *within* the lower bracket.

Table 1 about here

Increasing the inequality aversion parameter ρ raises the extent of redistribution, as indicated by lump sum transfers and general levels of tax rates, as we would again expect, but does not change the predominance of the convex tax structure.

The bracket limits consistently occur in a neighbourhood of the percentiles in which the wage distribution starts to rise steeply, again as we would expect from the theoretical conditions (27) defining the optimal brackets. Finally, in each case, holding the parameters constant, the value of the SWF under a piecewise linear tax is greater than that under a linear tax. The intuition expressed in Figures 1 and 2 earlier is confirmed, with for example in the convex cases, the lump sum and lower bracket tax rates falling significantly and upper bracket tax rates rising, also significantly.[28]

Comparing the results for the three distributions shows that reducing the degree of inequality in the underlying wage distribution reduces the general level of tax rates and lump sum transfers, i.e. reduces the extent of redistribution, but does not change the conclusions on tax structure. The globally optimal tax system continues to be piecewise linear and convex, except for the case in which the top percentiles have an elasticity that is an implausibly high multiple of that of the lower percentiles - thirty times as high in fact. In this case, redistribution is taking place within the lower bracket but hardly at all within the upper bracket, though the high lump sum tax on the upper bracket incomes corresponding to the high lower bracket tax rate helps fund a relatively large lump sum transfer, benefiting the very lowest wage types, as illustrated in Figure 2 earlier.

The main difference to the results that arises when we take the second set of distributions, as shown in Table 2, is that now we consistently obtain the nonconvex case when elasticities are constant. As elasticities rise, for example, from 0.1 to 0.3, optimal tax rates and the size of the transfer fall. Raising the inequality aversion parameter has the opposite effect. In each case the optimal bracket limit is again found in the neighbourhood of the sharp change in slope of the wage distribution, which occurs at around the 20th percentile.

Table 2 about here

It is interesting to note just how low the upper bracket tax rates are, and how sharply the extent of redistribution falls, as compared to the results given in Table 1. In terms of the theoretical characterisation of the optimum in conditions (35)-(37), we can say that there is in this case little need to redistribute *within* the subset of upper bracket workers, which will tend to keep the corresponding tax rate low, while the relatively high tax rate in the lower bracket, because it is a lump sum tax on the upper bracket, raises most of the revenue that is required to finance the lump sum payment.

Within this type of optimal tax model, the adverse incentive effects of the high rate of tax on low wage individuals' labour supplies are not given great weight, because they contribute relatively little to total tax revenue. These wage earners are compensated for low earnings by the lump sum pay-

[28] Note that all changes in parameters are revenue neutral, since the tax revenue requirement is constant at zero in all cases. Imposing a positive revenue requirement raises the general level of taxes without qualitatively changing the results on structure.

ment, funded primarily by the lump sum tax on higher wage earners. This is confirmed by the result that when the inequality aversion parameter ρ increases, the lower bracket *mtr* also increases significantly, while the lump sum payment is also sharply higher.

In a broader perspective this aspect of the results can be seen as a weakness of this type of optimal tax model. It ignores the longer term issues presented by having a class of low wage workers facing at the margin very high tax disincentives to work, even though they are compensated in consumption terms by a lump sum transfer. Among other things, this can lead to the inter-generational transmission of negative attitudes to work and acquisition of labour market qualifications. It is for this reason that we prefer to work in terms of the first set of distributions, while arguing that a broader set of policies is required to deal with the welfare of households which are at the bottom of the wage distribution for reasons of ill-health or low human capital.

5 Conclusions

Given its significance in practice, the piecewise linear tax system seems to have received disproportionately little attention in the literature on optimal income taxation. This paper contributes a simple and transparent analysis of its main characteristics. An important result is that, contrary to the results presented by Slemrod et al. (1994), for currently relevant empirical wage distributions the optimal tax structures consistently show marginal rate progressivity, giving what we have called here the convex case.

We have considered formally only the two bracket case, but it is easy to see how this can be extended to an arbitrary number of brackets. It is theoretically possible that some portions of the tax system might be convex and some nonconvex, in a way that depends on the characteristics of the wage distribution, the income distributional preferences of the tax policy maker and the way in which labour supply elasticities vary with wage type. We would argue however that the problems presented by very low-wage and long-term unemployed workers are best addressed through specific policies directed at these groups, rather than through the design of the general tax structure.

