FINANCING EDUCATION USING OPTIMAL REDISTRIBUTIVE TAXATION

by

Craig Brett and John A. Weymark

Financing Education Using
Optimal Redistributive Taxation*

by

Craig Brett

*Department of Economics,
Royal Holloway, University of London,
Egham TW20 0EX, United Kingdom*
(e-mail: craig.brett@rhul.ac.uk)

and

John A. Weymark

*Department of Economics,
Box 1819, Station B, Vanderbilt University,
Nashville, TN 37235, U.S.A.*
(e-mail: john.weymark@vanderbilt.edu)

First version: June 2000
This version: May 2001

Abstract

"Financing Education Using Optimal Redistributive Taxation"

by

Craig Brett and John A. Weymark

In this article, the joint use of an income tax and public provision of education as instruments to achieve the government's distributional objectives is considered. Individuals differ in innate labour productivity and in aptitude to acquire skills through education. Actual labour productivity depends on both innate skill and the amount of education received. Using a generalized version of the Mirrlees tax problem that incorporates these features, qualitative properties of an optimal tax schedule are investigated.

Journal of Economic Literature classification numbers: D82, H21, H52.

1. Introduction

Publicly-provided education is primarily financed from general tax revenues. Typically, elementary and secondary education is provided free of charge, and the tuition for post-secondary education only covers a small fraction of the cost. As a consequence, government expenditures on education can be, and are, used to pursue redistributive goals. In this article, we consider the joint use of an income tax and public provision of education as instruments to achieve the government's distributional objectives. This is an issue that has already received considerable attention in the optimal nonlinear income tax literature. What distinguishes our analysis from previous work is the use of two parameters to characterize individuals. Individuals differ in innate labour productivity w and in aptitude to acquire skills through education a. An individual's actual labour productivity depends on both his innate skill and the amount of education received. An individual with a higher aptitude for education needs to spend less time to achieve a given amount of education. In earlier studies, only the innate labour productivity parameter has been considered.

In order to focus on the redistributive aspects of education, we only consider the role that education has to play in augmenting human capital, as reflected in labour productivity. Publicly-provided education has other uses as well. For example, it can be used to expand educational opportunities to individuals who are constrained by capital market imperfections from borrowing against future earnings. See, for example, Barham, Boadway, Marchand, and Pestieau (1995). Public education may be used to help offset underinvestment in education as the result of the time inconsistency of individual choices, as in Boadway, Marceau, and Marchand (1996). If private investment in education is unobservable, a moral hazard problem arises if future earnings and, hence, tax liabilities depend on the amount of education attained. Publicly-provided education may be used to help alleviate this moral hazard problem. See, for example, Wilson (1999). Education may also serve as a signal of labour productivity without affecting productivity, as in Arrow (1973) and Spence (1974). We do not consider the possibility of having privately-provided education, and therefore do not address issues such as the relative desirability of public versus private education, a topic considered by Grout (1983) from an optimal tax perspective.

Our model extends the standard Mirrlees (1971) formulation of the optimal nonlinear income tax problem by adding education as an additional

instrument and by having a second source of information asymmetry. Individual preferences depend on consumption, market labour, and time spent being educated. Because there are no capital market imperfections, we do not need to temporally separate education and work, and so, following Hare and Ulph (1979), we use a one-period model. All individuals are required to complete a basic amount of education and may undertake further discretionary education to improve their skills. The amount of compulsory education is exogenously determined.[1] As noted above, an individual's labour productivity depends on his innate skill w and on the amount of education received, and the labour required to attain a given educational standard depends on his aptitude for education a. Both the skill and educational-ability parameters are private information. There is a continuum of individuals, with the parameters w and a continuously distributed on a unit square. The government can only set taxes as a function of income. For simplicity, we assume that all education is provided free of charge. Given the tax schedule, individuals choose consumption, market labour, and time spent in education optimally. The production technology exhibits constant returns to scale. Our social planner (government) chooses the tax schedule to maximize a utilitarian social welfare function taking into account the optimal responses of the individuals to the schedule chosen and subject to the economy's overall resource constraint.

Our assumptions imply that the preferences over consumption and income, i.e., over the goods observeable by the planner, only depend on a single parameter θ (a function of w and a) that takes on values in an interval $[\underline{\theta}, \bar{\theta}]$. Further, these preferences satisfy the Mirrlees (1971) - Spence (1974) single-crossing property with respect to θ. The resource constraint expressed in terms of the publicly-observable variables includes a nonlinear term for the cost of education. Because of the presence of this term in the resource constraint, our problem fundamentally differs from the standard one-dimensional optimal tax problem.

The decisions made by individuals and by the government interact in a complex way. The incentives to work and to invest in education are affected by the choice of tax schedule. Individual educational choices, in turn, affect the distribution of wages, which, in part, determine the choice of an optimal tax schedule. In spite of these complex interactions, a number of clear-cut

---

[1]In a model in which individuals differ only in a single characteristic, Boadway and Marchand (1995) have investigated when some compulsory publicly-provided education is justified on redistributive grounds.

2

conclusions emerge from our analysis.

Because the amount of education demanded doesn't depend only on the aggregate parameter θ, it is not clear *a priori* if the same income-consumption pair should be assigned to individuals with the same value of θ. We show that it is in fact optimal to do so.

Because an increase in income increases the demand for education, the marginal rate of transformation between income and consumption is not a constant function of income. The marginal distortion faced by an individual is the difference between the marginal rate of transformation at his income level and his marginal rate of substitution. Only if the marginal rate of transformation is equal to 1 does the marginal distortion equal the marginal tax rate. The marginal rate of transformation and, hence, the marginal distortion is only well-defined for θ types who are not bunched. We demonstrate that it is optimal for all nonbunched individuals to face a positive marginal distortion.

At a solution to the planner's problem, we show that individuals of type $\bar{\theta}$ always have a nonnegative marginal tax rate and have a strictly positive marginal tax rate if they are bunched. If individuals of type $\underline{\theta}$ are not bunched with any other type, then the marginal tax rates are positive for all individuals in $(\underline{\theta}, \bar{\theta})$. If individuals of type $\underline{\theta}$ are bunched, the marginal tax rate may be negative (a marginal subsidy) in a neighbourhood of $\underline{\theta}$. Except for possibly at $\bar{\theta}$, the marginal tax rate is positive for individuals not bunched with individuals of type $\underline{\theta}$. Thus, our conclusions only partially mirror the standard results in the one-dimensional nonlinear income tax problem.

The productivity-enhancing role of education has also been analyzed in models of optimal nonlinear redistributive taxation with private information by Boadway and Marchand (1995), Grout (1983), Hare and Ulph (1979), Sheshinski (1971), Tuomala (1986), and Ulph (1977), among others. In all of these studies, individuals only differ in a single private characteristic—innate labour productivity. These models also differ from ours in a number of other respects. For example, Hare and Ulph (1979) and Sheshinski (1971) assume that individuals supply a fixed amount of labour and Tuomala (1986) has no resource cost to education. In some of these models, a uniform amount of public education is provided to everyone, whereas in other models, individuals choose their own levels of education. Many of these models permit individuals to pay for additional education. As is the case here, a major focus of this line of research has been to determine the signs of the marginal distortions and/or marginal tax rates.

3

Mirrlees (1976) considered optimal nonlinear taxation with multidimensional characteristics but, because of the complexity of his problem, was unable to say much about the properties of an optimal tax schedule. Our problem is more tractable because we are able to reduce it to a one-dimensional problem using the type aggregator θ. There are now a number of studies in which type aggregators are used to reduce a multidimensional screening problem to a one-dimensional problem in terms of the type aggregator. See, for example, Laffont, Maskin, and Rochet (1987), McAfee and McMillan (1988), Boadway, Marchand, Pestieau, and Racionero (1999), Shapiro (1999), and Rochet and Stole (2000). In these problems, a single-crossing property is satisfied and it is optimal for individuals with the same value for the type aggregator to receive the same allocation. In our problem, individuals with the same value of θ receive the same allocation of income and consumption (the goods that are publicly observable), but they do not receive the same amount of education.

