

# The Structure of the Optimal Income Tax in the Quasi-Linear Model

Nigar Hashimzade\*

University of Exeter

Gareth D. Myles†

University of Exeter and Institute of Fiscal Studies

October 1, 2004

## Abstract

Existing numerical characterizations of the optimal income tax have been based on a limited number of model specifications. As a result, they do not reveal which properties are general. We determine the optimal tax in the quasi-linear model under weaker assumptions than have previously been employed; in particular, we remove the assumption of a lower bound on the utility of zero consumption and the need to permit negative labor incomes. A Monte Carlo analysis is then conducted in which economies are selected at random and the optimal tax function constructed. The results show that in a significant proportion of economies the marginal tax rate rises at low skills and falls at high. The average tax rate is equally likely to rise or fall with skill at low skill levels, rises in the majority of cases in the centre of the skill range, and falls at high skills. These results are consistent across all the specifications we test.

---

\*Nigar Hashimzade, Department of Economics, University of Exeter, Exeter EX4 4RJ, UK, n.hashimzade@ex.ac.uk.

†Gareth D. Myles, Department of Economics, University of Exeter, Exeter EX4 4RJ, UK, g.d.myles@ex.ac.uk.

# 1 Introduction

An analysis of the optimal nonlinear income tax was first undertaken in the seminal paper of (Mirrlees 1971). That paper, and the many which followed, have determined the properties an optimal tax must possess (a survey of these results can be found in (Myles 1995)). These theoretical results provide only limited information. An important practical issue upon which they are silent is that of the progressiveness of the tax system. All developed countries utilize tax systems with increasing marginal and average tax rates. The theory has so far not fully resolved whether this is optimal. It is well-known that the marginal tax rate should be zero for the highest skill consumer which implies the tax function cannot have an everywhere-increasing marginal rate. But the results are silent on the behavior for the rest of the skill distribution and on the progressiveness of the average tax rate.

The limitations of the theoretical analysis have lead to the use of numerical simulations to investigate the optimal tax function. The most significant results have been obtained by combining a log-normal distribution of skill with either Cobb-Douglas ((Mirrlees 1971)) or CES utility ((Kanbur and Tuomala 1994)). These specifications have produced an optimal tax function for which the marginal tax rate either first rises with skill and then falls or is highest for the lowest skill and then falls. Which of these applies is dependent upon the degree of equity in social welfare and the standard deviation of the skill distribution (see (Kanbur and Tuomala 1994)). These qualitative properties have remained consistent in all the simulation results that have been reported. This consistency would suggest that the optimal income tax should always have these properties were it not for the very narrow range of model specifications that have been used to generate the results.

To make progress around the difficulty of limited specifications, this paper pursues a very different approach to the characterization of the optimal tax function. Our analysis begins with the observation that the structure of the optimal tax function is critically dependent upon the distribution of skills. This point was made clear by (Diamond 1998) who showed that tax rates may increase above the modal income for some skill distributions and (Myles 2000) where it was shown that a distribution of skill could be constructed which would support any chosen qualitative pattern of marginal tax rates. Consequently there can be no concept of a general structure for the optimal income tax function. Having said this, it may be the case that some structures can occur only for a small subset of possible skill distributions and that “most” economies do share a common structure. To obtain an insight into whether this claim is valid, our approach is to define an economy by its distribution of skills. We then make repeated random draws from the set of possible economies, compute the optimal tax function for each draw, and then study the “general” structure of the tax function. This procedure does not determine a typical tax function. Instead, what it does achieve is an evaluation of the proportion of economies for which the average and marginal tax rates will behave in a chosen manner. For example, our results can be used to predict the proportion of economies for which

the marginal rate of tax decreases as skill increases (it is very small). More importantly, looking at the pattern of results as a whole provides an insight into the most likely structure of optimal tax rates. We are not aware of any previous use of such Monte Carlo methods in tax theory.

The results are obtained by generalizing the quasi-linear model of Weymark (1986a, 1986b, 1987) in several ways. First, we allow for the possibility that there is more than one consumer with each skill level. This is very straightforward to implement. The second generalization is more significant and involves the removal of the assumption that the utility of a zero consumption level has a finite lower bound (the lower bound is set at zero in Weymark (1986a,b, 1987)). This assumption rules out many commonly-used utility functions (such as logarithmic utility) and has an undue influence upon the nature of optimal allocations. In addition, we also extend the model by removing the assumption that negative labor incomes are permissible. We do not find the possibility of negative labor incomes, as permitted in several previous contributions, to be appealing in the context of a model of labor supply. As our results show, these changes significantly affect the solution process for the model and the nature of optimal allocations. It should be stressed that although the results are developed here in the context of optimal income taxation, they are also applicable to nonlinear incentive schemes in general. The same generalizations can be applied to other applications of the quasi-linear model.

## 2 Quasi-Linearity and Tax Rates

This section briefly introduces the model of income taxation with quasi-linear utility. (Lollivier and Rochet 1983) applied this to a model with a continuum of consumers. The model with a finite numbers of consumers on which this paper is based is analyzed in detail in Weymark (1986a,b, 1987).

An *economy* is described by a vector  $E \equiv \{s_1, \dots, s_k, n_1, \dots, n_k\} \in \mathfrak{R}_+^k \times \mathfrak{R}_{++}^k$ , where  $s_i$  is a skill level and  $n_i$  the number of consumers with that level of skill. It is assumed that the vector  $\{s_1, \dots, s_k\}$  is ordered, so  $s_i < s_{i+1}$ . An *allocation* is a vector  $A \equiv \{z_1, \dots, z_k, x_1, \dots, x_k\}$  where  $z_i$  is the pre-tax income and  $x_i$  the consumption of a consumer with skill level  $s_i$ . With  $\ell$  denoting labor supply, all consumers have preferences described by the quasi-linear function

$$U = u(x) - \ell. \tag{1}$$

Throughout the paper, we impose the following assumption.

*Assumption 1:*  $u(x)$  is strictly monotonically increasing,  $\lim_{x_i \rightarrow 0} u'(x_i) = \infty$  and  $u''(x_i) < 0$ .

With a consumption level of  $x$ , the marginal rate of substitution ( $MRS_i$ ) between consumption and labor for a consumer with skill  $s_i$  is equal to  $1/u'(x) s_i$ , so that the condition of agent monotonicity applies. By definition, income, labor supply and skill are related by  $z = s\ell$ . So, a consumer with skill level  $s_i$  receiving pre-tax income  $z_i$  must supply labor  $\ell_i = z_i/s_i$ . Using this, given an allocation  $A$ , the utility achieved at skill level  $i$  is  $U_i \equiv u(x_i) - z_i/s_i$ .

The optimal allocation is chosen to maximize the welfare function  $W = \sum_{i=1}^k \mu_i n_i U_i$ , where the welfare weights satisfy  $\mu_i \geq 0$ . Any candidate for the optimal allocation must satisfy the incentive compatibility constraints

$$u(x_i) - \frac{z_i}{s_i} \geq u(x_j) - \frac{z_j}{s_j}, \quad \text{all } i, j, \quad (2)$$

and the feasibility constraint

$$\sum_{i=1}^k z_i = \sum_{i=1}^k x_i. \quad (3)$$

As (Weymark 1986a) noted, the tax function is kinked at the location of each consumer, so the marginal tax rate is not formally defined at these points. However, it is possible to take the gradient of the indifference curve as determining an implicit marginal rate of tax (which is equal to the left-derivative of the supporting function). For any allocation,  $A$ , the marginal tax rate ( $MTR_i$ ) facing a consumer with skill level  $s_i$  is

$$MTR_i := 1 - MRS_i = 1 - \frac{1}{u'(x_i) s_i}, \quad (4)$$

and the average tax rate

$$ATR_i := \frac{z_i - x_i}{z_i}. \quad (5)$$

### 3 Existing Analysis

In this section we adopt the assumptions of Weymark (1986*a,b*, 1987) concerning utility and income. Under these assumptions we derive a characterization of the optimal allocation (but taking into account that we allow for several consumers with each skill level). This is then employed to investigate the extent of bunching and the structure of optimal taxes.

The Weymark model is identified by two key assumptions on permissible income levels and the structure of the utility function. The first of these assumptions is that negative income levels are permitted (though consumption must be non-negative). The second is that there is a lower bound on utility at a consumption level of zero. We state the assumptions here but reserve discussion until later.

*Assumption 2:* (Possibility of negative incomes)  $A \in \mathfrak{R}^k \times \mathfrak{R}_+^k$ .

*Assumption 3:* (Lower bound on utility)  $u(0) = 0$ .

#### 3.1 Characterization of optimal allocation

The optimization that results when Assumptions 1-3 are imposed is termed Program I. This can be stated as:

$$\text{Program I: } \max_{\{A \in \mathfrak{R}^k \times \mathfrak{R}_+^k\}} W = \sum_{i=1}^k \mu_i n_i U_i \quad \text{subject to (2) and (3).}$$

Program I can be simplified by using two standard results. The proof of the first can be found in (Röell 1985).

