

The simple economics of bunching

Jan Boone* and Lans Bovenberg[†]

Dept. of Economics

Tilburg University

P.O. Box 90153

5000 LE Tilburg

The Netherlands

19-2-2005

Abstract

This paper models unemployment as a binding non-negativity constraint on hours worked in a standard optimal non-linear income tax problem with quasi-linear preferences. We show that bunching of workers resulting from this binding constraint provides a more convincing description of the bottom of the labor market than bunching due to violation of the second-order condition for individual optimization. In particular, with the least skilled working zero hours, revenue requirements affect marginal tax rates, consumption, the bunching interval, and the marginal cost of public funds. Although a binding non-negativity constraint destroys the closed form solution of optimal marginal tax rates, the optimal tax problem can be characterized in a two-dimensional diagram in which comparative statics can be performed in straightforward fashion.

Keywords: non-linear income tax, bunching, unemployment, optimal taxation, quasi-linear preferences

JEL code: H 2, J 2

1. Introduction

With general preferences, optimal non-linear taxes are difficult to characterize. In order to obtain more intuition for the determinants of the optimal non-linear income tax, therefore, a substantial literature (see, e.g., Boadway, Cuff and Marchand (2000), Ebert (1992), Weymark

*CentER, TILEC, Netspar, Tilburg University, ENCORE, UvA, IZA and CEPR. email: j.boone@uvt.nl. Jan Boone gratefully acknowledges financial support from NWO, KNAW and VSNU through a Vernieuwingsimpuls grant.

[†]CentER, Netspar, TILEC, Tilburg University and CEPR. email: a.l.bovenberg@uvt.nl

(1986, 1987), and Lollivier and Rochet (1983)) has turned to quasi-linear preferences in leisure since these preferences allow for closed-form solutions of the standard optimal non-linear income tax problem. The resulting closed-form solutions can allow for bunching due to violation of the second-order condition for individual optimization. In this manner, bunching of unskilled workers in low-productivity jobs as well as positive marginal tax rates at the bottom of the labor market can be modelled. However, this particular way of representing the trade offs policymakers face in setting optimal tax policy for low skilled labor suffers from a number of disadvantages. First of all, it does not incorporate unemployment, even though in reality redistribution may discourage unskilled workers from working positive hours. Second, the revenue requirements of the government do not affect marginal tax rates, consumption, the bunching interval, and the shadow cost of public revenues. This does not seem a convincing description.

In view of these drawbacks, this paper elaborates on another type of bunching. This bunching occurs if the non-negativity constraint on hours worked (and gross incomes) is binding for some households. It necessarily occurs at the bottom of the labor market and implies that workers with least skills are in fact unemployed. Moreover, in the presence of this type of bunching, public revenue requirements do impact both the bunching interval and the structure of marginal tax rates.

Also Boadway, Cuff and Marchand (2000) introduce bunching due to a binding non-negativity constraint for hours worked in a model with quasi-linear preferences in leisure. We extend their analysis in three ways. Most importantly, we show that the solution can be characterized graphically by an upward sloping curve ('the government budget constraint') and a downward sloping curve ('labor supply'). Thus, although a binding non-negativity constraint on hours worked no longer allows for a closed-form solution, the solution can still be characterized rather simply while comparative statics can be conducted in a rather straightforward manner. In this way, our analysis helps to provide intuition for the shape of the optimal non-linear income tax, something that is difficult to come by in both optimal tax models with general preferences and the solutions with a binding non-negative constraint on hours worked in Boadway, Cuff and Marchand (2000). Indeed, a model with quasi-linear preferences in leisure but with a binding non-negativity constraint on hours worked strikes a balance between, on the one hand, analytical tractability and, on the other hand, a realistic model of the bottom of the labor market with more convincing comparative static results. As a second extension of Boadway, Cuff and Marchand (2000), we establish that second-order condition for individual optimization excludes some solutions. This allows us to sign marginal tax rates and impose an upper bound on the marginal utility cost of public funds. Finally, we incorporate more general social welfare functions by including so-called rank-order weights, which allow governments to attach higher welfare weights to the least skilled. This enables us to explore how the welfare weights of a non-utilitarian government impact the optimal tax structure.

