Soc Choice Welfare (1986) 3:213–232

Social Choice and Welfare

Bunching Properties of Optimal Nonlinear Income Taxes*

John A. Weymark

Department of Economics, University of British Columbia, Vancouver, B.C., V6T 1Y2 Canada and Department of Economics, Northwestern University, Evanston, Illinois 60201, USA

Received September 19, 1984 / Accepted May 26, 1986

Abstract. Bunching is said to occur if individuals with different characteristics receive the same commodity bundle. This article analyzes bunching in a finite population optimal nonlinear income tax problem. Several easily-computed sufficient conditions for the optimality of particular bunching patterns as well as a simple necessary and sufficient condition for the optimal allocation to exhibit no bunching are presented. In addition, a characterization of the optimal allocation is provided. Is is shown that the bunching pattern obtained by S. Lollivier and J.-C. Rochet is a consequence of a convexity condition which is automatically satisfied in their continuum model but which is not generally satisfied in a finite model.

1. Introduction

Bunching is said to occur if individuals with different characteristics receive the same commodity bundle. In standard nonlinear income tax models, first-best allocations typically result in complete separation; i.e. bunching does not occur at first-best optima. However, as recognized in Mirrlees' [8] pioneering article, bunching may well be optimal when second-best constraints are taken into account. In the Mirrlees problem, second-best considerations arise because the government has only incomplete information about the characteristics of each individual in society. Thus, incentives must be provided if consumers are to reveal their true characteristics. In some circumstances, the benefits to be gained from obtaining this information are outweighed by the negative impact the incentive structure has on the labour supply of high-skilled individuals; as a consequence, it is desirable to have some individuals bunched together.

* Discussions with Steven Matthews and Dilip Mookherjee and the comments of two anonymous referees have been extremely helpful. The hospitality of the Center for Mathematical Studies in Economics and Management Science at Northwestern University and the research support of the Social Sciences and Humanities Research Council of Canada are gratefully acknowledged.

Most work to date on second-best income taxation assumes that there is a continuum of individuals and also assumes that these individuals differ only in a single characteristic (or adopts an assumption that effectively limits differences to a single dimension). In the Mirrlees [8] problem, this characteristic is productive ability. Bunching in these models is associated with corners (kinks) in the income tax schedule.

The Mirrlees [8] optimal income tax problem is a particular example of a general class of problems containing incentive-compatibility (self-selection) constraints. In general, the presence of these constraints greatly complicates the analysis. However, in some circumstances, if a quasilinearity assumption is adopted, the original problem can be replaced by a relatively simple reduced-form problem involving only a subset of the decision variables. It appears that Mussa and Rosen [9], in their analysis of a monopolist setting a price schedule while possessing only imperfect information about the demand for the firm's product line, were the first to discover this phenomenon. This result has been independently established in other contexts, notably by Myerson [10] in his study of optimal auction design. Guesnerie and Laffont [3] apply similar methods to an abstract self-selection model, thereby synthesizing results found in a number of particular applications. Lollivier and Rochet [6] demonstrate that the Mirrlees tax problem with quasilinear preferences has the appropriate structure that permits such a simplification; in their reduced-form problem they optimize only the amount of the consumption good allocated to each individual, rather than both the amounts of consumption and labour.

These reduced-form problems are particularly appropriate for studying bunching, and have been used for that purpose by Baron and Myerson [1], Guesnerie and Laffont [3], Lockwood [5], Lollivier and Rochet [6], Maskin and Riley [7], Mussa and Rosen [9], and Myerson [10], among others, in a variety of contexts. Mussa and Rosen [9], for example, develop a condition that identifies which consumers should be bunched together. The Mussa-Rosen condition provides an implicit characterization of the optimal bunching pattern. However, Guesnerie and Laffont [3] argue that this approach is operational and they outline an algorithm designed to determine which sets of individuals satisfy conditions of the sort introduced by Mussa and Rosen.[1]

Lollivier and Rochet [6] are able to obtain more concrete conclusions concerning bunching in the Mirrlees income tax problem by making more restrictive assumptions. Working with a continuum of individuals, they assume that (i) everyone has common quasilinear preferences, (ii) the social-welfare function is weighted utilitarianism with the weights satisfying a weak redistributive assumption, and (iii) skills are uniformly distributed. With these assumptions, Lollivier and Rochet demonstrate that all bunching, if any, should be with the lowest-skilled

[1] In a continuum model, Seade [12] obtains necessary conditions which must be satisfied at the solution to a quite general one-dimensional income tax problem, conditions which explicitly account for the possibility that bunching may be optimal. Seade's conditions extend the partial results available in Mirrlees [8]. Unfortunately, the simplest form of Seade's conditions relating to the presence of bunching employ the optimal values for the multipliers associated with the incentive-compatibility and production constraints, which makes Seade's conditions difficult to operationalize. The Mussa-Rosen [9] condition does not utilize any multipliers. Additional discussion of bunching in a continuum version of Mirrlees' problem may be found in Seade [11].

individual. More precisely, there is a skill level such that all individuals with abilities less than this value are bunched together and all individuals with abilities higher than this value receive distinct commodity bundles.

Determining whether it is optimal to have bunching is formally related to the problem of whether it is desirable to have a binding nonnegativity constraint on consumption for some individual. The main purpose of this article is to provide a detailed discussion of both of these phenomena in a finite population optimal nonlinear income tax problem. A secondary purpose of this article is to demonstrate that an explicit solution to the optimal income tax problem studied here can be obtained by utilizing a finite version of a technique developed by Myerson [10] in his study of optimal auction design. As maintained hypotheses, the assumptions employed by Lollivier and Rochet [6] to study bunching with a continuum of consumers are adopted. In sharp contrast to their result, for a finite population these assumptions are compatible with any conceivable pattern of bunching that does not involve the highest-skilled individual.

In Sect. 2, the model is introduced and background results from Weymark [13] are presented. In that section, a vector of adjusted wage rates is defined; the adjustments reflect the externalities which result from the information asymmetry. In Sect. 3, it is shown that a number of results concerning bunching follow directly from the sign and monotonicity properties of this vector, including a necessary and sufficient condition for the optimal allocation to exhibit no bunching. In Sect. 4, these adjusted wage rates are used to provide a characterization of the optimal allocation, and thus a characterization of the optimal bunching pattern. In Sect. 5, it is shown that if the vector of adjusted wage rates satisfies a certain convexity condition, then any bunching which is optimal must be at the bottom of the skill distribution. Furthermore, it is shown that this convexity condition is satisfied if the difference in wage rates between adjacent individuals is a constant. This property of the skill distribution is automatically satisfied with a uniform skill distribution in the continuum, which accounts for the simplicity of the bunching pattern observed by Lollivier and Rochet [6]. In Sect. 6, the consequences of dropping the assumption that skills are uniformly distributed is briefly considered. A few concluding remarks appear in Sect. 7. All proofs appear in the Appendix.

2. The Model and Background Results

The model considered here is a finite population version of the one introduced by Mirrlees [8] with the additional restriction that preferences are quasilinear. In the Mirrlees model, consumers have identical preferences but differ in ability.

