HOW OPTIMAL NONLINEAR INCOME TAXES CHANGE WHEN THE DISTRIBUTION OF THE POPULATION CHANGES
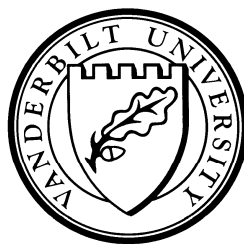
by

Craig Brett and John A. Weymark

How Optimal Nonlinear Income Taxes Change
When the Distribution of the Population Changes*

by

Craig Brett

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

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)

April 2010

Abstract

"How Optimal Nonlinear Income Taxes Change When the Distribution of
the Population Changes"

by

Craig Brett and John A. Weymark

The impacts of changing the number of individuals of a particular skill
level on the solutions to two versions of the finite population optimal nonlin-
ear income tax problem are investigated. In one version, preferences are
quasilinear-in-leisure. For this version, it is shown that it is possible to
sign the directions of change in everyone's optimal consumptions and op-
timal marginal tax rates. In the other version, preferences are quasilinear-
in-consumption. For this version, it is shown that is possible to sign the
directions of change in everyone's optimal before-tax incomes and optimal
marginal tax rates. Moreover, the directions of change in the optimal marginal
tax rates are the same for the two specifications of preferences.

Journal of Economic Literature classification numbers: D82, H21.

Keywords and phrases: asymmetric information, comparative statics, opti-
mal income taxation.

1. Introduction

Following the work of Mirrlees (1971), the theory of optimal nonlinear income taxation has investigated the properties of optimal income tax schedules when a government's limited information about taxpayers prevents the use of non-distorting lump-sum taxes. In the canonical model, the government lacks information about labor market skills. While the distribution of skills is commonly known, the government is unable to observe the skills of particular individuals. Distortions arise as a part of the cost of acquiring the tax-relevant skill information. Comparative static analyses of optimal nonlinear income taxes attempt to gauge how changes in the underlying structure of the economy affect optimal taxes and their associated distortions. Because of the nonconvexities introduced by the incentive constraints, there are relatively few comparative static results available for these kinds of problems when there are more than two skill types.[1]

In this article, we contribute to the small, but growing, literature on this issue. In particular, we derive new comparative static results about how optimal nonlinear income taxes and the associated allocations change when the distribution of skills changes. Changes in the demographic structure of the population may arise for a number of reasons. For example, the skill distribution may change due to mobility between jurisdictions, as in Hamilton and Pestieau (2005); it may change over time due to the augmentation of skills through education, as in Brett and Weymark (2003); or it may differ between identifiable subgroups of the population who, because of this tagging, may face different tax schedules, as in Boadway and Pestieau (2007).

We focus our attention on models with a discrete skill distribution and an arbitrary finite number of skill types. As noted by Brett and Weymark (2008a), it is possible to perturb discrete skill distributions in two ways. One can change the support of the distribution, say by increasing the skill level of individuals of a particular type. Comparative static analyses of this sort have been carried out by Brett and Weymark (2008a) and Simula (2010). Alternatively, one can keep the set of possible skill levels fixed and change the number of individuals with some skill level. This second kind of comparative static analysis is the focus of this article. Hamilton and Pestieau (2005) and Boadway and Pestieau (2007) have addressed some aspects of this issue,

---

[1]For similar reasons, there are relatively few comparative static results available for other kinds of asymmetric information problems when there are more than two types.

but only for the case of two skill types. We show in this article that two-type worlds are not rich enough to capture all of the interactions between demographics and optimal income taxes that are of interest.

Because of the technical challenges inherent in the optimal income tax problem, existing comparative static results for nonlinear taxes are available only under restrictive assumptions concerning individual preferences. Weymark (1987) and Brett and Weymark (2008a,b) derive comparative static results for many changes in the economic environment for the case of preferences that are quasilinear-in-leisure. Hamilton and Pestieau (2005) also consider these kinds of preferences. Boadway and Pestieau (2007) and Simula (2010) consider preferences that are quasilinear-in-consumption. We consider both kinds of quasilinearity, using the Weymark model for quasilinear-in-leisure preferences and a generalization of the Simula model for quasilinear-in-consumption preferences.

In the models used by Simula and Weymark, the government maximizes a weighted sum of individual utilities subject to an economy-wide resource constraint and incentive compatibility constraints. The social welfare weights are chosen so as to satisfy standard redistributive criteria, like those described by Guesnerie and Seade (1982), Röell (1985), and Hellwig (2007), but are otherwise arbitrary. The restrictions on the welfare weights guarantee that the adjacent downward incentive compatibility constraints bind (in the terminology of Guesnerie and Seade, there is a strict monotonic chain to the left). Having this pattern of binding incentive constraints allows both Simula and Weymark to derive tractable reduced-form optimization problems that can be used to describe the choices of some of the variables of interest. Weymark's reduced-form can be used to compute the optimal after-tax incomes and the optimal marginal tax rates, whereas Simula's reduced-form can be used to compute optimal before-tax incomes (consumptions) and optimal marginal tax rates.

At an optimum, individuals of all types (except the most highly skilled) face a positive marginal tax rate. Equivalently, their before-tax income (labor supply) is distorted downward in the sense that a one unit increase in both the before- and after-tax income of such an individual would make that individual better off. This proposed increase does not violate the resource constraint, but it is not optimal because it violates a self-selection constraint. In particular, it would make it attractive for individuals of the next highest type to pretend to be of that type. Satisfaction of this self-selection contraint can be restored if the proposed increase is combined with transfers to indi-

viduals of the next highest type. These additional transfers might destroy incentive compatibility for the still more highly skilled, so additional transfers might be needed to obtain a feasible allocation. The total of all such transfers constitute the information rents received by higher types. These rents make the cost of increasing the before-tax income of individuals more than just the additional after-tax income that workers receive. Moreover, the social cost of these rents determines the size of the optimal marginal tax rate of the type being considered. When preferences are quasilinear (in either sense), it is possible to derive explicit formulae for the social costs of the information rents and to deduce how the distribution of skills affects these costs. These formulae play an important role in our comparative static analyses.

For both the Simula and the Weymark models, we obtain results on the effects of an increase in the number of individuals of a particular skill type for every variable appearing in their respective reduced-form problems. In particular, we are able to sign the direction of change for every optimal marginal income tax rate. Moreover, we show that these signs are the same for both models.

