Optimal Nonlinear Income Taxes with Compensation

Robin Boadway, Craig Brett and Laurence Jacquet

UNIVERSITÉ
de Cergy-Pontoise

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

November 2015

Optimal Nonlinear Income Taxes with Compensation

Robin Boadway, Craig Brett and Laurence Jacquet

Department of Economics,
Queen's University,
Kingston ON, K7L 3N6, Canada
(email: boadwayr@econ.queensu.ca)

Department of Economics,
Mount Allison University,
144 Main Street, Sackville NB, E4L 1A7, Canada
(e-mail: cbrett@mta.ca)

THEMA University of Cergy-Pontoise,
33 Boulevard du Port,
95011 Cergy-Pontoise CEDEX, France
(email: laurence.jacquet@u-cergy.fr)

November 7, 2015

Abstract

This paper proposes a new objective function as a compromise between principles of compensation and responsibility in the design of optimal income taxes. It characterizes both first-best and second-best tax schedules. Results are interpreted in terms of both the model's primitives and taxable income elasticities.

1 Introduction

Dworkin (1981) argues that there is a cut between the characteristics of a person's environment, genetic and social, for which a person should not be held responsible for, and those for which she should be. This distinction and Dworkin's responsibility-sensitive arguments have been largely discussed in the egalitarian literature in philosophy but has also been exploited in economics. For instance, based on this cut between characteristics people are responsible for and those for which they are not, Roemer (1993, 1998), Fleurbaey (2008), and Fleurbaey and Maniquet (2009, 2011) propose distinct normative approaches to the compensation-responsibility dyad. As a first contribution, this paper offers a discussion of this literature in the light of optimal income taxation theory. This allows us to highlight the advantages and limits of previous approaches. As standard in this framework, individuals are heterogeneous in skill (for which they are assumed to be not responsible) and have heterogeneous preferences for labor (for which they are assumed to be responsible).

This paper investigates how the optimal income tax schedule should be designed when taking into consideration the compensation-responsibility couple. Building up on our survey of the previous approaches on this issue, we hope to offer a new and interesting avenue. Fleurbaey and Maniquet (2005, 2006, 2007, 2015) use an ordinal approach and build up social ordering functions which do not require any utility comparison or any cardinal measurement of utility. However, it is not always possible to characterize the tax schedule over the entire income distribution. In this paper, we aim at developing an approach to fully characterize tax policies when compensation and responsibility matter. Therefore, we assume that individual utility levels are measurable and comparable and can be aggregated in a social welfare function. This allows us to derive the optimal tax formula valid along the entire income distribution and the subjacent economic intuitions.

These assumptions are also used in Roemer (1993, 1998). Roemer's approach requires that individual welfare be equalized across skills for those people who have the same preferences for leisure and so belong to the same preference group. At the same time, there would be no redistribution between preference groups. However, the policy instruments that are available do not generally allow the government to achieve such an equality, and even if they are available the redistributive policy that would achieve it would generally depend upon the preference group. This calls for some compromise and Roemer (1998) proposes that the redistributive policy should be chosen so as to maximize the weighted sum of utilities of the worst-off individuals within each preference group, the weights being the shares in the population of the preference groups. After Fleurbaey (2008), this objective function is called the mean of mins and the way of aggregating across preference groups is called utilitarian reward. Implicitly, the latter provides an ethical criterion of what comprises the proper reward to effort. This paper builds up on Roemer's approach regarding the compensation principle since it assumes a weighted sum of utilities of the worst-off individuals within each preference group. However, differing from Roemer's approach, we will

1

allow the government to have varying attitudes towards persons of different preferences by possibly attaching distinct weights on individuals with distinct preferences for leisure.

We study how the tax schedule is modified with these weights and we also relate the weights we use to those (implicitly) used in Fleurbaey and Maniquet's works. Cuff (2000) is the first to explicitly link social weights to preference for leisure using alternative choices of cardinal utilities. Boadway et al. (2002) and Choné and Laroque (2010) show how it can be optimal to have binding upward incentive constraints (hence, negative marginal tax rates) when high preferences for leisure are associated with small social weights. In this paper, we assume a continuum of productivity levels and two values of preferences for leisure and a new objective function. We then show how changes in the binding constraints take place and, hence, the sign of the marginal tax rates can also change. We also write the tax formula in terms of income levels. The optimal tax schedule is then expressed as a function of elasticities, income densities and social marginal welfare weights applied to earnings (e.g. Saez (2001); Saez and Stantcheva (2015)).

The paper is organized as follows. In Section 2, we provide a survey of the literature. Section 3 describes the model. In Section 4, we assume that all individuals characteristics are observed by the government and characterize the optimal tax schedule. Section 5 assumes that the government cannot observe individual levels of skills and of preference for leisure (industriousness). We analyze the constraints face by the government under this informational setting and characterize the second-best optimum. All proofs that are not in the core of the paper can be found in the appendix.

2 Previous literature

In this section, we outline approaches used to obtain social orderings when individuals differ both in their skills and their preferences for leisure or work. An individual's type is a pair (β, w), where w is a skill parameter and β is a taste parameter that measures "industriousness". The skill parameter w is also the individual's wage rate. The higher is β, the less costly (in terms of utility) is work to an individual. We rely heavily on the work of Fleurbaey and Maniquet, summarized especially in Fleurbaey and Maniquet (2009, 2011). Next, we present and motivate the approach we use in this paper. We also motivate our approach relative to those of Roemer (1993, 1998) and Saez and Stantcheva (2015).

Fleurbaey and Maniquet (2009) obtain what they refer to as social ordering functions (SOFs) based on limited information assumptions and a set of minimal axioms. First, the basic information they use about individuals is restricted to individual preference orderings. Welfare is neither measurable nor comparable among individuals. This information is more restrictive than required for standard social welfare functions, but less restrictive than assumed in Arrovian social choice where only preferences over actual bundles are known to the planner. Second, SOFs must satisfy the Pareto principle, either in its strong or weak form. Third, some transfer axioms, anal-

ogous to the Pigou-Dalton transfer principle, are used to invoke fairness or equity principles. As well, some subsidiary regularity axioms are imposed, such as independence and autonomy, which are relatively non-controversial. The fairness axioms indicate in what circumstances transfers of resources yield a more preferred social ordering. Broadly speaking, they relate to two sorts of transfers: those between persons with the same preferences but different skills, and those between persons of the same skills but different preferences. The former are motivated by the principle of compensation: persons ought to be compensated for differences in access to resources arising from skill differences, which are judged to be beyond their control. According to the latter, they ought to be responsible for differences in outcomes resulting from their preferences: the principle of responsibility.

The restriction of information to preference orderings combined with the Pareto principle and the fairness-related transfer axioms leads to SOFs of the leximin sort. However, depending on the fairness axioms used, the ranking of individuals of different skills and preferences can differ. What makes the analysis particularly challenging from a normative perspective is that the principles of compensation and responsibility are in fundamental conflict: both cannot be fully satisfied at the same time, see, e.g., Fleurbaey (1994) and Bossert (1995) and the graphical proof in Boadway (2012a, p. 213-215). Some compromises must be made, and that is reflected in the fairness axioms chosen. Fleurbaey and Maniquet consider cases where compensation is fully satisfied, and responsibility is weakened, and vice versa. We focus attention on allocations that satisfy full compensation, that is, those in which all persons of a given preference type are on the same indifference curve. We now consider, in turn, two ways of relaxing full responsibility while keeping full compensation. Both ways have been proposed by Fleurbaey (2008) and Fleurbaey and Maniquet (2009, 2011).

A first way of weakening the responsibility principle relies on respecting what Fleurbaey and Maniquet define as (i) the equal-preference transfer axiom and (ii) laissez-faire selection. According to the equal-preference transfer axiom, if two individuals have the same preferences but distinct earnings, it is a social improvement (or at least not a worsening) to transfer some income from the richer to the poorer individual. This corresponds with the principle of compensation. Laissez-faire selection is a weakened version of the principle of responsibility and requires that if everyone has the same skill, the laissez-faire must be the most preferred outcome in the social ordering. Fleurbaey and Maniquet show that these two axioms help to characterize what they call the wage-equivalent leximin SOF (R^{wlex}).

According to this SOF, the well-being index of an individual at a given allocation is measured by the slope of the ray from the origin (the so-called implicit budget) tangential to the indifference curve on which the individual's allocation lies (hence the need for the assumption that the government knows individual preference orderings). This well-being index is called equivalent wage or skill. Agents whose well-being indices are lower are naturally disadvantaged and are considered to be the worst-off ones. As emphasized in Boadway (2012b) constructing implicit budget lines through

the origin is driven entirely by the need to satisfy laissez-faire selection (hence, to achieve full responsibility) in the (unlikely) event that all people actually did have the same skills. This implies that R^{wlex} favors industrious low-skilled agents over lazy low-skilled agents, even though laissez-faire eschews such favoritism. Moreover, responsibility is not satisfied for either skill-type: individuals with the same skill but different preferences are on different budget lines.