The analysis also provides an interesting perspective on tax policy in a number of countries over the past few decades, in particular in the US, UK and Australia. Cuts in tax rates at the top have been funded by higher tax rates over the range of low-to-middle incomes, and our analysis suggests that, given the substantial increases in wage inequality that have also taken place over this period, this policy can only be explained either by assumptions of unrealistically high values of earnings elasticities at the top relative to those lower down the distribution, or by strong preferences of the "social planner" for redistributing income to the already well-off.

The question of the *optimal number* of brackets is left open. Note, however, that we are not trying to find the best piecewise linear approximation to a known nonlinear tax function that is optimal in the sense of Mirrlees, in that it separates all wage types and offers each a marginal tax rate optimal for its type. Rather, we start from the position that it is practical only to pool all wage types. Given the complexity of the situation which faces the planner, in which the multi-dimensionality of the type-space rules out the practical derivation of a Mirrlees-optimal tax function, this may be the only feasible approach to designing real-world tax systems.

Appendix A

Note first that, since $S'[v(w)]$ is strictly decreasing in w, the first order condition in (24) implies that there exists a $\mathring{w} \in (w_0, w_1)$ such that $S'[v(w)] - \lambda \gtreqqless 0$ according as $w \lesseqqgtr \mathring{w}$, for all $w \in [w_0, w_1]$. Then, since $f(w) > 0$, we have

$$\int_{w_0}^{\mathring{w}} \{S'[v(w)] - \lambda\} f(w) dw = 1/2 \tag{40}$$

and

$$\int_{\mathring{w}}^{w_1} \{S'[v(w)] - \lambda\} f(w) dw = -1/2 \tag{41}$$

It then follows that

$$\int_{w_0}^{\tau} \{S'[v(w)] - \lambda\} f(w) dw > 0 \text{ for all } \tau \in [w_0, w_1) \tag{42}$$

and

$$\int_{\tau}^{w_1} \{S'[v(w)] - \lambda\} f(w) dw < 0 \text{ for all } \tau \in (w_0, w_1] \tag{43}$$

We make use of the result in (48) below.

Next, again from (24) we have that the numerator in (25) can be written as

$$\int_{C_0} (\frac{S'[v(w)]}{\lambda} - 1)[\phi(t_1^*, w) - \hat{y}^*]f(w)dw \tag{44}$$

which we have to show is negative. Recall that $C_0 = [w_0, \tilde{w})$. Then:

(a) $\tilde{w} \leq \mathring{w}$:

If $\tilde{w} < \mathring{w}$, we have $S'[v(\mathring{w})] - \lambda > 0$ and, since $\phi(t_1^*, w) - \hat{y}^* < 0$, we have the result immediately. (At $\tilde{w} = \mathring{w}$ of course the expression is zero).

Note in this case that we have

$$\int_{C_0} (\frac{S'[v(w)]}{\lambda} - 1)\phi(t_1^*, w)f(w)dw > 0 \tag{45}$$

Thus, in the absence of the second effect of t_1, acting as a lump sum tax on the consumers in $C_1 \cup C_2$, the lower bracket tax rate would be negative. Lower wage types would receive both a lump sum subsidy and a wage subsidy and these are totally funded by the upper bracket tax rate t_2. However, the fact that t_1 *is* a lump sum tax on wage types in $C_1 \cup C_2$ *always* makes this tax rate positive.

(b) $\tilde{w} > \mathring{w}$:

Since $(S'[v(w)]-\lambda)f(w) > 0$ and $\hat{y}^* - \phi(t_1^*, w) > \hat{y}^* - \phi(t_1^*, \mathring{w})$ over $[w_0, \mathring{w})$, we have

$$\int_{w_0}^{\mathring{w}} \{S'[v(w)]-\lambda\}[\hat{y}^*-\phi(t_1^*, w)]f(w)dw > \int_{w_0}^{\mathring{w}} \{S'[v(w)]-\lambda\}[\hat{y}^*-\phi(t_1^*, \mathring{w})]f(w)dw \tag{46}$$