**Lemma 1** (*Guesnerie and Seade/Röell*) *The incentive compatibility constraints form a monotonic chain to the left.*

The implication of this lemma is that the incentive compatibility constraints can be reduced to a set of equalities relating the allocation of each consumer to that of the consumer with the next lowest level of skill. These equalities can be written as

$$z_{i+1} = z_i + s_{i+1}[u(x_{i+1}) - u(x_i)], \quad i = 1, \dots, k-1. \quad (6)$$

The next lemma relates the monotonicity of the allocation and the incentive compatibility constraints. The proof of this result and all those that follow is given in the Appendix.

**Lemma 2** *If an allocation  $A \in \mathfrak{R}^k \times \mathfrak{R}_+^k$  satisfies  $0 \leq x_1 \leq x_2 \leq \dots \leq x_k$  and (6), then:*

- (i)  $z_1 \leq z_2 \leq \dots \leq z_k$ ;
- and
- (ii) *the allocation satisfies incentive compatibility.*

Collecting the reduced set of incentive compatibility constraints (6) and the resource constraint gives the system

$$\begin{bmatrix} n_1 & n_2 & n_3 & \cdots & n_{k-1} & n_k \\ 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & \vdots \\ \vdots & & & & & 0 \\ 0 & \dots & \dots & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^k n_i x_i \\ -s_2 [u(x_2) - u(x_1)] \\ \vdots \\ -s_k [u(x_k) - u(x_{k-1})] \end{bmatrix}. \quad (7)$$

Solving for the income levels associated with each level of skill provides

$$\begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix} = \frac{1}{\sum_{i=1}^k n_i} \begin{bmatrix} 1 & \sum_{i=2}^k n_i & \sum_{i=3}^k n_i & \cdots & n_k \\ 1 & -n_1 & \sum_{i=3}^k n_i & & n_k \\ 1 & -n_1 & -[n_1 + n_2] & & n_k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -n_1 & -[n_1 + n_2] & & -\sum_{i=1}^{k-1} n_i \end{bmatrix} \begin{bmatrix} \sum_{i=1}^k n_i x_i \\ -s_2 [u(x_2) - u(x_1)] \\ \vdots \\ -s_k [u(x_k) - u(x_{k-1})] \end{bmatrix}. \quad (8)$$

Let  $N = \sum_{i=1}^k n_i$ . Then the income of a consumer with skill level  $i$ ,  $i = 1, \dots, k$ , can be written as

$$z_i = \frac{1}{N} \left[ \sum_{j=1}^k n_j x_j + \sum_{j=1}^k [a_{i,j+1} s_{j+1} - a_{i,j} s_j] u(x_j) \right], \quad (9)$$

with

$$\begin{aligned}
a_{i,1} &= 0, \quad i = 2, \dots, k \\
a_{i,j} &= \begin{cases} \sum_{q=j}^k n_q, & i = 1, \dots, j-1 \\ -\sum_{q=1}^{j-1} n_q, & i = j, \dots, k \end{cases}, \quad j = 2, \dots, k \\
a_{i,k+1} &= 0, \quad i = 1, \dots, k.
\end{aligned} \tag{10}$$

Combining these results, Program I can be expressed as a simpler problem involving only the choice of the consumption allocation. This choice is subject to the single constraint that the consumption allocation forms a non-decreasing sequence. Once the consumption allocation has been chosen, the income allocation is defined by (9).

Let  $X := \{x_1, \dots, x_k\} \in \mathfrak{R}_+^k$ . The precise statement of the reduced form of Program I is

$$\text{Program I: } \max_{\{X \in \mathfrak{R}_+^k\}} \sum_{i=1}^k \mu_i n_i U_i, \quad \text{subject to } 0 \leq x_1 \leq x_2 \leq \dots \leq x_k,$$

with  $z_i$  given by (9). The necessary conditions for this problem can be found by forming the Lagrangean

$$\mathcal{L} = \sum_{i=1}^k \left[ \mu_i n_i \left( u(x_i) - \frac{z_i}{s_i} \right) + \nu_i (x_i - x_{i-1}) \right], \tag{11}$$

where  $x_0 := 0$ . The first-order conditions for the choice of the consumption levels are

$$\frac{\partial \mathcal{L}}{\partial x_j} = \mu_j n_j u'(x_j) - \sum_{i=1}^k \frac{\mu_i n_i}{s_i} \frac{\partial z_i}{\partial x_j} + \nu_j - \nu_{j+1} = 0, \quad j = 1, \dots, k, \tag{12}$$

$$\nu_1 > 0 \text{ if } x_1 = 0, \quad \nu_j > 0 \text{ if } x_j = x_{j-1}, \quad j = 2, \dots, k, \tag{13}$$

and  $\nu_{k+1} := 0$  as a notational convenience.

At this point a new notation is introduced that draws a closer parallel with the work of Weymark. To do this define

$$\beta_j := \mu_j n_j - \frac{1}{N} \sum_{i=1}^k \frac{a_{i,j+1} s_{j+1} - a_{i,j} s_j}{s_i} \mu_i n_i, \quad j = 1, \dots, k, \tag{14}$$

where  $a_{i,j}$  is defined in (10) and

$$\Theta := \frac{1}{N} \sum_{i=1}^k \frac{\mu_i n_i}{s_i}. \tag{15}$$

Using this notation, the necessary conditions (12) can be rewritten as

$$\beta_j u'(x_j) = \Theta n_j - \nu_j + \nu_{j+1}, \quad j = 1, \dots, k. \tag{16}$$

The necessary conditions (16), alongside the complementary slackness conditions (13), form the complete set of conditions describing the optimal consumption allocation. Given the consumption allocation, the definition of the income allocation then follows from (9).

### 3.2 Bunching

The nature of the allocations that can arise under Assumptions 1-3 can be best understood by investigating the conditions under which bunching will not occur and the patterns of bunching that can occur. The following results extend those of (Weymark 1986a) to allow for the variable number of consumers with each skill level.

The central result that characterizes when there is no-bunching is the following. Note that this result permits  $\beta_1 < 0$  but requires all other  $\beta$ s to be positive.

**Lemma 3** *There is no bunching if and only if  $\frac{\beta_1}{n_1} < \frac{\beta_2}{n_2} < \dots < \frac{\beta_k}{n_k}$  and  $\beta_2 > 0$ .*

The next result characterizes the marginal tax rate when there is no bunching and  $\beta_1 > 0$ .

**Proposition 1** *If  $\frac{\beta_1}{n_1} < \frac{\beta_2}{n_2} < \dots < \frac{\beta_k}{n_k}$  and  $\beta_1 > 0$ , the marginal tax rate at skill level  $s_i$  is given by  $MTR_i = 1 - \frac{\beta_i}{\Theta_{n_i s_i}}$ .*

It should be observed that under the conditions of the proposition  $MTR_i$  is dependent only upon the distribution of population across skill levels,  $\{s_1, \dots, s_k, n_1, \dots, n_k\}$ . Hence given an economy,  $E$ , that satisfies the conditions required of  $\frac{\beta_j}{n_j}$ , the same pattern of marginal tax rates will hold regardless of the preferences represented by the utility of consumption  $u(x)$ . This result was first noted and employed by (Myles 2000).

The next result shows there is bunching of the lowest-skill consumers if the initial values of  $\beta_j$  are negative. In addition, each consumer is allocated a consumption level of zero.

**Lemma 4** *If  $\beta_1 < 0, \dots, \beta_j < 0$  but  $\beta_{j+1} > 0$ , then  $x_j = 0$  for  $i = 1, \dots, j$  and  $x_{j+1} > 0$ .*

Using this result on bunching at the bottom, it is possible to characterize the tax rate that the bunched consumers face.

**Proposition 2** *If  $\beta_1 < 0, \dots, \beta_j < 0$  but  $\beta_{j+1} > 0$ , then the marginal tax rate at skill level  $s_i$ ,  $i = 1, \dots, j$ , is given by  $MTR_i = 1$ .*

The proposition shows that when there is bunching of the lowest-skill consumers at a zero consumption level, it remains possible to calculate the marginal tax rate of these consumers without specifying the utility function  $u(x)$ . The value of the marginal tax rate is due to the assumption that  $u(0) = 0$  forcing

the indifference curves to be asymptotic to the pre-tax income axis as consumption tends to zero. In this circumstance, any allocation that assigns a zero consumption level can be supported only by a consumption function which is flat.

The next lemma describes the outcome when the conditions for no bunching are violated. This result is the equivalent of Proposition 6 in (Weymark 1986a).