Recently, Saez and Stantcheva (2015) proposed deriving optimal income tax rates from the application of social marginal welfare weights (Saez (2001)) directly to income levels rather than to utility levels. These so-called generalized social welfare weights reflect society's views for justice and can depend on characteristics of individuals earning that level of income, e.g. their industriousness and skill.[1] The wage-equivalent leximin SOF R^{wlex} of Fleurbaey and Maniquet can also be reformulated in terms of generalized social marginal welfare weights. Since R^{wlex} focuses exclusively on individuals with the lowest skill level $w_{\min}$ who are the most industrious, i.e. who work full time, the generalized social marginal welfare weights are concentrated at their level of income. This level of income is $w_{\min}$ since Fleurbaey and Maniquet assume that labor ℓ is bounded, $0 \leq \ell \leq 1$.[2] At any income level beyond $w_{\min}$, the weights are zero (Saez and Stantcheva (2015); Fleurbaey and Maniquet (2015)). The optimal tax system therefore maximizes the net transfers to agents with the lowest skill level $w_{\min}$ who work full-time. The optimal marginal tax rate is negative for individuals whose incomes are below the income of agents with the lowest skill who work full-time and positive for incomes above (Fleurbaey and Maniquet, 2011, Theorem 11.5).

A second way Fleurbaey and Maniquet propose for relaxing the responsibility principle relies on respecting what they call the at-least-as-industrious transfer axiom, or the $\succsim^{MI}$-equal-skill transfer axiom, along with equal-preference transfer (full compensation). The $\succsim^{MI}$-equal-skill transfer states that if agents have identical skills but are on distinct budget lines, transfers from the richer to the poorer would be socially preferred only if the richer are more industrious (i.e., would choose to work more if their budget lines were the same). This therefore reduces the reward for being industrious. The so-called w_{min}-equivalent leximin SOF ($R^{w_{\min}lex}$) favors the lazy low-skilled types because transfers between people of the same skill are sanctioned only if they go from the industrious to the non-industrious.[3] This seems arbitrary since it penalizes some preferences. The $w_{\min}$-equivalent leximin SOF $R^{w_{\min}lex}$ puts full weight on those with $w_{\min}$ who receive the smallest net transfer from the government and zero (generalized) social marginal welfare weights are zero on all incomes above w_{min} (Fleurbaey and Maniquet (2015)). This leads to an optimal tax system with

---

[1]It is however not always straightforward to derive these generalized weights by income level as discussed in Fleurbaey and Maniquet (2015).

[2]The skill level $w_{\min}$ of the lowest skilled individuals who work full time therefore also stands for their income.

[3]See Boadway (2012b) for a discussion about this.

zero marginal tax rates in the income range $[0, w_{\min}]$. Therefore all individuals with income in this range receive the same transfer. The optimal tax system maximizes this transfer and has a positive marginal tax rate for income levels above $w_{\min}$.

Fleurbaey and Maniquet's approach replaces interpersonal welfare requirements with axioms about fair access to resources to derive clear-cut SOFs, but there is a price to pay. First, their orderings are of the leximin type that can preclude signing the optimal marginal income tax schedule except at the very bottom of the skill distribution in some cases (see e.g., Fleurbaey and Maniquet (2006, 2007, 2015)). Second, their approach still needs to rely on interpersonal comparisons of well-being indices. Third, their analysis depends on identifying the worst-off person and this leads to arbitrarily favoring one preference type, as previously explained. In this paper, we propose another way of satisfying compensation while relaxing responsibility. Contrary to Fleurbaey and Maniquet's approach, we assume that individual utilities are comparable and we will rely on a social welfare function. This will allow us to sign the optimal marginal tax rates along the entire skill distribution. Under full information, compensation will be satisfied thanks to a constraint we will incorporate into our welfare maximization problem: In a given preference group, everyone reaches the same utility level, regardless of skill levels. However, under asymmetric information, compensation may not be fully satisfied because of information and incentive compatibility constraints. The government then maximizes a weighted sum of the minimal utility level in each preference group.

Our approach can therefore also be related to Roemer (1993, 1998) who suggests maximizing a weighted average of the minimal utilities across individuals having the same tastes, the weights being the shares in the population of the preference groups. According to Roemer (1993, 1998), the difference between the utility levels of individuals at the same value of effort, but in different skill types, is ethically unjustifiable. In choosing an allocation, these differences should ideally be eliminated analogous to the compensation principle of Fleurbaey (2008) and Fleurbaey and Maniquet (2009). And if compensation is not fully achievable, Roemer suggests adopting the maximin criterion. Since the policies that allow compensation to be achieved (or a maximin formulation of it) in each preference group generally differ across preference group, Roemer proposes the compromise policy that averages over preference groups. Roemer's criterion is then a population-weighted average of the minimum utilities achieved in each preference type. As a result, Fleurbaey (2008) calls Roemer's criterion the mean-of-mins criterion. We will build on Roemer's work by choosing the maximin criterion in each preference group.

Moreover, according to Roemer (1993, 1998) it is perfectly acceptable if those who have higher industriousness (expend a higher degree of effort) reach higher outcomes (utility levels). The utility level should be a non-decreasing function of industriousness (β), which happens when simply aggregating across preference groups. After Fleurbaey (2008), this is called the rewards-to-effort principle and Roemer's objective function is an example of what Fleurbaey (2008) calls utilitarian reward.

In contrast to Roemer's approach, we will allow the government to have varying attitudes towards persons of different preferences by attaching distinct weights on individuals with distinct preferences for leisure. When the weight attached to industrious β_h-individuals $(1 - \gamma)$ is larger than the one attributed to β_l-individuals (γ), we will satisfy the rewards-to-effort principle. The weights we will use can be related to those used in Fleurbaey and Maniquet: The weight on the industrious individuals $(1 - \gamma)$ is implicitly higher with R^{wlex} than with $R^{w_{\min}lex}$.

Our approach allows us to capture and analyze the idea and effects of responsibility by simultaneously allowing the government to transfer income from one preference group to the other. This type of redistribution is largely unavoidable in a second-best environment we consider below. Changes in the weights will explicitly reflect the extent to which responsibility is more closely satisfied for one preference group or for the other.

3 Model

From now on, we assume that w is continuously distributed with a bounded support $[\underline{w}, \overline{w}]$ and that β takes on one of two values, $\beta_l < \beta_h$. We denote the share of β_l-type individuals by α. The distribution of w might depend on β and we let $f_l(w)$ and $f_h(w)$ be the (conditional) densities of w for β_l- and β_h-types, respectively. We assume throughout that people are responsible for their tastes for work β but not for their skills w. Utility depends on consumption, x, and labor supply and is represented by

$$\tilde{u}(x, \ell; \beta) = v(x) - h\left(\frac{\ell}{\beta}\right). \tag{1}$$

The function v is increasing, strictly concave and well-behaved. The function h is increasing, strictly convex and well-behaved. Using the usual substitution for before-tax income, $y = w\ell$, utility can be written as

$$U(x, y; w, \beta) = v(x) - h\left(\frac{y}{w\beta}\right). \tag{2}$$

Much of the literature focuses on quasilinear forms, and the separable form nests the two forms of quasi-linearity. In fact, this form can be made quasilinear-in-consumption by introducing $z = v(x)$. For future reference, define a function ϕ by

$$\phi = v^{-1}; \quad \text{that is,} \quad x = \phi(z) \longleftrightarrow z = v(x). \tag{3}$$

Because v is increasing and strictly concave, ϕ is increasing and strictly convex. Using this transformation of utility, we obtain

$$u(z, y; w, \beta) = z - h\left(\frac{y}{w\beta}\right). \tag{4}$$

6

4 First-best economy

As a benchmark, this section considers an ideal world where the government has the luxury to condition taxation on skills w and preference parameters β_i. Therefore, individual taxation can be written as $T(w, \beta_i) = y(w, \beta_i) - x(w, \beta_i)$. The tax policy aims to satisfy equal-preference transfer (full compensation for skill) but compromises on equal-skill transfer. The ideal is then to give the same level of well-being to those with the same preferences, irrespective of their skill levels. This amounts to equalizing utility levels in each group of preferences, i.e.:

$$v\left(x\left(w, \beta_i\right)\right) - h\left(\frac{y\left(w, \beta_i\right)}{w\beta_i}\right) = c_i \quad i = l, h \tag{5}$$

where c_i $(i := l, h)$ are scalars.

The tax authority maximizes a weighted sum of the minimal utility levels in each preference group

$$\gamma c_l + (1 - \gamma) c_h \tag{6}$$

with $\gamma \in [0, 1]$ and without any restriction on the weights (γ and $1 - \gamma$) that are assigned to each preference type. Government policy is purely redistributive, so we posit the budget constraint as

$$\alpha \int_{\underline{w}}^{\overline{w}} \left[y\left(w, \beta_l\right) - x\left(w, \beta_l\right)\right] f_l\left(w\right) dw$$

$$+ (1 - \alpha) \left[\int_{\underline{w}}^{\overline{w}} \left[y\left(w, \beta_h\right) - x\left(w, \beta_h\right)\right]\right] f_h(w) \, dw = 0. \tag{7}$$

The government maximizes (6) subject to the constraints of equal utility per preference group (5) and the budget constraint (7).