There are a fixed finite number of consumers, indexed by $i=1,\ldots,N$ where $N\geq 2$, and two commodities, consumption and labour. Person i's consumption and labour supply are denoted by c_i and l_i, respectively. Differences in ability are reflected in the wage rates w_i. The economy has a constant-returns-to-scale technology, so the wage rates are fixed. Following Lollivier and Rochet [6], skills are assumed to be positive and uniformly distributed. Without loss of generality, it is assumed that all consumers receive a distinct wage rate. Individuals are indexed in terms of the wage they receive, so

$$0 < w_1 < \ldots < w_N \ . \tag{1}$$

Person i's (before-tax) income is

$$y_i := w_i l_i \qquad i = 1, \ldots, N \tag{2}$$

which is also i's labour supply in efficiency units. A *commodity bundle* for person i is a vector $(y_i, c_i) \in \mathbf{R} \times \mathbf{R}_+$.[2]

Individuals have common preferences given by the quasilinear utility function u: $\mathbf{R} \times \mathbf{R}_+ \to \mathbf{R}$ where

$$u(l_i, c_i) := v(c_i) - \gamma l_i \ . \tag{3}$$

It is assumed that v is a continuously differentiable, strictly increasing, strictly concave function with $v(0) = 0$, $v'(0) = \infty$, and $\lim_{r \to \infty} v'(r) = 0$. The parameter γ is assumed to be positive.

Substituting (2) into (3) yields the utility function $\tilde{U}^i$: $\mathbf{R} \times \mathbf{R}_+ \to \mathbf{R}$ defined in terms of income and consumption where

$$\tilde{U}^i(y_i, c_i) := v(c_i) - \gamma y_i / w_i \qquad i = 1, \ldots, N \ . \tag{4}$$

While the utility functions defined in (4) are the appropriate cardinalizations for making interpersonal comparisons, it is more convenient to work with monotone transforms of these functions, U^i: $\mathbf{R} \times \mathbf{R}_+ \to \mathbf{R}$ where

$$U^i(y_i, c_i) := w_i v(c_i) - \gamma y_i \qquad i = 1, \ldots, N \ . \tag{5}$$

Thus, in terms of income and consumption, the preferences of different individuals vary in a systematic fashion; higher ability individuals have flatter indifference curves. As a consequence, the indifference curves in income-consumption space of different individuals intersect only once. This property of preferences is known in the self-selection literature as the *single-crossing property*.

An *allocation* is a vector $a := (y, c)$ where $y := (y_1, \ldots, y_N)$ is an *income vector* and $c := (c_1, \ldots, c_N)$ is a *consumption vector*. The *social welfare function* W: $\mathbf{R}^N \times \mathbf{R}^N_+ \to \mathbf{R}$ is of the weighted utilitarian form,

$$W(a) := \Sigma_i \mu_i \tilde{U}^i(y_i, c_i) \tag{6}$$

or, equivalently,

$$W(a) = \Sigma_i \lambda_i U^i(y_i, c_i) \tag{7}$$

where

$$\lambda_i := \mu_i / w_i \qquad i = 1, \ldots, N \ . \tag{8}$$

[2] Neither Guesnerie and Laffont [3] nor Lollivier and Rochet [6] consider a binding constraint on the sign or magnitude of y_i. Ignoring these constraints simplifies the discussion. Seade [11] considers this issue in detail using a continuum of abilities. An important feature of Seade's model is that the lowest-skilled individual receives a zero wage.

The vector of *skill-normalized welfare weights* $\lambda := (\lambda_1, \ldots, \lambda_N)$ is assumed to be strictly positive and monotone decreasing,

$$0 < \lambda_N < \ldots < \lambda_1 \ , \tag{9}$$

and is normalized to sum to N,

$$\Sigma_i \lambda_i = N \ . \tag{10}$$

Condition (9) is a weak redistribution assumption which, for example, is satisfied if (6) is the utilitarian social welfare function.

In this model, there is no loss of generality in considering only an aggregate production constraint, which also serves as the materials balance constraint in the model. The technology exhibits constant returns to scale,

$$p\Sigma_i c_i \leq \Sigma_i y_i \quad \text{with} \quad \Sigma_i c_i \geq 0 \ , \tag{11}$$

where $p > 0$. Letting income be the numeraire, p is the price of the consumption good.

The government knows the distribution of wages (skills) and the functional form of the utility function; it can also observe the actual pretax income of each individual but is not able to observe the wage or hours worked of any particular individual. The government's role is limited to setting an anonymous tax schedule. The set of allocations which can be generated by anonymous tax schedules when consumers optimize over the corresponding budget set is equivalent to the set of allocations which satisfy the *self-selection constraints*,

$$U^i(y_i, c_i) \geq U^i(y_j, c_j) \qquad \text{for all} \quad i,j \ . \tag{12}$$

Thus, the optimal nonlinear income tax problem can be stated in the following *direct form*.

Problem I. Choose an allocation $a \in \mathbf{R}^N \times \mathbf{R}^N_+$ to maximize the social welfare function (7) subject to the allocation satisfying the production constraint (11) and the self-selection constraints (12).

Because preferences satisfy the single-crossing property, any allocation which satisfies the self-selection constraints must have both consumption and income nondecreasing in ability. This result is well-known in the literature, see Cooper [2] for example, so it is stated here without proof.

Proposition 1. *A necessary condition for an allocation $a \in \mathbf{R}^N \times \mathbf{R}^N_+$ to satisfy the self-selection constraints is*

$$(y_1, c_1) \leq (y_2, c_2) \leq \ldots \leq (y_N, c_N) \tag{13}$$

with

$$(y_{i-1}, c_{i-1}) \ll (y_i, c_i) \quad \textit{if} \quad (y_{i-1}, c_{i-1}) \neq (y_i, c_i) \qquad i \geq 2 \ . \tag{14}$$

The main result in Weymark [13] establishes the equivalence of Problem I with a reduced-form problem which has only the consumption vector c as a variable. By Proposition 1, if an allocation a satisfies the self-selection constraints, the consumption vector c must be in

$$C:=\{c\in\mathbf{R}^N \mid c_i \geq c_{i-1}, i=1,\ldots,N\} \ , \tag{15}$$

where

$$c_0:=0 \ . \tag{16}$$

Define

$$\beta_i:=w_i+\left(i-\sum_{h=1}^{i}\lambda_h\right)(w_{i+1}-w_i) \ , \tag{17}$$

where w_{N+1} is an arbitrary number. Let $\beta:=(\beta_1,\ldots,\beta_N)$. It is natural to call β_i an *adjusted wage rate*. In view of the normalization rule, $\beta_N=w_N$. Because the welfare weights λ_i are positive and declining in ability, $\beta_i<w_i$ for all $i\neq N$. Using these new variables, define $\mathscr{W}: C\rightarrow\mathbf{R}$ by

$$\mathscr{W}(c):=\Sigma_i\beta_i v(c_i)-\gamma p\Sigma_i c_i \ . \tag{18}$$

A detailed economic interpretation of (18) may be found in Weymark [13]. The *reduced-form optimal nonlinear income tax problem* is Problem II.

Problem II. Choose a consumption vector $c\in C$ to maximize (18).

The main results in Weymark [13] are summarized in Theorem 1.