In addition, since $(S'[v(w)]-\lambda)f(w) < 0$ and $0 < \hat{y}^* - \phi(t_1^*, w) < \hat{y}^* - \phi(t_1^*, \mathring{w})$ over $(\mathring{w}, \tilde{w})$, we have

$$\int_{\mathring{w}}^{\tilde{w}} \{S'[v(w)]-\lambda\}[\hat{y}^*-\phi(t_1^*, w)]f(w)dw > \int_{\mathring{w}}^{\tilde{w}} \{S'[v(w)]-\lambda\}[\hat{y}^*-\phi(t_1^*, \mathring{w})]f(w)dw \tag{47}$$

Adding these two inequalities gives

$$\int_{w_0}^{\tilde{w}} \{S'[v(w)]-\lambda\}[\hat{y}^*-\phi(t_1^*, w)]f(w)dw > [\hat{y}^*-\phi(t_1^*, \mathring{w})] \int_{w_0}^{\tilde{w}} \{S'[v(w)]-\lambda\}f(w)dw > 0 \tag{48}$$

where the last inequality follows from applying (42) above with $\tilde{w} \equiv \tau$. This then gives

$$\int_{C_0} (\frac{S'[v(w)]}{\lambda} - 1)[\phi(t_1^*, w) - \hat{y}^*]f(w)dw < 0 \tag{49}$$

as required.

References

[1] P F Apps and R Rees, 2009, *Public Economics and the Household*, Cambridge: Cambridge University Press.

[2] P F Apps and R Rees, 2010, "Australian Family Tax Reform and the Targeting Fallacy", *Australian Economic Review*, 43 (2), 153-175.

[3] P F Apps and R Rees, 2011, "Two Extensions to the Theory of Optimal Income Taxation", mimeo.

[4] A B Atkinson, T Piketty and E Saez, 2011, "Top Incomes in the Long Run of History", *Journal of Economic Literature,* 49(1), 3-71.

[5] R Boadway, "The Mirrlees Approach to the Theory of Economic Policy", *International Tax and Public Finance*, 5, 67-81.

[6] B Dahlby, 1998, "Progressive Taxation and the Marginal Social Cost of Public Funds", *Journal of Public Economics*, 67, 105-122.

[7] B Dahlby, 2008, *The Marginal Cost of Public Funds*, The MIT Press, Cambridge, Mass.

[8] P Diamond and E Saez, 2011, "The Case for a Progressive Tax: From Basic Research to Policy Recommendations". *Journal of Economic Perspectives* (forthcoming)

[9] J A Mirrlees, 1971, "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, 38, 175-208.

[10] S Pudney, 1989, *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes,* Basil Blackwells, Oxford.

[11] E Sadka, 1976, "On Income Distribution, Incentive Effects and Optimal Income Taxation, *Review of Economic Studies,* 261-267.

[12] E Saez, 2001, "Using Elasticities to Derive Optimal Tax Rates", *Review of Economic Studies*, 68, 205-229.

[13] E Sheshinski, 1972, “The Optimal Linear Income Tax”, *Review of Economic Studies*, 39, 297-302.

[14] E Sheshinski, 1989, "Note on the Shape of the Optimum Income Tax Schedule", *Journal of Public Economics*, 40, 201-215.

[15] J Slemrod, S Yitzhaki, J Mayshar and M Lundholm, 1994, "The Optimal Two-Bracket Linear Income Tax", *Journal of Public Economics*, 53, 269-290.

[16] N Stern, 1976, "On the Specifiation of Models of Optimum Income Taxation", *Journal of Public Economics,* 6, 123-162.

[17] M Strawczinsky, 1988, Social Insurance and the Optimum Piecewise Linear Income Tax", *Journal of Public Economics*,, 69, 371-388.

[18] H R Varian, 1980, "Redistributive Taxes as Social Insurance", *Journal of Public Economics*,, 141(1), 49-68.