For simplicity in the exposition, we can neglect income effects on labor supply by using the following utility function $U(x, y) = \ln(\hat{u}(w, \beta))$ where $\hat{u}(w, \beta) = x(w, \beta) - (1/2)(y/w\beta)^2$.[4] Therefore, the constraints (5) which equalize utility levels in each group of preferences can be rewritten as

$$\ln\left(\hat{u}\left(w, \beta_l\right)\right) = \ln\left(x\left(w, \beta_l\right) - \frac{\left[y\left(w, \beta_l\right)\right]^2}{2\left(\beta_l w\right)^2}\right) = c_l \tag{8}$$

$$\ln\left(\hat{u}\left(w, \beta_h\right)\right) = \ln\left(x\left(w, \beta_h\right) - \frac{\left[y\left(w, \beta_h\right)\right]^2}{2\left(\beta_h w\right)^2}\right) = c_h. \tag{9}$$

The optimal tax schedule is then easily derived. The Lagrangian can be written as

$$L(.) = \gamma c_l + (1 - \gamma) c_h + \int_{\underline{w}}^{\overline{w}} \lambda_l\left(w\right) \left[\ln\left(x\left(w, \beta_l\right) - \frac{\left[y\left(w, \beta_l\right)\right]^2}{2\left(\beta_l w\right)^2}\right) - c_l\right] dw$$

---

[4]This form is an increasing transformation of a special case of the form used in (2).

$$+ \int_{\underline{w}}^{\overline{w}} \lambda_h \left(w \right) \left[\ln \left(x \left(w, \beta_h \right) - \frac{\left[y \left(w, \beta_h \right) \right]^2}{2 \left(\beta_h w \right)^2} \right) - c_h \right] dw$$

$$+ \eta \left\{ \alpha \int_{\underline{w}}^{\overline{w}} \left[y \left(w, \beta_l \right) - x \left(w, \beta_l \right) \right] f_l \left(w \right) dw \right.$$

$$\left. + \left(1 - \alpha \right) \left[\int_{\underline{w}}^{\overline{w}} \left[y \left(w, \beta_h \right) - x \left(w, \beta_h \right) \right] \right] f_h \left(w \right) dw. \right\}$$

where $\lambda_i \left(w \right)$ ($i := l, h$) are the Lagrangian multipliers associated with the two equalities in (5) and η is the multiplier associated with the budget constraint (7). The necessary conditions are the budget constraint (7), (8)–(9), and the first-order conditions with respect to c_l, c_h, $x \left(w, \beta_l \right)$, $x \left(w, \beta_h \right)$, $y \left(w, \beta_l \right)$ and $y \left(w, \beta_h \right)$, i.e.:

$$\gamma - \int_{\underline{w}}^{\overline{w}} \lambda_l \left(w \right) dw = 0 \tag{10}$$

$$1 - \gamma - \int_{\underline{w}}^{\overline{w}} \lambda_h \left(w \right) dw = 0 \tag{11}$$

$$\frac{\lambda_l \left(w \right)}{\widehat{u} \left(w, \beta_l \right)} - \eta \alpha f_l \left(w \right) = 0 \tag{12}$$

$$\frac{\lambda_h \left(w \right)}{\widehat{u} \left(w, \beta_h \right)} - \eta \left(1 - \alpha \right) f_h \left(w \right) = 0 \tag{13}$$

$$\frac{-\lambda_l \left(w \right)}{\widehat{u} \left(w, \beta_l \right)} \frac{y \left(w, \beta_l \right)}{\left(\beta_l w \right)^2} + \eta \alpha f_l \left(w \right) = 0 \tag{14}$$

$$\frac{-\lambda_h \left(w \right)}{\widehat{u} \left(w, \beta_h \right)} \frac{y \left(w, \beta_h \right)}{\left(\beta_h w \right)^2} + \eta \left(1 - \alpha \right) f_h \left(w \right) = 0. \tag{15}$$

From (12) and (14), we have:

$$y \left(w, \beta_l \right) = \left(\beta_l w \right)^2 \tag{16}$$

and from (13) and (15), we obtain:

$$y \left(w, \beta_h \right) = \left(\beta_h w \right)^2. \tag{17}$$

From these two equations we have the following lemma.

Lemma 1. *Under full information, more industrious people earn more at a given skill level.*

From (14) and (8) we obtain

$$x \left(w, \beta_l \right) = \kappa_l + \frac{\left(w \beta_l \right)^2}{2} \tag{18}$$

where $\kappa_l = e^{c_l}$. This leads to the following lemma that highlights the progressivity of the tax function among agents having the same preferences.

8

Lemma 2. *Under full information, the individual lump-sum tax is increasing with skill, within each preference group.*

Proof. Subtracting (18) from (16) allows the tax function to be written as

$$T(w, \beta_i) = \frac{(w\beta_i)^2}{2} - \kappa_i.$$

Therefore, $dT(w, \beta_i)/dw = w\beta_i^2 > 0$. $\square$

In the first-best economy, we know that compensation is (fully) satisfied but we are also interested in the implicit transfers between preference groups. More precisely, the following lemma highlights the impact of the welfare weights γ on the (implicit) taxes between groups of distinct preferences.

Lemma 3. *Under full information, the total (implicit) tax paid by a preference group is decreasing in the weight assigned to this group in the objective function.*

Proof. The implicit amount of tax revenue collected from individuals of preference type β_l is denoted:

$$T_l = \alpha \int_{\underline{w}}^{\overline{w}} [y(w, \beta_l) - x(w, \beta_l)] f_l(w) \, dw. \tag{19}$$

This implicit revenue level T_l may take on any sign. From (7), the total tax liability of the β_h-type group is $-T_l$. This implies that the implicit amount of tax revenue collected from individuals of preference type β_h is:

$$(1 - \alpha) \int_{\underline{w}}^{\overline{w}} [y(w, \beta_h) - x(w, \beta_h)] f_h(w) \, dw = -T_l, \tag{20}$$

Using $\widehat{u}(w, \beta_l) = x(w, \beta_l) - (1/2)(y/w\beta_l)^2$, (18) and (16) yield $\widehat{u}(w, \beta_l) = \kappa_l$. Substituting (18) and (10) into (12) implies

$$\gamma = \kappa_l \eta \alpha. \tag{21}$$

Likewise, from (11), (13), (15) and (9), we obtain

$$1 - \gamma = \kappa_h \eta (1 - \alpha). \tag{22}$$

and $\widehat{u}(w, \beta_h) = \kappa_h$. Substituting (21) and (22) into the budget constraint (7) determines η, κ_l and κ_h as functions of exogenous variables of the model:

$$\eta = \frac{2}{\alpha E_1(w^2) \beta_l^2 + (1 - \alpha) E_2(w^2) \beta_h^2}$$

$$\kappa_l = \frac{\gamma [\alpha \beta_l^2 E_1(w^2) + (1 - \alpha) \beta_h^2 E_2(w^2)]}{2\alpha}$$

9

$$\kappa_h = \frac{(1-\gamma)\left[\alpha\beta_l^2 E_1\left(w^2\right) + (1-\alpha)\beta_h^2 E_2\left(w^2\right)\right]}{2\left(1-\alpha\right)},$$

where $E_i\left(w^2\right) = \int_{\underline{w}}^{\overline{w}} w^2 f_i\left(w\right)dw$ $(i := l, h)$ is the mean of w^2 for β_i-agents.

Substituting the above expression for κ_l as well as (18), (16) into (19) yields

$$T_l = \alpha\left(1-\gamma\right)\frac{\beta_\ell^2}{2}E_l\left(w^2\right) - \gamma\left(1-\alpha\right)\frac{\beta_h^2}{2}E_h\left(w^2\right). \tag{23}$$

From the previous equation, we directly see that T_l, the total tax paid by the β_l-preference group, is decreasing in γ, the welfare weight on this group. In particular, when $\gamma = 0 : T_l = \alpha\beta_l^2 E_l\left(w^2\right)/2 > 0$ i.e. the group of β_l-preferences pays an implicit positive tax towards the β_h-group. When $\gamma = 1 : T_l = -\left(1-\alpha\right)\beta_h^2 E_h\left(w^2\right)/2 < 0$ that is, the group of β_l-preferences receives an implicit (positive) transfer from β_h-group.[5]

□

We just studied how the weights in our objective function modifies the implicit tax between preference groups. We can now look at the impact of these weights on the (common) utility level in each preference group. Equations (21) and (22) can be rewritten as

$$\frac{\kappa_h}{\kappa_l} = \frac{(1-\gamma)\,\alpha}{\gamma\left(1-\alpha\right)}. \tag{24}$$

From (24), we see that at any w, the more industrious agents obtain a larger (lower) utility level than the less industrious agents if and only if $(1-\gamma)\,\alpha > (<)\gamma\left(1-\alpha\right)$. For a given α, the higher the weight on the more industrious people $(1-\gamma)$, the larger their (relative) utility level. We recover equal utility levels, which is more in the spirit of Roemer's work, if $\alpha = \gamma$.

5   Second-best economy

The above analysis is based on the government adopting a tax function that depends on all individual characteristics. We now turn to an economy where taxation can only be conditioned on the before-tax income y. In this second-best world, the government can observe before-tax earnings but not individual skills w or preferences β_i. The government is then unable to equalize the utility levels within each preference group, as required by (5). Equal utility constraints (5) are rewritten as some minimal utility requirements

$$v\left(x\left(w,\beta_i\right)\right) - h\left(\frac{y\left(w,\beta_i\right)}{w\beta_i}\right) \geq c_i \quad i = l, h. \tag{25}$$

---

[5]When the distributions of skills are identical in both preferences groups, i.e. $\alpha = 1/2$ and $f_1\left(w\right) = f_2\left(w\right)$, we obtain that T_l takes the sign of $T_l = (1-\gamma)\beta_l^2 - \beta_h^2\gamma$. In the case where the government gives the same weight to both preference groups, $\gamma = 1 - \gamma$, it gives $T_l < 0$ so that the β_l-preference group receives a positive transfer while the β_h-preference group pays a (positive) tax. This illustrates that fully satisfying compensation somewhat favors the group of the less industrious workers. Responsibility is not fully satisfied.