Theorem 1. *If $a^*=(y^*,c^*)$ is a solution to Problem I, then c^* is a solution to Problem II. If c^* is a solution to Problem II, then $(g(c^*),c^*)$ is a solution to Problem I where $g: C\rightarrow\mathbf{R}^N$ is given by*

$$g_1(c):=\frac{1}{N}\left\{p\Sigma_i c_i-\frac{1}{\gamma}\sum_{h=2}^{N}(N+1-h)w_h[v(c_h)-v(c_{h-1})]\right\} \tag{19}$$

and

$$g_i(c):=g_1(c)+\frac{1}{\gamma}\sum_{h=2}^{i}w_h[v(c_h)-v(c_{h-1})] \qquad i=2,\ldots,N \ . \tag{20}$$

If a^* is a solution to Problem I, (i) it is production efficient, i.e. satisfies (11) *with equality, and (ii) all the adjacent downward incentive constraints are binding, i.e.*

$$U^i(y_i,c_i)=U^i(y_{i-1},c_{i-1}) \qquad i=2,\ldots,N \ . \tag{21}$$

From a computational perspective, the reduced-form problem has many attractive features. The vector β which appears in the objective function can be calculated using simple arithmetic operations. Similarly, only simple computations are required to calculate the optimal income vector $g(c^*)$ once the optimal consumption vector c^* has been determined. The results presented in subsequent sections suggest that practical procedures are available for the computation of c^*.

Bunching occurs if two consumers with different characteristics receive the same commodity bundle. Here, the only characteristic that distinguishes individuals is the wage rate and, by assumption, all individuals have distinct wage rates. To analyze the optimal bunching pattern it is only necessary to consider the reduced-form problem, as Proposition 1 implies that individuals i and j $(i \neq j)$ are bunched *if and only if* $c_i = c_j$. Furthermore, because c must be monotone nondecreasing, i and j $(j > i+1)$ are bunched if and only if they are bunched with all $k \in \{i+1, \ldots, j-1\}$. Consequently, the problem of determining the optimal bunching pattern reduces to the problem of determining which, if any, of the constraints defining C are binding at the solution to Problem II. Thus, a consumption vector is in C^0, the interior of C, if and only if there is no bunching and all individuals have a positive level of consumption. Henceforth, the first constraint in (15) is referred to as the *nonnegativity constraint* and the remaining constraints in (15) are referred to as the *monotonicity constraints*.

3. Bunching and the Sign and Monotonicity Properties of the Adjusted Wage Rates

If the objective function $\mathscr{W}$ in Problem II is maximized subject only to the constraint that consumption is nonnegative, i.e. the monotonicity constraints on consumption are ignored, the solution is $\hat{c} := (\hat{c}_1, \ldots, \hat{c}_N)$ where

$$\hat{c}_i := \begin{cases} v'^{-1}(\gamma p/\beta_i) & \text{if} \quad \beta_i > 0 \\ 0 & \text{if} \quad \beta_i \leq 0 \end{cases} \quad i = 1, \ldots, N \; . \tag{22}$$

Obviously, if $\hat{c}$ is monotone nondecreasing, $\hat{c}$ is also the solution to Problem II. Because v is strictly concave, the monotonicity properties of $\hat{c}$ are closely related to the monotonicity properties of β. Provided both β_i and β_j are positive, $\hat{c}_i \geq \hat{c}_j$ if and only if $\beta_i \geq \beta_j$. Furthermore, $\hat{c}_i = 0$ if and only if β_i is nonpositive. Thus, there is no bunching and everyone receives positive consumption if the adjusted wage rates are positive and strictly increasing in ability. The main theorem in this section demonstrates that these restrictions on β are also necessary for the optimal consumption vector to be in C^0. In this section it is also shown that a number of other easily-verifiable restrictions on the sign and monotonicity properties of β partially characterize the optimal bunching pattern. These results are used in the next section to provide a complete characterization of the solution to Problem II.

If $\beta_i \leq 0$, then $\partial \mathscr{W}(c)/\partial c_i = \beta_i v'(c_i) - \gamma p < 0$.[3] Consequently, if $\beta_i \leq 0$ it is optimal for person i to bunch with person $i-1$ (or receive zero consumption if $i = 1$).

Proposition 2. *If c^* is a solution to Problem II and if $\beta_i \leq 0$, then $c_i^* = c_{i-1}^*$.*

Turning to the monotonicity properties of β, suppose β_{i-1} and β_i are positive with $\beta_{i-1} \geq \beta_i$. From (22), $\hat{c}_{i-1} \geq \hat{c}_i$ which suggests that it is optimal for these individuals to be bunched together. Intuitively, if $\beta_{i-1} > 0$, $\beta_i > 0$, and $\beta_{i-1} \geq \beta_i$, it is desirable to transfer consumption from i to $i-1$ if $c_i > c_{i-1}$ since $i-1$ receives a higher weight in the reduced-form objective function than i.[4]

---

3 $\mathscr{W}$ is strictly concave if and only if $\beta_i > 0$ for all i. Lollivier and Rochet [6] adopt assumptions on the distribution of skills that ensure that the adjusted wage rates in their reduced-form problem are all positive.

4 The proof of Proposition 3 is omitted as Proposition 3 is a special case of a more general result established in the next section.

Proposition 3. *If c^* is a solution to Problem II and if $\beta_{i-1} \geq \beta_i > 0$ with $i \neq 1$, then $c_i^* = c_{i-1}^*$.*

In a more general model, Guesnerie and Seade [4] establish that the highest-skilled individual should not be bunched with anyone. In the model considered here, their result can be generalized. Proposition 4 presents sufficient conditions for the individuals at the upper end of the skill distribution not to be bunched with anyone and for these individuals to receive positive consumption at a solution to Problem II. These conditions state that the adjusted wage rates of these individuals should be positive and increasing in ability and, in addition, that no one in the rest of the population should have a higher adjusted wage than any of these high-skilled individuals. Furthermore, if i is such an individual, then $\hat{c}_i$ defined in (22) is i's optimal consumption.

Proposition 4. *If c^* is a solution to Problem II, if either $k = N$ or*

$$0 < \beta_k < \ldots < \beta_N \ , \tag{23}$$

and if either $k = 1$ or

$$\beta_k > \beta_i \quad \text{for all} \quad i \in \{1, \ldots, k-1\} \ , \tag{24}$$

then

$$c_{i-1}^* < c_i^* \quad \text{for all} \quad i \in \{k, \ldots, N\} \tag{25}$$

and

$$c_i^* = \hat{c}_i > 0 \quad \text{for all} \quad i \in \{k, \ldots, N\} \ . \tag{26}$$

The hypotheses of Proposition 4 are automatically satisfied if $k = N$ since $w_N = \beta_N > \beta_j$ for all $j \neq N$. Thus Guesnerie and Seade's [4] result that the highest-skilled individual should not be bunched with anyone is a special case of Proposition 4.

Corollary 1. *If c^* is a solution to Problem II, then $c_N^* > c_{N-1}^*$ and $c_N^* = \hat{c}_N$.*

The hypotheses of Proposition 4 are also satisfied if β_i is strictly increasing in ability. While in general Proposition 4 is consistent with quite complicated bunching patterns at the lower end of the skill distribution, if the β_i are strictly increasing, all bunching, if any, should be with the lowest-skilled individual and all individuals who are bunched together should receive zero consumption.

Corollary 2. *If β is strictly increasing, the solution to Problem II is $\hat{c}$ defined in* (22), *so all i with $\beta_i \leq 0$ receive zero consumption and all i with $\beta_i > 0$ are not bunched with anyone.*

Some insight into Proposition 4 may be gained by noting that if (23) and (24) are satisfied, then $0 < \hat{c}_i$ for all $i \in \{k, \ldots, N\}$ and $\hat{c}_k > \hat{c}_i$ for all $i \notin \{k, \ldots, N\}$. Thus $\hat{c}$ satisfies the ith constraint in (15) with a strict inequality for $i \geq k$. The additive separability of $\mathscr{W}$ guarantees that $\{\hat{c}_k, \ldots, \hat{c}_N\}$ are also the optimal values for these variables when the monotonicity constraints are imposed.