The general character of our results can be illustrated by considering an increase in the number of individuals, n_k, of a particular type, say k, other than the highest skilled when preferences are quasilinear-in-leisure. An increase in n_k does not change the size of the information rents associated with an increase in the consumption of type k individuals. It does, however, increase the number of type k individuals facing the distortion caused by these rents. As a result, an increase in n_k provides the government with an incentive to reduce these distortions. Therefore, the optimal marginal tax rate for individuals of type k falls, which induces an increase in their consumption.[2]

An increase in n_k also affects the optimal marginal tax rates faced by individuals of other types. For a type $i < k$, individuals of type k are receivers of information rents when the consumption of type i individuals increases. An increase in n_k increases the aggregate amount of information rents because they must now be paid to more individuals. However, and perhaps unexpectedly, the social marginal cost of the rents may actually decrease. The rents are a source of utility to individuals of type k. Thus, the benefits associated with the rents increase. The costs of the information

---

[2]There are additional effects, but we show that the effect described here dominates.

rents are borne by all individuals in society. If individuals of type k have a social welfare weight greater than the average social welfare weight, then the additional benefits outweigh the additional costs and the net social marginal cost of the information rents decreases. In this case, the marginal tax rate faced by individuals of type i decreases, with the consequence that their consumption increases so as to increase these information rents. For a type $j > k$, individuals of type k help finance the information rents associated with increases in the consumption of type j individuals. If individuals of type k have a social welfare weight greater than the average social welfare weight, their tax contributions weigh more heavily in the computation of the social cost of the information rents. In this case, an increase in n_k causes the net social marginal cost of the rents to increase. Consequently, the marginal tax rate faced by individuals of type j increases and their consumption is decreased so as to decrease the information rents.[3]

In Section 2, we present a general framework that encompasses the models of both Simula and Weymark. We carry out our comparative analyses for the Weymark and Simula models, respectively, in Sections 3 and 4. Section 5 contains concluding remarks. Our proofs are gathered in an Appendix.

2. The Framework

The economy consists of H individuals, who are partitioned into N types according to their labor productivities. The number of individuals of type i is denoted by n_i. The productivity of this type of individual is given by w_i. Types are ordered such that $w_1 < w_2 < \cdots < w_N$. The production sector is assumed to exhibit constant returns to scale and the labor market is perfectly competitive. With these assumptions, the before-tax labor income of an individual of type i is given by

$$y_i = w_i l_i, \tag{2.1}$$

where l_i is his labor supply

The government chooses a tax schedule without the ability to observe individual skill types, w_i, or labor supply, l_i. It can observe before-tax income, y_i, and is assumed to be able to set any anonymous tax schedule it wishes using income as the tax base. Without loss of generality, we can set the price

---

[3]The preceding discussion also applies to quasilinear-in-consumption preferences except that changes in before-tax incomes are substituted for changes in consumptions.

of the consumption good equal to 1, so an individual's consumption, c_i, is simply his after-tax income.

Individuals have preferences over c and l represented by a common cardinally significant utility function $\tilde{u}(c, l)$. Because of differences in skills, preferences over the observable variables c and y are type-specific, given by

$$\bar{u}_i(c, y) = \tilde{u}\left(c, \frac{y}{w_i}\right). \tag{2.2}$$

For the Weymark (1986, 1987) model, we multiply (2.2) by w_i (a type-specific monotonic transformation) to obtain a utility function $u_i(c, y)$. See (3.2) below. For the Simula (2010) model, we simply set $u_i = \bar{u}_i$.

An allocation $(c_1, y_1, \ldots, c_N, y_N)$ is a list of the consumptions and before-tax incomes of the N types of individuals. Because taxation is anonymous, revealed preference implies that when individuals optimize on their common budget set, the resulting allocation satisfies the self-selection (i.e., incentive compatibility) conditions

$$u_i(c_i, y_i) \geq u_i(c_j, y_j), \quad i, j = 1, \ldots, N. \tag{2.3}$$

The taxation principle [cf. Guesnerie (1995)] implies that the set of allocations that can be obtained by individuals maximizing subject to some anonymous tax schedule is equivalent to the set of allocations satisfying (2.3). Thus, the government can be viewed as directly choosing the consumption levels and before-tax incomes of each type of individual subject to (2.3) rather than indirectly determining the allocation through the choice of the tax schedule.

The government employs a weighted utilitarian social welfare function of the form

$$W(u_1, \ldots, u_N) = \sum_{i=1}^{N} \lambda_i n_i u_i, \tag{2.4}$$

where the λs are social welfare weights that satisfy

$$\lambda_1 > \lambda_2 > \cdots > \lambda_N > 0. \tag{2.5}$$

This linear social welfare function may be viewed as a local linear approximation to a quasiconcave social welfare function at a solution to the optimal tax problem. By assuming that the social welfare weights are declining in type, the optimal tax problem is redistributive in the sense of Guesnerie and Seade (1982); i.e., at the solution to the optimal tax problem, it would be

desirable to transfer resources from higher-skilled to lower-skilled individuals in the absence of the self-selection constraints. In the Weymark model, if the social welfare function is defined in terms of the untransformed utilities $\bar{u}_i$, then the welfare weight applicable to type i is $w_i \lambda_i$. Classical utilitarianism requires these $w_i \lambda_i$ welfare weights to be the same for all i.

The government chooses an allocation to maximize (2.4) subject to the self-selection constraints (2.3) and a material balance constraint of the form

$$\sum_{i=1}^{N} n_i c_i \leq \sum_{i=1}^{N} n_i y_i. \tag{2.6}$$

By Walras' Law, the latter constraint is equivalent to requiring that the government not run a deficit.

The quasilinear preferences used by Simula and Weymark satisfy the usual single-crossing property. As a consequence, because the welfare weights are declining in type, in the government's optimization problem it is possible to replace the self-selection constraints (2.3) with the requirements that the adjacent downward incentive compatibility constraints bind,

$$u^i(c_i, y_i) = u^i(c_{i-1}, y_{i-1}), \quad i = 2, \ldots, N, \tag{2.7}$$

and that the individual consumptions and pre-tax incomes are nondecreasing in type,

$$(c_1, y_1) \leq (c_2, y_2) \leq \cdots \leq (c_N, y_N).^4 \tag{2.8}$$

Furthermore, at a solution to the optimal tax problem, the material balance constraint (2.6) binds.

Both Simula and Weymark find it convenient to normalize the welfare weights so that their sum is equal to the total number of individuals, i.e., so that the average welfare weight is equal to 1. Formally,

$$\sum_{i=1}^{N} n_i \lambda_i = \sum_{i=1}^{N} n_i = H. \tag{2.9}$$

Because the social welfare function is homogeneous of degree one in the welfare weights, this normalization has no effect on the optimal allocation. The normalization (2.9) adds considerably to the tractability of the analyses

---

[4] See Matthews and Moore (1987) for a general treatment of the conditions under which the self-selection constraints can be reformulated in this way.

of Simula and Weymark. However, it makes the comparative static analysis of changes in the composition of the population somewhat more difficult to determine because if the number of individuals of, say, type k increases, it may be necessary to rescale all of the welfare weights so as to re-establish (2.9).