Full compensation is therefore not guaranteed, in contrast with the full information economy.

It is clear from (4) that labor supply behavior is determined by the type-aggregator

$$\theta = w\beta. \tag{26}$$

We assume that $\underline{w}\beta_l < \overline{w}\beta_h$, so that the support of θ is the interval $[\underline{\theta}, \overline{\theta}] = [\underline{w}\beta_l, \overline{w}\beta_h]$.[6] With this assumption in mind, we define $\theta_1 = \underline{w}\beta_h$ and $\theta_2 = \overline{w}\beta_l$ and note that $\underline{\theta} < \theta_1 < \theta_2 < \overline{\theta}$. The conditional and unconditional density functions of the random variable θ are given by

$$g_1(\theta|\beta = \beta_l) = f_l\left(\frac{\theta}{\beta_l}\right)\frac{1}{\beta_l}, \quad \theta \in [\underline{\theta}, \theta_2]; \tag{27}$$

$$g_1(\theta|\beta = \beta_h) = f_h\left(\frac{\theta}{\beta_h}\right)\frac{1}{\beta_h}, \quad \theta \in [\theta_1, \overline{\theta}]; \tag{28}$$

$$g(\theta) = \begin{cases} \frac{\alpha}{\beta_l}f_l\left(\frac{\theta}{\beta_l}\right), & \theta \in [\underline{\theta}, \theta_1), \\ \frac{\alpha}{\beta_l}f_l\left(\frac{\theta}{\beta_l}\right) + \frac{1-\alpha}{\beta_h}f_h\left(\frac{\theta}{\beta_h}\right), & \theta \in [\theta_1, \theta_2], \\ \frac{1-\alpha}{\beta_h}f_h\left(\frac{\theta}{\beta_h}\right), & \theta \in (\theta_2, \overline{\theta}]. \end{cases} \tag{29}$$

where the Jacobian of the transformation $(\beta, \theta) \mapsto (\beta, w(\beta, \theta))$ has been used to state the conditional density functions and where the unconditional density function $g(\theta)$ is equal to $\alpha g_1(\theta|\beta = \beta_l) + (1 - \alpha)g_2(\theta|\beta = \beta_h)$ in the considered interval of θ's.

The taxation authority sets an anonymous income taxation schedule that determines the tax liability as a function of before-tax income y. In this way, after-tax income, x, and its transformation, z, are also determined. Individuals choose their most preferred allocation from the taxation schedule, denoted by $T(y)$. It is clear that, when faced with such an anonymous tax schedule, individuals with the same value of θ make the same choices. The individual of type θ maximizes (4) subject to $x(\theta) = y(\theta) - T(y(\theta))$, yielding the first-order condition:[7]

$$1 - T'(y(\theta)) = \phi'(z(\theta))h'\left(\frac{y(\theta)}{\theta}\right)\frac{1}{\theta} \tag{30}$$

where (3) and (26) have been used. It is straightforward to show that the planner cannot differentiate among individuals of the same θ. The choices made by individuals can be described as a schedule of allocations $(z(\theta), y(\theta))$. By revealed preference, these choices are incentive compatible.

The technical analysis of incentive compatibility across θ-types is standard. Define

$$V(\theta) = z(\theta) - h\left(\frac{y(\theta)}{\theta}\right). \tag{31}$$

---

[6]Our analysis can be modified in straightforward ways to accommodate the other case.

[7]We are restricting attention to continuous piecewise differentiable allocation functions $z(\theta)$ and $y(\theta)$.

Incentive compatibility requires

$$V(\theta) = \max_{\theta'} z(\theta') - h\left(\frac{y(\theta')}{\theta}\right). \tag{32}$$

The first-order (envelope) condition for incentive compatibility is

$$V'(\theta) = h'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2}. \tag{33}$$

Because h is increasing, (33) implies that utility is increasing in θ whenever incentive compatibility is satisfied. As a result, utility is increasing in w for fixed β and utility is increasing in β for fixed w. The second-order condition for incentive compatibility is that $y(\theta)$ is increasing.

We will use optimal control techniques to analyze the government's decision problem. It is helpful to follow the procedure set out by Ebert (1992) to reformulate the monotonicity constraint. To that end, define the variable $a(\theta)$ to be the derivative of $y(\theta)$. We will treat $a(\theta)$ as a control variable and $y(\theta)$ as a state variable with associated flow constraint

$$y'(\theta) = a(\theta). \tag{34}$$

The monotonicity constraint on $y(\theta)$ can be re-expressed as

$$a(\theta) \geq 0, \quad \forall \theta. \tag{35}$$

The notation set out above allows us to rewrite budget constraint (19) in the form

$$\int_{\underline{\theta}}^{\bar{\theta}} \delta_1(\theta)[y(\theta) - \phi(z(\theta))]f_l\left(\frac{\theta}{\beta_l}\right)\frac{\alpha}{\beta_l}d\theta$$
$$+ \int_{\underline{\theta}}^{\bar{\theta}} \delta_2(\theta)[y(\theta) - \phi(z(\theta))]f_h\left(\frac{\theta}{\beta_h}\right)\frac{1-\alpha}{\beta_h}d\theta = 0 \tag{36}$$

where

$$\delta_1(\theta) = \begin{cases} 1, & \theta \in [\underline{\theta}, \theta_2] \\ 0, & \theta \in (\theta_2, \bar{\theta}]. \end{cases} \tag{37}$$

and where

$$\delta_2(\theta) = \begin{cases} 1, & \theta \in [\theta_1, \bar{\theta}] \\ 0, & \theta \in [\underline{\theta}, \theta_1). \end{cases} \tag{38}$$

5.1 The Optimal Tax Problem

The planner's problem is to choose $(z(\theta), y(\theta))$ to maximize some function of the utility levels $V(\theta)$, subject to the budget constraints (36) and the incentive compatibility constraint (33), along with the requirement that $y(\theta)$ be increasing.

The planner's problem can be formulated as an optimal control problem with the level of utility $V(\theta)$ as a state variable. Its flow equation is the incentive compatibility condition (33). The definition of $V(\theta)$, namely (31), must also be included as a constraint on the optimization problem. This particular constraint must hold for all θ. In addition to $a(\theta)$, $z(\theta)$ is a control variable.

The following table sets out all the variables and constraints in the problem, and give notation for costate variables and multipliers needed to state the necessary conditions for an optimum for this problem.

Table 1: Outline of the Control Problem

Control Variables			
$a(\theta); z(\theta)$			
State Variables			
Variable	Flow Constraint	Boundary Conditions	Co-State
$V(\theta)$	(33)	free	$\kappa(\theta)$
$y(\theta)$	(34)	free	$\sigma(\theta)$
Constraints			
Name	Equation	Type	Multiplier(s)
Monotonicity	(35)	for all θ	$\rho(\theta)$
Budget	(36)	isoperimetrical	η
Utility	(31)	for all θ	$\xi(\theta)$

6 Analysis of the second-best

Like in first-best, the planner maximizes the weighted sum of the utilities of the least-skilled individuals (in each preference group), i.e. (6). From the incentive-compatibility constraints (the utility being increasing in w for fixed β), criterion (6) can be rewritten as

$$\gamma V(\underline{\theta}) + (1 - \gamma)V(\theta_1), \quad \gamma \in [0,1]. \tag{39}$$

When $\gamma = 1$, the planner maximizes $V(\underline{\theta})$; when $\gamma = 0$, the planner maximizes $V(\theta_1)$.

Given (33),

$$V(\theta_1) = \int_{\underline{\theta}}^{\theta_1} h'\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} d\theta + V(\underline{\theta}), \tag{40}$$

and it is possible to re-write (39) as

$$V(\underline{\theta}) + (1 - \gamma)\delta_0(\theta) \int_{\underline{\theta}}^{\bar{\theta}} h'\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} d\theta, \tag{41}$$

where

$$\delta_0(\theta) = \begin{cases} 1, & \theta \in [\underline{\theta}, \theta_1) \\ 0, & \theta \in [\theta_1, \bar{\theta}] \end{cases} \tag{42}$$

This objective function is of the usual form encountered in optimal control problems, apart from the additional term in the objective associated with the initial value of the state.[8]

Proposition 1. *For any θ level where there is no bunching, the optimal tax schedule satisfies:*

$$\frac{T'\left(y(\theta)\right)}{1 - T'\left(y\left(\theta\right)\right)} =$$

$$\left(1 + \frac{h''(y(\theta)/\theta)y(\theta)}{h'(y(\theta)/\theta)\theta}\right) \frac{1}{\theta q(\theta)} \frac{1}{\phi'(z(\theta))} \left[\int_\theta^{\bar\theta} \phi'(z(n))q(n)dn - \frac{(1-\gamma)\delta_0(\theta)}{\eta}\right] \quad (43)$$

with

$$\eta = \frac{1}{\int_{\underline\theta}^{\bar\theta} \phi'(z(\theta))q(\theta)\, d\theta} \quad and \quad (44)$$

$$q\left(\theta\right) = \delta_1(\theta)f_l\left(\frac{\theta}{\beta_l}\right)\frac{\alpha}{\beta_l} + \delta_2(\theta)f_h\left(\frac{\theta}{\beta_h}\right)\frac{1-\alpha}{\beta_h}, \quad (45)$$

where $q(\theta)$ is the density of population at θ.