Combining Proposition 2, 3, and 4 yields necessary and sufficient conditions (a) for there are to be no bunching at an optimum and (b) for none of the constraints in (15) to bind, i.e. for the optimal solution to be in C^0.

Theorem 2. *If c^* is a solution to Problem II, then (a) c^* exhibits no bunching if and only if*

$$\beta_1 < \ldots < \beta_N \quad \textit{with} \quad \beta_2 > 0 \tag{27}$$

and (b) $c^* \in C^0$ if and only if

$$0 < \beta_1 < \ldots < \beta_N \ . \tag{28}$$

If (27) or (28) is satisfied, then $c^* = \hat{c}$ where $\hat{c}$ is defined in (22).

It is trivial to determine whether the hypotheses of the propositions presented in this section are satisfied given the vector of adjusted wage rates β, and calculating β only requires simple arithmetic operations. In particular, if it is optimal to have no bunching, this fact will be immediately apparent from the vector β. Furthermore, when there is no bunching, it is optimal to set c_i equal to $\hat{c}_i$ and it is a straightforward exercise to determine $\hat{c}_i$. For example, if $v(c_i) = 2c_i^{1/2}$ and (28) is satisfied, person i's optimal consumption is $\gamma p/\beta_i$. Once the optimal consumption vector has been determined, again only simple arithmetic operations are required to determine the optimal income vector; the relevant formulae are given in (19) and (20).[5]

4. Characterizing the Optimal Consumption Vector

In this section a method is presented for determining the optimal consumption vector and, thus, for completely characterizing the optimal pattern of bunching. Propositions 2 and 3 provide sufficient conditions for it to be optimal for a pair of individuals to be bunched together. Before stating the characterization theorem, it is useful to generalize these two propositions by providing sufficient conditions for it to be optimal for groups of individuals to be bunched together.

Proposition 5. *If c^* is a solution to Problem II and if there exist $j, k \in \{1, \ldots, N\}$ with $j \leq k$ such that*

$$\sum_{i=h}^{k} \beta_i \leq 0 \quad \textit{for all} \quad h \in \{j, \ldots, k\} \ , \tag{29}$$

then

$$c_i^* = c_{j-1}^* \quad \textit{for all} \quad i \in \{j, \ldots, k\} \ . \tag{30}$$

Proposition 5 generalizes Proposition 2. Starting with person k and working down the skill distribution to person j, if the cumulative sums of the adjusted wage

[5] In the absence of bunching, (19), (20), and (22) provide explicit formulae for the optimal allocation. In Weymark [14] these expressions are used to develop some comparative static properties of the optimal allocation with respect to variations in the exogenous parameters γ, p, and λ.

rates are all nonpositive, all of these individuals should be bunched together with person $j-1$ (or receive zero consumption if $j=1$). Thus, for example, if $\beta_{i+1} \leq 0$ and $\beta_i + \beta_{i+1} \leq 0$, it is optimal for $i-1$, i, and $i+1$ to bunch together even if $\beta_i > 0$; the negative value for β_{i+1} outweighs the moderating influence associated with a positive value for β_i.

Proposition 3 states that it is optimal for i to bunch with $i-1$ if $\beta_{i-1} \geq \beta_i > 0$. If this condition is satisfied, it is also true that β_{i-1} is at least as large as the average adjusted wage of these two individuals, $(\beta_{i-1}+\beta_i)/2$. This observation leads to the following generalization of Proposition 3.

Proposition 6. *If c^* is a solution to Problem II and if there exist $j, k \in \{1, \ldots, N\}$ with $j<k$ such that*

$$\frac{\sum_{i=j}^{h} \beta_i}{(h+1-j)} \geq \frac{\sum_{i=j}^{k} \beta_i}{(k+1-j)} > 0 \quad \textit{for all} \quad h \in \{j, \ldots, k\} \ , \tag{31}$$

then

$$c_i^* = c_j^* \quad \textit{for all} \quad i \in \{j, \ldots, k\} \ . \tag{32}$$

Starting with person j and working up the skill distribution to person k, if the cumulative averages of the adjusted wage rates are all positive and at least as large as the average adjusted wage of the whole group, then all of these individuals should be bunched together.

For a vector $\xi = (\xi_1, \ldots, \xi_N)$, define the function $f_\xi : [0, N] \rightarrow \mathbf{R}$ by setting

$$f_\xi(r) := \xi_{i-1} + (r+1-i)(\xi_i - \xi_{i-1}) \quad \text{for all} \quad r \in [i-1, i] \ , \quad i = 1, \ldots, N \ , \tag{33}$$

where, as a convention ξ_0 is set equal to zero. With a slight abuse of terminology, f_ξ is called the graph of ξ. Informally, f_ξ is obtained by plotting (0,0) and (i, ξ_i) for $i=1, \ldots, N$ and connecting consecutive points by straight-line segments.

Let $\tau := (\tau_1, \ldots, \tau_N)$ where

$$\tau_i := \sum_{h=1}^{i} \beta_i \quad i = 1, \ldots, N \ ; \tag{34}$$

τ_i is the cumulated total adjusted wage of the first i individuals. The graph of τ, f_τ, is used to determine the solution to the reduced-form income tax problem. Let $\bar{f}_\tau : [0, N] \rightarrow \mathbf{R}$ be the convex hull of f_τ,

$$\bar{f}_\tau(r) := \text{Min} \ \{\zeta f_\tau(r_1) + (1-\zeta) f_\tau(r_2) | \zeta \in [0,1], \ r_1, r_2 \in [0, N], \ \text{and} \quad \zeta r_1 + (1-\zeta) r_2 = r\} \ . \tag{35}$$

The function $\bar{f}_\tau$ is the highest convex function which lies nowhere above f_τ. It is clear that

$$\beta_i = f_\tau(i) - f_\tau(i-1) \qquad i = 1, \ldots, N \ . \tag{36}$$

By analogy, $\bar{f}_\tau$ can be used to define a new vector $\bar{\beta}$ by setting

$$\bar{\beta}_i := \bar{f}_\tau(i) - \bar{f}_\tau(i-1) \qquad i=1,\ldots,N \ . \tag{37}$$

Since $\bar{f}_\tau$ is a convex function, $\bar{\beta}$ is nondecreasing. Letting $\bar{\tau}$ be the vector of cumulated adjusted wages corresponding to $\bar{\beta}$, it is easy to verify that $\bar{f}_\tau = \bar{f}_{\bar{\tau}}$ and $f_{\bar{\tau}} = \bar{f}_{\bar{\tau}}$.

Theorem 3.[6] *There exists a unique solution c^* to Problem II given by*

$$c_i^* := \begin{cases} v'^{-1}(\gamma p/\bar{\beta}_i) & \text{if } \bar{\beta}_i > 0 \\ 0 & \text{if } \bar{\beta}_i \leq 0 \end{cases} \qquad i=1,\ldots,N \tag{38}$$

where $\bar{\beta}$ is defined in (37).