**Lemma 5** *Assume that consumers  $j$  to  $j + m$ ,  $m \geq 0$ , are bunched. Consumer  $j + m + 1$  will also be bunched with  $j$  to  $j + m$  if  $\frac{\beta_{j+m+1}}{n_{j+m+1}} \leq \frac{\sum_{t=0}^m \beta_{j+t}}{\sum_{t=0}^m n_{j+t}}$ .*

It should be noted that this result also implies the condition for bunching to begin. Bunching will start if

$$\frac{\beta_j}{n_j} > \frac{\beta_{j+1}}{n_{j+1}}, \quad (17)$$

in which case  $j + 1$  will be bunched with  $j$ . Although the proposition shows when bunching will begin, it should be stressed that it does not determine the structure of the bunching intervals. Put another way, the result provides the sufficient condition for adding another skill level to the set that are bunched but it does not state the necessary condition. Consequently the result does not give a complete characterization of bunching. At present, no conditions that are necessary and sufficient are known but we have developed a computational procedure that worked in all cases to which it was applied.

The behavior of the marginal tax rate for bunched consumers is described in the next result. This shows that the marginal tax rate increases with skill across the set of bunched consumers.

**Proposition 3** *Assume that consumers  $i = j, \dots, j + m$  are bunched. Then the marginal tax rate faced by  $i$  is  $MTR_i = 1 - \frac{\bar{\beta}}{s_i \Theta}$ ,  $\bar{\beta} := \frac{\sum_{i=j}^{j+m} \beta_i}{\sum_{i=j}^{j+m} n_i}$ .*

The *MTR* is determined by the weighted average value of  $\beta_i$  for the consumers who are bunched and is independent of the utility of consumption. Because of agent monotonicity, the *MTR* increases with skill at the bunched allocation.

### 3.3 Implementation

In this section we explore the implementation of the model. What this shows is that bunching occurs in a high proportion of economies so that the characterization with no-bunching is of limited interest. Solving the bunching problem, we provide an implementation of the model that illustrates the structure of optimal taxes. However, for the reasons we detail in the following sub-section, we believe the assumptions under which this is derived are inappropriate so do not place a great emphasis upon this.

The analysis in the previous section has shown that there will be no-bunching only if a very particular pattern of the  $\beta_j$ s emerge. The discussion in (Weymark



1986a) noted that the values of  $\beta_j$  for an economy are especially sensitive to small changes in the set of skill levels. This raises the issue of what patterns of  $\beta_j$ s can emerge. We shed some light upon this by considering a special case in which the values of  $\beta_j$  can be related simply to the number of skill levels and the difference between successive skill levels. The sensitivity of the  $\beta_j$ s to small changes in the set of skill levels is reflected in similar sensitivity of the structure of the tax function to similar small changes. To overcome this sensitivity and obtain a general picture of the tax function we conduct a Monte Carlo analysis, choosing economies randomly and averaging. The details of our methodology are described later.

An interesting pattern for the  $\beta$  values can be obtained by assuming that the skills levels  $j$  and  $j + 1$  are related by  $s_{j+1} = (1 + \rho)s_j$ . If  $\mu_i = n_i = 1$ , then it is possible to solve to find

$$\beta_j = s_1 (1 + \rho)^{j-1} \left[ \left(\frac{1}{k}\right) \sum_{i=1}^k \frac{1}{s_1 (1 + \rho)^{i-1}} (1 + \rho)^j - \sum_{i=1}^j \frac{1}{(1 + \rho)^{i-1}} \rho \right]. \quad (18)$$

Now let the highest and lowest skill levels be related by  $s_k = ms_1$ , so that  $m$  measures the ratio of the highest skill to the lowest skill. Since we also have  $s_k = (1 + \rho)^k s_1$ , we can solve to write

$$\rho = m^{1/k-1} - 1, \quad (19)$$

so that the constant of proportionality,  $\rho$ , can be adjusted to keep  $m$  constant as  $k$  is changed. The results from applying these calculations are plotted in the following figures.

Figure 1 plots  $\beta_j$  for  $k = 5$ . In this case  $\beta_i$  is negative for  $i = 1, \dots, 4$ . It decreases until  $i = 3$  then increases. In Figure 2,  $k$  is raised to 10. This increases the values of  $\beta$  for the lowest-skill levels but the negativity remains. When  $k$  is increased to 100 in Figure 3,  $\beta_i$  is positive until  $i = 10$ , becomes negative, and then becomes positive at  $i = 70$ .

What does this exercise show? First it emphasizes that even for this restrictive pattern of skill levels there need be no monotonicity in the  $\beta$ s. Secondly, it makes clear that negative  $\beta$  values can be a frequent problem. For the lower values of  $k$ , the fact that the initial  $\beta$  values are negative implies the welfare function would be unbounded if utility did not have a lower bound. Although increasing  $k$  removes the negativity of  $\beta$  for the lowest skills, it remains negative for intermediate skills. In every case, there would be extensive bunching with this pattern of skills. These figures show only one possible outcome; others can be achieved by using (18) with different values of  $m$ .

Table 1 reports the results of a numerical analysis of bunching. The skill levels are drawn at random from the interval  $[1, 10]$  with the number of skills given by  $k$ . Four methods of determining the number of consumers with each skill are employed. A detailed description of these methods is reserved until Section 5. For the present it is sufficient to observe that the proportion of

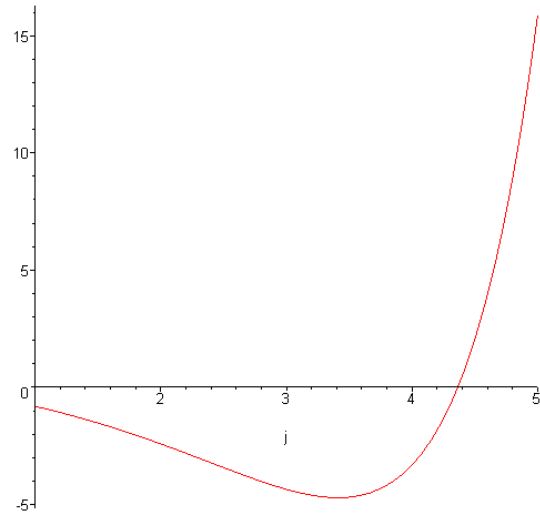


Figure 1: Parameters  $m = 10, k = 5$

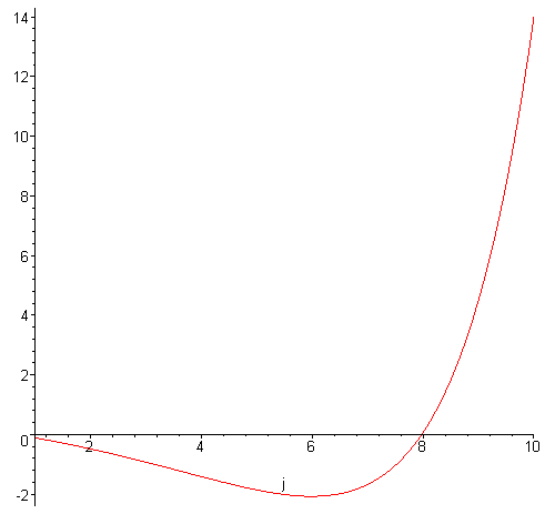


Figure 2: Parameters  $m = 50, k = 10$

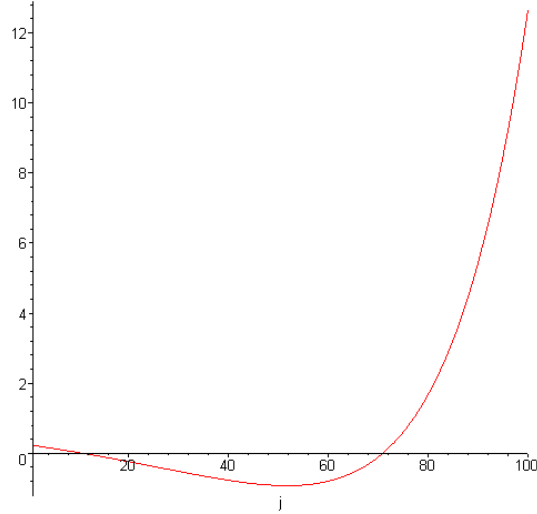


Figure 3: Parameters  $m = 50$ ,  $k = 100$

economies with no-bunching tends rapidly to zero as  $k$  increases. Hence the characterization of no-bunching applies to an empty set of economies for even a fairly small number of skill levels.

$n$	$P/k$	Uniform	Log-normal	Chi-square
$k = 5$	0.302	0.265	0.669	0.421
$k = 10$	0.001	0.002	0.148	0.006
$k = 20$	0.0	0.0	0.0	0.0

Table 1: Proportion of Economies with No-Bunching

The results in Table 2 report the proportion of economies for which the *MTR* and *ATR* increase between two consecutive skill levels. Hence the first number in the top row, 0.624, shows that in 62.4% of economies the *MTR* was higher for the second skill level than the first. The second number, 0.712, shows that it was higher in 71.2% of economies for the third skill level than the second. Reading the results in this way, it can be seen that in the majority of economies the *MTR* initially rises with skill and then falls. Surprisingly, the *ATR* behaves in the same manner. These descriptions apply to both methods of selecting the number of consumers with each skill. Additionally, when  $n = P/k$ ,  $z_1 < 0$  in 93.7% of economies and  $x_1 = 0$  in 30%. The corresponding numbers for  $n \sim \chi^2$  were 87.1% and 25.7%. Negative incomes are therefore a frequent occurrence. Zero consumption levels are less frequent, but still not unusual.