For concreteness, consider a small increase ε_k in n_k. Following this change, all of the λs are multiplied by a constant κ so as to re-establish (2.9). Let the adjusted weights be

$$\bar{\lambda}_i = \kappa \lambda_i, \quad i = 1, \ldots, N. \tag{2.10}$$

Note that if $\kappa = 1$, no renormalization is necessary. To re-establish the normalization, it must be the case that

$$(n_k + \varepsilon_k)\bar{\lambda}_k + \sum_{i \neq k} n_i \bar{\lambda}_i = H + \varepsilon_k. \tag{2.11}$$

Substituting (2.10) into (2.11) and rearranging gives

$$\kappa = \frac{H + \varepsilon_k}{\sum_{i=1}^{N} n_i \lambda_i + \lambda_k \varepsilon_k} = \frac{H + \varepsilon_k}{H + \lambda_k \varepsilon_k}. \tag{2.12}$$

Therefore,

$$\bar{\lambda}_i = \lambda_i \left[\frac{H + \varepsilon_k}{H + \lambda_k \varepsilon_k} \right], \quad i = 1, \ldots, N. \tag{2.13}$$

Hence,

$$\frac{\partial \bar{\lambda}_i}{\partial \varepsilon_k} = \lambda_i \left[\frac{H - \lambda_k H}{(H + \lambda_k \varepsilon_K)^2} \right], \quad i = 1, \ldots, N. \tag{2.14}$$

Evaluating (2.14) at $\epsilon_k = 0$ and simplifying yields

$$\frac{\partial \lambda_i}{\partial n_k} := \frac{\partial \bar{\lambda}_i}{\partial \varepsilon_k} \bigg|_{\varepsilon_k = 0} = \frac{\lambda_i(1 - \lambda_k)}{H}, \quad i = 1, \ldots, N. \tag{2.15}$$

It follows from (2.15) that every (re-normalized) λ_i changes in the same direction when n_k increases. Moreover, the direction of change in the λs is determined by the initial welfare weight assigned to individuals of type k. If this initial welfare weight is greater (less) than the average welfare weight of 1, then the λs must be scaled downward (upward) following an increase in n_k in order to re-establish (2.9). The preceding discussion is summarized in the following lemma.

Lemma 1. *An increase in n_k accompanied by a proportionate rescaling of all the λs so as to re-establish the normalization (2.9) results in:*

(i) *an increase in λ_i, $i = 1, \ldots, N$, if $\lambda_k < 1$;*

(ii) *a decrease in λ_i, $i = 1, \ldots, N$, if $\lambda_k > 1$;*

(iii) *no change in λ_i, $i = 1, \ldots, N$, if $\lambda_k = 1$.*

3. The Weymark Model

Weymark (1986, 1987) considers the optimal nonlinear income tax problem when individuals have quasilinear-in-leisure preferences of the form

$$\bar{u}_i^W(c, y) = v(c) - \gamma \frac{y}{w_i}, \tag{3.1}$$

where v is a twice continuously differentiable, increasing, and strictly concave function with $v(0) = 0$, $v'(0) = \infty$, and $\lim_{r \to \infty} v'(r) = 0$. The marginal disutility of labor, γ, is a positive constant. For convenience, he uses the linear-in-wages transformation of (3.1) given by

$$u_i^W(c, y) = w_i v(c) - \gamma y. \tag{3.2}$$

Weymark's assumptions concerning the form of the utility function allow him to derive a reduced-form optimal tax problem that can be used to derive the optimal consumptions of the N types of individuals. To state this optimization problem, we introduce a reduced-form welfare weight β_i for type i, defined by setting

$$\beta_i = \begin{cases} n_i w_i + \sum_{j=i+1}^{N} n_j \lambda_j (w_{i+1} - w_i) - \sum_{j=i+1}^{N} n_j (w_{i+1} - w_i), & i \neq N, \\ n_i w_i, & i = N. \end{cases} \tag{3.3}$$

It follows from (2.5) and (2.9) that $\beta_i < n_i w_i$ for all $i \neq N$. We reproduce Weymark's result in the following lemma.

Lemma 2. [Weymark (1986, Problem 5, p. 212)] *The optimal consumptions in the Weymark model are found by solving*

$$\max_{c_1, \ldots, c_N} \sum_{i=1}^{N} \beta_i v(c_i) - \gamma \sum_{i=1}^{N} n_i c_i \tag{3.4}$$

8

subject to

$$0 \leq c_1 \leq c_2 \leq \cdots \leq c_N. \tag{3.5}$$

If the government had full information about the individuals' types, then a self-financed marginal increase in c_i changes the total utility of individuals of type i by $n_i w_i v'(c_i) - \gamma n_i$. Such an increase in c_i is socially beneficial if this expression is positive. As noted by Weymark (1986), the reduced-form welfare weight β_i shows how the term $n_i w_i$ that multiplies $v'(c_i)$ in this calculation must be adjusted to take account of the self-selection constraints. The second term in the top line of (3.3) measures the social value of the increases in utility that must be given to individuals more skilled than those of type i in order to re-establish the binding adjacent downward incentive compatibility constraints following an increase in c_i. The third term measures the resource cost (in social welfare terms) of the changes in utility needed to re-establish the binding materials balance constraint following the adjustments captured by the second term. The latter adjustments are accomplished by changing each person's before-tax income by a common amount and, hence, because preferences are quasilinear in labor, they do not affect the pattern of binding self-selection constraints. When $i = N$, increasing c_i violates no self-selection constraint, and so β_N is simply $n_N w_N$.[5]

To facilitate the comparison of the Simula and Weymark models, it is useful to rewrite the objective function in Lemma 2 in terms of the information rents received by each individual. Let

$$R_i^W(c_i) = (w_{i+1} - w_i)v(c_i), \quad i = 1, \ldots, N, \tag{3.6}$$

where $w_{N+1} = w_N$. $R_i^W(c_i)$ measures the difference between the utility obtained by an individual of type $i+1$ and the utility obtained by an individual of type i at (c_i, y_i). It is the information rent obtained by type $i+1$ individuals when the type $i+1$ adjacent downward incentive compatibility constraint is binding.[6] The assumptions on v imply that R_i^W is an increasing concave function of c_i.

---

[5]See Weymark (1986, pp. 210–211) for a more detailed discussion of the reduced-form welfare weights.

[6]While there are no type $N + 1$ individuals, some of the subsequent expressions are simpler to state if we set $R_N^W(c_N)$ identically equal to zero.

Let α_i be given by

$$\alpha_i = \begin{cases} \sum_{j=i+1}^{N} (n_j \lambda_j - n_j), & i \neq N, \\ 0, & i = N. \end{cases} \tag{3.7}$$

By (2.5) and (2.9), $\alpha_i < 0$ for all $i \neq N$. Using (3.6) and (3.7), the objective function in Lemma 2 can be rewritten as

$$\sum_{i=1}^{N} n_i w_i v(c_i) - \gamma \sum_{i=1}^{N} n_i c_i + \sum_{i=1}^{N} \alpha_i R_i^W(c_i). \tag{3.8}$$

The social costs of the self-selection constraints are now isolated in the last term in (3.8). For each type $i \neq N$, the information rent that must be provided to prevent a type $i+1$ individual from mimicking someone of type i is given a negative weight of α_i. Hence, the absolute value of α_i measures the net social cost of each additional unit of information rent that needs to be allocated to individuals of type $i+1$ when the consumpton bundle of type i is changed.

As Lemma 3 demonstrates, the influence of an increase in the number of individuals of one skill type, say k, on β_i depends on the relationship between i and k and on the sign of $(1 - \lambda_k)$.