An important implication of Theorem 3 is that the optimal bunching pattern depends only on the vector β, which in turn depends on the vector of welfare weights λ and the vector of ability levels w. Given the maintained hypotheses that preferences are quasilinear and the technology exhibits constant returns to scale, the values of the parameters γ and p and the functional form of v influence the optimal allocation but not the optimal bunching pattern.

While Theorem 3 presents an explicit solution to the reduced-form optimal nonlinear income tax problem, the computation of this solution requires the computation of the vector $\bar{\beta}$. While, in general, there is no simple formula to determine $\bar{\beta}$ from β, it is not difficult to devise algorithms which can compute $\bar{\beta}$. Consequently, it is not unrealistic to calculate the solution to an optimal nonlinear tax problem of the sort considered here, provided N is not so large that running the algorithm to determine $\bar{\beta}$ becomes prohibitively expensive.

If β is nondecreasing, then f_τ is a convex function and $\beta = \bar{\beta}$. This observation and Theorem 3 imply that the optimal solution to the reduced-form income tax problem is the consumption vector $\hat{c}$ defined in (22) if β is nondecreasing. In Corollary 2, β is assumed to be strictly increasing, so the special bunching properties identified there can be interpreted as resulting from the convexity of the corresponding function f_τ.

To illustrate how Theorem 3 can be used to study bunching, it is useful to consider an example. Let $\beta = (2, -6, 3, -1, 6, -1, 1, 7)$. For this β, $\tau = (2, -4, -1, -2, 4, 3, 4, 11)$ and $\bar{\beta} = (-2, -2, 1, 1, 2, 2, 2, 7)$. The graph of τ and its convex hull are shown in Fig. 1. Thus it is optimal for (a) individuals 1 and 2 to be bunched together and receive zero consumption, (b) individuals 3 and 4 to be bunched together, (c) individuals 5, 6, and 7 to be bunched together, and (d) individual 8 to be separated from everybody else.

The average adjusted wage of all individuals in $\{h,\ldots,i\}$ where $h < i$ is given by the slope of a line joining the point $(h-1, \tau_{h-1})$ to (i, τ_i). Consequently, if $\bar{\beta}_i$ is positive, $\bar{\beta}_i$ is the average adjusted wage of all individuals who should be bunched with person i. Furthermore, $\bar{\beta}_i$ can be thought of as a social shadow wage rate. If person i chooses consumption and labour supply to maximize the utility function (3) subject to the budget constraint $pc_i \leq \bar{\beta}_i l_i$, the optimal consumption is c_i^* given by (38).

---

6 The inspiration for Theorem 3 comes from Myerson [10] and Baron and Myerson [1].

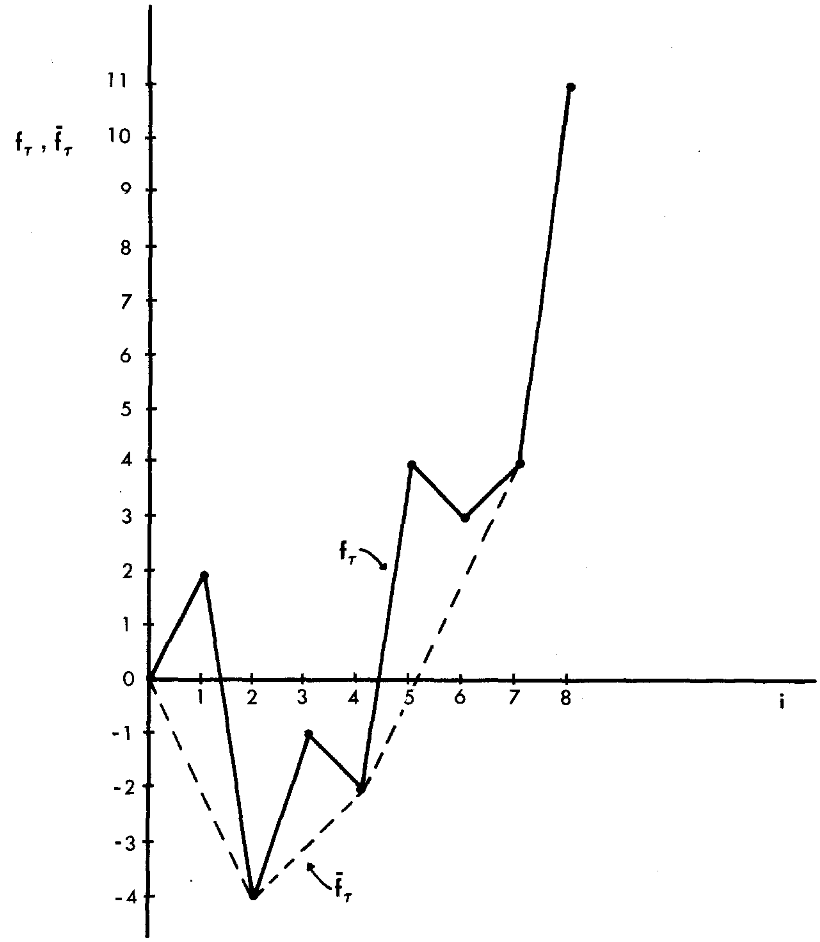

Fig. 1

5. Convexity of the Adjusted Wage Distribution

All of the results presented here employ three maintained assumptions: (i) all individuals have common quasilinear preferences, (ii) the skill-normalized welfare weights are declining in ability, and (iii) there are an equal number of people with each skill level, where for convenience this number has been set equal to one. Working with a continuum of skills, Lollivier and Rochet [6] demonstrate that with these assumptions all bunching, if any, should be with the lowest-skilled individual. The results presented in the previous sections show that this conclusion is not valid with a finite population.

Corollary 2 presents one sufficient condition, namely that β be strictly increasing, for all bunching, if any, to be with the lowest-skilled individual. In this section it is shown that this pattern of bunching is also optimal if the graph of β, i.e. f_β, is a convex function on $[1, N]$. Furthermore, if the gaps between consecutive wage levels are evenly-spaced, this convexity condition is satisfied. This spacing property is automatically satisfied by a continuum of skill levels, which accounts for Lollivier and Rochet's findings.

It is natural to call an N-vector ξ *convex* if and only if the corresponding graph f_ξ is a convex function on $[1, N]$. Convexity of ξ is equivalent to requiring the vector $(\xi_2 - \xi_1, \ldots, \xi_N - \xi_{N-1})$ to be nondecreasing. If $\xi_N > \xi_j$ for all $j \neq N$, as is the case for β, and if ξ is convex, then there exists an integer $i_\xi \in \{2, \ldots, N-1\}$ such that

$$\begin{aligned} &\xi_i \leq \xi_{i-1} \quad \text{for all} \quad i < i_\xi \ , \quad i \geq 2 \\ &\text{and} \\ &\xi_i > \xi_{i-1} \quad \text{for all} \quad i > i_\xi \ . \end{aligned} \tag{39}$$

If ξ is strictly increasing, ξ is convex with $i_\xi = 2$.

Define the vector $\alpha := (\alpha_1, \ldots, \alpha_N)$ by setting

$$\begin{aligned} \alpha_i &:= \tau_i / i \\ &\equiv \sum_{h=1}^{i} \beta_h / i \end{aligned} \qquad i = 1, \ldots, N \ . \tag{40}$$

Thus α_i is the average adjusted wage of all individuals with wage rates less than or equal to w_i. The fact that $\beta_N > \beta_j$ for all $j \neq N$ implies that $\alpha_N > \alpha_j$ for all $j \neq N$ as well.