$n = P/k$	<i>MTR</i>	0.624	0.712	0.663	0.628	0.561	0.493	0.459	0.363	0
	<i>ATR</i>	0.612	0.543	0.503	0.506	0.518	0.516	0.452	0.304	0.088
$n \sim \chi^2$	<i>MTR</i>	0.608	0.632	0.573	0.542	0.514	0.460	0.394	0.291	0
	<i>ATR</i>	0.589	0.514	0.550	0.564	0.586	0.544	0.392	0.234	0.076

Table 2: Proportion of Increases in *MTR* and *ATR*  
 $k = 10, 1 < s < 10, P = 10000, \ddagger \text{ sim.} = 1000$

### 3.4 Discussion

It is important to fully comprehend the consequences of the assumptions upon utility and income in order to appreciate why a relaxation is needed. We begin by discussing the assumption placed on utility but eventually conclude that it is interlinked with the possibility of negative income.

Consider first the assumption that there is a lower bound upon utility, specifically that  $u(0) = 0$ . This assumption is not a usual one and it rules out many standard utility functions, including the class of isoelastic functions  $u(x) = \frac{x^{1-\varepsilon}}{1-\varepsilon}$  for  $\varepsilon \geq 1$  ( $= \ln(x)$  for  $\varepsilon = 1$ ). The implication of this assumption is that consumption levels of zero are possible in the optimum allocation. Hence if a set of skill levels generate a value of  $\beta_1 < 0$ , the solution to Program I' has  $x_1 = 0$ . We have already noted that this outcome arises in numerous cases, so if interpreted literally the model frequently generates optimal allocations where some consumers receive no consumption.

The next step is to consider what happens if the assumption of a finite lower bound is relaxed. An alternative assumption would be that  $\lim_{x \rightarrow 0} u(x) = -\infty$  as applies, for example, with logarithmic utility. Under this assumption, if  $\beta_1 < 0$  then  $\lim_{x \rightarrow 0} \beta_1 u(x) = \infty$  and the value of social welfare is unbounded as the consumption of the lowest skill consumer tends to zero. Since this cannot be a sensible solution to the allocation problem, there must be some further features of such an allocation that render it unacceptable. The following example illustrates what these features are.

**Example 1** Consider an economy with two consumers. Consumer 1 has skill level  $s_1 = 1$  and consumer 2 has skill level  $s_2$ . Both have preferences  $U = \ln(x) - \ell$ . Assuming  $\mu_1 = \mu_2 = 1$ , the value of  $\beta_1$  can be written as

$$\beta_1 = 1 - \frac{s_2}{2} \left( 1 - \frac{1}{s_2} \right), \quad (20)$$

so that  $\beta_1 < 0$  when  $s_2 > 3$ . Solving the optimization problem gives

$$x_1 = \begin{cases} \frac{1 - \frac{s_2}{2} \left( 1 - \frac{1}{s_2} \right)}{\frac{1}{2} \left( 1 + \frac{1}{s_2} \right)} & \text{if } \frac{1 - \frac{s_2}{2} \left( 1 - \frac{1}{s_2} \right)}{\frac{1}{2} \left( 1 + \frac{1}{s_2} \right)} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

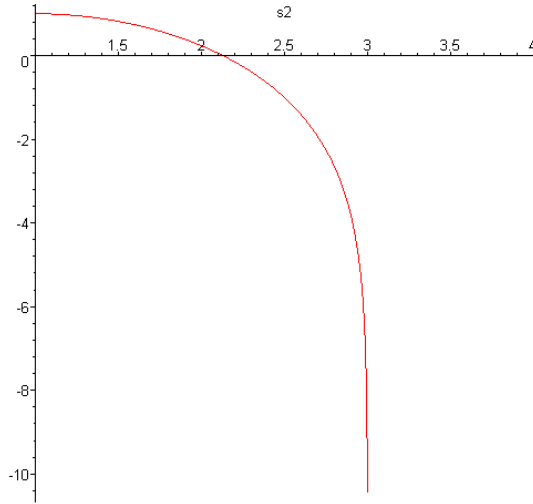


Figure 4: Income of Low-Skill Consumer

Clearly,  $x_1 = 0$  if  $s_2 \geq 3$ , which corresponds to the range in which  $\beta_1 < 0$ .

Now consider incomes and welfare. A plot of  $z_1$  as a function of  $s_2$  is given in Figure 4 which shows that  $z_1$  tends to  $-\infty$  as  $s_2 \rightarrow 3$ . The behavior of  $z_2$  is the converse - it tends to  $+\infty$  as  $s_2 \rightarrow 3$ . For  $s_2 > 3$ ,  $z_1, z_2$  and  $W$  are undefined. Hence the solution to Program I' for  $u(x) = \ln(x)$  is only defined for  $s_2 < 3$ . If the skill levels are too diverse there is no optimal allocation.

As the next section will make clear, the solution method has to be significantly modified to accommodate negative values of  $\beta_i$  - the solution in general is not to simply assign a zero consumption level. The analysis we present does not require any assumption to be placed upon  $u(0)$  which is clearly more in line with standard analysis.

The second assumption was to allow negative income levels, that is to permit optimal allocations that satisfy  $z_i < 0$ . As explained by (Weymark 1986a), previous authors ((Guesnerie and Laffont 1984) and (Lollivier and Rochet 1983)) also did not restrict incomes to be positive. Allowing incomes to be negative does not seem to be a natural assumption in the context of the model of income taxation. The income of consumer  $i$  is defined by  $z_i := s_i \ell_i$ . Since the level of skill is non-negative, with  $s_i \geq 0$ , income can only be negative if labor supply is negative,  $\ell_i < 0$ , - but it is difficult to see what interpretation can be given to this. Consequently, we conclude that in a model of income taxation the constraint that labor income is non-negative must be imposed and the solution for the optimal allocation adapted to incorporate this constraint.

A simple example is sufficient to illustrate the effect that the non-negative income requirement has upon the optimal allocation.

**Example 2** Assume an economy with four consumers who have skill levels  $s_1 = 1.3516, s_2 = 9.6729, s_3 = 15.4798$  and  $s_4 = 17.9347$ . These generate values of  $\beta_1 = -3.8265, \beta_2 = 0.2313, \beta_3 = 3.2747$  and  $\beta_4 = 4.3206$ . Solving Program I' allowing negative incomes, Theorem 2 in (Weymark 1986a) would imply no bunching at the optimum. With the utility function  $u(x) = x^{1/2}$  the optimal consumption plan would be  $x_1 = 0, x_2 = 0.2305, x_3 = 46.1962$  and  $x_4 = 80.4176$ . These consumption levels are supported by the income vector  $z_1 = -30.3956, z_2 = -25.719, z_3 = 72.0295$  and  $z_4 = 110.9623$ , so that two of the four consumers are allocated a significantly negative labor income.

Using the results we derive in Section 4, the optimal allocation if income is constrained to be non-negative is given by  $x_1 = 10.6327, x_2 = 10.6327, x_3 = 53.1735$  and  $x_4 = 80.4176, z_1 = 0, z_2 = 0, z_3 = 62.4028$  and  $z_4 = 92.4494$ . At this allocation, consumers 1 and 2 are both bunched with a zero income level and identical levels of consumption. Their consumption levels are also significantly above those when income can be negative.

The contrast between the allocation with non-negativity and that when negative incomes are allowed shows that it is much more than just a simplification. When non-negativity is imposed the solution changes dramatically and the result on no-bunching no longer applies.

Finally, we have noted how the structure of the utility function is important when some  $\beta$ s are negative. The example also illustrates a further fact: negative betas and non-negative incomes are closely inter-linked. By this it is meant that there is a direct connection between the finding of a negative value of  $\beta_1$  and zero income as a binding constraint. This will become clear when we formally characterize the optimal allocation with non-negative incomes.

## 4 Non-Negative Incomes

This section summarizes the derivation of the solution to the quasi-linear optimal income tax problem when a non-negativity constraint is placed upon income. We begin with the statement of the basic optimization and then gradually refine this to a form that allows comparison with the earlier results. Section 5 then describes the properties of the tax function that emerge from an implementation of this solution.

The analysis described in Section 3 allows income to be negative. We have already described the consequences of this assumption and the reasons for wishing to reject it. We now restrict the choice set to non-negative incomes, so that  $A \in \mathbb{R}_+^k \times \mathbb{R}_+^k$ . In addition we want to remove the assumption that the utility of zero consumption has a lower bound. Hence we drop Assumption 3 and replace Assumption 2 by

*Assumption 2'*: (Non-negativity of income)  $A \in \mathbb{R}_+^k \times \mathbb{R}_+^k$ .