Lemma 3. *Suppose that there is a marginal increase in n_k accompanied by a proportionate rescaling of the λs so as to re-establish (2.9). Then:*

(i) $\dfrac{\partial \beta_i}{\partial n_k}$ *has the same sign as $(1 - \lambda_k)$ if $k < i$ and $i < N$.*

(ii) $\dfrac{\partial \beta_i}{\partial n_k}$ *has the opposite sign to $(1 - \lambda_k)$ if $k > i$.*

(iii) $\dfrac{\partial \beta_i}{\partial n_k} = 0$ *if $k < N$ and $i = N$.*

(iv) $\dfrac{\partial \beta_i}{\partial n_k} > 0$ *if $k = i = N$.*

If $k < i$ and $i < N$, then changes in n_k influence β_i solely through changes, if any, in the rescaled λs. If, for example, $(1 - \lambda_k)$ is positive, then by Lemma 1, an increase in n_k leads to an increase in the rescaled

λs. In this case, the social value of the utility transfers to individuals of types greater than i required to re-establish the binding adjacent downward incentive compatibility constraints following an increase in c_i increases and, therefore, so does β_i.

If $k > i$, when n_k increases, there are more of this type of individual whose utilities are increased to re-establish the binding adjacent downward incentive compatibility constraints following an increase in c_i, and this contributes to social welfare. Further, because there are more of these individuals, the resource cost of restoring materials balance following the incentive compatibility adjustments also increases. In other words, in addition to the effect on β_i due to the rescaling of the λs described above, when $k > i$, there is a direct effect on β_i of increasing n_k due to the increase in the social value and resource cost of these individuals. This increase in social value is proportional to λ_k, the welfare weight assigned to individuals of type k, whereas the increased resource cost is proportional to the average welfare weight of all of the individuals, which is equal to 1. If $(1 - \lambda_k)$ is positive, then the increment to marginal social cost is greater than the increment to marginal social benefit when only the direct effect of increasing n_k is considered, which results in a decrease in β_i. We thus see the direct effect of changing n_k on β_i when $k > i$ has the opposite sign to $(1 - \lambda_k)$, whereas the effect due to the rescaling of the λs has the same sign as $(1 - \lambda_k)$. Part (ii) of Lemma 3 shows that the direct effect dominates the rescaling effect.

As we have seen, $\beta_N = n_N w_N$. The last two parts of Lemma 3 follow directly from this observation.[7]

For the rest of this section, we assume that the optimal consumptions satisfy $0 < c_1 < \cdots < c_N$; i.e., the nonnegativity and monotonicity constraints on consumption do not bind. When this assumption is satisfied, these inequalities remain strict following a marginal change in any of the n_k.[8] With this assumption, the first-order conditions for a solution to the consumption optimization problem described in Lemma 2 are

$$\beta_i v'(c_i) = \gamma n_i \quad i = 1, \ldots, N. \tag{3.9}$$

The left-hand side of (3.9) is the marginal net benefit (inclusive of the adjust-

---

[7]When $k = i < N$, the presence of the term $n_i w_i$ in the definition of β_i results in the sign of $\frac{\partial \beta_i}{\partial n_k}$ in the first two parts of of Lemma 3 being indeterminate.

[8]Minor qualifications to our comparative static results are needed if the pattern of binding constraints (3.5) in the reduced-form optimization problem in Lemma 2 is altered when n_k changes.

ments needed to restore binding adjacent downward incentive compatibility and materials balance) of a marginal increase in c_i, while the right-hand side is the direct marginal resource cost of that same change. It follows from (3.9) that

$$c_i = v'^{-1}\left(\frac{\gamma n_i}{\beta_i}\right) = \varphi\left(\frac{\gamma n_i}{\beta_i}\right), \quad i = 1, \ldots, N. \tag{3.10}$$

Because v is strictly concave, the function φ implicitly defined in (3.10) is decreasing.

A tax schedule consistent with an allocation satisfying (2.7) is typically nondifferentiable. Thus, marginal tax rates are only implicitly defined by the difference between producer and consumer prices (which are, respectively, 1 and the marginal rate of substitution) at an individual's consumption bundle. In light of (3.1), the implicit marginal tax rate for labor income applicable to type i is given by

$$\tau_i^W = 1 - \frac{\gamma}{w_i v'(c_i)}, \quad i = 1, \ldots, N. \tag{3.11}$$

For fixed γ and w_i, the implicit marginal tax rate varies inversely with consumption. It then follows from (3.9) that the optimal implicit marginal tax rates are given by

$$\tau_i^W = 1 - \frac{\beta_i}{n_i w_i}, \quad i = 1, \ldots, N. \tag{3.12}$$

Equations (3.3) and (3.12) imply that the standard properties of optimal taxes apply: individuals of type N face a zero implicit marginal tax rate and all other types of individuals face positive implicit marginal tax rates.

The comparative static responses of the optimal consumptions and the optimal implicit marginal tax rates to an increase in the number of individuals of any skill type can be deduced from the closed-form expressions (3.10) and (3.12). These responses are characterized in following theorem.

Theorem 1. *Suppose that there is a marginal increase in n_k accompanied by a proportionate rescaling of the λs so as to re-establish (2.9).*

(i) *If $k < N$, then the optimal c_k increases and the optimal τ_k^W decreases.*

(ii) *If $\lambda_k > 1$, then the optimal c_i decreases and the optimal τ_i^W increases for all $i = k+1, \ldots, N-1$ and the optimal c_i increases and the optimal τ_i^W decreases for all $i = 1, \ldots, k-1$.*

(iii) *If $\lambda_k < 1$, then the optimal c_i increases and the optimal τ_i^W decreases for all $i = k+1, \ldots, N-1$ and the optimal c_i decreases and the optimal τ_i^W increases for all $i = 1, \ldots, k-1$.*

(iv) *If $\lambda_k = 1$, then the optimal c_i and the optimal τ_i^W do not change for all $i \neq k, N$.*

(v) *Neither the optimal c_N nor the optimal τ_N^W change.*

In order to understand the economic forces at work, we first consider the case in which $k < N$ and analyze the effect of increasing n_k on the consumption and implicit marginal tax rate of this type of individual. If $\lambda_k = 1$, an increase in n_k has no effect on the welfare weights. Consequently, the only effect of n_k on β_k is the direct welfare effect. Specifically, by (3.3), a marginal increase in n_k increases β_k by w_k. By (3.9), the net marginal benefit of a marginal increase in c_k therefore increases by $w_k v'(c_k)$. It also follows from (3.9) that a marginal increase in n_k causes the resource cost of a marginal increase in c_k to increase by γ (the marginal disutility of labor) because the extra consumption is given to more individuals of this type. Because the optimal implicit marginal tax rate is positive, $w_k v'(c_k) > \gamma$. Hence, an increase in n_k causes the marginal benefit of increasing c_k to increase by more than the marginal cost. This provides an incentive to increase c_k (and thereby reduce τ_k^W) when n_k increases. Loosely speaking, as n_k increases, the tax wedge between the gross wage paid to and the net-of-tax wage received by individuals of type k becomes more costly to society. The government's optimal response to this increased social cost is to reduce the size of this tax wedge.