Lemma 1. *If β is convex, α is also convex with $i_\alpha \geq i_\beta$.*

Lemma 1 is analogous to a result in the theory of the firm; if a firm's marginal cost curve is convex, then so is its average cost curve. If, furthermore, the marginal cost curve is ∪-shaped, so is the average cost curve and the output for which minimum average cost is attained exceeds the output for which marginal cost is minimized. If marginal cost is always increasing, so is average cost, which corresponds to having $i_\alpha = i_\beta = 2$.

Let i_0 be the lowest-skilled individual such that the weights $(\beta_{i_0}, \ldots, \beta_N)$ are positive and strictly increasing and such that β_{i_0} is larger than the average adjusted wage of all lower-skilled individuals. Formally, i_0 is the solution to:

$$\begin{aligned} &\text{Min } i \\ &\text{subject to} \quad \text{(i)} \quad i \in \{2, \ldots, N\} \\ &\qquad\qquad\quad \text{(ii)} \quad 0 < \beta_i < \beta_{i+1} < \ldots < \beta_N \ , \\ &\qquad\qquad\quad \text{(iii)} \quad \beta_i > \alpha_{i-1} \ . \end{aligned} \tag{41}$$

Since $\beta_N > 0$ and $\beta_N > \beta_j$ for all $j \neq N$, the constraints in (41) are satisfied with $i = N$, so i_0 is well-defined. If β is convex, it is easy to verify that of all $i \geq i_\alpha$, i_0 is the lowest-skilled individual with $\beta_i > 0$.

Proposition 7 demonstrates that if β is convex, any bunching which is optimal must involve the lowest-skilled individual. More precisely, if β is convex, (i) it is optimal to have all $i < i_0$ bunched with the lowest-skilled individual and (ii) it is optimal for all $i \geq i_0$ not to be bunched with anyone. Thus, if $i_0 = 2$, bunching is not desirable. In contrast to Corollary 2, individuals bunched together may receive a positive amount of the consumption good.

Proposition 7. *If c^* is the solution to Problem II and if β is convex, then*

$$c_i^* = \hat{c}_i > 0 \quad \textit{for all} \quad i \geq i_0 \tag{42}$$

and

$$c^* = \hat{s} \quad \textit{for all} \quad i < i_0 \tag{43}$$

where

$$\hat{s} := \begin{cases} v'^{-1}(\gamma p/\alpha_{i_0-1}) & \textit{if} \quad \alpha_{i_0-1} > 0 \\ 0 & \textit{if} \quad \alpha_{i_0-1} \leq 0 \end{cases} \tag{44}$$

and where $\hat{c}_i$ is defined in (22) *and i_0 is defined in* (41).

From a computational perspective, two features of Proposition 7 are worth noting. First, it is very simple to determine if β is convex. Second, if β is convex, the integer i_0, i.e. the lowest-skilled individual not bunched with person one, is easy to identify.

If β is convex, $\hat{c}_i$ is the optimal consumption for each $i \geq i_0$, i.e. these individuals should receive the consumption that is optimal for them at an unconstrained maximum of $\mathscr{W}$. A striking feature of this result is that this conclusion is compatible with having $\hat{c}_j > \hat{c}_{i_0}$ for some $j < i_0$. At first glance, this last inequality would suggest that it is optimal for i_0 to bunch with j. However, the convexity properties of β and the definition of i_0 guarantee that the β_i are nonincreasing for $i \in \{j, \ldots, i_0 - 1\}$, pushing the average of the β_i's for this group sufficiently below β_j that c_j^* is lower than $\hat{c}_{i_0}$.

The definition of the adjusted wage β_i contains a term involving the difference in wages received by consecutive individuals in the skill distribution. The skill distribution is defined to be *evenly-spaced* if and only if there exists a Δ such that $w_i - w_{i-1} = \Delta$ for all $i \in \{2, \ldots, N\}$. Evenly-spaced wage distributions result in β being convex.

Lemma 2. *If the skill distribution is evenly-spaced, β is convex.*

An immediate implication of Lemma 2 and Proposition 7 is Theorem 4.

Theorem 4. *If the skill distribution is evenly-spaced, at the solution to Problem II there is a wage $\bar{w}$ such that it is optimal for all individuals receiving a wage less than $\bar{w}$ to be bunched together and it is optimal for all individuals receiving a wage greater than $\bar{w}$ not to be bunched with anyone.*

Theorem 4 accounts for the special bunching pattern observed by Lollivier and Rochet [6]. With a continuum of abilities, skills are automatically evenly-spaced. Thus, with a uniform distribution, the bunching pattern described in Theorem 4 must result. Consequently, it appears that much of the complexity of the finite case (with a uniform distribution) is due to the fact that skills need not be evenly-spaced.

6. Nonuniform Skill Distributions

A further complication is introduced if the assumption that each person has a different skill level is dropped. Assuming that consumers with the same skill level have the same welfare weight, in Weymark [13] it is shown that it is optimal for all consumers with a given skill level to receive the same commodity bundle. Consequently, the analysis can proceed by either indexing each individual in society or by indexing the skill levels.

If the former approach is adopted and individuals are indexed in increasing order of ability with ties broken arbitrarily, Problem II remains the appropriate reduced-form income tax problem and the characterization of its solution presented in Theorem 3 continues to apply. With the possibility of multiple consumers with the same skill level, the results presented in the previous sections can be interpreted to be theorems concerning the desirability of having distinct individuals receive the same commodity bundle rather than results directly concerned with bunching, i.e. having individuals with distinct *characteristics* receive the same commodity bundle. With this interpretation, however, a number of the theorems are vacuous (in particular, the propositions in Sect. 5), because the vector β cannot be strictly increasing nor can it be convex (unless the only skill level with multiple consumers is the lowest).

To study bunching, it is more natural to index skill levels rather than individuals. The corresponding reduced-form problem may be found in Weymark [13]. The results in Sect. 3 and Propositions 5 and 6 in Sect. 4 have straightforward analogues in this formulation of the problem. However, to compute the optimal solution, it is still necessary to compute the convex hull of the graph of β with each individual indexed separately, so, for this purpose, working with the skill levels only complicates the analysis. The results in Sect. 5, when reinterpreted in terms of skill levels, do not generalize to nonuniform skill distributions.

7. Concluding Remarks

A few general conclusions concerning the bunching properties of the solution to a finite population optimal nonlinear income tax problem satisfying the maintained assumptions adopted here emerge from the previous sections. First, in contrast to the continuum result obtained by Lollivier and Rochet [6], all bunching need not be at the bottom of the skill distribution. Second, the continuum bunching pattern is obtained with a finite population if skills are evenly-spaced, a property that is necessarily satisfied by a uniform distribution in the continuum, or, more generally, if the vector of adjusted wage rates is convex. Third, the optimal bunching pattern depends only on the vector of adjusted wage rates β (and, hence, on the welfare weights λ and the skill levels w), and not on the other parameters of the model. In particular, many features of the optimal bunching pattern are directly related to the sign and monotonicity properties of β. Fourth, it is possible to present an explicit, albeit indirect, characterization of the solution to the optimal income tax problem. Taken together, these results suggest that if individuals' preferences are not too dissimilar and if income effects are not too strong, much can be determined about the bunching properties of the allocations that result when nonlinear income taxes are set optimally.