We can still apply Lemmas 1 and 2 to express the problem in terms of consumption. The optimization problem for the choice of the allocation now

becomes

$$\text{Program II: } \max_{\{X \in \mathbb{R}_+^k\}} \sum_{i=1}^k \mu_i n_i \left( u(x_i) - \frac{z_i}{s_i} \right), \quad (22)$$

subject to

$$\begin{aligned} z_i &\geq 0, \quad i = 1, \dots, k, \\ 0 &\leq x_1 \leq x_2 \leq \dots \leq x_k, \end{aligned}$$

and  $z_i$  defined by (9).

The necessary conditions for this problem can be found by forming the Lagrangean

$$\mathcal{L} = \sum_{i=1}^k \left[ \mu_i n_i \left( u(x_i) - \frac{z_i}{s_i} \right) + \lambda_i z_i + \nu_i (x_i - x_{i-1}) \right], \quad (23)$$

where  $x_0 := 0$ . The first-order conditions for the choice of the consumption levels are

$$\frac{\partial \mathcal{L}}{\partial x_j} = \mu_j n_j u'(x_j) - \sum_{i=1}^k \frac{\mu_i n_i}{s_i} \frac{\partial z_i}{\partial x_j} + \sum_{i=1}^k \lambda_i \frac{\partial z_i}{\partial x_j} + \nu_j - \nu_{j+1} = 0, \quad j = 1, \dots, k, \quad (24)$$

$$\lambda_i > 0 \text{ if } z_i = 0, \quad \nu_1 > 0 \text{ if } x_1 = 0, \quad \nu_j > 0 \text{ if } x_j = x_{j-1}, \quad j = 2, \dots, k, \quad (25)$$

and  $\nu_{k+1} := 0$  as a notational convenience.

Equation (24) can be rearranged as

$$\mu_j n_j u'(x_j) = \sum_{i=1}^k \left( \frac{\mu_i n_i}{s_i} - \lambda_i \right) \frac{\partial z_i}{\partial x_j} - \nu_j + \nu_{j+1} \quad (26)$$

Using (9) in (26) we obtain the following set of equations for the  $x_i$ s,  $\nu_i$ s and  $\lambda_i$ s:

$$\begin{aligned} &u'(x_j) \left[ \mu_j n_j - \frac{1}{N} \sum_{i=1}^k \left( \frac{\mu_i n_i}{s_i} - \lambda_i \right) (a_{i,j+1} s_{j+1} - a_{i,j} s_j) \right] \\ &= \frac{1}{N} \sum_{i=1}^k \left( \frac{\mu_i n_i}{s_i} - \lambda_i \right) - \nu_j + \nu_{j+1}, \quad j = 1, \dots, k, \end{aligned} \quad (27)$$

and

$$\lambda_i > 0 \text{ if } z_i = 0, \quad \nu_1 > 0 \text{ if } x_1 = 0, \quad \nu_j > 0 \text{ if } x_j = x_{j-1}, \quad j = 2, \dots, k. \quad (28)$$

Using the notation introduced earlier, the necessary condition can be rewritten as

$$u'(x_j) \left[ \beta_j + \sum_{i=1}^k \tilde{\lambda}_i [a_{i,j+1} s_{j+1} - a_{i,j} s_j] \right] = \Theta n_j - \sum_{i=1}^k \tilde{\lambda}_i - \nu_j + \nu_{j+1}, \quad (29)$$

with  $\tilde{\lambda} \equiv \lambda/N$ . Hence,

$$u'(x_j) = \frac{\hat{\Theta}_j}{\hat{\beta}_j}, \quad j = 1, \dots, k, \quad (30)$$

where

$$\hat{\Theta}_j = \Theta n_j - \sum_{i=1}^k \tilde{\lambda}_i - \nu_j + \nu_{j+1}, \quad \hat{\beta}_j = \beta_j + \sum_{i=1}^k \tilde{\lambda}_i [a_{i,j+1} s_{j+1} - a_{ij} s_j].$$

The necessary conditions (30), alongside the complementary slackness conditions (28) and the definition of the income allocation (9), form the complete set of conditions describing the optimal consumption allocation.

When none of the constraints is binding we obtain the interior no-bunching solution,

$$u'(\hat{x}_j) = \frac{\Theta n_j}{\beta_j}, \quad j = 1, \dots, k. \quad (31)$$

This corresponds to the solution of Program I', but only if at the optimum all income levels are non-negative. The important point to note is that there is nothing in the model which relates the allocation  $\{\hat{x}\}$  satisfying (31) to the non-negativity of the solution of (9). Moreover, note that the structure of the vector  $\{\beta_1, \dots, \beta_k\}$  reveals nothing about the nature of the solution. If any of the income levels are constrained at zero, so at least one  $\lambda_i > 0$ , then nothing can be inferred from (30) about the structure of bunching at the optimum. Hence the characterizations of bunching in Section 3 do not apply when a binding non-negativity constraint is added.

We now focus on the case where the non-negativity constraints on income are binding. The optimization problem can be solved in two equivalent ways. The first approach is to directly solve the system of equations (26), the complementary slackness conditions and the definition of the income allocation. This is what is implemented in the numerical section.

There is an alternative statement of the optimization program that draws a closer link with the solution of Section 3. To introduce this, define

$$\beta_{j,t} := \mu_j n_j - \frac{1}{N} \sum_{i=t+1}^k \frac{a_{i,j+1} s_{j+1} - a_{i,j} s_j}{s_i} \mu_i n_i, \quad j = 1, \dots, k, \quad (32)$$

and

$$\Theta_t = \frac{1}{N} \sum_{i=t+1}^k \frac{\mu_i n_i}{s_i}. \quad (33)$$

Also, let the function  $x = \eta(x_{t+1}, \dots, x_k)$  be defined implicitly by

$$0 = \frac{1}{N} \left[ \sum_{j=1}^t n_j x + \sum_{j=t+1}^k n_j s_{t+1} u(x) + \sum_{j=t+1}^k (n_j x_j + [a_{i,j+1} s_{j+1} - a_{ij} s_j] u(x_j)) \right]. \quad (34)$$

The alternative optimization program is described in the following theorem.



**Theorem 1** *If allocation  $\{\hat{z}, \hat{x}\}$ , with*

$$\hat{z}_i = 0 \quad \text{for } i = 1, \dots, t \quad \text{and} \quad \hat{z}_i > 0 \quad \text{for } i = t + 1, \dots, k, \quad (35)$$

*is optimal for Program II, then  $\{\hat{z}, \hat{x}\}$  is optimal for Program II' defined by*

$$\text{Program II':} \quad \max_{\{x_{t+1}, \dots, x_k\}} \sum_{i=1}^k [\beta_{i,t} u(x_i) - \Theta_t n_i x_i], \quad (36)$$

*subject to*

$$\begin{aligned} x_i &= \eta(x_{t+1}, \dots, x_k), \quad i = 1, \dots, t, \\ 0 &\leq x_1 = x_2 = \dots = x_t < x_{t+1} \leq x_{t+2} \leq \dots \leq x_k. \end{aligned}$$

The first-order conditions for Program II' are given by

$$\beta_{j,t} u'(x_j) - \Theta_t n_j + \nu_j - v_{j+1} + \sum_{i=1}^t (\beta_{i,t} u'(x_i) - \Theta_t n_i + \nu_i - v_{i+1}) \frac{\partial \eta}{\partial x_j} = 0, \quad (37)$$

for  $j = t + 1, \dots, k$ . Each of these conditions involves the term  $\partial \eta / \partial x_j$  which captures the extent to which the consumption levels of the constrained consumers must be adjusted to maintain their incomes at zero.

The value of this characterization of the optimum is two-fold. First it shows that the values of  $\beta_j$  do not characterize the solution. This emphasizes why the analysis of bunching with negative incomes does not extend. Second it shows that it is the value of  $\beta_{j,t}$  in conjunction with the  $\partial \eta / \partial x_j$  that matters. No simple statements can be made to link the structure of the consumption allocation with the  $\beta_j$ .

Although Program II' has the same solution as Program II there is an important distinction. For Program II' it must be known in advance which of the  $z_j$ s will be zero. Therefore it can only be used to find the solution in an iterative manner. That is, use  $\beta_j$  first to find if solution with non-negative incomes exists. If it does not, then use  $\beta_{j,1}$ . If this does not generate a non-negative set of incomes - it will guarantee  $z_1 = 0$  but not necessarily  $z_2, \dots$  - then try  $\beta_{j,2}$ . All  $\beta_{j,t}$  generating non-negative solutions must then be contrasted to find that which generates the highest welfare level.