When $\lambda_k \neq 1$, an increase in n_k also affects the (rescaled) welfare weights. As we have already noted, changes in these weights affect the social value of the utility changes of higher-skilled individuals required to re-establish the binding adjacent downward incentive compatibility constraints following an increase in c_k. The change in this social value might reinforce or attenuate the direct welfare effect described in the preceding paragraph. As the proof of Theorem 1 demonstrates, the direct welfare effect of an increase in n_k on c_k and τ_k^W (i.e., the direct tax wedge effect) always outweighs the effect that operates through the rescaling of the welfare weights.

The only channel through which an increase in the number of individuals of skill-type k can affect the consumption or implicit marginal tax rate faced by individuals of type i ($i \neq k, N$) is through changes in the reduced-form

13

welfare weight β_i. Parts (ii)–(iv) of Theorem 1 follow directly from this observation. If, for example, an increase in n_k induces an increase in β_i, then the marginal benefit of c_i increases and the government optimally increases c_i and reduces τ_i^W. The circumstances under which such an change occurs are described in our discussion of Lemma 3.

It is always optimal for the highest-skilled individuals to face a zero marginal tax rate. Because preferences are quasilinear in income, there is only one consumption level compatible with this restriction and, hence, both c_N and τ_N^W are completely insensitive to the demographic structure of the economy.

Theorem 1 provides a way to determine the sign of the comparative static response of every individual's optimal consumption and optimal marginal tax rate to an increase in the number of individuals of any type.[9] For any pair of distinct types i and k with $i \neq N$, the direction of change in the optimal values of c_i and τ_i^W in response to an increase in n_k depend on the sign of $(1 - \lambda_k)$ and on which of i and k is the greater. However, because $\lambda_1 > 1$ and $\lambda_N < 1$, it is possible to provide more succinct descriptions of the comparative statics of this model than in Theorem 1 when it is the number of individuals at one end of the skill distribution that changes.

Corollary 1.

(i) *An increase in n_1 results in* (a) *an increase in the optimal value of c_1, a decrease in the optimal values of $c_2, \ldots, c_{N-1}$, and no change in the optimal value of c_N and* (b) *a decrease in the optimal value of τ_1^W, an increase in the optimal values of $\tau_2^W, \ldots, \tau_{N-1}^W$, and no change in the optimal value of τ_N^W.*

(ii) *An increase in n_N results in* (a) *a decrease in the optimal values of $c_1, \ldots, c_{N-1}$ and no change in the optimal value of c_N and* (b) *an increase in the optimal values of $\tau_1^W, \ldots, \tau_{N-1}^W$ and no change in the optimal value of τ_N^W.*

An increase in n_1 reduces the normalized social welfare weights attached to all other types of individuals. As we have seen in Lemma 3, as a consequence, the reduced-form welfare weight β_i decreases for all i for which $1 < i < N$. Thus, for such i, the social marginal benefit of c_i decreases,

---

[9]For the reasons discussed in Weymark (1987), it is not, in general, possible to provide comparative static results for changes in before-tax incomes.

resulting in a decrease in the optimal value of c_i. On the other hand, an increase in n_N increases the normalized social welfare weights and decreases the reduced-form welfare weights of all other types of individuals, resulting in a decrease in their optimal consumptions.

Theorem 1 can also be used to determine how the optimal consumptions and marginal tax rates change when the numbers of more than one skill type change provided that the responses for each skill type reinforce each other. If they do not, then the comparative statics depend on the values of the parameters that characterize the economy, in which case the formulae provided in the proof of Theorem 1 can be used to determine the comparative static responses to such changes in the composition of the population.

4. The Simula Model

Simula (2010) considers the optimal nonlinear income tax problem when consumers have quasilinear-in-consumption preferences of the form

$$u_i^S(c, y) = \eta c - h\left(\frac{y}{w_i}\right), \tag{4.1}$$

where the function h is thrice continuously differentiable, increasing, and strictly convex with $h(0) = 0$, $h'''(r) \geq 0$ for all r, and $\lim_{r \to \infty} h'(r) = \infty$. The marginal utility of consumption, η, is a positive constant.

The information rent accruing to a type $i+1$ individual when his adjacent downward incentive compatibility constraint is binding is now

$$R_i^S(y_i) = h\left(\frac{y_i}{w_i}\right) - h\left(\frac{y_i}{w_{i+1}}\right), \quad i = 1, \ldots N, \tag{4.2}$$

where, as before, $w_{N+1} = w_N$. Convexity of h implies that R_i^S is increasing in y_i. The assumption on the third derivative of h implies that R_i^S is convex in y_i.

Using these expressions for the information rents, it is possible to derive a reduced-form optimization problem for the Simula model that can be used to determine the government's optimal choice of before-tax incomes. This problem is presented in following lemma, which generalizes Simula (2010, Problem 2) by allowing for the number of individuals of each type to be nonconstant.

15

Lemma 4. *The optimal before-tax incomes in the Simula model are found by solving*

$$\max_{y_1,\ldots,y_N} \sum_{i=1}^{N} n_i \left[\eta y_i - h\left(\frac{y_i}{w_i}\right) \right] + \sum_{i=1}^{N} \alpha_i R_i^S(y_i) \qquad (4.3)$$

subject to

$$0 \le y_1 \le y_2 \le \cdots \le y_N. \qquad (4.4)$$

With full information about types, it is socially beneficial to increase the consumption of each individual of type i by a self-financed marginal increase in y_i if the derivative of the term in square brackets in (4.3) for type i is positive. To take account of the self-selections constraints, $\alpha_i \frac{dR_i^S(y_i)}{dy_i}$ is first added to this derivative before determining the sign of the resulting expression. For $i = N$, no adjustment is required because an increase in y_N violates no self-selection constraint. For any other i, the adjustment is negative because $\frac{dR_i^S(y_i)}{dy_i} > 0$ and $\alpha_i < 0$. Thus, as in (3.8), the social costs of the self-selection constraints are captured by the last term in the reduced-from objective function.

In the Simula model, following an increase in y_i, the incomes of higher types are first adjusted to restore the binding adjacent downward incentive compatibility constraints and then everyone's consumption is changed by a common amount so as to restore budget balance. Because preferences are quasilinear in consumption, the latter changes preserve the pattern of binding self-selection constraints. It is straightforward to use these observations to interpret the individual terms in the expression for α_i along the lines that was done above for β_i.[10]

The following lemma can be established using the argument that was used to establish Lemma 3.[11]

Lemma 5. *Suppose that there is a marginal increase in n_k accompanied by a proportionate rescaling of the λs so as to re-establish (2.9). Then:*

(i) $\dfrac{\partial \alpha_i}{\partial n_k}$ *has the same sign as* $(1 - \lambda_k)$ *if $k \le i$ and $i < N$.*

---

[10]See Simula (2010) for an alternative interpretation of the α weights and the information rents in his model.

[11]The only difference in the proof is that $(w_{i+1} - w_i)$ is replaced by 1 and that the proof for $k < i$ now also applies for $k = i$.

(ii) $\dfrac{\partial \alpha_i}{\partial n_k}$ *has the opposite sign to* $(1 - \lambda_k)$ *if* $k > i$.

(iii) $\dfrac{\partial \alpha_i}{\partial n_k} = 0$ *if* $k \leq N$ *and* $i = N$.