Appendix

Proof of Proposition 4. Recall that $\beta_N > 0$ and $\beta_N > \beta_i$ for all $i \neq N$. First, it is shown that $c_i^* \leq \hat{c}_N$ for all i. On the contrary, suppose there exists an i such that $c_i^* > \hat{c}_N$. Let j be the smallest such i. By construction, $c_j^* > c_{j-1}^*$, so it is feasible to marginally decrease c_j. But

$$\partial \mathscr{W}(c^*)/\partial c_j = \beta_j v'(c_j^*) - \gamma p$$

$$\leq \beta_N v'(c_j^*) - \gamma p$$

(since $\beta_j < w_j < w_N = \beta_N$ if $j \neq N$)

$$< \beta_N v'(\hat{c}_N) - \gamma p$$

(since $c_i^* > \hat{c}_N$)

$$= 0$$

by (22). Thus, a marginal decrease in c_j increases $\mathscr{W}$, a contradiction.

If $c_N^* < \hat{c}_N$, then

$$\partial \mathscr{W}(c^*)/\partial c_N = \beta_N v'(c_N^*) - \gamma p > 0$$

from (22) and the assumption that $c_N^* < \hat{c}_N$. Again, this conclusion contradicts the optimality of c^*, so $c_N^* = \hat{c}_N$.

To show that person N is not bunched with anyone, suppose the contrary. Let j be the lowest-skilled individual bunched with person N. By construction, $c_j^* > c_{j-1}^*$ and it is feasible to marginally decrease c_j. But

$$\partial \mathscr{W}(c^*)/\partial c_j = \beta_j v'(c_j^*) - \gamma p < 0$$

since $\beta_j < \beta_N$, $c_j^* = c_N^* = \hat{c}_N$, and $\beta_N v'(\hat{c}_N) - \gamma p = 0$. Hence, marginally decreasing c_j increases $\mathscr{W}$, a contradiction. This argument establishes Proposition 3 if $k = N$.

If $k < N$, the proof proceeds by induction. Suppose $h \in \{k, \ldots, N-1\}$ and (25) and (26) hold for all $i > h$. A straightforward adaptation of the above argument establishes first that $c_i^* \leq c_h^*$ for all $i \leq h$, second that $c_h^* = \hat{c}_h$, and third that $c_h^* > c_i^*$ for all $i < h$.

Proof of Proposition 5. Suppose $h \in \{j, \ldots, k\}$ and $c_i^* = c_h^*$ for all $i \in \{h, \ldots, k\}$. Contrary to the theorem, suppose $c_h^* > c_{h-1}^*$. It is feasible to change the consumption of all $i \in \{h, \ldots, k\}$ by a common infinitesimal amount $\delta < 0$. The resulting change in the value of $\mathscr{W}$ is

$$\delta \left\{ \sum_{i=h}^{k} \beta_i v'(c_i^*) - \gamma p(k+1-h) \right\} = \delta \left\{ v'(c_h^*) \sum_{i=h}^{k} \beta_i - \gamma p(k+1-h) \right\}$$

which is positive since $\delta < 0$, $k \geq h$, and (29) is satisfied, contradicting the optimality of c^*.

Proof of Proposition 6. From (31), for all $h \in \{j+1,\dots,k\}$

$$(k+1-j)\sum_{i=j}^{h-1}\beta_i \geq (h-j)\sum_{i=j}^{k}\beta_i \ ,$$

so

$$(k+1-h)\sum_{i=j}^{h-1}\beta_i-(h-j)\sum_{i=h}^{k}\beta_i \geq 0 \ . \tag{A.1}$$

Suppose $c_i^*=c_h^*$ for all $i \in \{h,\dots,k\}$ where $h \in \{j+1,\dots,k\}$ but $c_h^* > c_{h-1}^*$. If $\beta_h \leq 0$, by Proposition 2, there is a contradiction, so suppose $\beta_h > 0$. It is feasible to marginally decrease c_i by $\delta/(k+1-h)$ for all $i \in \{h,\dots,k\}$ and to marginally increase c_i by $\delta/(h-j)$ for all $i \in \{j,\dots,h-1\}$ where $\delta > 0$. Such a change keeps total consumption constant. The resulting change in $\mathscr{W}$ is

$$\frac{\delta}{(h-j)}\sum_{i=j}^{h-1}\beta_i v'(c_i^*)-\frac{\delta}{(k+1-h)}\sum_{i=h}^{k}\beta_i v'(c_i^*)$$
$$\geq \frac{\delta v'(c_{h-1}^*)}{(h-j)}\sum_{i=j}^{h-1}\beta_i-\frac{\delta v'(c_h^*)}{(k+1-h)}\sum_{i=h}^{k}\beta_i \ . \tag{A.2}$$

To verify this inequality, first note that the second terms on each side of the inequality are identical because of the assumption that $c_i^*=c_h^*$ for all $i \in \{h,\dots,k\}$. If $\beta_i>0$ for all $i \in \{j,\dots,h-1\}$, the inequality then follows from the concavity of v and the monotonicity of c^*. From (31) it follows that $\beta_j>0$ and $\sum_{i=j}^{h-1}\beta_i>0$, so Proposition 2 can be used to establish this inequality if $\beta_i \leq 0$ for some $i \in \{j+1,\dots,h-1\}$.

The first term on the right-hand-side of (A.2) is strictly positive. If $\sum_{i=h}^{k}\beta_i \leq 0$, the last term is nonpositive, which makes the overall change in $\mathscr{W}$ positive, a contradiction. If $\sum_{i=h}^{k}\beta_i > 0$, since v is concave and $c_h^* > c_{h-1}^*$ the right-hand-side of (A.2) is strictly larger than

$$\delta v'(c_{h-1}^*)\left[\frac{\sum_{i=j}^{h-1}\beta_i}{(h-j)}-\frac{\sum_{i=h}^{k}\beta_i}{(k+1-h)}\right] ,$$

which is positive by (31), a contradiction. Thus, if $c_i^*=c_h^*$ for all $i \in \{h,\dots,k\}$ where $h \in \{j+1,\dots,k\}$, then $c_h^*=c_{h-1}^*$. Since the antecedent in this statement is trivially valid if $h=k$, the proposition is established.

Proof of Theorem 3. By Corollary 1, any solution to Problem II must be contained in $\bar{C} := \{c \in C \mid c_N=\hat{c}_N\}$, which is a compact set. As $\mathscr{W}$ is continuous, the Weierstrass Theorem implies that a solution to Problem II exists.

Since $\bar{f}_\tau$ is a convex function, by construction $\bar{\beta}$ is nondecreasing and, thus, the c^* given by (38) satisfies the nonnegativity and monotonicity constraints on c. Since $\bar{\beta}$ is nondecreasing, the set of individuals sharing a common value for $\bar{\beta}_i$ are indexed by consecutive integers. Suppose $\{j,\ldots,k\}$ is such a set of individuals, i.e. $\bar{\beta}_i=\bar{\beta}_{i'}$, for all $i,i'\in\{j,\ldots,k\}$ and $\bar{\beta}_i\neq\bar{\beta}_{i'}$ if $i\in\{j,\ldots,k\}$ and $i'\notin\{j,\ldots,k\}$. Since $\bar{f}_\tau$ is the highest convex function which lies nowhere above f_τ, by construction

$$\frac{\sum_{i=j}^{h}\beta_i}{(h+1-j)}\geq\frac{\sum_{i=j}^{k}\beta_i}{(k+1-j)}=\bar{\beta}_h \quad \text{for all} \quad h\in\{j,\ldots,k\}\ . \tag{A.3}$$

If $\bar{\beta}_h>0$, by Proposition 6 all $h\in\{j,\ldots,k\}$ should be bunched together. If $\bar{\beta}_h\leq 0$, then (29) holds and the same conclusion follows from Proposition 5. Hence, all individuals with the same value for $\bar{\beta}_i$ should be bunched together.