## 5 Implementation

We have already noted on several occasions that the tax function is closely dependent upon the structure of an economy described in the vector  $E$ . To use the model to obtain an insight into what can be viewed as the typical pattern of the tax function we employ a Monte Carlo analysis. This involves making a repeated random draw from a set of possible economies and solving Program II for each draw. The pattern of results can then be analyzed for regularities in the pattern of behavior.

More specifically, we are able to determine the proportion of cases in which the non-negative income constraint is binding and the proportion in which bunching occurs. These values illustrate the extent to which allowing negative incomes is a harmless simplification. In fact, it will be seen that the proportion of cases in which the non-negative income constraint binds can be above 80%. These results, though, remain incidental to the major objective.

The major objective is to try and understand the qualitative properties of the optimal tax function in terms of progressiveness. How we do this is as follows. The major qualitative property of the tax function is whether the marginal (or, equally, the average) rate of tax increases as we progress from one skill level to the next. We know that this cannot be true for all skill distributions since the results in (Diamond 1998) and (Myles 2000) provide skill distributions where the marginal rate falls. It may still be true, though, that these cases are in the minority amongst the set of economies. Following this reasoning, the value that interests us is the proportion of economies in which the marginal (or average) tax rate is greater for skill level  $j + 1$  than it is for skill level  $j$ . If, for instance, we find that this is true in 80% of economies we can argue that marginal rate progressiveness is a typical property of the optimal tax function.

The method employed in the Monte Carlo analysis is to specify a permissible skill interval  $S := [s_\ell, s_u]$  and the number of skill levels,  $k$ . We then make a random selection using a uniform distribution of  $k$  skill levels from  $S$ . Two different methods are then employed to determine the number of consumers with each skill level. The first method is to fix a total potential population,  $P$ . For each skill level,  $s_i$ , the number of consumers with that skill,  $n_i$ , is selected by making a random draw from a uniform distribution over  $[1, P/k]$ . Because each  $n_i$  is an independent draw, this method places no *a priori* structure on the relation between the numbers with different skills.

The second method places more structure upon the  $n_i$ . Three different structures are used. The simplest is to assume that the population is the same for all skill levels, so for all  $i = 1, \dots, k$ ,  $n_i = P/k$ . The next method is to assume that the population for each skill is determined by a log-normal distribution, with the mean and variance chosen to match that of the uniform distribution over  $S$ . Then if  $f(s_i)$  is the probability density at skill  $s_i$ ,  $n_i$  is given by  $n_i = f(s_i)P$ . Hence this constructs a discrete approximation to the log-normal. The final method is to replace the log-normal distribution by a  $\chi^2$  distribution, with the degrees of freedom chosen to equate the mean of the distribution with the mean of the uniform distribution.

The first set of results describe the incidence of zero incomes and of bunching of the lowest-skill consumers. Table 3 reports results for the utility functions  $U = \ln(x) - \ell$  and  $U = x^{1/2} - \ell$  for the four different methods of selecting the values of  $n$ . In both cases the log-normal distribution has the lowest proportion of zero incomes and bunching. The non-negativity constraint on the income of the lowest skill type is binding in a majority of cases, with the proportion rising above 95% for the uniform selection case. The table also shows that there is a great frequency of bunching of consumption at the lower end. The frequency with which zero incomes arise demonstrates that allowing negative incomes will

generate an incorrect optimal allocation in the majority of cases. The results for other combinations of parameter values are similar.

Distribution	$n = P/k$	$U(1, P/k)$	$\chi^2$	Log-normal
$z_1 = 0$	0.860	0.803	0.793	0.660
$x_2 = x_1$	0.788	0.785	0.652	0.432

Table 3: Incidence of bunching at lower end: logarithmic utility  
 $k = 10, s_l = 1, s_u = 10, P = 10000, \# \text{ sim.} = 1000$

Distribution	$n = P/k$	$U(1, P/k)$	$\chi^2$	Log-normal
$z_1 = 0$	0.950	0.957	0.901	0.781
$x_2 = x_1$	0.874	0.883	0.777	0.517

Table 4: Incidence of bunching at lower end: square root utility  
 $k = 10, s_l = 1, s_u = 10, P = 10000, \# \text{ sim.} = 1000$

The results of implementing the model for the marginal tax rate and the average tax rate are displayed in Figures 5 to 12. These figures present the proportion of economies sampled for which the tax rate increases at each level of skill. The interpretation of a high proportion at skill level  $i$  is that the tax rate (either  $MTR$  or  $ATR$ ) increases for most economies between skill level  $i$  and skill level  $i+1$ . So for Figure 5 the marginal tax rate is higher for consumers with skill level 4 than for consumers with skill level 3 in just over 80% of economies with square root utility. It should be noted that the proportion of increases of  $MTR$  at skill level 9 is always zero since the  $MTR$  is non-negative, must be zero for skill level 10 (the highest skill) and skill level 10 is never bunched with skill level 9.

The figures for the  $MTR$  show a good consensus between the different specifications. The proportion of increases is close to 1 at low skill levels, so as we move up the skill distribution from the lower end the  $MTR$  will increase for the average case. However, the proportion of increases declines monotonically with the level of skill. Decreases in the marginal rate become predominant above the mean level of skill, although the possibility of increases remains significant for all but skill level 9. Although the interpretation is different, the picture provided by these results is very similar to that of other computations of optimal tax rates - especially those in the original paper of (Mirrlees 1971). It is possible to select economies where the  $MTR$  behaves in a different way, but these will be in a distinct minority. For example, the Diamond case of a U-shaped tax structure (5 decreases of  $MTR$  followed by 3 increases) occurs in  $1.47 \times 10^{-6}$  per cent of economies for square root utility and  $n = P/k$ .

The behavior of the  $ATR$  is more surprising than that of the  $MTR$ . In all cases, the proportion of increases first rises with skill and then falls. On average,

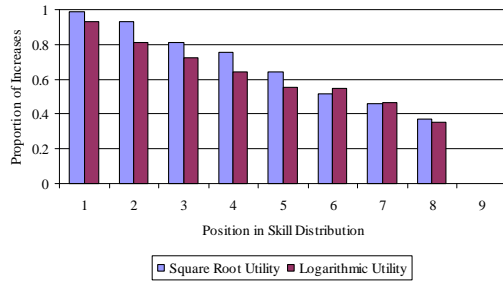


Figure 5: Increases in  $MTR$  for  $n = P/k$

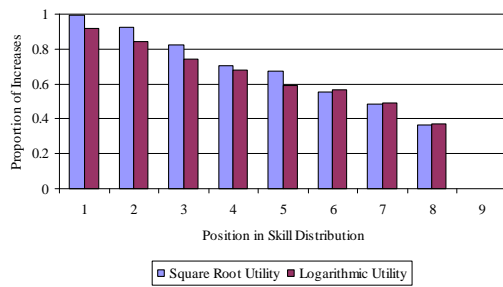


Figure 6: Increases in  $MTR$  for  $n \sim U(1, P/k)$

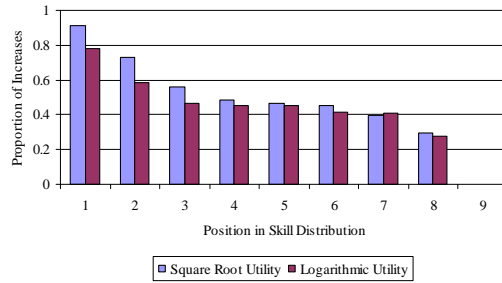


Figure 7: Increases in  $MTR$  for  $n \sim \text{Log} - \text{normal}$

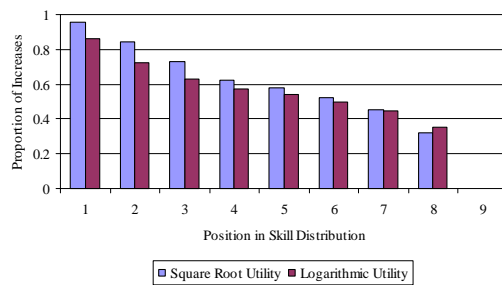


Figure 8: Increases in  $MTR$  for  $n \sim \chi^2$

the  $ATR$  must be highest at mean levels of skill. Towards the upper end of the skill distribution the  $ATR$  falls in a very high proportion of economies. This result is described as surprising since it is in contrast to the previous simulations in which the  $ATR$  increased throughout the skill distribution. This finding illustrates how previous simulations based on a single specification of the economy can fail to provide the “typical” picture.