We set out our model in the next section. In Section 3, we show how to convert our multidimensional problem to a simpler one-dimensional problem using the type aggregator θ. In Section 4, this one-dimensional problem is reformulated as an optimal control problem. In Section 5, we investigate the signs of the marginal distortions and marginal tax rates at a solution to the planner's problem. A final section provides some concluding remarks. A number of technical results are presented in the Appendix.

2. The model

The economy is populated by individuals who differ in innate labour productivity w. This innate skill can be augmented by education. All individuals receive a basic education from the state and may supplement this compulsory education with more advanced training. The amount of compulsory education is held fixed in our analysis, and is therefore suppressed in our notation. The amount of voluntary education acquired is e, where e is a scalar-valued measure of educational attainment. An individual with innate skill level w who acquires education in the amount e has labour productivity $h(w, e)$. As in Tuomala (1986), we assume that

$$h(w, e) = we. \tag{2.1}$$

Thus, h is increasing in both arguments and is proportional to the amount of noncompulsory education acquired.

An individual with labour productivity h supplies h units of labour in efficiency units per unit of time. There is a single consumption good that is produced using labour in efficiency units as the sole input. The technology exhibits constant returns to scale. The consumption good is our numeraire and we choose the units of measurement so that one unit of labour in efficiency units produces one unit of output. The labour market is perfectly competitive, so an individual's wage is equal to the marginal (and average) product of labour h. Before-tax income y is thus given by

$$y = h(w, e)l = wel, \tag{2.2}$$

where l is the amount of labour supplied to the market.

Individuals also differ in their aptitudes for acquiring skills through education. An individual with aptitude a who devotes s units of time to education obtains

$$e = s^{\frac{1}{2}} a^{\frac{1}{2}} \tag{2.3}$$

units of education. The source of differences in aptitude could be genetic or due to differences in social background, such as parental inputs into the education process. We assume only that the underlying causes do not exert a concomittant influence on the underlying skill, w.

Preferences are defined over consumption x, labour supply l, and time spent in formal education s. Because the consumption good is our numeraire, x is also after-tax income.[2] We assume that an individual with aptitude a has a utility function of the form

$$U^a(x, l, s) = 2x^{\frac{1}{2}} - \left(l + \frac{s^{\frac{3}{2}}}{a^{\frac{1}{2}}} \right)^2. \tag{2.4}$$

This specification of preferences is restrictive, but the restrictions are easy to interpret. Individuals with the same aptitude for education have the same preferences for consumption, market labour, and time devoted to education. Moreover, it is possible to convert time spent in education into a labour-time equivalent such that everyone has the same preferences for consumption and the total time spent working and educating oneself. Given the education technology, an individual who acquires e units of education regards the time spent educating himself as equivalent (in utility terms) to spending

$$\frac{se}{a} = \frac{s^{\frac{3}{2}}}{a^{\frac{1}{2}}} \tag{2.5}$$

---

[2]Henceforth, we refer to x as consumption and y as income.

units of time in the labour market. Thus, the ratio s/a is the labour-time equivalent of the time required to obtain an additional unit of education for an individual of aptitude a and (2.4) describes preferences of the form

$$U(x,t) = 2x^{\frac{1}{2}} - t^2, \tag{2.6}$$

where t is the total effective time spent on market labour and education.[3]

Using (2.2) and (2.3), we can rewrite (2.4) as

$$\tilde{U}^{wa}(x,y,s) = 2x^{\frac{1}{2}} - \left(\frac{y}{ws^{\frac{1}{2}}a^{\frac{1}{2}}} + \frac{s^{\frac{3}{2}}}{a^{\frac{1}{2}}} \right)^2. \tag{2.7}$$

All education is provided free of charge and is financed from general tax revenues. As a consequence, for given (x,y), s is chosen to maximize $\tilde{U}^{wa}(x,y,s)$. Because of the additive separability of the utility function, the solution only depends on y. It is given by

$$s^w(y) = \left(\frac{y}{3w} \right)^{\frac{1}{2}}.[4] \tag{2.8}$$

For a given income, individuals with the same innate labour productivities choose to spend the same time educating themselves, regardless of their aptitude for education. This is the main behavioural implication of the assumptions made on preferences and technology. It is clear from (2.2) and (2.3) that the market time required to generate a given income decreases with a if s is held constant. It is natural that an individual with a high aptitude for education chooses to the spend some of this "extra" time on leisure, thereby substituting for time spent being educated. Counterbalancing this effect is the decrease in the opportunity cost of education as aptitude for education increases. A special feature of our model is that these two effects completely

---

[3]Individuals with higher aptitudes find the time spent being educated less onerous. The parameter a is thus both a utility parameter and a productivity parameter. This double role for a is not without precedent in models of asymmetric information. In the Spence (1974) education model, labour productivity and the costs of acquiring education are perfectly correlated.

[4]We assume that the values of the parameters are such that all of the optimization problems we consider have interior solutions. In particular, we assume that everyone wants to supplement the compulsory schooling with further education and that our implicit upper bound on time spent working and in education is not binding.

offset each other. However, it should be borne in mind that, in general, individuals with the same innate skill level, but different aptitudes for education, will not choose to work the same amount, and so will have different incomes.

Using the optimized value of s found in (2.8), an individual of type (w, a) has a utility function in terms of x and y given by

$$\bar{U}^{wa}(y, x) = 2x^{\frac{1}{2}} - \frac{y^{\frac{3}{2}}}{\theta}, \tag{2.9}$$

where

$$\theta = \frac{w^{\frac{3}{2}}a}{b^2} \tag{2.10}$$

and $b = 3^{\frac{1}{4}} + 3^{-\frac{3}{4}}$. From (2.10), we see that θ is an increasing, strictly quasi-concave function of w and a. Simple computations show that θ is distributed on $T = [\underline{\theta}, \bar{\theta}] = [b^{-2}, 2^{\frac{5}{2}}b^{-2}]$. Thus, although individuals are identified by two characteristics, w and a, their preferences over consumption and income can be summarized by the single parameter θ. Following Shapiro (1999), we refer to θ as a *type aggregator*.

It is more convenient to work with the following monotonic transformation of $\bar{U}^{wa}$,

$$V^{wa}(y, x) = V^{\theta}(y, x) = 2\theta x^{\frac{1}{2}} - y^{\frac{3}{2}}. \tag{2.11}$$

The marginal rate of substitution for an individual of type θ is

$$MRS^{\theta}(y, x) = \frac{3}{2\theta}(xy)^{\frac{1}{2}}, \tag{2.12}$$

which is decreasing in θ. Therefore, in terms of x and y, preferences satisfy the Mirrlees (1971) - Spence (1974) single-crossing property with respect to θ. For a given θ, if y is increased, x must also increase if utility is to be held constant. Hence, the marginal rate of substitution is increasing in y along an indifference contour of V^{wa}.

Substituting (2.8) into (2.3), using the definition of θ, and letting $c = 3^{\frac{1}{4}}/b$, we obtain

$$e^{wa}(y) = \frac{y^{\frac{1}{4}}\theta^{\frac{1}{2}}}{wc}, \tag{2.13}$$

the optimal amount of education for a type (w, a) individual with income y.

There is a continuum of individuals distributed according to the continuous probability density function f with support $S = [1, 2] \times [1, 2]$. The

7

function f is known to our social planner, who can also observe each individual's choice of x and y. The planner cannot observe w, a, l, and s. In addition, we assume that either e is not observable or that the planner does not have the authority to base taxes on e. The planner is thus limited to choosing a nonlinear tax schedule that depends only on income. Because consumption is equal to after-tax income, we can equivalently think of the planner as choosing consumption as a function of income. Although this function only implicitly specifies taxes, we refer to it as a tax schedule.