Part (i) also includes the case of $k = i$ because the effect of changes in y_i on the utility of an individual of type i are not included in α_i. The intuition for these results is essentially the same as the intuition for the results in Lemma 3.

Henceforth, we assume that the nonnegativity and monotonicity conditions (4.4) are all non-binding. The government's reduced-form optimization problem is then the unconstrained maximization of (4.3). The first-order necessary conditions for the optimum are

$$n_i h' \left(\frac{y_i}{w_i} \right) \frac{1}{w_i} - \alpha_i \frac{dR_i^S(y_i)}{dy_i} = \eta n_i, \quad i = 1, \ldots, N. \tag{4.5}$$

The left-hand side of (4.5) describes the marginal costs of an increase in y_i. There are two sources of marginal cost. Because individuals of type i now work more, they suffer a loss in utility. This is captured by the first term on the left-hand side of (4.5). In addition, because the information rent $R_i^W(y_i)$ is increasing in y_i, the social cost associated with this rent also increases with an increase in y_i. These costs are accounted for by the second term on the left-hand side of (4.5). The right-hand side of (4.5) is the marginal benefit of an increase in y_i. Because a one-unit increase in y_i produces n_i units of output, such a change increases the total utility of individuals of type i by ηn_i. According to (4.5), at the optimal before-tax income of a type i individual, the marginal benefit of an increase in y_i is equal to its marginal cost.

When preferences have the functional form given in (4.1), the implicit marginal tax rates are given by

$$\tau_i^S = 1 - h' \left(\frac{y_i}{w_i} \right) \frac{1}{\eta w_i}, \quad i = 1, \ldots, N. \tag{4.6}$$

It follows directly from (4.5) and (4.6) that the optimal implicit marginal income tax rates (which satisfy the standard properties) are

$$\tau_i^S = -\frac{\alpha_i}{\eta n_i} \frac{dR_i^S(y_i)}{dy_i}, \quad i = 1, \ldots, N. \tag{4.7}$$

17

For fixed η and w_i, τ_i^S and y_i move in opposite directions.

The comparative static responses of the optimal incomes and the optimal implicit marginal tax rates to an increase in the number of individuals of any skill type are characterized in Theorem 2.

Theorem 2. *Suppose that there is a marginal increase in n_k accompanied by a proportionate rescaling of the λs so as to re-establish* (2.9).

   (i) *If $k < N$, then the optimal y_k increases and the optimal τ_k^S decreases.*

   (ii) *If $\lambda_k > 1$, then the optimal y_i decreases and the optimal τ_i^S increases for all $i = k+1, \ldots, N-1$ and the optimal y_i increases and the optimal τ_i^S decreases for all $i = 1, \ldots, k-1$.*

   (iii) *If $\lambda_k < 1$, then the optimal y_i increases and the optimal τ_i^S decreases for all $i = k+1, \ldots, N-1$ and the optimal y_i decreases and the optimal τ_i^S increases for all $i = 1, \ldots, k-1$.*

   (iv) *If $\lambda_k = 1$, then the optimal y_i and the optimal τ_i^S do not change for all $i \neq k, N$.*

   (v) *Neither the optimal y_N nor the optimal τ_N^S change.*

A marginal increase in n_k perturbs both sides of the optimality condition (4.5) for $i = k$. The right-hand side rises by η, the marginal utility of the output produced by an additional person of type k. The first term on the left-hand side increases by $h'\left(\frac{y_k}{w_k}\right)\frac{1}{w_k}$, the marginal disutility of before-tax income, now experienced by one more individual of type k. For types $k < N$, the optimal implicit marginal tax rate is positive, which implies that $\eta > h'\left(\frac{y_k}{w_k}\right)\frac{1}{w_k}$. Thus, the total utility of individuals of type k increases when n_k increases. Therefore, in the absence of changes in the marginal costs of information rents (as would be the case if $\lambda_k = 1$), a marginal increase in n_k augments the net social value of y_k and the government has an incentive to increase y_k and reduce τ_k^S. Once again, the government wishes to reduce a tax wedge when that wedge applies to more individuals. Typically (i.e., when $\lambda_k \neq 1$), a change in n_k also affects the marginal information rents, but as we have shown in Theorem 2, these effects never completely offset the tax-wedge effect.

It is apparent from (4.7) that a marginal increase in n_k can affect the optimal marginal income tax rate faced by individuals of type i ($i \neq k$) only

through its affect, if any, on α_i. Suppose, for example, that an increase in n_k induces an increase in the absolute value of α_i. Then the marginal cost of the additional information rents associated with an increase in y_i goes up. This gives the government an incentive to reduce y_i. Accordingly, τ_i^S is increased. Because Lemma 5 identifies the effects of an increase in any n_k on any α_i, in Theorem 2 we are able to provide a complete account of the effects of an increase in n_k on the solutions to government's reduced-form optimization problem.

Theorem 2 is strikingly similar to Theorem 1.[12] Indeed, the two theorems are identical with respect to their statements concerning the effects of an increase in n_k on implicit marginal tax rates. The similarities between the the theorems also provides a measure of confidence in the results being due to fundamental economic forces, rather than being artifacts of specific functional forms. We have already discussed at length how, in both frameworks, an increase in n_k amplifies the effects of the tax wedge for individuals of type k, thereby providing a reason to reduce this wedge. In both models, the changes in the consumption bundle of individuals of type k as a consequence of increasing n_k impact type i ($i \neq k$) through changes in the information rents accruing to individuals of type $k+1$ and in the social cost α_i of these rents for individuals of type i. If the information rents become more costly at the margin, then the distortions increase.

5. Conclusion

An individual can be viewed in one of three lights in an optimal nonlinear income tax problem. First, he can be viewed as person whose labor supply is distorted downward due to information asymmetries. Second, he can be seen as the receiver of information rents, thereby contributing to the distortion of less productive individuals. Third, he can be regarded as a contributor to the financing of the information rents paid to individuals of higher types. Our results suggest that the effect of the number of individuals of a particular skill-type on optimal marginal income tax rates depends on which light is switched on in a particular comparative static experiment. Once a viewpoint is adopted, however, the results are unambiguous.

Quasilinearity of preferences is a strong assumption. It is unlikely that some of our more exact results, for example the invariance of the consump-

---

[12]It is straigthforward to state an exact analogue of Corollary 1 for the Simula model.

19

tion or before-tax income of individuals of type i to changes in n_k $(i \neq k)$ when $\lambda_k = 1$, hold for more general preferences. It might be reasonably conjectured, however, that the marginal distortions applied to these types will not change if type k individuals have the average social welfare weight. Moreover, one might expect the underlying forces uncovered by our analysis to operate in all finite-population optimal nonlinear income tax models.

Some of the insights of our analysis may carry over into related screening problems, such as nonlinear pricing. It is natural to ask how a monopolist might adjust its optimal pricing schedule when the distribution of consumer preferences changes. Our results suggest that an increase in the number of individuals of a particular taste type would induce a decrease in the marginal distortion associated with that type. It would be of interest to test this intuition in, for example, the model of Maskin and Riley (1984).