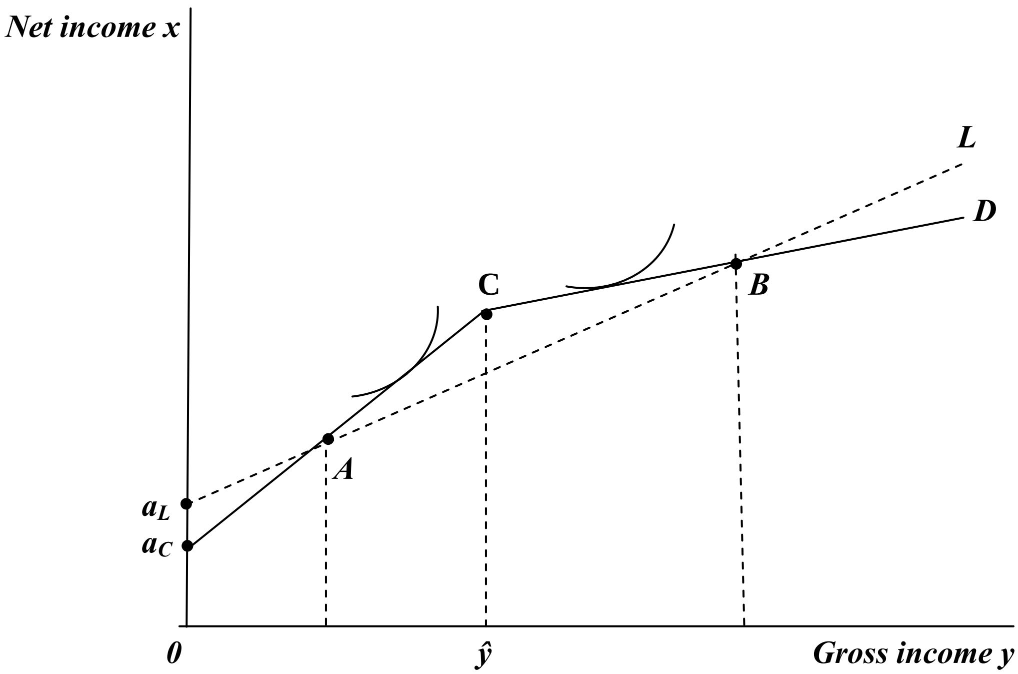

Figure 1 Linear and convex piecewise linear tax systems

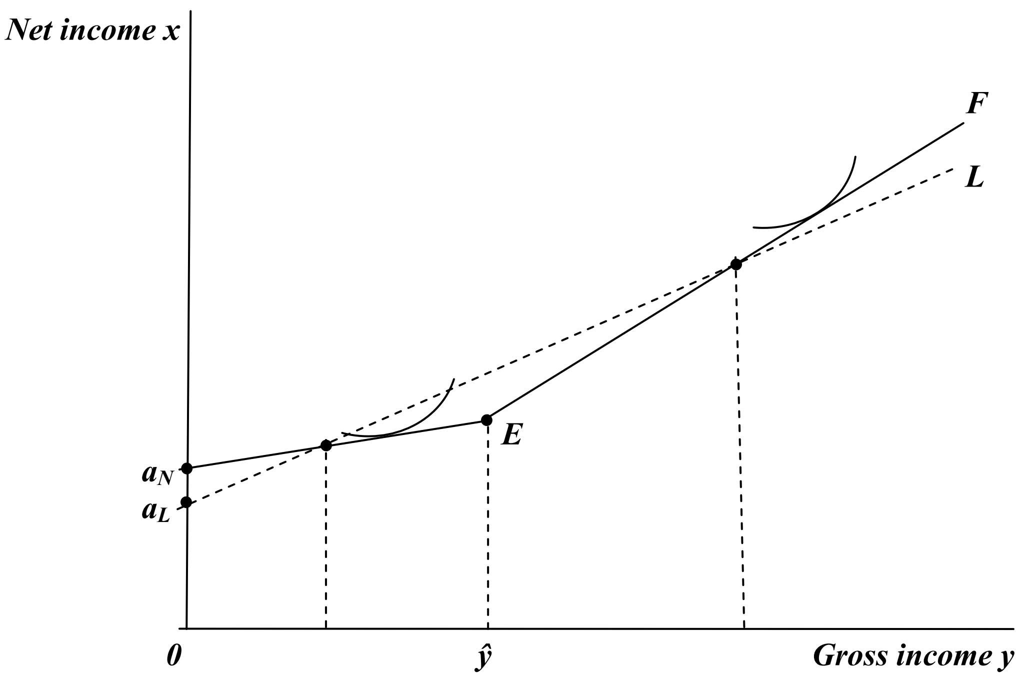

Figure 1 Linear and nonconvex piecewise linear tax systems

Figure 3a Wage Distribution Set 1

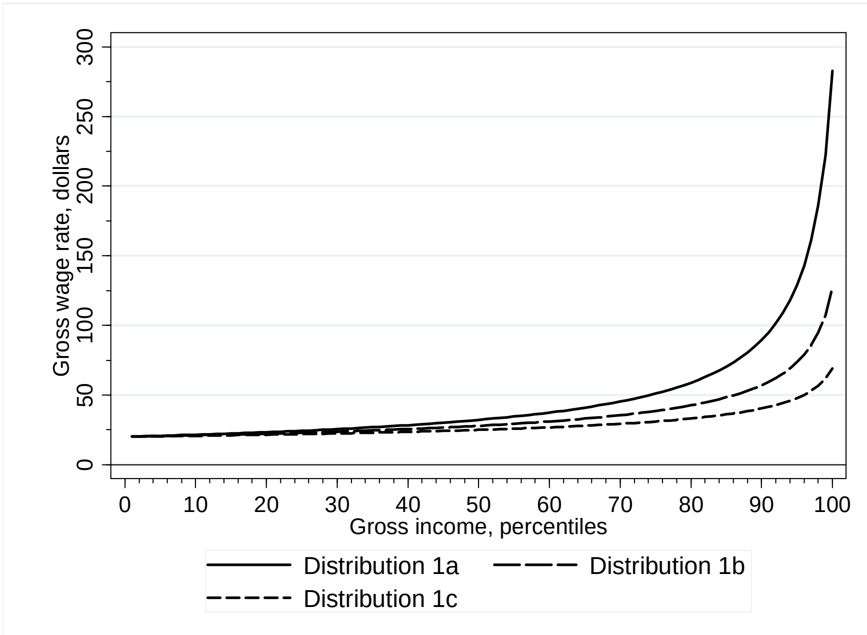

Figure 3b Wage Distribution Set 2

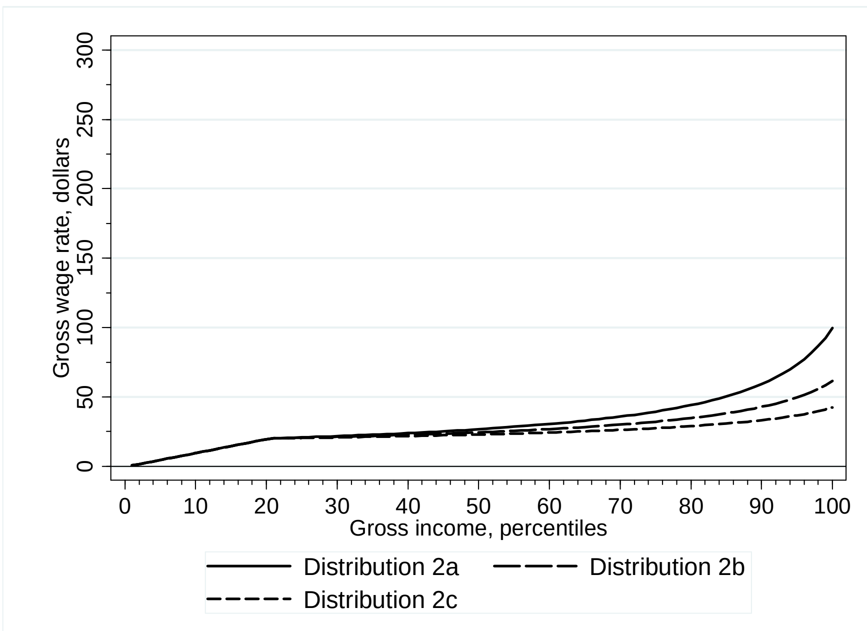

Table 1 Wage Distribution Set 1

ρ	Elasticity ε	Linear			Piecewise linear				
		mtr	Transfer	SWF/1000	mtr1	mtr2	Bracket*	Transfer	SWF/1000
Distribution 1a									
0.1	0.01	0.66	62986	15807	0.41	0.71	60	49563	15817
	0.1	0.29	27773	14698	0.13	0.34	62	20303	14712
	0.2	0.21	20580	13858	0.09	0.25	62	15619	13871
	0.3	0.18	18176	13250	0.08	0.21	64	14174	13262
	0.01/0.3	0.22	23661	15875	0.56	0.18	79	47060	15909
0.2	0.01	0.75	71355	29889	0.52	0.79	61	58724	29913
	0.1	0.41	38545	27502	0.20	0.46	61	28529	27553
	0.2	0.31	29568	25761	0.14	0.36	62	22412	25814
	0.3	0.27	26329	24495	0.11	0.31	62	19989	24549
	0.01/0.3	0.32	33656	29671	0.69	0.28	77	58672	29772
Distribution 1b									