The rest of this paper is structured as follows. Section 2 introduces the model and sets out the optimal tax problem. Section 3 characterizes the closed-form solution without bunching and provides intuition for this recursive solution in which public revenue requirements affect neither marginal tax rates nor the marginal cost of public funds. Section 4 turns to bunching on

account of violation of second-order incentive compatibility. It shows that this type of bunching does not address some of the unrealistic implications of the model without bunching. Section 5, therefore, introduces bunching originating in a binding non-negativity constraint on hours worked. This bunching not only provides a more realistic description of the bottom of the labor market but also allows for a rather simple, intuitive characterization of an optimal tax problem in which the least skilled workers are unemployed. Section 6 concludes. The appendix, finally, provides the proofs of the various lemmas and propositions contained in the main text.

2. The model

We consider an economy which is populated with agents featuring homogenous preferences but heterogeneous skills. A worker of ability (or skill or efficiency level) n working y hours (or providing y units of work effort) supplies ny efficiency units of homogeneous labor. With constant unitary labor productivity, these efficiency units are transformed in the same number of units of output. We select output as the numeraire. The before-tax wage per hour is thus given by exogenous skill n . Hence, overall gross output produced by a worker of skill n , $z(n)$, amounts to $z(n) = ny(n)$. Since workers collect only labor income, this gross output $z(n)$ corresponds to the gross (i.e. before-tax) labor income earned by a worker of that skill n . The density of agents of ability n is denoted by $f(n)$, which is differentiable. $F(n)$ represents the corresponding cumulative distribution function. The support of the distribution of abilities is given by $[n_0, n_1]$. In line with the optimal income tax literature, the government is assumed not to be able to observe skills n but to know the distribution function $f(n)$ and observe before-tax income of each individual $z(n)$.

Workers share the following quasi-linear utility function over consumption x and hours worked (or work effort) y

$$u(x, y) = v(x) - y,$$

where $v(x)$ is increasing and strictly concave: $v'(x) > 0, v''(x) < 0$ for all $x \geq 0$. Furthermore, $v(0) = 0, \lim_{x \downarrow 0} v'(x) = \infty$ and $\lim_{x \rightarrow +\infty} v'(x) = 0$. The concavity of $v(\cdot)$ implies that agents are risk averse and thus want to obtain insurance against the risk of a low earning capacity n . The specific cardinalization of the utility function affects the distributional preferences of a utilitarian government. In particular, the concavity of $v(\cdot)$ implies that a utilitarian government aims to fight poverty. In other words, such a government wants to insure agents against the risk of a low consumption level.

As in Lollivier and Rochet (1983), Weymark (1987), Ebert (1992), and Boadway, Cuff and Marchand (2000), utility is linear in work effort y and separable in work effort and consumption x . This has a number of important implications. First, consumption x is not affected by income effects. A higher average tax rate thus induces households to raise work effort y rather than to cut consumption x . Second, the single-crossing (or sorting) property is met, implying that the incentive compatibility constraints can be replaced by (much simpler) monotonicity conditions on $x(\cdot)$ and $z(\cdot)$ (see below). Third, the specific quasi-linear utility function allows for a closed-form solution of the standard optimal income tax problem. Fourth, a utilitarian government

cares only about aggregate work effort in the economy. Such a government thus aims at an equal distribution of consumption (i.e. the alleviation of poverty) rather than an equal distribution of work effort over the various agents.