Faced with an anonymous tax schedule, an individual of type (w, a) chooses the combination of consumption, labour supply, and time spent acquiring education that maximizes $U^a(x, l, s)$. This is equivalent to choosing consumption and income to maximize $V^{wa}(y, x)$ subject to the tax schedule, with the level of education determined by (2.13). Let $x(w, a)$ and $y(w, a)$ denote the optimized values of consumption and income and let

$$e(w, a) = e^{wa}(y(w, a)) = \frac{y(w, a)^{\frac{1}{4}} \theta^{\frac{1}{2}}}{wc}. \tag{2.14}$$

As is standard in nonlinear tax problems, we can equivalently think of the planner as selecting a menu of observable outcomes $\{x(w, a), y(w, a)\}$ based on the reports of the private information (w, a) subject to the incentive (self-selection) constraints

$$V^{wa}(y(w, a), x(w, a)) \geq V^{wa}(y(w', a'), x(w', a')), \quad \forall (w, a), (w', a') \in S. \tag{2.15}$$

In addition to these incentive constraints, the planner must satisfy the resource constraint

$$\int_1^2 \int_1^2 (x(w, a) + qe(w, a)) f(w, a) dw da + C \leq \int_1^2 \int_1^2 y(w, a) f(w, a) dw da, \tag{2.16}$$

where C is the resources (valued in terms of the numeraire) required to finance the compulsory part of education, and q is the price (again in terms of the numeraire) of a unit of discretionary education.[5] Resources spent on discretionary education are a function of educational outcomes, e, rather than a direct function of s. How much time is spent on education outside the classroom varies from one individual to the next. The opportunity cost of this time is not reflected in the resource cost of education.

---

[5]Alternatively, C can be interpreted as a fixed government revenue requirement, perhaps necessary to finance other areas of public spending.

We assume that the planner wants to maximize the utilitarian social welfare functional

$$\int_1^2 \int_1^2 \bar{U}^{wa}(y(w,a), x(w,a)) f(w,a) \, dw \, da. \qquad (2.17)$$

By adopting this specification, we are assuming that the planner views the cardinalization of utility given by (2.9) as being the appropriate one for making interpersonal comparisions. Using (2.11), the planner's objective function may be rewritten as

$$\int_1^2 \int_1^2 \frac{1}{\theta} V^{wa}(y(w,a), x(w,a)) f(w,a) \, dw \, da. \qquad (2.18)$$

3. A reparameterization of the planner's problem

As formulated in Section 2, the planner must solve an optimal nonlinear income tax problem with two-dimensional characteristics. In this section, we show that this multidimensional screening problem can be reduced to a considerably simpler one-dimensional problem using the type aggregator θ.

From the self-selection constraints, it is clear that the income-consumption pair assigned to individuals whose characteristics are aggregated into the same value of θ must lie on the same indifference contour of their shared preferences for these goods. Because the demand for education depends on w, θ, and y, one might expect that it is optimal to have y depend on both w and θ so as to minimize the cost of educating the type θ individuals. Theorem 1 demonstrates that it is in fact optimal to allocate all individuals of type θ the same income-consumption pair.

Theorem 1 *For all $(w,a), (w',a') \in S$ for which $\theta = \theta'$, it is optimal to have $(y(w,a), x(w,a)) = (y(w',a'), x(w',a'))$ at a solution to the planner's problem.*

The proof of this theorem appears in the Appendix. The basic idea of the proof is as follows. All possible values of θ can be obtained for individuals for whom a equals 1 or 2. For a given value of a, say $a = 1$, if the tax schedule is continuous at $\bar{w}$ and individuals of types $(\bar{w}, 1)$ and $(\bar{w}, \bar{a})$ have the same value of θ, then they must receive the same income-consumption pair, otherwise there must be an individual of type $(w, 1)$ with w close to $\bar{w}$ who prefers the income-consumption pair assigned to the type $(\bar{w}, \bar{a})$ individual to his own.

The values of θ for which the tax schedule is not continuous in w when a is equal to 1 or 2 is a set of measure zero. The conclusion of the theorem holds in this case as well because the resource constraint is satisfied no matter what these individuals receive.[6]

Making the obvious identification, we can therefore regard the planner as choosing a menu of outcomes $\{x(\theta), y(\theta)\}$. The incentive compatibility conditions (2.15) can then be rewritten as

$$V^\theta(y(\theta), x(\theta)) \geq V^\theta(y(\theta'), x(\theta')), \quad \forall \theta, \theta' \in T. \tag{3.1}$$

It is useful to define
$$V(\theta) = V^\theta(y(\theta), x(\theta)), \tag{3.2}$$

which measures the utility of an individual of type θ at θ's assigned allocation.

Each $(w, a) \in S$ can be identified with a $(w, \theta) \in [1, 2] \times T$, with θ given by (2.10). The variables w and θ are distributed according to the probability density
$$g(w, \theta) = f(w, a(w, \theta))J(w, \theta). \tag{3.3}$$

where $J(w, \theta)$ is the Jacobian of the transformation
$$(w, \theta) \mapsto (w, a(w, \theta)) = (w, \theta w^{-\frac{3}{2}} b^2).^7 \tag{3.4}$$

In order to compute integrals, the domain S must be reparameterized. For each $\theta \in T$, let $\underline{w}(\theta)$ and $\bar{w}(\theta)$ denote, respectively, the lowest and highest values of w along the iso-θ contour in S. The marginal density of θ is given by
$$G(\theta) = \int_{\underline{w}(\theta)}^{\bar{w}(\theta)} g(w, \theta)dw. \tag{3.5}$$

Note that $\underline{w}(\underline{\theta}) = \bar{w}(\underline{\theta})$, $\underline{w}(\bar{\theta}) = \bar{w}(\bar{\theta})$, and $\underline{w}(\theta) \neq \bar{w}(\theta)$ for all other $\theta \in T$. Hence, $G(\underline{\theta}) = G(\bar{\theta}) = 0$ and $G(\theta) > 0$ for $\theta \neq \underline{\theta}, \bar{\theta}$.

In terms of the transformed variables, the objective function (2.18) of the planner is
$$\int_{\underline{\theta}}^{\bar{\theta}} \xi(\theta)V(\theta)d\theta, \tag{3.6}$$

where
$$\xi(\theta) = \int_{\underline{w}(\theta)}^{\bar{w}(\theta)} \frac{g(w, \theta)}{\theta}dw. \tag{3.7}$$

---

[6]Our proof strategy requires that the set of values for either w or a is a continuum.

[7]The Jacobian is $J(w, \theta) = b^2 w^{-\frac{3}{2}}$.

$\xi(\theta)$ is the aggregate welfare weight assigned to the type θ individuals.

Using (2.13), the resource constraint (2.16) can be rewritten as

$$\int_{\underline{\theta}}^{\bar{\theta}} \Big[(y(\theta) - x(\theta))G(\theta) - \frac{q}{c}\theta^{\frac{1}{2}}y(\theta)^{\frac{1}{4}}W(\theta) \Big] d\theta \geq C, \tag{3.8}$$

where

$$W(\theta) = \int_{\underline{w}(\theta)}^{\bar{w}(\theta)} \frac{1}{w}g(w,\theta)dw. \tag{3.9}$$

Note that $W(\underline{\theta}) = W(\bar{\theta}) = 0$ and $W(\theta) > 0$ for all other $\theta \in T$.

The planner's problem has thus been reduced to choosing the menu $\{x(\theta), y(\theta)\}$ to maximize (3.6) subject to (3.1) and (3.8). This is a one-dimensional problem, with individuals characterized by the type aggregator θ. Further, preferences satisfy the standard single-crossing property in terms of this parameter. What is nonstandard is the presence of the cost of education in the resource constraint. In addition to being nonlinear, this term is integrated using a different density function than the other terms in this constraint.[8] A further difference with what is typically assumed in one-dimensional models is that the density is zero at the endpoints of the distribution. As we shall see, these differences result in an optimal tax schedule that is qualitatively different from the ones found when the type space is one-dimensional.