The factors determining the optimal tax structure are presented in Equation (43) in the vein of Diamond (1998) (see Equation (10) p.86) or Saez (2001) (see Equation (25), p.227 in the appendix). However, major differences with the usual welfarist optimal tax schedule appear due to the equality of opportunity criterion. These differences should become apparent as we proceed.

The first component, $1 + [h''(y(\theta)/\theta)y(\theta)] / [h'(y(\theta)/\theta)\theta]$, is a measure of the elasticity of labor supply[9] and as such reflects an efficiency effect. The loss of revenue from increasing the marginal tax rate is larger, the larger is the efficiency effect.

The second component, $1/[\theta q\left(\theta\right)]$, can be called the density effect. It indicates that the optimal marginal tax rate is lower the higher the density of people at type aggregator θ. With single-peaked type-aggregator distributions, this term always decreases before the mode. Beyond the mode, it will either increase or decrease depending on how rapidly $q\left(\theta\right)$ falls with θ.

---

[8]This is very much like a "scrap value" problem in which value is assigned to some function of the terminal state. The problem is a special case of type of problem covered by Seierstad and Sydsæter (1987, Theorem 5, p. 185).

[9]This term can be rewritten as $[1 + \ell h''(\ell/\beta)/h'(\ell/\beta)]$ is equal to $[1 + e^u(\theta_n)]/e^c(\theta_n)$ where $e^c(\theta_n)$ and $e^u(\theta_n)$ are the compensated and uncompensated elasticities of labor supply, respectively. More precisely, using (30), $e^c(\theta_n)$ and $e^u(\theta_n)$ satisfy

$$e^c(\theta_n) = \frac{h'(\ell/\beta)}{(h''(\ell/\beta) - w_n^2 v''(x))\,\ell} > 0 \quad and \quad e^u(\theta_n) = \frac{h'(\ell) + v''(x)\theta_n^2\ell}{(h''(\ell) - \theta_n^2 v''(x))\ell}$$

where $\theta_n = \theta(1 - T'(y(\theta)))$ is the after-tax wage rate.

14

The last component, $(1/\phi'(z(\theta)))\left[\int_\theta^{\bar\theta} \phi'(z(n))q(n)\,dn - ((1-\gamma)\delta_0(\theta))/\eta\right]$, captures the influence of social preferences for income redistribution. It is the product of the marginal utility of consumption, $v'(x(.))$, i.e. $1/\phi'(z(\theta))$ from (3), and the gain in tax revenue from decreasing the marginal utility of everyone above θ by one unit net of the loss in welfare. Intuitively, suppose we reduce the utility of everyone above θ by a marginal unit (so that the FOIC constraints are still satisfied in that range). The gain in increased revenue is $\phi'(z(\theta))$ per person. Thus, the integral represents the gain in tax revenue of a marginal reduction in utility above θ, and depends on the number of people above θ. There is also a loss in welfare. The utilities of the concerned agents fall and the government takes this decrease into account for the utility level at $\theta = \theta_1$ only, since $V(\theta_1)$ is taken into account in the objective function (41).[10] More precisely, welfare is reduced by $(1-\gamma)$, evaluated at the marginal cost of public funds η, for individuals whose $\theta = \theta_1$, when determining the marginal tax rate at $\theta \in [\underline\theta, \theta_1)$ (so that $\delta_0(\theta) = 1$ in formula (43)) and welfare is not affected when looking at the marginal tax rate at $\theta \in [\theta_1, \bar\theta)$.

We now begin the process of determining the sign of the implicit marginal tax rates. Because h is increasing and convex, as a consequence of (43), an individual of type θ faces an implicit marginal tax rate $T'(y(\theta))$ of the same sign as

$$\left[\int_\theta^{\bar\theta} \phi'(z(n))q(n)\,dn - [(1-\gamma)\delta_0(\theta)]/\eta\right]. \tag{46}$$

Therefore, from (42), since $\delta_0(\theta)$ takes the value zero except when $\theta \in [\underline\theta, \theta_1)$, negative marginal tax rates never prevail for $\theta \in [\theta_1, \bar\theta]$. More precisely, for $\theta \in (\theta_1, \bar\theta)$, we have $\int_\theta^{\bar\theta} \phi'(z(n))q(n)\,dn > 0$ from (A.3), (A.4) and (A.8) and $\delta_0(\theta) = 0$ hence $T'(y(\theta)) > 0$. Moreover, at $\theta = \bar\theta$: $\int_\theta^{\bar\theta} \eta(n)\,dn = 0$ hence $T'(y(\bar\theta)) = 0$. These results are summarized in the following Lemma.

Lemma 4. *The implicit marginal tax rate faced by individuals with types strictly between θ_1 and $\bar\theta$ is positive. Moreover, individuals of type $\bar\theta$ face a zero implicit marginal tax rate.*

$T'(y(\bar\theta)) = 0$ is a standard result in the framework with heterogeneity in skills only (Sadka (1976), Seade (1977)): raising the marginal tax rate at the top above zero is suboptimal because it would distort the labor supply decision of the highest earner but would raise no revenue.

As the next proposition shows, the situation is a little more complicated for lower types. See the Appendix for its proof and for all subsequent proofs.

[10]The utility level $V(\underline\theta)$ is also taken into account into objective function (41), by $\gamma V(\underline\theta)$, however, it will never explicitly appear in tax formula (43). This can be easily explained as follows. Determining the sign of $T'(y(\theta))$ at $\theta = \underline\theta$ requires assuming a marginal reduction in utility of everyone above $\theta = \underline\theta$. Therefore, the utility of people at $\theta = \underline\theta$, $V(\underline\theta)$, is not affected and does not imply any welfare loss from those people.

15

Proposition 2.

(i) *If $\gamma = 1$ then the implicit marginal tax rate faced by all individuals, except those of type $\bar{\theta}$, is positive.*

(ii) *If $\gamma = 0$ then all unbunched individuals of types strictly between $\underline{\theta}$ and θ_1 face a negative implicit marginal tax rate, while $\underline{\theta}$-types face a zero marginal tax rate.*

Intuitively, if $\gamma = 1$, only people at the very bottom of the θ-distribution matter for the social planner. Therefore, provided the incentive compatibility constraints, this is equivalent to maximin hence, marginal tax rates are positive also at $\theta = \underline{\theta}$ (see Boadway and Jacquet (2008)).

If $\gamma = 0$ then only individuals characterized by $\theta = \theta_1$ matter for the social planner. Social weights are nil everywhere except at $\theta = \theta_1$. In this context, marginal tax rates are negative for people with $\theta \in (\underline{\theta}, \theta_1)$. The sign of the marginal tax rates indicate how the incentive constraints bind (see Boadway et al. (2002)). Assume an increase dT' in the marginal tax rate on the interval $[\theta - d\theta, \theta]$, which reduces consumption by $-dT'd\theta$ for θ in $(\theta, \bar{\theta}]$. When $\gamma = 0$, increasing the marginal tax rate at $\theta \in (\underline{\theta}, \theta_1)$, the gain in tax revenue $\int_\theta^{\bar{\theta}} \phi'(z(n))q(n)\, dn dT'd\theta$ valued at the marginal cost of public funds η is lower than the reduction in welfare (i.e. in $V(\theta_1)$), $(1 - \gamma)dT'd\theta = dT'd\theta$, since the gain in tax revenue is such that

$$\eta \int_\theta^{\bar{\theta}} \phi'(z(n))q(n)\, dn dT'd\theta < \eta \int_{\underline{\theta}}^{\bar{\theta}} \phi'(z(n))q(n)\, dn dT'd\theta = dT'd\theta$$

from (A.9). Redistribution lowers net incomes of types below θ (compared to laissez-faire) and increases those of type above θ_1. The marginal tax rates are then negative in $(\underline{\theta}, \theta_1)$ and the incentive constraints in this interval bind upwards.

Moreover, the marginal tax rate is zero at the very bottom if $\gamma = 0$ and there is no bunching at the bottom. At $\underline{\theta}$, because of the incentive constraints (33), a change in $V(\underline{\theta})$ induces a rise in V for all θ levels. This implies mechanical effects that do not necessarily cancel out at each skill level. However, because the government optimally selects $V(\underline{\theta})$, the aggregation of mechanical effects over all skill levels cancel out. There is therefore no rationale for distorting earnings at the bottom. Hence, in the absence of bunching, the lowest earnings are not distorted and the optimal marginal tax rate at the very bottom is nil.

If $\gamma \in (0,1)$, then implicit marginal tax rates are negative for types just above $\underline{\theta}$ and positive for types just below θ_1. In this range, the social planner faces countervailing incentives for any type in $(\underline{\theta}, \theta_1)$. It wishes to redistribute both downward from such a type to the benefit of $\underline{\theta}$ and upward to the benefit of θ_1.

The following proposition rearranges the first-order conditions for the government's problem in Proposition 1 to obtain a characterization of the optimal marginal

16

tax rates in terms of "sufficient statistics". This approach consists in focusing on empirical combinations of the primitives of the model that can be estimated using data, rather than considering the full economic structure (Chetty, 2009), e.g., Piketty (1997), Saez (2001), Diamond and Saez (2011), Piketty and Saez (2013), Hendren (2014), Golosov et al. (2014)). The sufficient statistics we need are: $\varepsilon(y)$, the elasticity of before-tax income y with respect to the so-called retention rate $1 - T'(y)$; $\xi(y)$, the derivative of before-tax income with respect to a lump-sum change in income; $m(y)$ and $M(y)$, income density and cumulative distribution; and an expression for the social welfare weights. Because the types above and below θ_1 are treated differently in the social objective function, income levels above and below $y(\theta_1)$ are treated asymmetrically. In order to state our results we need to define

$$\iota(y(\theta_1)) = \begin{cases} 1, & \text{if } y \leq y(\theta_1); \\ 0, & \text{if } y > y(\theta_1). \end{cases} \tag{47}$$

Throughout this exercise, we limit our attention to segments of the income tax schedule without bunching. We assume also that the allocation is such that y is differentiable in θ and that the tax function T is twice differentiable everywhere in earnings.