A worker determines his work effort. Instead of working with work effort $y(n)$ and consumption $x(n)$ as the instruments of the worker, we write the utility function in terms of gross income (or output) $z(n) \equiv ny(n)$ and net income (or consumption) $x(n)$. Utility of type n is then written as $u(n) \equiv v(x(n)) - z(n)/n$. The utility of a type n agent is determined by type n 's choice of gross income z :

$$u(n) = \max_z \left\{ v\left(z - \tilde{T}(z)\right) - \frac{z}{n} \right\}, \quad (1)$$

where $\tilde{T}(z)$ denotes the tax schedule as a function of gross income z . We can write $T(n) = \tilde{T}(z(n))$ since type n chooses gross income $z(n)$ in equilibrium. The envelope theorem yields the first-order incentive compatibility constraint

$$u'(n) = \frac{z(n)}{n^2}. \quad (2)$$

The following lemma shows that the second-order condition for the agents' optimal choice of consumption and gross income implies that consumption and gross income are non-decreasing in type n . The inequalities in the lemma are therefore called the second-order incentive compatibility constraints.

Lemma 1 *The second-order condition for individual optimization is satisfied if and only if*

$$\begin{aligned} z'(n) &\geq 0, \\ x'(n) &\geq 0, \end{aligned} \quad (3)$$

while $z'(n) = 0$ if and only if $x'(n) = 0$.

As a last constraint on individual optimization, labor supply and therefore before-tax income should be non-negative:

$$z(n) \geq 0. \quad (4)$$

The government maximizes a weighted average of agents' utility:

$$W \equiv \int_{n_0}^{n_1} u(n)f(n)\phi(n)dn.$$

We normalize the rank-order weights $\phi(n)$ such that $\int_{n_0}^{n_1} f(n)\phi(n) = 1$, and assume $\phi'(n) \leq 0$.¹ The government is utilitarian if the rank-order weights are constant, i.e. $\phi(n) = 1$ for all n .

¹The rank-order weights depend on ability n rather than utility $u(n)$. This approach, which involves non-welfarists elements, allows us to derive a closed-form solution for the standard optimal tax problem. Atkinson (1995) defends this assumption by noting that empirical measures of inequality are based on the distribution of gross wages n rather than utilities.

This is the usual assumption adopted in the literature on optimal non-linear income taxation in the presence of preferences that are quasi-linear in leisure (see Lollivier and Rochet (1983), Weymark (1987), Ebert (1992), and Boadway, Cuff, and Marchand (2000)²). If the welfare weights are declining (i.e. $\phi'(n) < 0$), the government is concerned about the distribution of not only consumption but also leisure (or work effort).

The government has to respect the following budget constraint

$$\int_{n_0}^{n_1} f(n) T(n) dn = E, \quad (5)$$

where E represents exogenously given exhaustive government expenditure, and $T(n) \equiv z(n) - x(n)$ denotes the tax paid by type n .

In optimizing social welfare, the government faces four constraints: the first-order and second-order incentive compatibility constraints (2) and (3), the non-negativity constraint on gross incomes (4), and the government budget constraint (5). Instead of $x(n)$, we employ $u(n)$ as a control variable in order to facilitate the inclusion of first-order incentive compatibility (2) into our optimization problem. To incorporate the second-order incentive compatibility constraints, we introduce a non-negative variable $\omega(n) \equiv z'(n)$ determining how fast $z(n)$ rises with ability n . We thus arrive at the following optimization problem

$$\max_{\substack{u(\cdot), z(\cdot), \\ \omega(\cdot) \geq 0}} \int_{n_0}^{n_1} \left\{ \begin{array}{l} u(n)\phi(n)f(n) - \lambda_u(n) \left[u'(n) - \frac{z(n)}{n^2} \right] \\ - \lambda_z(n) [z'(n) - \omega(n)] \\ + \lambda_E f(n) [T(n) - E] - \delta(n) [0 - z(n)] \end{array} \right\} dn, \quad (6)$$

where $T(n) \equiv z(n) - x(n) = z(n) - v^{-1} \left(u(n) + \frac{z(n)}{n} \right)$. $\lambda_u(n)$ and $\lambda_z(n)$ represent the Lagrange multipliers of the first-order and second-order incentive compatibility constraints, λ_E stands for the multiplier of the government budget constraint, and $\delta(n)$ is the Lagrange multiplier of the non-negativity constraint on before-tax income.