4. The planner's optimal control problem

In this section, we recast the planner's problem as a optimal control problem, in much the same way that Brito and Oakland (1977), Ebert (1992), Diamond (1998), and Boadway, Cuff, and Marchand (2000), among others, treat the standard nonlinear tax problem.[9]

In order to use optimal control techniques to analyze the planner's problem, we need to reformulate the self-selection constraints (3.1). From the proof of Theorem 1, we know that (3.1) implies that both $x(\cdot)$ and $y(\cdot)$ are nondecreasing and, hence, continuous almost everywhere. Necessary and sufficient conditions for (3.1) to hold are provided in Lemma 1.

---

[8]Nonlinear resource constraints appear in, for example, Mirrlees (1971) and Brito and Oakland (1977), but in their models each term in the constraint is integrated using the same density function.

[9]There is also a long tradition, going back to Mirrlees (1971, 1976), of using a calculus of variations approach to analyze nonlinear taxation problems.

Lemma 1 *The incentive compatibility conditions* (3.1) *are satisfied if and only if* $V(\cdot)$ *is convex and for almost all* $\theta \in T$,

$$V_\theta(\theta) = 2x(\theta)^{\frac{1}{2}}. \tag{4.1}$$

Moreover, $V(\cdot)$ *is convex if and only if* $x(\cdot)$ *is nondecreasing.*

Proof. This lemma follows from Propositions 1 and 2 in Rochet (1987). To apply Rochet's results, note that $2x^{\frac{1}{2}}$ is the derivative of $V^\theta(y, x)$ with respect to θ, $V^\theta(\cdot)$ is a quasilinear function of x and $y^{\frac{3}{2}}$, and the first term of $V^\theta(\cdot)$ is linear in θ. ∎

In view of Lemma 1, the planner's optimal nonlinear taxation problem can be restated as that of choosing a menu of actions $\{x(\theta), y(\theta)\}$ with $x(\cdot)$ nondecreasing to maximize (3.6) subject to (3.8) and (4.1).

Because the incentive compatibility conditions impose a constraint on the derivative of $V(\theta)$, which is a function of $x(\theta)$ [see (4.1)], it is convenient to treat both $V(\theta)$ and $x(\theta)$ as state variables. We assume that $x_\theta(\cdot)$ is piecewise continuous and, hence, that $x(\cdot)$ and $V(\cdot)$ are continuous and piecewise differentiable. Because the self-selection constraints imply that $x(\cdot)$ is continuous almost everywhere, our differentiability assumption is relatively weak.

We define the variable $z(\theta) = x_\theta(\theta)$ and treat it as a control variable. This allows us to write the monotonicity condition as the inequality constraint $z(\theta) \geq 0$. The problem has another control variable, $y(\theta)$.[10]

Formally, the planner's optimal control problem is

$$\max_{y(\cdot), z(\cdot), x(\cdot), V(\cdot)} \quad \int_{\underline{\theta}}^{\bar{\theta}} \xi(\theta) V(\theta) d\theta \tag{4.2}$$

subject to (3.8), (4.1), and the additional constraints

$$x_\theta(\theta) = z(\theta); \tag{4.3}$$

$$z(\theta) \geq 0; \tag{4.4}$$

$$V(\theta) = 2\theta x(\theta)^{\frac{1}{2}} - y(\theta)^{\frac{3}{2}}. \tag{4.5}$$

The last constraint is just the definition of $V(\theta)$.

---

[10]Because preferences are not quasilinear in y, we cannot employ the techniques of Lolliver and Rochet (1983) and Boadway, Cuff, and Marchand (2000) to eliminate this variable from the problem.

In order to state the Hamiltonian for this program, we associate costate variables $\kappa(\theta)$ and $\delta(\theta)$ with the state variables $V(\theta)$ and $x(\theta)$, respectively. We associate with (4.4) and (4.5), respectively, the multipliers $\rho(\theta)$ and $\pi(\theta)$. The multiplier for the resource constraint (3.8) is η, the marginal cost of public funds. Thus, the Hamiltonian can be written as

$$
\begin{aligned}
\mathcal{H}(\theta) = \quad & \xi(\theta)V(\theta) + \kappa(\theta)2x(\theta)^{\frac{1}{2}} + \pi(\theta)[V(\theta) - 2\theta x(\theta)^{\frac{1}{2}} + y(\theta)^{\frac{3}{2}}] + \\
& \eta\Big[(y(\theta) - x(\theta))G(\theta) - \tfrac{q}{c}\theta^{\frac{1}{2}}y(\theta)^{\frac{1}{4}}W(\theta)\Big] + \delta(\theta)z(\theta) + \rho(\theta)z(\theta).
\end{aligned}
\tag{4.6}
$$

The necessary conditions for optimality include:

$$
\mathcal{H}_y(\theta) = \frac{3}{2}\pi(\theta)y(\theta)^{\frac{1}{2}} + \eta\Big[G(\theta) - \frac{q}{4c}\theta^{\frac{1}{2}}y(\theta)^{-\frac{3}{4}}W(\theta)\Big] = 0;
\tag{4.7}
$$

$$
\mathcal{H}_z(\theta) = \delta(\theta) + \rho(\theta) = 0;
\tag{4.8}
$$

$$
\mathcal{H}_x(\theta) = [\kappa(\theta) - \theta\pi(\theta)]x(\theta)^{-\frac{1}{2}} - \eta G(\theta) = -\delta_\theta(\theta);
\tag{4.9}
$$

$$
\mathcal{H}_V(\theta) = \xi(\theta) + \pi(\theta) = -\kappa_\theta(\theta);
\tag{4.10}
$$

where (4.9) and (4.10) hold except at points of discontinuity in the controls.

Because there are no constraints on the state variables, the costate variables must be continuous.[11] A discontinuity in the derivative of a costate variable can only occur at a point of non-differentiability in the controls.

Furthermore, the following transversality conditions must be satisfied:

$$
\kappa(\underline{\theta}) = \kappa(\bar{\theta}) = 0;
\tag{4.11}
$$

$$
\delta(\underline{\theta}) = \delta(\bar{\theta}) = 0.
\tag{4.12}
$$

The inequality constraints require that the following complementary slackness conditions be satisfied:

$$
\rho(\theta)z(\theta) = 0, \quad \rho(\theta) \geq 0, \quad \forall \theta \in T;
\tag{4.13}
$$

$$
\eta\left(\int_{\underline{\theta}}^{\bar{\theta}}\Big[(y(\theta) - x(\theta))G(\theta) - \frac{q}{c}\theta^{\frac{1}{2}}y(\theta)^{\frac{1}{4}}W(\theta)\Big]d\theta - C\right) = 0, \quad \eta \geq 0.
\tag{4.14}
$$

---

[11]Because the monotonicity constraint imposes restrictions on the derivative of the state variable $x(\cdot)$, we have been able to reformulate it as a constraint on the control variable $z(\cdot)$. Thus, the planner's problem has no pure state constraints. In such circumstances, the necessary conditions for optimality include continuity of the costate variables. See Seierstad and Sydsæter (1977, Note 5, pp. 375–376) for a discussion of this issue.

We note in passing that no further restrictions are imposed on the signs of $\pi(\cdot)$, $\delta(\cdot)$, and $\kappa(\cdot)$. In the next section, we use (4.7)–(4.14) to deduce some further properties of these functions.

The individuals with types in the set $\Theta \subseteq T$ are θ-*bunched* if they are all allocated the same income-consumption pair. Bunching occurs when the second-order condition (4.4) is binding. Because $x(\cdot)$ is continuous and non-decreasing, a bunching region Θ must be a closed interval. We assume that T is not a single bunching interval at the solution to the planner's problem. The continuity of $x(\cdot)$ (and, hence, of $y(\cdot)$) and the curvature and monotonicity properties of the individual preferences imply that in the interior of T, discontinuities in $x_\theta(\cdot)$ only occur at the endpoints of bunching regions. It is at these points that the costate variables need not be differentiable. Let N be the set of θ-types in $(\underline{\theta}, \bar{\theta})$ who are not θ-bunched.