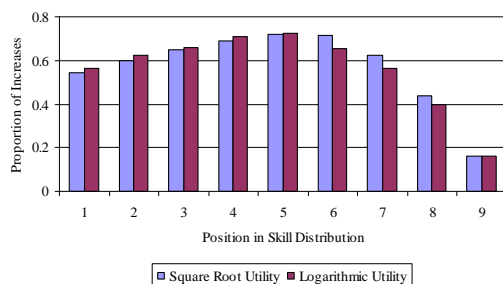


Figure 9: Increases in  $ATR$  for  $n = P/k$

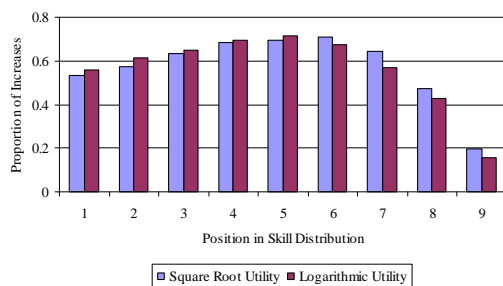


Figure 10: Increases in  $ATR$  for  $n \sim U(1, P/k)$

## 6 Conclusions

Ever since the publication of Mirrlees’ initial work on optimal income taxation it has been apparent that numerical analysis is necessary to supplement the limited

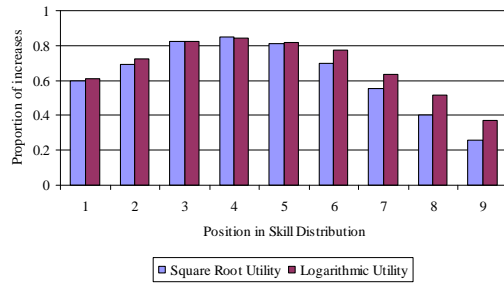


Figure 11: Increases in  $ATR$  for  $n \sim \text{Log} - \text{normal}$

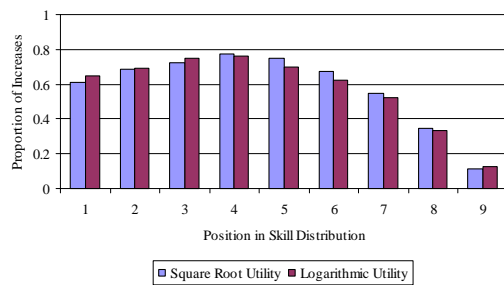


Figure 12: Increases in  $ATR$  for  $n \sim \chi^2$

theoretical results. The primary difficulty with the use of numerical methods is typically the limited number of model specifications that can be tested. This makes it difficult to distinguish the properties of a numerical solution that are general from those which are specific. It is clear that the structure of the optimal income tax function is dependent upon the distribution of skill in the economy. Hence, the fact that existing simulations have focussed upon the log-normal distribution is a major limitation.

The quasi-linear model apparently offers a route out of these limitations. Although it restricts the structure of preferences, the quasi-linear model simplifies the characterization of the optimal allocation. This is particularly true if there is no bunching of consumers. In such cases, the marginal tax rate is independent of the utility of consumption and is determined entirely by the skill distribution and the social welfare weights. This seems to promise a framework in which the link between the skill distribution and the optimal tax function can be explicitly investigated. Unfortunately, this promise is not realized: the assumptions supporting the analysis are excessively restrictive and very few economies satisfy the no-bunching condition.

To implement the quasi-linear model we have extended previous work in order to characterize the optimal allocation under a set of less restrictive assumptions. In particular, we have removed the assumption of a lower bound on the utility of zero consumption and have removed that assumption that labor incomes can be negative. Using our characterization of the optimum we have conducted a Monte Carlo analysis. A random process has been used to select economies and the optimal income tax calculated. We have then studied the frequency with which alternative patterns of marginal and average tax rates arise. By doing so, we aimed to reveal which properties occur with a high frequency. We are not aware of such methodology having been previously applied to a problem in tax theory.

The results demonstrate that although the model is capable of generating a variety of qualitative patterns of marginal tax rates, there is a good consensus between alternative specifications of the average structure of the tax function. This average structure is characterized by a monotonic reduction in the proportion of economies for which the marginal rate of tax rises as the level of skill increases. Hence on average the picture emerges of a marginal rate of tax which declines with skill.

The behavior of the average rate of tax is different. At low levels of skill, the proportion of increases is increasing in skill until it reaches a maximum around the mean of the skill distribution. Decreases in the average rate of tax then occur for the majority of economies. This reveals the most common tax function to be one where the average rate of tax initially increases and then falls. The observation that the average rate of tax may fall at the upper end of the skill distribution is not a result that has previously been observed.



## 7 Appendix

### *Proof of Lemma 2*

The first part of the result follows directly from using (6). Utility is strictly increasing in  $x$  and skill is non-negative. Therefore, if the  $x_i$ s form a monotonic sequence, so must the  $z_i$ s. The second part of the result is a standard application of agent monotonicity.

### *Proof of Lemma 3*

(i) Assume that  $\frac{\beta_1}{n_1} < \frac{\beta_2}{n_2} < \dots < \frac{\beta_k}{n_k}$  and  $\beta_2 > 0$ . Using  $u'' < 0$ , the fact that  $\frac{\beta_2}{n_2} < \dots < \frac{\beta_k}{n_k}$  implies there exists a strictly increasing sequence of  $x_j$ s that solve

$$\frac{\beta_j}{n_j} u'(x_j) - \Theta = 0, \quad j = 2, \dots, k. \quad (38)$$

Moreover, since  $\beta_2 > 0$  it follows from Assumption 3 that  $x_2 > 0$ . Finally, since  $\frac{\beta_1}{n_1} < \frac{\beta_2}{n_2}$  it must be the case that  $x_2 > x_1 \geq 0$ .

(ii) Assume that there is no bunching so  $x_{j+1} > x_j$ ,  $j = 1, \dots, k-1$  and  $x_1 \geq 0$ . The consumption allocation must satisfy the necessary conditions for the program

$$\frac{\beta_1}{n_1} u'(x_1) - \Theta + \frac{\nu_1}{n_1} = 0, \quad \nu_1 x_1 = 0, \quad \nu_1 \geq 0, \quad (39)$$

$$\frac{\beta_j}{n_j} u'(x_j) - \Theta = 0, \quad j = 2, \dots, k. \quad (40)$$

Since  $u'' < 0$ , it follows that  $\frac{\beta_{j+1}}{n_{j+1}} > \frac{\beta_j}{n_j}$ . Furthermore, since  $x_2 > x_1 \geq 0$ ,  $\beta_2 > 0$ .

### *Proof of Proposition 1*

It follows from the Lemma 3 that if there is no bunching then  $\frac{\beta_1}{n_1} < \frac{\beta_2}{n_2} < \dots < \frac{\beta_k}{n_k}$ . These inequalities and the assumption that  $\beta_1 > 0$  imply  $\nu_j = 0 \forall j$ . The necessary conditions (16) then reduces to

$$\frac{\beta_j}{n_j} u'(x_j) - \Theta = 0, \quad j = 1, \dots, k. \quad (41)$$

Substituting from this condition into (4), the marginal tax rate at skill level  $j$  can be derived as

$$MTR_j = 1 - \frac{\beta_j}{\Theta n_j s_j}. \quad (42)$$

*Proof of Lemma 4*

If  $x_i = 0$  for  $i = 1, \dots, j$  the allocation must solve

$$\beta_1 u'(0) - n_1 \Theta + \nu_1 - \nu_2 = 0, \quad (43)$$

$$\beta_i u'(0) - n_i \Theta + \nu_i - \nu_{i+1} = 0, \quad i = 2, \dots, j-1, \quad (44)$$

$$\beta_j u'(0) - n_j \Theta + \nu_j = 0, \quad (45)$$

with  $\nu_i > 0$  for  $i = 1, \dots, j$ . Solving (43) to (45)

$$\nu_i = \sum_{\ell=i}^j [n_\ell \Theta - \beta_\ell u'(0)]. \quad (46)$$

Clearly,  $\nu_i > 0$  if  $\frac{\beta_1}{n_1} < 0, \dots, \frac{\beta_j}{n_j} < 0$ , hence these will be bunched.

But  $j+1$  cannot be bunched given the assumption on  $u'(0)$  since  $\nu_{j+1} = [n_{j+1} \Theta - \beta_{j+1} u'(0)] > 0$  would imply  $\beta_{j+1} \leq 0$ .

*Proof of Proposition 2*

Under Assumption 1,  $\lim_{x_i \rightarrow 0} \frac{1}{u'(x_i) s_i} = 0$ , so  $\lim_{x_i \rightarrow 0} MTR_i = 1$ .