Appendix

Proof of Lemma 3. Let $i < N$ and $k < i$. Then β_i is given by the top line of (3.3). Differentiation yields

$$\frac{\partial \beta_i}{\partial n_k} = \sum_{j=i+1}^{N} n_j \frac{\partial \lambda_j}{\partial n_k}(w_{i+1} - w_i). \tag{A.1}$$

Substituting (2.15) into (A.1) yields

$$\frac{\partial \beta_i}{\partial n_k} = \sum_{j=i+1}^{N} n_j \frac{\lambda_j(1 - \lambda_k)}{H}(w_{i+1} - w_i). \tag{A.2}$$

Part (i) of Lemma 3 follows from (A.2).

Now, suppose that $k > i$. Differentiating (3.3) yields

$$\frac{\partial \beta_i}{\partial n_k} = \sum_{j=i+1}^{N} n_j \frac{\partial \lambda_j}{\partial n_k}(w_{i+1} - w_i) + (\lambda_k - 1)(w_{i+1} - w_i). \tag{A.3}$$

Substituting (2.15) into (A.3) and rearranging yields

$$\frac{\partial \beta_i}{\partial n_k} = \left[\sum_{j=i+1}^{N} n_j \lambda_j - H \right] \frac{(1 - \lambda_k)}{H}(w_{i+1} - w_i). \tag{A.4}$$

20

By the normalization rule (2.9), the term in square brackets in (A.4) is negative, from which Part (ii) of Lemma 3 follows.

The other parts of the lemma follow directly from the bottom line of (3.3). □

Proof of Theorem 1. Let $k < N$. Differentiating (3.10) yields

$$\frac{\partial c_k}{\partial n_k} = \frac{\varphi'(\cdot)\gamma}{\beta_k^2}\left[\beta_k - n_k\frac{\partial \beta_k}{\partial n_k}\right]. \tag{A.5}$$

Because φ is decreasing, the right-hand side of (A.5) has the opposite sign to the term in square brackets in (A.5). Using (2.15) and (3.3),

$$\beta_k - n_k\frac{\partial \beta_k}{\partial n_k} = \left[\sum_{j=k+1}^{N}[n_j\lambda_j - n_j] - n_k\sum_{j=k+1}^{N}\frac{n_j\lambda_j(1-\lambda_k)}{H}\right](w_{k+1}-w_k). \tag{A.6}$$

Because the λs are decreasing and satisfy (2.9), (A.6) is negative if $\lambda_k \leq 1$. Then, by (A.5), $\frac{\partial c_k}{\partial n_k} > 0$.

Now, suppose that $\lambda_k > 1$. By (2.9),

$$\sum_{j=k+1}^{N}[n_j\lambda_j - n_j] = -\sum_{j=1}^{k}[n_j\lambda_j - n_j]. \tag{A.7}$$

Substituting (A.7) into (A.6) and rearranging yields

$$\begin{aligned}
\beta_k &- n_k\frac{\partial \beta_k}{\partial n_k} \\
&= \left\{-\sum_{j=1}^{k-1}[n_j\lambda_j - n_j] - n_k(\lambda_k - 1) - n_k\frac{\sum_{j=k+1}^{N}n_j\lambda_j(1-\lambda_k)}{H}\right\}(w_{k+1} - w_k) \\
&= \left\{-\sum_{j=1}^{k-1}[n_j\lambda_j - n_j] + n_k(1-\lambda_k)\left[\frac{H - \sum_{j=k+1}^{N}n_j\lambda_j}{H}\right]\right\}(w_{k+1} - w_k)
\end{aligned} \tag{A.8}$$

Because the λs are decreasing and satisfy (2.9), the first term inside the braces in (A.8) is negative. Because of (2.9), the second term inside the braces is also negative when $\lambda_k > 1$. Thus, (A.8) is negative. Hence, by (A.5), $\frac{\partial c_k}{\partial n_k} > 0$ in this case as well. Part (i) of Theorem 1 then follows.

21

To establish Part (ii) of Theorem 1, suppose, once again, that $\lambda_k > 1$ and let $i \neq k$. Differentiating (3.10), we obtain

$$\frac{\partial c_i}{\partial n_k} = -\frac{\varphi'(\cdot)\gamma}{\beta_i^2}\frac{\partial \beta_i}{\partial n_k}. \tag{A.9}$$

Because φ is decreasing, $\frac{\partial c_i}{\partial n_k}$ has the same sign as $\frac{\partial \beta_i}{\partial n_k}$. Now let $i > k$ with $i \neq N$. By Lemma 3.(i), $\frac{\partial \beta_i}{\partial n_k} < 0$, so that c_i decreases (hence, τ_i^W increases) when n_k increases. On the other hand, if $i < k$, Lemma 3.(ii) implies that $\frac{\partial \beta_i}{\partial n_k} > 0$, and c_i increases (hence, τ_i^W decreases) when n_k increases.

The proof of Part (iii) of Theorem 1 is analogous to the proof of Part (ii). Part (iv) follows directly from (A.9) and Lemma 3.

For $i = N$, substituting (3.3) into (3.10) yields

$$c_N = \varphi\left(\frac{\gamma}{w_N}\right), \tag{A.10}$$

which is independent of any of the n_k. Part (v) then follows. $\qquad\square$

Proof of Lemma 4. We present only the derivation of the reduced-form objective function. The other aspects of the proof are identical to those presented by Simula (2010).

Let U_i be the utility of an individual of type i, measured according to (4.1), at a solution to the optimal nonlinear tax problem. Equation (2.7) implies that

$$U_i = \eta c_i - h\left(\frac{y_i}{w_i}\right) = \eta c_{i-1} - h\left(\frac{y_{i-1}}{w_i}\right), \quad i = 2, \ldots, N. \tag{A.11}$$

Adding and subtracting $h\left(\frac{y_{i-1}}{w_{i-1}}\right)$ on the right-hand side of (A.11) and using (4.2) implies that

$$U_i = U_{i-1} + R_{i-1}^S(y_{i-1}), \quad i = 2, \ldots, N. \tag{A.12}$$

Hence,

$$U_i = U_1 + \sum_{j=1}^{i-1} R_j^S(y_j), \quad i = 2, \ldots, N. \tag{A.13}$$

It follows from (A.13) that

$$\sum_{i=1}^{N} n_i U_i = \sum_{i=1}^{N} n_i U_1 + \sum_{i=2}^{N}\left[n_i \sum_{j=1}^{i-1} R_j^S(y_j)\right]. \tag{A.14}$$

22

Reversing the order of the double-sum on the right-hand side of (A.15) and simplifying yields

$$\sum_{i=1}^{N} n_i U_i = H U_1 + \sum_{i=1}^{N-1} \left(\sum_{j=i+1}^{N} n_j \right) R_i^S(y_i). \qquad \text{(A.15)}$$

On the other hand, by (4.1),

$$\sum_{i=1}^{N} n_i U_i = \sum_{i=1}^{N} n_i \eta c_i - \sum_{i=1}^{N} n_i h \left(\frac{y_i}{w_i} \right). \qquad \text{(A.16)}$$