5. Signing the marginal distortions and marginal tax rates

In this section, for each type θ, we consider the sign of the marginal income tax rate at the solution to the planner's problem. For $\theta \in N$, we also consider the sign of the marginal distortion between this type's marginal rate of substitution and the marginal rate at which income can be transformed into consumption along the economy's resource constraint.[12] In the standard nonlinear tax model with a constant-returns-to-scale technology, there is no distinction between the marginal tax rate and the marginal distortion. In our model, the marginal rate of transformation is not constant; a one-unit increase in income induces extra investment in education, which needs to be financed by the government. This link between before-tax income and publicly-financed education introduces a nonlinearity into the government's resource constraint.

The marginal tax rate of a type θ individual is given by

$$MTR(\theta) = 1 - MRS(\theta) = 1 - \frac{3(x(\theta)y(\theta))^{\frac{1}{2}}}{2\theta}, \tag{5.1}$$

where

$$MRS(\theta) = MRS^\theta(y(\theta), x(\theta)). \tag{5.2}$$

---

[12] At kinks in the tax schedule, the marginal rate of transformation is not well-defined.

14

This definition is equivalent to the one obtained by taking one minus the slope of the tax schedule whenever this slope is well-defined. *A priori*, an individual's marginal tax rate could be of either sign.

For all $\theta \in N$, the marginal rate of transformation is computed by taking the derivative of $x(\theta)$ with respect to $y(\theta)$ along the boundary of the resource constraint (3.8). Thus,

$$MRT(\theta) = \frac{G(\theta) - \frac{q}{4c}\theta^{\frac{1}{2}}y(\theta)^{-\frac{3}{4}}W(\theta)}{G(\theta)}, \quad \forall \theta \in N. \tag{5.3}$$

The marginal distortion is the difference between the marginal rate of transformation and the marginal rate of substitution,

$$MD(\theta) = \frac{G(\theta) - \frac{q}{4c}\theta^{\frac{1}{2}}y(\theta)^{-\frac{3}{4}}W(\theta)}{G(\theta)} - \frac{3(x(\theta)y(\theta))^{\frac{1}{2}}}{2\theta}, \quad \forall \theta \in N. \tag{5.4}$$

Comparing (5.1) with (5.4), we see that for all $\theta \in N$, the size of the marginal distortion is less than the marginal tax rate.

If individuals of type θ are not bunched, then (4.8) and the complementary slackness condition (4.13) imply that $\delta(\theta) = 0$. As this is true for some neighbourhood of θ, we also have $\delta_\theta(\theta) = 0$. It then follows from (4.7) and (4.9) that, at an optimum,

$$\frac{3(x(\theta)y(\theta))^{\frac{1}{2}}}{2\theta} = \frac{\eta[G(\theta) - \frac{q}{4c}\theta^{\frac{1}{2}}y(\theta)^{-\frac{3}{4}}W(\theta)]}{\eta G(\theta) - \kappa(\theta)x(\theta)^{-\frac{1}{2}}}, \quad \forall \theta \in N, \tag{5.5}$$

provided that the denominator on the right-hand-side of (5.5) is nonzero.[13] The left-hand-side of (5.5) is positive. Hence, the numerator on the right-hand-side of (5.5) is positive if the denominator is positive. Consequently, agents of type $\theta \in N$ face a positive marginal distortion if $\kappa(\theta) < 0$.

In the full-information solution, we would not have the two incentive constraints, (4.1) and (4.4), and $\kappa(\theta)x(\theta)^{-\frac{1}{2}}$ would not appear in the optimality condition (5.5). This condition would then say that the marginal rate of substitution should equal the marginal rate of transformation; i.e., there are no marginal distortions.

In order to sign the distortions, we need to investigate the properties of $\kappa(\cdot)$. Substituting (4.10) into (4.9) and multiplying by $x(\theta)^{\frac{1}{2}}$ yields

$$\kappa(\theta) + \theta\kappa_\theta(\theta) = \eta G(\theta)x(\theta)^{\frac{1}{2}} - \theta\xi(\theta) - \delta_\theta(\theta)x(\theta)^{\frac{1}{2}}. \tag{5.6}$$

---

[13]Below we show that this is in fact the case.

The left-hand-side of (5.6) is the differential of $\theta\kappa(\theta)$. Thus,

$$\theta\kappa(\theta) = \int_{\underline{\theta}}^{\theta} [\eta G(\nu)x(\nu)^{\frac{1}{2}} - \nu\xi(\nu) - \delta_{\nu}(\nu)x(\nu)^{\frac{1}{2}}]d\nu. \qquad (5.7)$$

From (3.5) and (3.7), $\xi(\theta) = G(\theta)/\theta$. Thus, (5.7) implies

$$\theta\kappa(\theta) = \int_{\underline{\theta}}^{\theta} [(\eta x(\nu)^{\frac{1}{2}} - 1)G(\nu) - \delta_{\nu}(\nu)x(\nu)^{\frac{1}{2}}]d\nu. \qquad (5.8)$$

By (4.11), $\kappa(\bar{\theta}) = 0$. Evaluating (5.8) at $\bar{\theta}$ and solving for η yields

$$\eta = \frac{1 + \int_{\underline{\theta}}^{\bar{\theta}} \delta_{\nu}(\nu)x(\nu)^{\frac{1}{2}}d\nu}{\int_{\underline{\theta}}^{\bar{\theta}} x(\nu)^{\frac{1}{2}}G(\nu)d\nu}. \qquad (5.9)$$

Lemma A.1 in the Appendix shows that $\int_{\theta_1}^{\theta_2} \delta_{\nu}(\nu)x(\nu)^{\frac{1}{2}}d\nu = 0$ on any bunching interval $[\theta_1, \theta_2]$. We have already seen that $\delta(\theta) = 0$ when θ is not bunched. Hence, the second term in the numerator of (5.9) is 0, and thus

$$\eta = \frac{1}{\int_{\underline{\theta}}^{\bar{\theta}} x(\nu)^{\frac{1}{2}}G(\nu)d\nu}. \qquad (5.10)$$

We have assumed that $x(\theta)$ is positive for all θ. Therefore, the marginal cost of public funds is positive. Moreover,

$$\frac{1}{\int_{\underline{\theta}}^{\bar{\theta}} x(\nu)^{\frac{1}{2}}G(\nu)d\nu} \leq \frac{1}{[\int_{\underline{\theta}}^{\bar{\theta}} x(\nu)^{-\frac{1}{2}}G(\nu)d\nu]^{-1}} = \int_{\underline{\theta}}^{\bar{\theta}} x(\nu)^{-\frac{1}{2}}G(\nu)d\nu.^{14} \qquad (5.11)$$

Thus, the shadow value of government revenue is less than the average marginal utility of consumption, according to the cardinalization of utility given by (2.9), at a second-best optimum.[15]

Because $x(\cdot)$ is nondecreasing, so is $\eta x(\cdot)^{\frac{1}{2}} - 1$. The continuity of $x(\cdot)$ and (5.10) imply that there is a maximal interval $[\tilde{\theta}, \hat{\theta}]$ in which $\eta x(\theta)^{\frac{1}{2}} = 1$ for all $\theta \in [\tilde{\theta}, \hat{\theta}]$. This interval may be a single point, in which case $\tilde{\theta} = \hat{\theta}$.

---

[14]It follows from Jensen's inequality that $\left[\int_{\underline{\theta}}^{\bar{\theta}} x(\nu)^{-\frac{1}{2}}G(\nu)d\nu\right]^{-2} \leq \left[\int_{\underline{\theta}}^{\bar{\theta}} x(\nu)^{\frac{1}{2}}G(\nu)d\nu\right]^{2}$. See Marshall and Olkin (1979, p. 455). Squaring both sides of this inequality yields (5.11) after some manipulation.

[15]Sandmo (1998) reports that a similar relationship holds in a finite-type model in which a public good is financed using linear taxation.