0.1	0.01	0.53	36896	11526	0.28	0.60	54	25261	11530
	0.1	0.18	12525	10640	0.08	0.23	61	8220	10644
	0.2	0.12	8434	9874	0.05	0.16	61	5685	9877
	0.3	0.10	7117	9254	0.04	0.13	61	4831	9258
	0.01/0.3	0.14	10448	11264	0.44	0.11	80	28065	11276
0.2	0.01	0.64	44435	21803	0.38	0.70	54	32122	21817
	0.1	0.27	18570	20005	0.13	0.34	61	12607	20025
	0.2	0.20	13792	18504	0.08	0.24	59	8749	18522
	0.3	0.17	11809	17292	0.07	0.21	62	7837	17310
	0.01/0.3	0.22	16212	21206	0.54	0.17	78	34563	21250
Distribution 1c									
0.1	0.01	0.37	20885	9339	0.18	0.45	55	12579	9341
	0.1	0.09	5080	8595	0.04	0.13	62	3034	8597
	0.2	0.06	3398	7919	0.03	0.08	64	2155	7919
	0.3	0.05	2844	7354	0.02	0.07	65	1634	7355
	0.01/0.3	0.07	4115	9012	0.26	0.05	81	13944	9015
0.2	0.01	0.49	27600	17676	0.27	0.57	55	17876	17683
	0.1	0.15	8409	16225	0.07	0.20	58	5121	16231
	0.2	0.1	5614	14927	0.04	0.14	60	3232	14932
	0.3	0.08	4507	13845	0.04	0.12	63	3062	13850
	0.01/0.3	0.13	7588	17036	0.38	0.09	79	20321	17049

* Income percentile of bracket point

Table 2 Wage Distribution Set 2