Let $\{\beta_1^0,\ldots,\beta_T^0\}$ be the T distinct values for $\bar{\beta}_i$, $i=1,\ldots,N$ written in increasing order and let n_t be the number of individuals with $\bar{\beta}_i=\beta_t^0$. Constraining all individuals who have the same value for $\bar{\beta}_i$ to receive the same consumption and using the equality in (A.3), the reduced-form objective function (18) may be written as

$$\sum_{t=1}^{T} n_t\beta_t^0 v(c_t^0)-\gamma p\sum_{t=1}^{T} n_t c_t^0 \tag{A.4}$$

where c_t^0 is the common consumption level of the individuals with the t^{th} largest distinct value for $\bar{\beta}_i$. The solution (expressed in terms of the original labelling of individuals) to the maximization of (A.4) subject only to the constraint that each c_t^0 be nonnegative is unique and is given by (38). As noted above, this consumption vector is in C, which completes the proof.

Proof of Lemma 1. Convexity of β is equivalent to

$$\beta_{i+1}+\beta_{i-1}-2\beta_i\geq 0 \qquad i=2,\ldots,N\ . \tag{A.5}$$

Analogously, convexity of α is equivalent to

$$\alpha_{i+1}+\alpha_{i-1}-2\alpha_i\geq 0 \qquad i=2,\ldots,N\ . \tag{A.6}$$

Let

$$X_i:=2\sum_{h=1}^{i-1}\beta_h+(2-i-i^2)\beta_i+(i^2-i)\beta_{i+1} \qquad i=2,\ldots,N\ .$$

Note that

$$\begin{aligned}X_i&=i(i-1)\sum_{h=1}^{i+1}\beta_h+(i+1)i\sum_{h=1}^{i-1}\beta_h-2(i+1)(i-1)\sum_{h=1}^{i}\beta_h\\&=(i+1)(i-1)i(\alpha_{i+1}+\alpha_{i-1}-2\alpha_i)\ .\end{aligned}$$

Thus, (A.6) holds if and only if $X_i\geq 0$ for all $i=2,\ldots,N$. Induction is used to establish this result. Solving for X_{j+1} in terms of X_j,

$$\begin{aligned}X_{j+1}&=2\sum_{h=1}^{j}\beta_h-(j^2+3j)\beta_{j+1}+(j^2+j)\beta_{j+2}\\&=X_j+(j^2+j)(\beta_j+\beta_{j+2}-2\beta_{j+1})\ .\end{aligned}$$

Since (A.5) holds, $X_{j+1} \geq 0$ if $X_j \geq 0$. As

$$X_2 = 2(\beta_3 + \beta_1 - 2\beta_2) \ ,$$

which is nonnegative by (A.5), α is convex. That $i_\alpha \geq i_\beta$ is a straightforward consequence of the relationship between average and marginal values.

Proof of Proposition 7. First, it is shown that $c_i^* = c_1^*$ for all $i < i_0$. This conclusion is trivial if $i_0 = 2$, so consider the case where $i_0 > 2$. Since β, and hence α, is convex, for all $i \in \{2, \ldots, i_{\alpha-1}\}$, $\alpha_i \leq \alpha_{i-1}$. Thus, (A.3) is satisfied for $j = 1$ and $k = i_{\alpha-1}$. The argument following (A.3) implies that all $i \in \{1, \ldots, i_{\alpha-1}\}$ should be bunched together. If $i_0 \neq i_\alpha$, for all $i \in \{i_\alpha, \ldots, i_0 - 1\}$, $\beta_i \leq 0$. By Proposition 2, all of these individuals should bunch with $i_{\alpha-1}$. Hence $c_i^* = c_1^*$ for all $i < i_0$.

Constraining the reduced-form problem to have $c_1 = \ldots = c_{i_0-1}$, it can be rewritten as:

Choose $(s, c_{i_0}, \ldots, c_N)$ to maximize

$$(i_0 - 1)\alpha_{i_0-1} v(s) + \sum_{h=i_0}^{N} \beta_h v(c_h) - \gamma p \left[(i_0 - 1)s + \sum_{h=i_0}^{N} c_h \right] \tag{A.7}$$

subject to

$$0 \leq s \leq c_{i_0} \leq \ldots \leq c_N \ , \tag{A.8}$$

where s denotes the common value of $\{c_1, \ldots, c_{i_0-1}\}$. Ignoring the monotonicity constraints, the solution is $(\hat{s}, \hat{c}_{i_0}, \ldots, \hat{c}_N)$. This vector also satisfies (A.8) because, by the definition of i_0, $\alpha_{i_0-1} < \beta_{i_0} < \ldots < \beta_N$. Furthermore, $\hat{c}_i > 0$ for all $i > i_0$ since $\beta_{i_0} > 0$.

Proof of Lemma 2. Substituting $\Delta > 0$ for $w_{i+1} - w_i$ in (17) implies

$$(\beta_{i+1} - \beta_i) - (\beta_i - \beta_{i-1}) = \Delta(\lambda_i - \lambda_{i+1}) \qquad i = 2, \ldots, N \ . \tag{A.9}$$

As λ_i is strictly decreasing in i, the expression in (A.9) is positive, so β is convex.

References

1. Baron DP, Myerson RB (1982) Regulating a monopolist with unknown costs. Econometrica 50:911–930
2. Cooper R (1984) On allocative distortions in problems of self-selection. Rand J Econ 15:568–577
3. Guesnerie R, Laffont J-J (1984) A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm. J Public Econ 25:329–369
4. Guesnerie R, Seade J (1982) Nonlinear pricing in a finite economy. J Public Econ 17:157–179
5. Lockwood B (1985) Non-linear pricing and the exclusion of consumers. Econ Lett 18:313–316
6. Lollivier S, Rochet J-C (1983) Bunching and second-order conditions: A note on optimal tax theory. J Econ Theory 31:392–400
7. Maskin E, Riley J (1984) Monopoly with incomplete information. Rand J Econ 15:171–196
8. Mirrlees JA (1971) An exploration in the theory of optimum income taxation. Rev Econ Studies 38:175–208
9. Mussa M, Rosen S (1978) Monopoly and product quality. J Econ Theory 18:301–317

10. Myerson RB (1981) Optimal auction design. Math Oper Res 6:58–73
11. Seade JK (1977) On the shape of optimal tax schedules. J Public Econ 7:203–235
12. Seade J (1980) Optimal non-linear policies for non-utilitarian motives. In: Collard D, Lecomber R, Slater M (eds) Income distribution: The limits to redistribution. Scientechnica, Bristol, pp 53–68
13. Weymark JA (1986) A reduced-form optimal nonlinear income tax problem. J Public Econ (in press)
14. Weymark JA (1985) Comparative static properties of optimal nonlinear income taxes. Discussion Paper No. 85–02, Department of Economics, University of British Columbia (revised 1986)