In order to gain some insight about the marginal distortions, it is useful to consider first the case in which the second-order condition (4.4) is everywhere non-binding. In this circumstance, agents of different θ-types receive different allocations. Hence, $N = (\underline{\theta}, \bar{\theta})$ and $\tilde{\theta} = \hat{\theta}$.

In this special case, (5.8) reduces to

$$\theta\kappa(\theta) = \int_{\underline{\theta}}^{\theta} (\eta x(\nu)^{\frac{1}{2}} - 1)G(\nu)d\nu. \tag{5.12}$$

By assumption, $x(\cdot)$ is increasing. Thus, the term in parentheses in the integrand of (5.12) is also increasing. By (4.11), the integral in (5.12) is 0 when evaluated at $\theta = \bar{\theta}$. Because $G(\theta)$ is positive for $\theta \neq \underline{\theta}, \bar{\theta}$, it follows that $\theta\kappa(\theta) < 0$ for all $\theta \in (\underline{\theta}, \tilde{\theta}]$. To sign the distortion term for higher values of θ, we consider

$$\frac{d}{d\theta}[\theta\kappa(\theta)] = (\eta x(\theta)^{\frac{1}{2}} - 1)G(\theta). \tag{5.13}$$

For any $\theta > \hat{\theta} = \tilde{\theta}$, the right-hand-side of (5.13) is positive. Thus, $\theta\kappa(\theta)$ is increasing over this range, reaching 0 at $\bar{\theta}$. Hence, $\theta\kappa(\theta) < 0$ for all $\theta \in (\tilde{\theta}, \bar{\theta})$.

We have shown that $\kappa(\theta) < 0$ for all $\theta \neq \underline{\theta}, \bar{\theta}$, from which we conclude that all of these individuals have a positive marginal distortion when there is no bunching.[16] Furthermore, these individuals face a positive marginal tax rate because $MTR(\theta) > MD(\theta)$. Continuity of $x(\cdot)$ then allows us to infer that the marginal distortions and the marginal tax rates at the endpoints of the θ distribution are nonnegative.

Provided there is no bunching at the bottom of the θ distribution, the possibility of θ-bunching does not change these conclusions significantly. If there is bunching at $\underline{\theta}$, we are not able to rule out the possibility that individuals with sufficiently low values of θ have a negative marginal tax rate; i.e., at the margin, they have an income subsidy.

If there is θ-bunching, the optimal tax schedule consists of closed intervals of bunching alternating with open intervals over which $x(\cdot)$ is strictly increasing. As we have seen, outside the bunching intervals, both $\rho(\cdot)$ and $\delta(\cdot)$ vanish. By (5.5), the marginal distortion is positive for any non-bunched θ if $\kappa(\theta) < 0$. Lemma A.1 allows us to conclude that $\int_{\underline{\theta}}^{\theta} \delta_\nu x(\nu)^{\frac{1}{2}}d\nu = 0$ for any such θ. Thus, for $\theta \in N$, (5.12) may be used to determine the sign of $\kappa(\theta)$,

---

[16]Combined with our earlier finding that $\eta > 0$, knowing that $\kappa(\theta) < 0$ for $\theta \neq \underline{\theta}, \bar{\theta}$ allows us to confirm that the denominator on the right-hand-side of (5.5) is positive. A similar argument can be used to sign this denominator if there is bunching.

the same as when there is no bunching at all. Lemma A.3 in the Appendix proves that $\kappa(\theta) < 0$ for all $\theta \in N$. Hence, even when there is bunching, all unbunched individuals in the interior of T face a positive marginal distortion.

Theorem 2 *At a solution to the planner's problem, the marginal distortion is positive for all $\theta \in N$.*

Because the size of the marginal distortion is less than the marginal tax rate for all $\theta \in N$, we can conclude from Theorem 2 that all unbunched individuals in the interior of the θ distribution face a positive marginal tax rate. Although the marginal distortion is not well-defined on bunching intervals, we can nevertheless sign everyone's marginal tax rate if $\underline{\theta}$ is not θ-bunched. If $[\underline{\theta}, \theta^*]$ is a bunching interval, we can sign the marginal tax rate for all types $\theta \geq \theta^*$.

Consider any bunching interval $[\theta_1, \theta_2]$ with $\theta_1 > \underline{\theta}$. The limit of $MD(\theta)$ as we approach θ_1 from below is the expression in (5.4) evaluated at θ_1.[17] Because the marginal distortion is positive for all $\theta \in N$, this limit is nonnegative. Comparing the expression for this limit with (5.1), we conclude that the marginal tax rate at θ_1 is positive. On a bunching interval, the allocation of income and consumption is constant, so the marginal tax rate is increasing in θ [see (5.1)]. Hence, the marginal tax rate is positive on $[\theta_1, \theta_2]$.

If there is a bunching interval $[\underline{\theta}, \theta^*]$, by taking the limit of $MD(\theta)$ as we approach θ^* from above, an analogous argument shows that the marginal tax rate at θ^* is positive. However, because the marginal tax rate is increasing in θ on $[\underline{\theta}, \theta^*]$, we are unable, in general, to sign the marginal tax rates on $[\underline{\theta}, \theta^*)$. Continuity ensures that there is a neighbourhood of θ^* in which the marginal tax rate is positive, but there could be a critical value of $\theta < \theta^*$ below which individuals have a marginal subsidy.

If $\bar{\theta}$ (resp. $\underline{\theta}$) is not θ-bunched, we cannot use the limiting argument used above to conclude that the marginal tax rate for this type is positive. The problem is that $W(\underline{\theta}) = W(\bar{\theta}) = 0$. However, because the marginal tax rate is positive for all $\theta \in N$, the continuity of $x(\cdot)$, $y(\cdot)$, and $MRS(\cdot)$ imply that the marginal tax rate at $\bar{\theta}$ (resp. $\underline{\theta}$) is nonnegative.

We note in passing that the self-selection constraints imply that x is an increasing function of y, ruling out marginal tax rates in excess of 100%.

This discussion is summarized in Theorem 3.

---

[17]This limit should not be interpreted as the marginal distortion at θ_1 because of the nondifferentiability of the resource constraint at this point.

Theorem 3 *At a solution to the planner's problem:*

(a) *If $\underline{\theta}$ is not θ-bunched, then (i) the marginal tax rate is positive for all $\theta \in (\underline{\theta}, \bar{\theta})$, (ii) the marginal tax rate at $\underline{\theta}$ is nonnegative, and (iii) the marginal tax rate at $\bar{\theta}$ is nonnegative if $\bar{\theta}$ is not θ-bunched and positive if it is.*

(b) *If $[\underline{\theta}, \theta^*]$ is a bunching interval, then (i) the marginal tax rate is positive for all $\theta \in [\theta^*, \bar{\theta})$ and (ii) the marginal tax rate at $\bar{\theta}$ is nonnegative if $\bar{\theta}$ is not θ-bunched and positive if it is.*

(c) *The marginal tax rate is increasing in θ within a bunching interval.*

Some of these conclusions differ from what is found in the standard nonlinear income tax problem with a constant-returns-to-scale technology.[18] In the standard model, the highest type is not bunched with anyone and has a zero marginal tax. Anyone else who is not bunched with the lowest type has a positive marginal tax. If the lowest type is not bunched, has a positive labour productivity, and earns a positive income, then this type's marginal tax rate is zero. If there is a bunching interval at the bottom of the type distribution, then the marginal tax rate is positive at the top of this interval and it is nonnegative on the rest of this interval.[19]

One of the key assumptions used to derive the properties of the optimal tax schedule at the endpoints of the type distribution in the standard model is that the density function is strictly positive for these types. This is not the case here. While the density function $f(\cdot)$ for the two-dimensional characteristics space S is positive everywhere, the density $G(\theta)$ is 0 when θ is either $\underline{\theta}$ or $\bar{\theta}$.[20] If the type space were changed in such a way that the density is positive at an endpoint of the θ distribution, then standard arguments would allow us to conclude that the marginal distortion for this type is 0 in the absence of bunching. It would then follow that the corresponding marginal tax rate is positive because the marginal cost of educating these individuals is positive.