ρ	Elasticity ε	Linear			Piecewise linear				
		mtr	Transfer	SWF/1000	mtr1	mtr2	Bracket*	Transfer	SWF/1000
Distribution 2a									
0.1	0.1	0.24	15237	9744	0.46	0.21	20	21213	9748
	0.2	0.17	10937	9101	0.36	0.15	20	15854	9106
	0.3	0.13	8546	8592	0.32	0.12	20	13393	8596
0.2	0.1	0.34	21283	18289	0.56	0.31	20	27287	18304
	0.2	0.25	15761	17013	0.48	0.23	20	21657	17029
	0.3	0.21	13411	16007	0.43	0.19	20	18554	16024
Distribution 2b									
0.1	0.1	0.18	9397	7991	0.45	0.14	20	17335	7999
	0.2	0.12	6321	7413	0.33	0.09	20	12241	7421
	0.3	0.1	5321	6940	0.31	0.07	20	10771	6947
0.2	0.1	0.28	14429	15035	0.56	0.21	20	22163	15064
	0.2	0.2	10335	13911	0.48	0.15	20	17783	13943
	0.3	0.16	8339	12996	0.43	0.12	20	15083	13027
Distribution 2c									
0.1	0.1	0.15	6827	6961	0.44	0.08	20	15592	6971
	0.2	0.10	4575	6434	0.35	0.05	20	11919	6444
	0.3	0.08	3685	5997	0.30	0.04	20	9920	6006
0.2	0.1	0.23	10364	13119	0.51	0.12	21	18675	13158
	0.2	0.16	7219	12105	0.47	0.09	20	15928	12146
	0.3	0.13	5889	11266	0.42	0.07	20	13729	11308

* Income percentile of bracket point