Because the materials balance constraint (2.6) binds,

$$\sum_{i=1}^{N} n_i U_i = \sum_{i=1}^{N} n_i \eta y_i - \sum_{i=1}^{N} n_i h \left(\frac{y_i}{w_i} \right). \qquad \text{(A.17)}$$

Equating the right-hand sides of (A.15) and (A.17) implies

$$U_1 = \frac{1}{H} \left\{ \sum_{i=1}^{N} n_i \left[\eta y_i - h \left(\frac{y_i}{w_i} \right) \right] - \sum_{i=1}^{N-1} \left(\sum_{j=i+1}^{N} n_j \right) R_i^S(y_i) \right\}. \qquad \text{(A.18)}$$

By (A.13), social welfare is given by

$$\begin{aligned}
\mathcal{W} &= \left(\sum_{i=1}^{N} n_i \lambda_i \right) U_1 + \sum_{i=2}^{N} n_i \lambda_i \sum_{j=1}^{i-1} R_j^S(y_j) \\
&= \left(\sum_{i=1}^{N} n_i \lambda_i \right) U_1 + \sum_{i=1}^{N-1} \left(\sum_{j=i+1}^{N} n_j \lambda_j \right) R_i^S(y_i).
\end{aligned} \qquad \text{(A.19)}$$

Substituting (A.18) into (A.19) and using the normalization rule (2.9) yields

$$\mathcal{W} = \sum_{i=1}^{N} n_i \left[\eta y_i - h \left(\frac{y_i}{w_i} \right) \right] + \sum_{i=1}^{N-1} \left[\sum_{j=i+1}^{N} (n_j \lambda_j - n_j) \right] R_i^S(y_i), \qquad \text{(A.20)}$$

which is the objective function in (4.3). $\qquad \square$

Proof of Theorem 2. Let $k < N$. Implicit differentiation of (4.5) yields

$$\left[n_k h'' \left(\frac{y_k}{w_k} \right) \frac{1}{w_k^2} - \alpha_k \frac{d^2 R_k^S(y_k)}{dy_k^2} \right] dy_k =$$

$$\left[\eta - h' \left(\frac{y_k}{w_k} \right) \frac{1}{w_k} + \frac{dR_k^S(y_k)}{dy_k} \frac{\partial \alpha_k}{\partial n_k} \right] dn_k \quad \text{(A.21)}$$

23

Because h is strictly convex, R_k^S is convex, and $\alpha_k < 0$, the term in square brackets on the left-hand side of (A.21) is positive. Thus, by the implicit function theorem, $\frac{\partial y_k}{\partial n_k}$ exists and has the same sign as the term in square brackets on the right-hand side of (A.21). Using (4.5),

$$\eta - h'\left(\frac{y_k}{w_k}\right)\frac{1}{w_k} + \frac{dR_k^S(y_k)}{dy_k}\frac{\partial \alpha_k}{\partial n_k} = \left[\frac{\partial \alpha_k}{\partial n_k} - \frac{\alpha_k}{n_k}\right]\frac{dR_k^S(y_k)}{dy_k}. \tag{A.22}$$

Using (2.15) and (3.7), (A.22) implies

$$\eta - h'\left(\frac{y_k}{w_k}\right)\frac{1}{w_k} + \frac{dR_k^S(y_k)}{dy_k}\frac{\partial \alpha_k}{\partial n_k} = \\ \left[\sum_{j=k+1}^{N}\frac{n_j\lambda_j(1-\lambda_k)}{H} - \frac{\alpha_k}{n_k}\right]\frac{dR_k^S(y_k)}{dy_k}. \tag{A.23}$$

The term in square brackets on the right-hand side of (A.23) is clearly positive when $\lambda_k \leq 1$. When $\lambda_k > 1$, the argument used to establish that the first term on the right-hand side of (A.6) in the proof of Theorem 1 is negative implies that the term in square brackets on the right-hand side of (A.23) is positive because the latter is the former divided by $-n_k$. Because $R_k^S(y_k)$ is increasing in y_k, the right-hand sides of (A.21) and (A.23) are positive. Hence, $\frac{\partial y_k}{\partial n_k} > 0$, thereby establishing Part (i) of Theorem 2.

To establish Part (ii), first note that it follows directly from (4.7) that $\frac{\partial \tau_i^S}{\partial n_k}$ has the opposite sign to $\frac{\partial \alpha_i}{\partial n_k}$ when $i \neq k$. If $i > k$ and $i < N$, by Lemma 5.(i), $\frac{\partial \alpha_i}{\partial n_k} < 0$, so that τ_i^S increases when n_k increases. Because η and w_i are fixed, it then follows from (4.6) that y_i decreases. On the other hand, if $i < k$, Lemma 5.(ii) implies that $\frac{\partial \alpha_i}{\partial n_k} < 0$. Therefore, τ_i^S decreases and y_i increases when n_k increases.

The proof of Part (iii) is analogous to the proof of Part (ii). Part (iv) follows directly from (4.7) and Lemma 5.

Finally, let $k = N$. Substituting (3.7) into (4.5) yields

$$\eta = h'\left(\frac{y_i}{w_i}\right), \tag{A.24}$$

which is independent of any of the n_k. Part (v) then follows. $\qquad\square$

References

Boadway, R., Pestieau, P., 2007. Tagging and redistributive taxation. Annales d'Économie et de Statistique 83–84, 123–147.

Brett, C., Weymark, J. A., 2003. Financing education using optimal redistributive taxation. Journal of Public Economics 87, 2549–2569.

Brett, C., Weymark, J. A., 2008a. The impact of changing skill levels on optimal nonlinear income taxes. Journal of Public Economics 92, 1765–1771.

Brett, C., Weymark, J. A., 2008b. Public good provision and the comparative statics of optimal nonlinear income taxation. International Economic Review 49, 255–290.

Guesnerie, R., 1995. A Contribution to the Pure Theory of Taxation. Cambridge University Press, Cambridge.

Guesnerie, R., Seade, J., 1982. Nonlinear pricing in a finite economy. Journal of Public Economics 17, 157–179.

Hamilton, J., Pestieau, P., 2005. Optimal income taxation and the ability distribution: Implications for migration equilibria. International Tax and Public Finance, 29–45.

Hellwig, M., 2007. A contribution to the theory of optimal utilitarian income taxation. Journal of Public Economics 91, 1449–1477.

Maskin, E., Riley, J., 1984. Monopoly with incomplete information. Rand Journal of Economics 15, 171–196.

Matthews, S., Moore, J., 1987. Monopoly provision of quality and warranties: An exploration in the theory of multi-dimensional screening. Econometrica 55, 441–467.

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

Röell, A. A., 1985. A note on the marginal tax rate in a finite economy. Journal of Public Economics 28, 267–272.

Simula, L., 2010. Optimal nonlinear income tax and nonlinear pricing: Optimality conditions and comparative static properties. Social Choice and Welfare, forthcoming.

Weymark, J. A., 1986. A reduced-form optimal nonlinear income tax problem. Journal of Public Economics 30, 199–217.

Weymark, J. A., 1987. Comparative static properties of optimal nonlinear income taxes. Econometrica 55, 1165–1185.