The shadow values $\lambda_z(n)$ and $\delta(n)$ are associated with the two types of bunching considered here. $\lambda_z(n) < 0$ (implying $\omega(n) \equiv z'(n) = 0$) corresponds to the case in which $z(n)$ and $x(n)$ are constant over a range of skills. We call this *bunching due to violation of monotonicity*. Also the case $\delta(n) > 0$ implies that gross and net incomes are constant over a range of skills. In contrast to bunching on account of violation of monotonicity, however, gross incomes $z(n)$ are necessarily zero over this range so that utility is constant over the bunching interval (see (2) with $z(n) = 0$). This is called *$z = 0$ bunching*.

²The latter paper considers also a maxi-min objective function where the government cares only about the least able persons (i.e. agents with skill n_0). This is the special case of our formulation in which $\phi(n) = 0$ for $n > n_0$.

3. The case without bunching

As a benchmark, this section characterizes the case without bunching and provides intuition for the recursive, closed-form solution.

Throughout the paper, we use the function $G(\cdot)$ defined as follows

$$G(n) = \int_{n_0}^n \frac{\phi(t)}{t} f(t) dt.$$

Without any bunching, the optimal solution is characterized as follows.³

Lemma 2 *If $z(n_0) > 0$ and $\lambda_z(n) = 0$ for all $n \in [n_0, n_1]$, the solution to maximization problem (6) satisfies*

$$\lambda_E = G(n_1), \tag{7}$$

$$\tau(n) = \frac{\frac{G(n)}{G(n_1)} - F(n)}{nf(n)} \geq 0 \text{ for all } n \in [n_0, n_1], \tag{8}$$

$$v'(x(n)) = \frac{1}{n(1 - \tau(n))}, \tag{9}$$

$$u(n) = \frac{1}{n} \left(K - E + \int_{n_0}^n v(x(t)) dt \right), \tag{10}$$

$$z(n) = n(v(x(n)) - u(n)) = nv(x(n)) - \int_{n_0}^n v(x(t)) dt + E - K, \tag{11}$$

$$W = u(n_0)n_0G(n_1) + \int_{n_0}^{n_1} [G(n_1) - G(n)]v(x(n))dn, \tag{12}$$

where

$$K \equiv \int_{n_0}^{n_1} \{[tf(t) - (1 - F(t))]v(x(t)) - x(t)f(t)\} dt,$$

and the marginal tax rate for type n is defined as

$$\tau(n) \equiv \left. \frac{d\tilde{T}(z)}{dz} \right|_{z=z(n)}.$$

The marginal utility cost of government revenue λ_E depends only on the distribution of skills and the social welfare weights $\phi(n)$. The spending requirement E does not affect it. The reason is that utility is linear in work effort so that marginal utility costs do not rise if a higher level of government spending induces agents to work harder. More precisely, a uniform tax on all agents acts like a lump-sum tax, which yields only income effects and no substitution effects.

³This solution is found by assuming that the second-order condition $x'(n) \geq 0$ and the non-negativity constraint $z(n) \geq 0$ are met. If the solution implied by (9) violates (one of) these constraints, the solutions provided in the next sections become relevant.

With quasi-linear preferences, only labor supply responds to income effects. Hence, raising one additional euro of tax from each agent induces all agents to raise their gross incomes by a euro, while net incomes are unaffected. Since preferences are linear in leisure, the private utility costs of one additional unit of gross income do not depend on the level of leisure, but are inversely proportional to the skill level, $1/n$. Indeed, extracting a euro from a higher skilled agent imposes a lower effort cost than extracting the same euro from a lower skilled agent. The aggregate welfare effect on the social objective function, λ_E , corresponds to the weighted population average of these private welfare costs, i.e. $\lambda_E = \int_{n_0}^{n_1} \frac{\phi(n)}{n} f(n) dn = G(n_1)$.