---

[18]See, for example, Brito and Oakland (1977), Ebert (1992), and Seade (1977).

[19]Mirrlees (1997) has recently investigated how the properties of the optimal tax schedule at the bottom of the distribution depend on the thickness of the tail of the distribution. A novel feature of Mirrlees' model is his assumption that the marginal utility of work is positive for small amounts of labour.

[20]Although these restrictions on $G(\cdot)$ were derived using our assumption that the set of possible types is a square, this conclusion is quite robust. For example, in order for $G(\bar{\theta})$ to be positive, the upper boundary of the type space would have to intersect an iso-θ contour in a set of positive measure, which is implausible.

A striking feature of our results is that if there is bunching at the bottom of the θ-distribution, there may be an interval of low θ types who receive a marginal income subsidy.[21] A marginal income subsidy provides an incentive for individuals to work more. This increases the demand for education, which in turn raises these individuals' wage rates and, hence, their incomes.

Because education is provided free of tuition, we can also say something about the implicit subsidies to education of individuals who are of the same θ type but who differ in the underlying characteristics. For such individuals, from (2.10) and (2.13) we see that the amount of education received is decreasing in w (or, equivalently, increasing in a). Because individuals of the same θ type have the same income and pay the same amount of tax, the educational subsidy is increasing with aptitude for education and decreasing with respect to innate labour productivity.

6. Conclusion

Government provision of higher education is commonplace. While education is often available to all citizens, members of society vary in the extent to which they take it up. In particular, those who have a special aptitude for education and those for whom formal education makes a large contribution to market earnings are the most likely to acquire higher than average levels of higher education. These choices by individuals can have a profound effect on the distribution of market skills, which, in turn, shapes the desire and possibilities for redistribution across individuals. On the other hand, redistributive income taxes affect the returns to skill acquisition and, ultimately, educational choices. In this article, we have made a modest attempt to capture this rich set of economic interactions in a consistent way.

Despite the complex forces at work, we have been able to determine a number of qualitative properties of an optimal tax schedule. We have shown that all unbunched individuals face a positive marginal distortion. Further, except for possibly at the endpoints of the θ distribution, everyone is subject to a positive marginal tax if there is no bunching at the bottom of the distribution. If there is bunching at $\underline{\theta}$, this conclusion requires some qualification; some of the low-θ individuals in this bunching interval may have a marginal income subsidy. Individuals at the top of the θ distribution always have a

---

[21]In a two-type model, Stiglitz (1982) shows that the highest-type individual should receive a marginal income subsidy if the two types of labour are not perfect substitutes.

nonnegative marginal tax rate and have a strictly positive marginal tax rate if they are θ-bunched. The differences between our conclusions and those found with the standard Mirrlees model are accounted for by the extra term in the resource constraint introduced by the need to finance educational expenditures out of general tax revenues and by the fact that the density is 0 at the endpoints of the θ distribution.

We make no claim of absolute generality. Indeed, as is often necessary in problems with asymmetric information, we have placed quite a bit of structure on preferences and technology. However, the model is flexible enough to allow further investigation. For example, it is of interest to determine how the possibility of obtaining education privately affects our analysis. Our model could also be used to determine the effects of varying the level of required education. A more ambitious project would be to assess how the planner could use information on discretionary educational attainment, if available, to supplement the tax schedule. This would provide a basis for evaluating the rationale for a graduate tax in a second-best environment.

Appendix

Proof of Theorem 1. Let $\{x(w,a), y(w,a)\}$ be the menu of optimal income-consumption pairs. Set $a = 1$. With this restriction, θ is distributed on $T_1 = [b^{-2}, 2^{\frac{3}{2}}b^{-2}]$. For fixed a, the single-crossing property is satisfied with respect to w. Because the marginal rate of substitution is decreasing in w, the self-selection constraints (2.15) imply that both $x(\cdot, 1)$ and $y(\cdot, 1)$ must be nondecreasing on [1,2]. Because $x(\cdot, 1)$ and $y(\cdot, 1)$ are monotone functions on an interval, they are continuous almost everywhere on [1,2].

Suppose that these functions are continuous at $\hat{w} \in (1, 2)$ and let $\hat{\theta}$ be the corresponding value of θ. Consider any other type $(\bar{w}, \bar{a})$ whose value for the type aggregator is $\hat{\theta}$. Because types $(\hat{w}, 1)$ and $(\bar{w}, \bar{a})$ have the same preferences for income and consumption, the incentive compatibility constraints imply that $(x(\bar{w}, \bar{a}), y(\bar{w}, \bar{a}))$ lies on the type $\hat{\theta}$ indifference curve through $(x(\hat{w}, 1), y(\hat{w}, 1))$. If $(x(\bar{w}, \bar{a}), y(\bar{w}, \bar{a})) \neq (x(\hat{w}, 1), y(\hat{w}, 1))$, then $(x(\bar{w}, \bar{a}), y(\bar{w}, \bar{a}))$ lies above the part of the tax schedule designed for individuals of type $a = 1$. If $(x(\bar{w}, \bar{a}), y(\bar{w}, \bar{a})) \gg (x(\hat{w}, 1), y(\hat{w}, 1))$, then for $w > \hat{w}$ sufficently close to $\hat{w}$, a type $(w, 1)$ individual prefers $(x(\bar{w}, \bar{a}), y(\bar{w}, \bar{a}))$ to $(x(w, 1), y(w, 1))$, violating the self-section constraints. Analogously, if $(x(\bar{w}, \bar{a}), y(\bar{w}, \bar{a})) \ll (x(\hat{w}, 1), y(\hat{w}, 1))$, then for $w < \hat{w}$ sufficently close to $\hat{w}$,

a type $(w, 1)$ individual prefers $(x(\bar{w}, \bar{a}), y(\bar{w}, \bar{a}))$ to $(x(w, 1), y(w, 1))$. Hence, for $\theta \in T_1$, individuals with the same value of θ must be allocated the same income-consumption pair at points of continuity in the tax schedule.

Let Ω denote the set of values of w for which either there is a discontinuity in the tax schedule at the allocation designed for a type $(w, 1)$ individual or $w \in \{1, 2\}$ and let T_Ω be the corresponding set of θs. For any $\tilde{w} \in \Omega$, let $\tilde{\theta}$ be the θ value of someone of type $(\tilde{w}, 1)$. For all other types whose value for the type aggregator is $\tilde{\theta}$, reallocate their income-consumption pairs if necessary so that all individuals of type $\tilde{\theta}$ receive $(x(\tilde{w}, 1), y(\tilde{w}, 1))$. None of these individuals have had their utility changed as a result of this reallocation. Because no new allocations have been added to the tax schedule, the self-selection constraints are still satisfied after the reallocation. Because the set of individuals with types in T_Ω has zero measure, the reallocation satisfies the resource constraint and does not change the value of the planner's objective function. Hence, for $\theta \in T_1$, it is optimal to allocate the same income-consumption pair to individuals with the same value of θ.