Proposition 3. *Under our assumed regularity conditions,*

$$\frac{T'(y)}{1 - T'(y)} = \frac{1 - M(y)}{\varepsilon(y)ym(y)} \left[1 + \frac{\int_y^{y(\bar{\theta})} T'(n)\xi(n)m(n)dn}{1 - M(y)} - \frac{v'(x(\theta_1))(1 - \gamma)}{\eta(1 - M(y))} \iota(y(\theta_1)) \right]. \tag{48}$$

The first fraction on the right-hand side of (48) characterizes the revenue-maximizing (or maximin utility) tax schedule in the absence of income effects. This part of the expression reconciles efficiency effects of a small increase in the marginal tax rate at y with the revenue gains to be had. Notably, without income effects there is a lump-sum gain in tax revenue from all individuals with income greater than y. This gain is proportional to $1 - M(y)$. But when there are income effects in labor supply, the lump-sum increases in taxes paid by those earning more than y induce increases in their respective before-tax incomes, thereby further increasing tax revenue. These are captured by the addition in square brackets in (48). The subtraction captures the effect on social welfare of the lump-sum increase in taxes paid by earners above y. When $y > y(\theta_1)$ this term disappears, because the social welfare function places no weight on individuals with a type of above θ_1, hence it places no weight on incomes above that earned by those of type θ_1. When the marginal tax rate is increased for incomes below $y(\theta_1)$, type θ_1 individuals are among those who experience a lump-sum tax increase. Their loss of welfare, normalized by the shadow value of public funds, is given by final term in (48). This final term multiplied by $1 - M(y)$ is also called marginal social welfare weight applied to earnings.

7 Second-Best in the Quasi-Linear Case

In order to derive further insights into the optimal tax distortions, to study the effects of the welfare weight γ, and to examine possibility of bunching at the optimum, we consider the case of preferences that are quasi-linear in consumption. Mathematically, this is simply a special case of the analysis of the previous section obtained by setting the function v, and hence its inverse ϕ, to be the identity function. In this case, $\eta = 1$. The following Proposition then follows immediately from Proposition 3.

Proposition 4. *When preferences are quasi-linear in consumption, then for anywhere there is no bunching, the optimal tax schedule satisfies*

$$
\frac{T'(y)}{1 - T'(y)} = \begin{cases} \frac{1 - M(y)}{\varepsilon(y) y m(y)}, & y > y(\theta_1); \\[2ex] \frac{\gamma - M(y)}{\varepsilon(y) y m(y)}, & y < y(\theta_1). \end{cases} \tag{49}
$$

The top line of (49) is exactly the formula for optimal marginal tax rates under maximin. The government objective places zero weight on the utilities of anyone with $\theta > \theta_1$. Optimal taxation for these workers amounts to optimal revenue generation. In the absence of income effects — as is the case with quasi-linear utility — it does not matter if this tax revenue is transfered downward to individuals of type θ_0 (maximin) or directed downward to both θ_0 and θ_1 types.

The bottom line of (49) makes plain the countervailing incentives prevailing for types below θ_1. The further is γ from one, the more social weight is given to the θ_1 types, the more salient is the motive for upward distribution, and the lower is the optimal marginal tax rate. In the extreme case of $\gamma = 0$, optimal tax rates are negative for unbunched individuals in the interval $(y(\underline{\theta}), y(\theta_1))$. Indeed, when $\gamma = 0$ the optimal tax rates are identical to those described by Brett and Weymark (2015) in a selfishly-optimal model if the decision-maker should happen to be of type θ_1. They show that the optimal marginal tax rates in this range are identical to those that would obtain under maximax utilities. For $\theta \in (0,1)$, the tax rates lie strictly between the maximin optimum and this maximax outcome.

Proposition 4 also demonstrates that bunching will typically occur near θ_1. When preferences are quasi-linear in consumption, before-tax income $y(\theta)$ is strictly decreasing the marginal tax rate. For $\gamma < 1$, (49) demonstrates a discrete jump in the marginal tax rate at the income of type θ_1. Consequently, before-tax incomes jumps downward at $y(\theta_1)$, in violation of the second-order conditions for incentive compatibility. For this reason, the before-tax income schedule implied by Equation (49) must be ironed and a mass of types around θ_1 must be bunched. Alternatively, one can visualize a kink in the optimal tax schedule at $y(\theta_1)$. A positive mass of workers choose to locate at that kink. An exact description of the bunching region is not our primary concern. We do emphasize, however, that the formulas in Propositions 1 and 4 hold outside of any bunching regions.

8 Conclusion

Market income is determined by both ability and effort. When individuals differ in both their willingness to provide effort and ability there is bound to be heterogeneity among people who earn the same income. In second-best settings, governments are constrained to give equal treatment to these unequals. Whether this treatment is more fitting the "unable" or the "lazy" is one of the things we have studied in this paper. We have shown that an important determinant of optimal marginal tax rates is the relative social weight placed on responsibility. The more salient is responsibility, the lower is the marginal tax rate. Indeed, we uncover a motivation for marginal wage subsidies if reponsibility is sufficiently salient in the social objective. In some sense, an emphasis on responsibility is consistent with social transfers coming in the form of wage subsidies and perhaps serving as a "hand up."

Naturally, our results depend on the precise formulation of the social objective. We believe that the objective we consider embodies a reasonable and flexible compromise between the notions compensation and responsibility. At the very least, it provides a way of testing the robustness of the compensation-based maximin principle to the introduction of some elements of responsibility. In the second-best setting, this introduction leads to a standard countervailing incentive problem over an interval of the lowest incomes while not fundamentally changing the nature of tax schedules for high income earners. This has some resonance in policy debates. One often hears calls for wage subsidies for low-income workers, but rarely for the rich.

References

Boadway, R., 2012a. From Optimal Tax Theory to Tax Policy, Retrospective and Prospective Views. MIT Press, Cambridge, Massachussetts.

Boadway, R., 2012b. Review of Marc Fleurbaey and François Maniquet, A Theory of Fairness and Social Welfare (New York: Cambridge University Press, 2011), 239pp. Journal of Economic Literature 50, 517–522.

Boadway, R., Jacquet, L., 2008. Optimal marginal and average income taxation under maximin. Journal of Economic Theory 143, 425–441.

Boadway, R., Marchand, M., Pestieau, P., del mar Racionero, M., 2002. Optimal redistribution with heterogeneous preferences for leisiure. Jounal of Public Economic Theory 4, 475–498.

Bossert, W., 1995. Redistribution mechanisms based on individual characteristics. Mathematical Social Sciences 29, 1–17.

Brett, C., Weymark, J. A., 2015. Voting over selfishly optimal nonlinear income tax schedules. Mimeo, Mount Allison University.

Chetty, R., 2009. Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. Annual Review of Economics 1, 451–488.

Choné, P., Laroque, G., 2010. Negative marginal tax rates and heterogeneity. American Economic Review 100, 2532–2547.

Cuff, K., 2000. Optimality of workfare with heterogeneous preferences. Canadian Jour-

nal of Economics 33, 149–174.

Diamond, P., Saez, E., 2011. The Case for a Progressive Tax: From Basic Research to Policy Recommendations. Journal of Economic Perspectives 25, 165–190.

Diamond, P. A., 1998. Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. American Economic Review 88, 83–95.

Dworkin, R., 1981. What is equality? Part 2: Equality of resources. Philosophy and Public Affairs 10, 283–345.

Ebert, U., 1992. A reexamination of the optimal nonlinear income tax. Journal of Public Economics 49, 47–73.

Fleurbaey, M., 1994. On fair compensation. Theory and Decision 36, 277–307.

Fleurbaey, M., 2008. Fairness, Responsibility and Welfare. Oxford University Press, Oxford.

Fleurbaey, M., Maniquet, F., 2005. Fair social orderings when agents have unequal production skills. Social Choice and Welfare 24, 93–127.

Fleurbaey, M., Maniquet, F., 2006. Fair income tax. Review of Economic Studies 73, 55–83.

Fleurbaey, M., Maniquet, F., 2007. Help the low skilled or let the hardworking thrive? A study of fairness in optimal income taxation. Journal of Public Economic Theory 9, 467–500.

Fleurbaey, M., Maniquet, F., 2009. Compensation and responsibility. In: Arrow, K. J., Sen, A. K., Suzumura, K. (Eds.), Handbook of Social Choice and Welfare. Vol. 2. Elsevier North Holland.

Fleurbaey, M., Maniquet, F., 2011. A Theory of Fairness and Social Welfare. Cambridge: Cambridge University Press.

Fleurbaey, M., Maniquet, F., 2015. Optimal taxation theory and principles of fairness. CORE Working Paper 2015/5.

Golosov, M., Tsyvinski, Werquin, N., 2014. Dynamic tax reforms. NBER Working Papers 20780.

Hendren, N., 2014. The inequality deflator: Interpersonal comparisons without a social welfare function. NBER Working Papers 20351.