With λ_E depending only on the distribution of skills and the social welfare weights $\phi(n)$, also the marginal tax rates can be written in terms of these elements only. Rewriting equation (8) while using $\lambda_E = G(n_1)$, we obtain the following expression for the marginal tax rate:

$$\lambda_E \tau(n) n f(n) = \lambda_E (1 - F(n)) - (G(n_1) - G(n)). \quad (13)$$

The marginal tax rate at each skill is determined by trading off the efficiency gains of a lower marginal tax rate and the distributional costs of a more dispersed income distribution. More specifically, consider an increase of one unit of work effort by type n (i.e. $dy(n) = 1$), while keeping type n 's utility constant. With taxation driving a wedge between the social and private marginal value of work, more work effort generates additional government revenues $\tau(n)n$. Multiplying this with the utility value of government funds, λ_E , and the number of type n agents, $f(n)$, we arrive at the efficiency gain at the left-hand side of (13).

The right-hand side of this equation measures the distributional costs of higher work effort of type n . In particular, with agents of skill n earning higher gross incomes (at the same utility level), higher ability agents find it more attractive to mimic type n . To prevent these substitution effects, the government has to raise utility of all workers who are more skilled than type n by reducing gross incomes with one unit.⁴ The right-hand side of (13) stands for the costs in terms of the required additional government revenue $\lambda_E(1 - F(n))$ minus the utility benefits of the agents involved ($G(n_1) - G(n)$).

Expression (8) implies that marginal tax rates at the top and the bottom are zero (i.e. $\tau(n_0) = \tau(n_1) = 0$), while these rates are positive at interior skills (i.e. $\tau(n) > 0$ for $n_0 < n < n_1$).⁵ Two factors determine marginal tax rates in the interior. The first factor, the distributional benefits of a higher marginal tax rate (represented by the term $[G(n)/G(n_1) - F(n)]$) raises the marginal tax rate. This term is maximal at the unique 'critical' skill level n_c at which the welfare weight $\phi(n_c)/n_c$ equals the population average of these welfare weights

⁴The following two steps show that $z(\cdot)$ has to fall by one unit for all types above n in order to keep the incentive compatibility constraints satisfied. First, note that $dy(n) = 1$ implies $dz(n) = n$. Hence, in view of $u'(n) = \frac{z(n)}{n^2}$, $u(n)$ has to increase with an *additional* $\frac{1}{n}$ for the type slightly above n (i.e. $du'(n) = \frac{dz(n)}{n^2} = \frac{1}{n}$ with some abuse of notation). This is achieved by reducing z by one unit for the type slightly above n . Second, as regards the incentive compatibility constraints for all other types above n , a uniform decrease in gross incomes $z(n)$ with one unit for all $t > n$ leaves all these constraints $v(x(t)) - z(t)/t \geq v(x(t')) - z(t')/t$ (for $t', t > n$) unaffected. Hence, such a uniform decrease in z does not result in any substitution effects for types $t > n$.

⁵This is because the weight $\frac{\phi(n)}{n}$ is a declining function of n so that the average of these weights over the interval $[n_0, n]$, $G(n)/F(n)$, exceeds the average of these weights over the interval $[n_0, n_1]$, $G(n_1)$. $G(n)/F(n) > G(n_1)$ implies that the numerator of (8) is positive.

$\lambda_E = \int_{n_0}^{n_1} \frac{\phi(n)}{n} f(n) dn$. The government wants to redistribute resources to all agents below this critical skill level. The second factor determining the marginal tax rate is the productive capacity of agents at type n , $nf(n)$, in the denominator of (8). The higher this productive capacity, the larger are the efficiency costs associated with a high marginal tax rate and therefore the lower the marginal tax rate should be.⁶

Consumption of each skill depends only on the marginal wage rate $n(1 - \tau(n))$ (see (9)) and not on government spending requirements E , as consumption depends only on substitution effects and all income effects go into work effort. Expression (11) implies that additional public spending requirements are optimally financed by uniformly increasing gross incomes of all agents, i.e. $dz(n)/dE = 1$. The system is thus recursive. Consumption, marginal tax rates, and the marginal costs of public funds are determined independently from public spending, which affects work effort $z(n)$ required to meet resource and incentive constraints.