*Proof of Lemma 5*

Consumer  $j+m+1$  is bunched with  $j, \dots, j+m$  if and only if there is a solution of the form  $\{\tilde{x} > 0, \nu_{j+1} > 0, \dots, \nu_{j+m+1} > 0, \nu_{j+m+2} \geq 0\}$  to the system

$$\beta_j u'(\tilde{x}) - n_j \Theta - \nu_{j+1} = 0, \quad (47)$$

$$\beta_{j+\ell} u'(\tilde{x}) - n_{j+\ell} \Theta + \nu_{j+\ell} - \nu_{j+\ell+1} = 0, \quad \ell = 1, \dots, m \quad (48)$$

$$\beta_{j+m+1} u'(\tilde{x}) - n_{j+m+1} \Theta + \nu_{j+m+1} - \nu_{j+m+2} = 0 \quad (49)$$

Here it is assumed that consumer  $j$  is not bunched with consumer  $j-1$ . Solving the equations

$$\nu_{j+m+1} = \nu_{j+m+2} + \frac{\Theta n_{j+m+1}}{\sum_{t=0}^m \beta_{j+t}} \left[ \sum_{t=0}^m \beta_{j+t} - \frac{\beta_{j+m+1}}{n_{j+m+1}} \sum_{t=0}^m n_{j+t} \right]. \quad (50)$$

Hence

$$\frac{\beta_{j+m+1}}{n_{j+m+1}} \leq \frac{\sum_{t=0}^m \beta_{j+t}}{\sum_{t=0}^m n_{j+t}} \quad (51)$$

is a sufficient condition for  $\nu_{j+m+1} \geq 0$ . However, it is not a necessary condition: it is possible that  $\nu_{j+m+1} > 0$  when (51) does not hold, provided that  $\nu_{j+m+2}$  is strictly positive and large enough. In other words, when condition (51) does not hold either (i) consumer  $j+m+1$  is not bunched with consumers  $j, \dots, j+m$ , or (ii) consumer  $j+m+1$  is bunched with consumers  $j, \dots, j+m$  and  $j+m+2$  (and, perhaps,  $j+m+3, \dots$ ).

*Proof of Proposition 3*

The allocation for bunched consumers is defined by the solution to (47) to (49). Summing these conditions the optimal consumption level,  $\bar{x}$ , must satisfy

$$\sum_{i=j}^{j+m+1} \beta_i u'(\bar{x}) - \sum_{i=j}^{j+m+1} n_i \Theta = 0. \quad (52)$$

The tax rate can then be calculated as

$$MTR_i = 1 - \frac{\sum_{i=j}^{j+m} \beta_i}{s_i \Theta \sum_{i=j}^{j+m} n_i}. \quad (53)$$

*Proof of Theorem 1*

Substitute the constraints  $\hat{z}_i = 0$  for  $i = 1, \dots, t$  in the objective function (22) and take into account that (i) the system of incentive compatibility constraints plus the resource constraints now has  $k$  equations for  $k - t$  variables  $z_{t+1}, \dots, z_k$  and (ii) the incentive compatibility constraints (6) together with the condition (35) of the theorem imply  $x_1 = \dots = x_t < x_{t+1}$ . This imposes  $t$  restrictions on  $x_1, \dots, x_k$ . Without loss of generality we may assume  $x_1, \dots, x_t$  to be a function of  $x_{t+1}, \dots, x_k$ .

The derivatives of the resulting Lagrangean with respect to  $x_j$ ,  $j = t+1, \dots, k$ , are:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_j} &\equiv \mu_j n_j u'(x_j) - \sum_{r=t+1}^k \frac{\mu_r n_r}{s_r} \frac{\partial z_r}{\partial x_j} + \nu_j - \nu_{j+1} \\ &+ \left[ \sum_{i=1}^t \left( \mu_i n_i u'(x_i) - \sum_{r=t+1}^k \frac{\mu_r n_r}{s_r} \frac{\partial z_r}{\partial x_i} + \nu_i - \nu_{i+1} \right) \frac{\partial x_i}{\partial x_j} \right] = 0. \end{aligned}$$

Hence, the solution satisfies

$$\begin{aligned} \mu_j n_j u'(x_j) &= \sum_{r=t+1}^k \frac{\mu_r n_r}{s_r} \frac{\partial z_r}{\partial x_j} + \nu_j - \nu_{j+1} \\ &- \sum_{i=1}^t \left( \mu_i n_i u'(x_i) - \sum_{r=t+1}^k \frac{\mu_r n_r}{s_r} \frac{\partial z_r}{\partial x_i} + \nu_i - \nu_{i+1} \right) \frac{\partial x_i}{\partial x_j} \quad (54) \end{aligned}$$

with  $x_1 = \dots = x_t = \eta(x_{t+1}, \dots, x_k)$ . Now we will show that (54) can also be obtained as a solution of P1 under (35).

For this we rewrite the necessary conditions (24) of P1 as

$$\mu_i n_i u'(x_i) - \sum_{r=t+1}^k \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_i} + \nu_i - \nu_{i+1} = \sum_{r=1}^t \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_i} \quad (55)$$

for  $i = 1, \dots, k$ . Next, we take (55) for  $i = 1, \dots, t$  and multiply each of these  $t$  equations by  $\frac{\partial x_i}{\partial x_j}$  for  $j = t+1, \dots, k$ . This results in  $t(k-t)$  equations. Now we group these equations by  $j$  having therefore a group of  $t$  equations,  $i = 1, \dots, t$ , for each  $j = t+1, \dots, k$ . Next, we sum up left-hand sides and right-hand sides of  $t$  equations in each group:

$$\begin{aligned} & \sum_{i=1}^t \left[ \mu_i n_i u'(x_i) - \sum_{r=t+1}^k \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_i} + \nu_i - \nu_{i+1} \right] \frac{\partial x_i}{\partial x_j} \\ &= \sum_{i=1}^t \sum_{r=1}^t \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_i} \frac{\partial x_i}{\partial x_j} \end{aligned}$$

Changing the order of double summation we obtain the following for the right-hand side of the last equation:

$$\begin{aligned} \sum_{i=1}^t \sum_{r=1}^t \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_i} \frac{\partial x_i}{\partial x_j} &= \sum_{r=1}^t \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \sum_{i=1}^t \frac{\partial z_r}{\partial x_i} \frac{\partial x_i}{\partial x_j} \\ &= - \sum_{r=1}^t \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_j} \end{aligned}$$

where the last equality follows from (35). Hence, from the first  $t$  equations from the solution (55) of P1 under (35) we obtain

$$\sum_{r=1}^t \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_j} = - \sum_{i=1}^t \left[ \mu_i n_i u'(x_i) - \sum_{r=t+1}^k \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_i} + \nu_i - \nu_{i+1} \right] \frac{\partial x_i}{\partial x_j}$$

for  $j = t+1, \dots, k$ , with  $x_i = \eta(x_{t+1}, \dots, x_k)$  for  $i = 1, \dots, t$ . Finally, we substitute the above in the last  $(k-t)$  equations from the solution (55) of P1 to obtain

$$\begin{aligned} \mu_j n_j u'(x_j) &= \sum_{r=1}^t \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_j} + \sum_{i=t+1}^k \frac{\mu_i n_i}{s_i} \frac{\partial z_i}{\partial x_j} - \nu_j + \nu_{j+1} \\ &= - \sum_{i=1}^t \left[ \mu_i n_i u'(x_i) - \sum_{r=t+1}^k \left( \frac{\mu_r n_r}{s_r} - \lambda_r \right) \frac{\partial z_r}{\partial x_i} + \nu_i - \nu_{i+1} \right] \frac{\partial x_i}{\partial x_j} \\ &\quad + \sum_{i=t+1}^k \frac{\mu_i n_i}{s_i} \frac{\partial z_i}{\partial x_j} - \nu_j + \nu_{j+1} \end{aligned} \quad (56)$$

which is exactly the same as (54).

The proof is completed by observing that forming a Lagrangean from (36) and the constraints gives the necessary conditions

$$\beta_{j,t} u'(x_j) - \Theta_t n_j + \nu_j - \nu_{j+1} + \sum_{i=1}^t (\beta_{i,t} u'(x_i) - \Theta_t n_i + \nu_i - \nu_{i+1}) \frac{\partial \eta}{\partial x_j} = 0. \quad (57)$$

Using (32) and (33) shows that (56) and (54) are identical.

## References

- Diamond, P. A. (1998). Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rate. *American Economic Review* 88, 83–95.
- Guesnerie, R. and J.-J. Laffont (1984). A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm. *Journal of Public Economics* 25, 329–369.
- Kanbur, S. and M. Tuomala (1994). Inherent inequality and the optimal graduation of marginal tax rates. *Scandinavian Journal of Economics* 96, 275–282.
- Lollivier, S. and J.-C. Rochet (1983). Bunching and second-order conditions: A note on optimal tax theory. *Journal of Economic Theory* 31, 392–400.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income tax. *Review of Economic Studies* 38, 175–208.
- Myles, G. D. (1995). *Public Economics*. Cambridge: Cambridge University Press.
- Myles, G. D. (2000). On the optimal marginal rate of income tax. *Economic Letters* 66, 113–119.
- Röell, A. (1985). A note on the marginal tax rate in a finite economy. *Journal of Public Economics* 28, 267–272.
- Weymark, J. A. (1986a). Bunching properties of optimal nonlinear tax problem. *Social Choice and Welfare* 3, 213–232.
- Weymark, J. A. (1986b). A reduced-form optimal income tax problem. *Journal of Public Economics* 30, 199–217.
- Weymark, J. A. (1987). Comparative statics properties of optimal nonlinear income taxes. *Econometrica* 55, 1165–1185.