Jacquet, L., Lehmann, E., Van der Linden, B., 2013. Optimal redistributive taxation with both extensive and intensive responses. Journal of Economic Theory 148, 1770–1805.

Piketty, T., 1997. La redistribution fiscale face au chômage. Revue Française d'Economie 12, 157–201.

Piketty, T., Saez, E., 2013. Optimal Labor Income Taxation. Handbook of Public Economics 5, 391–414.

Roemer, J. E., 1993. A pragmatic theory of reponsibility and fairness. Philosophy and Public Affairs 22, 146–166.

Roemer, J. E., 1998. Equality of Opportunity. Harvard University Press, Cambridge.

Sadka, E., 1976. On income distribution, incentive effects and optimal income taxation. Review of Economic Studies 43, 261–267.

Saez, E., 2001. Using elasticities to derive optimal income tax rates. Review of Economic Studies 68, 205–229.

Saez, E., Stantcheva, S., 2015. Generalized marginal social weights for optimal tax theory. NBER Working Paper [Forthcoming at the American Economic Review] 18835.

Seade, J. K., 1977. On the shape of optimal tax schedules. Journal of Public Economics 7, 203–235.

Seierstad, A., Sydsæter, K., 1987. Optimal Control Theory with Economic Applications. North-Holland, Amsterdam.

Appendix

Proof of Proposition 1. Write the Hamilton-Lagrange function for the second-best problem as

$$
\mathcal{H} = [(1-\gamma)\delta_0(\theta) + \kappa(\theta)] \, h'\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} + \sigma(\theta)a(\theta) + \rho(\theta)a(\theta)
$$
$$
+ \eta \left[\delta_1(\theta)[y(\theta) - \phi(z(\theta))]f_l\left(\frac{\theta}{\beta_l}\right)\frac{\alpha}{\beta_l} + \delta_2(\theta)[y(\theta) - \phi(z(\theta))]f_h\left(\frac{\theta}{\beta_h}\right)\frac{1-\alpha}{\beta_h} \right]
$$
$$
+ \xi(\theta)\left[V(\theta) - z(\theta) + h\left(\frac{y(\theta)}{\theta}\right) \right].
$$
(A.1)

The necessary conditions for an optimum include:

$$
\mathcal{H}_a = \sigma(\theta) + \rho(\theta) = 0;
$$
(A.2)

$$
\mathcal{H}_z = -\eta \left[\delta_1(\theta)f_l\left(\frac{\theta}{\beta_l}\right)\frac{\alpha}{\beta_l} + \delta_2(\theta)f_h\left(\frac{\theta}{\beta_h}\right)\frac{1-\alpha}{\beta_h} \right]\phi'(z(\theta)) - \xi(\theta) = 0;
$$
(A.3)

$$
\mathcal{H}_V = \xi(\theta) = -\kappa'(\theta);
$$
(A.4)

$$
\mathcal{H}_y = \frac{(1-\gamma)\delta_0 + \kappa(\theta)}{\theta^2}\left[h''\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta} + h'\left(\frac{y(\theta)}{\theta}\right) \right] + \xi(\theta)h'\left(\frac{y(\theta)}{\theta}\right)\frac{1}{\theta}
$$
$$
+ \eta \left[\delta_1(\theta)f_l\left(\frac{\theta}{\beta_l}\right)\frac{\alpha}{\beta_l} + \delta_2(\theta)f_h\left(\frac{\theta}{\beta_h}\right)\frac{1-\alpha}{\beta_h} \right] = -\sigma'(\theta);
$$
(A.5)

The transversality conditions are:[11]

$$
\kappa(\underline{\theta}) = -1, \quad \text{and} \quad \kappa(\bar{\theta}) = 0;
$$
(A.6)

$$
\sigma(\underline{\theta}) = 0, \quad \text{and} \quad \sigma(\bar{\theta}) = 0.
$$
(A.7)

Moreover, because there are no pure state constraints in this control problem, the costate variables are continuous.

---

[11]Here, we use Seierstad and Sydsæter (1987, Theorem 5, p. 185).

When there is no bunching, the monotonicity constraint on $y(\theta)$ is not active. In this circumstance $\rho(\theta) = 0$. By (A.2), $\sigma(\theta) = 0$ as well.

From (45), (A.3), and (A.4) we obtain

$$\kappa'(\theta) = \eta q(\theta)\phi'(z(\theta)). \tag{A.8}$$

Integrating both sides between $\underline{\theta}$ and $\bar{\theta}$ and using the transversality conditions (A.6), this can be rewritten as

$$\eta \int_{\underline{\theta}}^{\bar{\theta}} \phi'(z(\theta))q(\theta)\, d\theta = 1. \tag{A.9}$$

The latter equation implies that η is positive. Moreover, from (A.6) and (A.8) we also have

$$\kappa(\theta) = -\eta \int_{\theta}^{\bar{\theta}} \phi'(z(n))q(n)\, dn \tag{A.10}$$

Now, from the optimizing behavior of the private agents

$$1 - T'(y(\theta)) = \phi'(z(\theta))h'\left(\frac{y(\theta)}{\theta}\right)\frac{1}{\theta}. \tag{A.11}$$

In the absence of bunching, the right-hand side of (A.5) is zero. Substituting (A.3) into this equation and rearranging yields

$$\phi'(z(\theta))h'\left(\frac{y(\theta)}{\theta}\right)\frac{1}{\theta}$$
$$= \frac{1}{\eta q(\theta)}\left\{\frac{(1-\gamma)\delta_0 + \kappa(\theta)}{\theta^2}\left[h''\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta} + h'\left(\frac{y(\theta)}{\theta}\right)\right] + \eta q(\theta)\right\}. \tag{A.12}$$

Equating the left-hand side of (A.11) with the right-hand side of (A.12) yields

$$T'(y(\theta)) = \left[\frac{-\kappa(\theta) - (1-\gamma)\delta_0}{\theta^2 \eta q(\theta)}\right]\left[h''\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta} + h'\left(\frac{y(\theta)}{\theta}\right)\right]. \tag{A.13}$$

Substituting (A.10) into (A.13) and dividing by (A.11) yields (43) in the proposition. $\square$

Proof of Proposition 2. As stated after Proposition 1, the marginal tax rate faced by individuals of type θ takes the same sign as $\int_{\theta}^{\bar{\theta}} \phi'(z(n))q(n)\, dn - (1-\gamma)\delta_0(\theta)/\eta$ from (43), hence the same sign as $\eta \int_{\theta}^{\bar{\theta}} \phi'(z(n))q(n)\, dn - (1-\gamma)\delta_0(\theta)$.

If $\gamma = 1$, then $\eta \int_{\theta}^{\bar{\theta}} \phi'(z(n))q(n)\, dn - (1-\gamma)\delta_0(\theta) = \eta \int_{\theta}^{\bar{\theta}} \phi'(z(n))q(n)\, dn > 0$ for all types in $[\underline{\theta}, \bar{\theta})$, thereby establishing part (i).

If $\gamma = 0$, note that

$$\eta \int_{\theta}^{\bar{\theta}} \phi'(z(n))q(n)\, dn - (1-\gamma)\delta_0(\theta) = \eta \int_{\theta}^{\bar{\theta}} \phi'(z(n))q(n)\, dn - 1, \quad \theta \in [\underline{\theta}, \theta_1). \tag{A.14}$$

From (A.10), (A.6) and (A.8) we know that $-1 < \kappa(\theta) = -\eta \int_\theta^{\bar\theta} \phi'(z(n))q(n)\,dn < 0 \;\forall \theta \in (\underline\theta, \bar\theta)$ hence

$$\eta \int_\theta^{\bar\theta} \phi'(z(n))q(n)\,dn - (1-\gamma)\delta_0(\theta) < 0 \quad \text{and} \quad -\kappa(\underline\theta) - (1-\gamma)\delta_0(\underline\theta) = 0. \quad \text{(A.15)}$$

Part (ii) of the proposition follows from (A.15). □

Proof of Proposition 3. The proof consists of re-expressing the various components of (43) in terms of elasticities and the distribution of y. We first present three claims, each one re-expressing a component of the right-hand side of (43). We need some notation and definitions in order to state these claims. We note that, when viewed as arising from individual choices from a tax schedule, before-tax income depends on θ, and on the tax function $T(y)$. First, we define the elasticity of before-tax income with respect to θ,[12]

$$\alpha(\theta) = \frac{\theta}{y(\theta)} \frac{\partial y}{\partial \theta}. \quad \text{(A.16)}$$

Next, we define the elasticity of before-tax income with respect to $\rho = 1 - T'(y)$ (we use ρ as a shorthand for the income-retention rate),

$$\varepsilon(y) = \frac{\rho}{y} \frac{\partial y}{\partial \rho}. \quad \text{(A.17)}$$

Finally, we denote the level of taxation by L and define the response of before-tax income to a lump-sum increase in the level of taxation by

$$\xi(y) = \frac{\partial y}{\partial L}. \quad \text{(A.18)}$$

<u>Claim 1</u>: At any unbunched θ,

$$\theta q(\theta) = \alpha(\theta)y(\theta)m(y(\theta)). \quad \text{(A.19)}$$

Because $y(\theta)$ is everywhere non-decreasing, $Q(\theta) = M(y(\theta))$. When there is no bunching, the distribution M is differentiable and

$$q(\theta) = m(y(\theta))\frac{dy(\theta)}{d\theta}. \quad \text{(A.20)}$$

Multiplying both sides of (A.20) by θ and multiplying the right-hand side by $y(\theta)/y(\theta)$ yields

$$\theta q(\theta) = y(\theta)m(y(\theta))\frac{\theta}{y(\theta))}\frac{dy(\theta)}{d\theta} = \alpha(\theta)y(\theta)m(y(\theta)). \quad \text{(A.21)}$$

---

[12]These elasticities account for the nonlinearity of the income tax schedule as the presence of $T''(y)$ in their denominators will testify in (A.26), (A.27), (A.30) and (A.35) (Jacquet et al. (2013)).