The rank-order weights do impact marginal tax rates, consumption levels and gross incomes. To derive the effects of changes in these weights, we consider the following family of rank-order weights indexed by $\alpha \in \mathbb{R}_+$.

Definition 1 *A rise in α is said to make the government more redistributive if*

$$\frac{d \ln \phi_\alpha(n)}{d\alpha} \text{ is decreasing in } n \quad (14)$$

while the normalization of the rank-order weights (i.e. $\int_{n_0}^{n_1} \phi_\alpha(n) f(n) dn = 1$) is maintained

$$\frac{d \left(\int_{n_0}^{n_1} \phi_\alpha(n) f(n) dn \right)}{d\alpha} = 0$$

As α goes up, the social planner increases the weight attached to low skilled agents relative to higher skilled agents, i.e.

$$\frac{d(\phi_\alpha(n) / \phi_\alpha(\bar{n}))}{d\alpha} = (\phi_\alpha(n) / \phi_\alpha(\bar{n})) \left\{ \frac{d \ln \phi_\alpha(n)}{d\alpha} - \frac{d \ln \phi_\alpha(\bar{n})}{d\alpha} \right\} > 0 \text{ if } n < \bar{n}.$$

The following two families of rank-order weights satisfy this definition. First, $\phi_\alpha(n) = \begin{cases} 1 + \frac{\alpha}{F(\bar{n})} & \text{if } n < \bar{n} \\ 1 - \frac{\alpha}{1-F(\bar{n})} & \text{if } n \geq \bar{n} \end{cases}$

for some value of $\bar{n} \in \langle n_0, n_1 \rangle$ and where $\alpha \in [0, 1 - F(\bar{n})]$. In this case, a higher weight α raises the weight attached to agents below skill \bar{n} at the expense of agents above \bar{n} . Second, define $n(\alpha) = n_1 - \alpha(n_1 - n_0)$ for $\alpha \in [0, 1]$. Then we consider the following family of rank-order weights $\phi_\alpha(n) = \begin{cases} \frac{1-\varepsilon}{F(n(\alpha))} & \text{if } n \in [n_0, n(\alpha)) \\ \frac{\varepsilon}{1-F(n(\alpha))} & \text{if } n \in [n(\alpha), n_1] \end{cases}$ for $\varepsilon > 0$ close to 0. A rise in α reduces the number of skills $n(\alpha)$ to which the government attaches a high rankorder weight, thereby increasing the level of that weight.

The following lemma characterizes the effects of a more redistributive government.

⁶For particular skill distributions, Diamond (1998) analytically investigates the shapes of the marginal tax rates for preferences that are quasi-linear in consumption. Boadway, Cuff and Marchand (2002) conduct a similar analysis for the quasi-linear preferences explored in our paper.

Lemma 3 *An increase in α reduces consumption $x(n)$ and increases the marginal tax rate $\tau(n)$ for all types $n \in [n_0, n_1]$. Moreover there exists $\tilde{n} \in \langle n_0, n_1 \rangle$ such that utility $u(n)$ increases for all $n \in [n_0, \tilde{n})$ while it decreases for all $n \in \langle \tilde{n}, n_1]$.*

A more redistributive government raises marginal tax rates in order to redistribute more resources from the high skilled to the low skilled. The substitution effects associated with the higher marginal tax rates reduce consumption for all workers. Nevertheless, the least skilled benefit from more redistribution. In particular, for these workers, lower working hours more than compensate for lower consumption levels.

4. Bunching due to violation of monotonicity constraint

This section shows that the closed-form solution survives violation of second-order incentive compatibility. It generalizes the results in Ebert (1992) and Boadway, Cuff, and Marchand (2000) to a non-utilitarian government. This section also demonstrates that bunching on account of violation of second-order incentive compatibility neither offers a good description of inactivity at the bottom of the labor market nor provides convincing comparative static results.