Now set $a = 2$. With this restriction, θ is distributed on $T_2 = [2b^{-2}, 2^{\frac{5}{2}}b^{-2}]$. Proceding as above, it follows that for $\theta \in T_2$, it is optimal to allocate the same income-consumption pair to individuals with the same value of θ. Because $T = T_1 \cup T_2$, the proof is complete. ∎

Lemma A.1 *At a solution to the planner's problem, if $[\theta_1, \theta_2]$ is a bunching interval, then $\int_{\theta_1}^{\theta_2} \delta_\nu(\nu) x(\nu)^{\frac{1}{2}} d\nu = 0$.*

Proof. By (4.8), $\delta(\nu) = 0$ if and only if $\rho(\nu) = 0$. Hence, by the complementary slackness condition (4.13), $\delta(\nu) \neq 0$ only on bunching intervals; that is, only on intervals for which $x(\cdot)$ is constant. Moreover, because $\delta(\cdot)$ is continuous, so is $\rho(\cdot)$. Thus, at the endpoints of any bunching interval, $\rho(\nu) = \delta(\nu) = 0$.[22] Let $\bar{x}$ denote the constant value of $x(\cdot)$ on this interval. By the Fundamental Theorem of Calculus,

$$\int_{\theta_1}^{\theta_2} \delta_\nu(\nu) x(\nu)^{\frac{1}{2}} d\nu = \bar{x}^{\frac{1}{2}} \int_{\theta_1}^{\theta_2} \delta_\nu(\nu) = \bar{x}^{\frac{1}{2}} [\delta(\theta_2) - \delta(\theta_1)] = 0, \qquad \text{(A.1)}$$

which completes the proof. ∎

Lemma A.2 *At a solution to the planner's problem, if $[\theta_1, \theta_2]$ is a bunching interval with $\theta_1 > \hat{\theta}$, then it is not the case that $\kappa(\theta_1) \geq 0$ and $\kappa(\theta_2) \leq 0$.*

---

[22]If either $\underline{\theta}$ or $\bar{\theta}$ is one of the endpoints of the bunching interval, we know the costate variable $\delta(\nu)$ is 0 by the transversality condition (4.12).

Proof. By (5.8),

$$\theta_2\kappa(\theta_2) - \theta_1\kappa(\theta_1) = \int_{\theta_1}^{\theta_2}[\eta x(\nu)^{\frac{1}{2}} - 1]G(\nu)d\nu - \int_{\theta_1}^{\theta_2}\delta_\nu(\nu)x(\nu)^{\frac{1}{2}}d\nu. \quad (A.2)$$

Let $\bar{x}$ denote the common value of $x(\cdot)$ on $[\theta_1, \theta_2]$. The last term on the right-hand-side of (A.2) is 0 by Lemma A.1. Hence, (A.2) implies that

$$\theta_2\kappa(\theta_2) - \theta_1\kappa(\theta_1) = [\eta\bar{x}^{\frac{1}{2}} - 1][\bar{G}(\theta_2) - \bar{G}(\theta_1)], \quad (A.3)$$

where

$$\bar{G}(\theta) = \int_{\underline{\theta}}^{\theta} G(\nu)d\nu, \quad \forall\theta \in T. \quad (A.4)$$

Because $\eta\bar{x}^{\frac{1}{2}} > 1$ when $\theta > \hat{\theta}$ and because $\bar{G}(\theta_2) > \bar{G}(\theta_1)$ when $\theta_2 > \theta_1$, the right-hand-side of (A.3) is positive. Thus,

$$\theta_2\kappa(\theta_2) > \theta_1\kappa(\theta_1). \quad (A.5)$$

If $\kappa(\theta_1) \geq 0$ and $\kappa(\theta_2) \leq 0$, then (A.5) would be violated. ∎

Lemma A.3 *At a solution to the planner's problem, $\kappa(\theta) < 0$ for all $\theta \in N$.*

Proof. We have established that (5.12) holds for all $\theta \in N$. Even though $x(\cdot)$ is now only nondecreasing, the argument used below (5.12) to show that $\kappa(\theta) < 0$ continues to apply if $\theta \leq \tilde{\theta}$.[23] We now consider $\theta > \hat{\theta}$. Contrary to the lemma, suppose that $\kappa(\theta) \geq 0$. The argument used when there is no bunching to show that the right-hand-side of (5.13) is positive is valid for this θ. Hence, we can conclude that $\kappa(\cdot)$ is increasing outside bunching intervals above $\hat{\theta}$. Because non-bunching intervals are open, there exists a $\theta' \in N$ with $\theta' > \theta$ for which $\kappa(\theta') > 0$. In Lemma A.2 we have shown that for any bunching region $[\theta_1, \theta_2]$ in $(\hat{\theta}, \bar{\theta}]$, if $\kappa(\theta_1) \geq 0$, then $\kappa(\theta_2) > 0$. Therefore, $\kappa(\bar{\theta}) > 0$, contradicting the transversality condition (4.11). ∎

References

Arrow, K.J., 1973. Higher education as a filter. Journal of Public Economics 2, 193–216.

---

[23]Note that θ can only equal $\tilde{\theta}$ in this argument if $\tilde{\theta}$ is not bunched.

Barham, V., Boadway, R., Marchand, M., Pestieau, P., 1995. Education and the poverty trap. European Economic Review 39, 1257–1275.

Boadway, R., Cuff, K., Marchand, M., 2000. Optimal income taxation with quasi-linear preferences revisited. Journal of Public Economic Theory 2, 435–460.

Boadway, R., Marceau, N., Marchand, M., 1996. Investment in education and the time inconsistency of redistributive tax policy. Economica 63, 171–189.

Boadway, R., Marchand, M., 1995. The use of public expenditures for redistributive purposes. Oxford Economic Papers 47, 45–59.

Boadway, R., Marchand, M., Pestieau, P., Racionero, M., 1999. Optimal redistribution with heterogeneous preferences for leisure. Unpublished manuscript, Center for Operations Research and Econometrics, Université Catholique de Louvain.

Brito, D.L., Oakland, W., 1977. Some properties of the optimal income-tax. International Economic Review 18, 407–423.

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

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

Grout, P., 1983. Imperfect information, markets and public provision of education. Journal of Public Economics 8, 113–121.

Hare, P.G., Ulph, D.T., 1979. On education and distribution. Journal of Political Economy 87, S193–S212.

Laffont, J.-J., Maskin, E., Rochet, J.-C., 1987. Optimal nonlinear pricing with two-dimensional characteristics. In: Groves, T., Radner, R., Reiter, S. (Eds.), Information, Incentives, and Economic Mechanisms. University of Minnesota Press, Minneapolis, pp. 256–266.

Lollivier, S., Rochet, J.-C., 1983. Bunching and second-order conditions: A note on optimal tax theory. Journal of Economic Theory 32, 392–400.

Marshall, A.W., Olkin, I., 1979. Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

McAfee, R.P., McMillan, J., 1988. Multidimensional incentive compatibility and mechanism design. Journal of Economic Theory 46, 335–354.

Mirrlees, J.A., 1971. An exploration in the theory of optimum income taxation. Review of Economic Studies 38, 175–208.

Mirrlees, J.A., 1976. Optimum tax theory: A synthesis. Journal of Public

Economics 7, 327–358.

Mirrlees, J.A., 1997. Optimal marginal taxes at low incomes. Unpublished manuscript, Faculty of Economics, University of Cambridge.

Rochet, J.-C. 1987. A necessary and sufficient condition for rationalizability in a quasi-linear context. Journal of Mathematical Economics 16, 191–200.

Rochet, J.-C., Stole, L.A., 2000. The economics of multidimensional screening. Unpublished manuscript, Graduate School of Business, University of Chicago.

Sandmo, A., 1998. Redistribution and the marginal cost of public funds. Journal of Public Economics 70, 365–382.

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

Shapiro, J., 1999. Income maintenance programs and multidimensional screening. Unpublished manuscript, Department of Economics, Princeton University.

Seierstad, A., Sydsæter, K., 1977. Sufficient conditions in optimal control theory. International Economic Review 18, 367–391.

Sheshinski, E., 1971. On the theory of optimal income taxation. Discussion Paper No. 172, Harvard Institute for Economic Research, Harvard University.

Spence, A.M., 1974. Market Signalling: Informational Transfer in Hiring and Related Screening Processes. Harvard University Press, Cambridge, MA.

Stiglitz, J.E., 1982. Self-selection and Pareto efficient taxation. Journal of Public Economics 17, 213–340.

Tuomala, M., 1986. On the optimal income taxation and educational decisions. Journal of Public Economics 30, 183–198.

Ulph, D., 1977. On the optimal distribution of income and educational expenditure. Journal of Public Economics 8, 341–356.

Wilson, A., 1999. Education policy: The case for government intervention with endogenous abilities. Unpublished manuscript, Department of Economics, Princeton University.