<u>Claim 2</u>: For all θ,

$$1 + \frac{h''(y(\theta)/\theta)y(\theta)}{h'(y(\theta)/\theta)\theta} = \frac{\alpha(\theta)}{\varepsilon(y)}. \tag{A.22}$$

To prove this claim, we must compute the comparative static properties of the solution to the agents' optimization problem

$$\max_{y} \quad v\left(y - T(y)\right) - h\left(\frac{y}{\theta}\right). \tag{A.23}$$

Using our defintion of ρ, the first-order condition associated with (A.23) is

$$v'\left(y - T(y)\right)\rho - h'\left(\frac{y}{\theta}\right)\frac{1}{\theta} = 0. \tag{A.24}$$

Implicitly differentiating (A.24) with respect to y and ρ yields

$$\left[\rho^2 v''\left(y - T(y)\right) - h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - v'\left(y - T(y)\right)T''(y)\right]dy \tag{A.25}$$
$$+ v'\left(y - T(y)\right)d\rho = 0,$$

so that

$$\frac{\partial y}{\partial \rho} = \frac{v'\left(y - T(y)\right)}{h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - \rho^2 v''\left(y - T(y)\right) + v'\left(y - T(y)\right)T''(y)}. \tag{A.26}$$

Using (A.24) to re-express the numerator of right-hand side of (A.26) yields

$$\varepsilon(y) = \frac{\rho}{y}\frac{\partial y}{\partial \rho} = \frac{h'\left(\frac{y}{\theta}\right)\frac{1}{\theta}}{y\left[h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - \rho^2 v''\left(y - T(y)\right) + v'\left(y - T(y)\right)T''(y)\right]}. \tag{A.27}$$

We now turn to the computation of $\alpha(\theta)$. Implicitly differentiating (A.24) with respect to y and θ yields

$$\left[\rho^2 v''\left(y - T(y)\right) - h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - v'\left(y - T(y)\right)T''(y)\right]dy \tag{A.28}$$
$$+ \left[h'\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} + h''\left(\frac{y}{\theta}\right)\frac{y}{\theta^3}\right]d\theta = 0,$$

so that

$$\frac{\partial y}{\partial \theta} = \frac{h'\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} + h''\left(\frac{y}{\theta}\right)\frac{y}{\theta^3}}{h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - \rho^2 v''\left(y - T(y)\right) + v'\left(y - T(y)\right)T''(y)}, \tag{A.29}$$

and

$$\alpha(\theta) = \frac{\theta}{y}\frac{\partial y}{\partial \theta} = \frac{h'\left(\frac{y}{\theta}\right)\frac{1}{\theta} + h''\left(\frac{y}{\theta}\right)\frac{y}{\theta^2}}{y\left[h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - \rho^2 v''\left(y - T(y)\right) + v'\left(y - T(y)\right)T''(y)\right]}. \tag{A.30}$$

Dividing (A.30) by (A.27) yields

$$\frac{\alpha(\theta)}{\varepsilon(y)} = \frac{h'\left(\frac{y}{\theta}\right)\frac{1}{\theta} + h''\left(\frac{y}{\theta}\right)\frac{y}{\theta^2}}{h'\left(\frac{y}{\theta}\right)\frac{1}{\theta}}. \tag{A.31}$$

Simplifying (A.31) gives (A.22).

<u>Claim 3</u>: For all θ,

$$\frac{1}{\phi'(z(\theta))}\left[\int_\theta^{\bar\theta}\phi'(z(n))q(n)\,dn - \frac{(1-\gamma)\delta_0(\theta)}{\eta}\right]$$
$$=[1-M(y)] + \int_y^{y(\bar\theta)} T'(n)\xi(n)m(n)\,dn - \frac{v'(x(\theta_1))(1-\gamma)}{\eta}\iota(y(\theta_1)). \tag{A.32}$$

To establish this claim, we modify the arguments of Saez (2001) and Jacquet et al. (2013, Appendix C) to account for our non-standard objective function. Before doing so, we need one more comparative static result.

Subclaim 3a: For all y,

$$\xi(y) = -y\varepsilon(y)\frac{v''(y-T(y))}{v'(y-T(y))}. \tag{A.33}$$

Implicitly differentiating (A.24) with respect to y and L yields

$$\left[\rho^2 v''(y-T(y)) - h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - v'(y-T(y))T''(y)\right]dy$$
$$-\rho v''(y-T(y))\,dL = 0, \tag{A.34}$$

so that

$$\xi(y) = \frac{\partial y}{\partial L} = -\frac{\rho v''(y-T(y))}{h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - \rho^2 v''(y-T(y)) + v'(y-T(y))T''(y)}. \tag{A.35}$$

On the other hand, (A.26) implies that

$$-y\varepsilon(y)\frac{v''(y-T(y))}{v'(y-T(y))} =$$
$$-y\frac{\rho}{y}\frac{v'(y-T(y))}{\left[h''\left(\frac{y}{\theta}\right)\frac{1}{\theta^2} - \rho^2 v''(y-T(y)) + v'(y-T(y))T''(y)\right]}\frac{v''(y-T(y))}{v'(y-T(y))} \tag{A.36}$$

Cancelling terms on the right-hand side of (A.36) and comparing with the right-hand side of (A.35) yields (A.33).

Now, let $J(\theta)$ be the left-hand side of (A.32), and recall that $v'(x(\theta)) = 1/\phi'(z(\theta))$. Differentiating the left-hand side of (A.32) yields

$$J'(\theta) = v''(x(\theta))\frac{dx}{d\theta}\frac{J(\theta)}{v'(x(\theta))} - v'(x(\theta))\phi'(z(\theta))q(\theta) - v'(x(\theta))\frac{(1-\gamma)}{\eta}D_{\theta_1}(\theta), \quad \text{(A.37)}$$

where $D_{\theta_1}(\theta)$ is the negative of the Dirac delta function centered at θ_1.[13]

It follows from (43) and (A.22) that

$$J(\theta) = \frac{T'(y)}{1 - T'(y)}\frac{\varepsilon(y)}{\alpha(\theta)}\theta q(\theta). \quad \text{(A.38)}$$

Upon substituting (A.38) into (A.37) and recognizing that v and ϕ are inverse functions, we have

$$J'(\theta) = \frac{v''(x(\theta))}{v'(x(\theta))}\frac{dx}{d\theta}\frac{T'(y)}{1 - T'(y)}\frac{\varepsilon(y)}{\alpha(\theta)}\theta q(\theta) - q(\theta) - v'(x(\theta))\frac{(1-\gamma)}{\eta}D_{\theta_1}(\theta). \quad \text{(A.39)}$$

Substituting (A.33) into (A.39) yields,

$$J'(\theta) = -\frac{\xi(y)}{\varepsilon(y)}\frac{dx}{d\theta}\frac{T'(y)}{1 - T'(y)}\frac{\varepsilon(y)}{\alpha(\theta)}\frac{\theta q(\theta)}{y(\theta)} - q(\theta) - v'(x(\theta))\frac{(1-\gamma)}{\eta}D_{\theta_1}(\theta). \quad \text{(A.40)}$$

We now use the fact that $(1 - T'(y))(dy/d\theta) = dx/d\theta$, to conclude that

$$J'(\theta) = -T'(y)\xi(y)\frac{\theta q(\theta)}{\alpha(\theta)y(\theta)}\frac{dy}{d\theta} - q(\theta) - v'(x(\theta))\frac{(1-\gamma)}{\eta}D_{\theta_1}(\theta). \quad \text{(A.41)}$$

Finally, by (A.19), we have

$$J'(\theta) = -T'(y)\xi(y)m(y(\theta))\frac{dy}{d\theta} - q(\theta) - v'(x(\theta))\frac{(1-\gamma)}{\eta}D_{\theta_1}(\theta). \quad \text{(A.42)}$$

The transversality conditions imply that $J(\bar\theta) = 0$ so that upon integration with respect to the transformed variable $y(\theta)$, we arrive at

$$J(\theta) = \int_y^{y(\bar\theta)} T'(n)\xi(n)m(y(n))dn + [1 - M(y)] + \int_y^{y(\bar\theta)} v'(y - T(y))\frac{(1-\gamma)}{\eta}D_{y(\theta_1)}(y). \quad \text{(A.43)}$$

Because $D_{y(\theta_1)}$ is the negative of the usual Dirac measure,

$$J(\theta) = \int_y^{y(\bar\theta)} T'(n)\xi(n)m(y(n))dn + [1 - M(y)] - v'(x(\theta_1))\frac{(1-\gamma)}{\eta}\iota(y(\theta_1)), \quad \text{(A.44)}$$

which establishes Claim 3.

The proposition follows from Claims 1–3 because substituting (A.19), (A.22), and (A.32) into (43) yields (48).

$\square$

---

[13] $\delta_0(\theta)$ has a derivative equal to zero everywhere except at θ_1, where it has a downward jump. The usual Dirac delta function represents the derivate of a function whose graph is two horizontal lines with an upward jump from one to the other.