As shown by Guesnerie and Laffont (1984) and Ebert (1992), the violation of the monotonicity condition on consumption $x(n)$ makes bunching optimal, that is, $z(n)$ and $x(n)$ are constant over a range of skills. Whereas Ebert (1992) and Fudenberg and Tirole (1991) include $\omega(n) \geq 0$ as an additional restriction with a Lagrange multiplier in the objective function, this is strictly speaking not necessary. The optimality conditions of Pontryagin (see, for instance, Kamien and Schwartz (1981)) imply that $\omega(n)$ is chosen for each n to solve

$$\max_{\omega} \lambda_z(n) \omega$$

From this it follows immediately that $\lambda_z(n) < 0$ implies $\omega(n) = 0$. Further, $\omega(n) > 0$ (and finite) can only happen if $\lambda_z(n) = 0$. Finally, if the problem is well defined then the case $\lambda_z(n) > 0$ can be excluded as it would imply $\omega(n) = +\infty$.

With $z'(n) \equiv \omega(n) = 0$, lemma 1 implies that $x'(n) = 0$. A so-called "ironing out" procedure yields the range of skills over which (gross and net) incomes are constant. The violation of monotonicity is due to rapidly rising marginal tax rates. Indeed, rising marginal tax rates are a necessary condition for this type of bunching to occur. Since the marginal tax rate is necessarily declining at the top, this bunching is not possible at the highest skill levels (see Ebert (1992)).

In contrast to $z = 0$ bunching, this type of bunching can happen not only at the bottom but also in the interior of the skill distribution. Since our analysis focuses on the bottom of the labor market, we do not consider bunching due to violation of monotonicity in the interior of the skill distribution.⁷ The following proposition generalizes the results in Boadway, Cuff, and Marchand (2000) to a non-utilitarian government.

⁷Generalizing the equations below for bunching at the bottom to the case of bunching in the interior is straightforward (see, e.g., Ebert (1992) or Fudenberg and Tirole (1991)).

Proposition 1 *The consumption path implied by equation (9) is non-monotone at the bottom (i.e. $x'(n_0) < 0$) if and only if⁸*

$$2n_0/\phi(n_0) < 1/\left[\int_{n_0}^{n_1} \frac{f(t)\phi(t)}{t} dt\right] = 1/G(n_1).$$

In this case, types $n \in [n_0, n_b]$ feature the same consumption level x_b and production level z_b with

$$\begin{aligned} x_b &= x(n_b), \\ z_b &= z(n_b), \end{aligned}$$

where n_b is determined by the following equation

$$f(n_b) \left[\frac{\bar{F}(n_b)}{F(n_b)} - \frac{G(n_b)n_b}{F(n_b)} \right] = G(n_1)F(n_b) - G(n_b), \quad (15)$$

with $\bar{F}(n_b) \equiv \int_{n_0}^{n_b} \phi(n)f(n)dn$. We find that

$$\tau(n_b) > 0$$

and consumption for types $n \in [n_b, n_1]$ is determined by

$$v'(x(n)) = \frac{f(n)}{\frac{G(n_1)-G(n)}{\lambda_E} + f(n)n - [1 - F(n)]}, \quad (16)$$

where

$$\lambda_E = G(n_1).$$

Production of types $n \in [n_b, n_1]$ is determined by

$$z(n) = nv(x(n)) - n_bv(x(x_b)) - \int_{n_b}^n v(x(t)) dt + E - K,$$

where

$$\begin{aligned} K &\equiv \int_{n_0}^{n_b} \{[tf(t) - [1 - F(t)]]v(x_b) - f(t)x_b\} dt + \\ &\int_{n_b}^{n_1} \{[tf(t) - [1 - F(t)]]v(x(t)) - f(t)x(t)\} dt. \end{aligned}$$

⁸To find the effects of a more redistributive government, we write this condition as $\int_{n_0}^{n_1} \frac{f(t)}{t} \frac{\phi_\alpha(t)}{\phi_\alpha(n_0)} dt < \frac{1}{2n_0}$, where $\phi_\alpha(n)$ is defined in definition 1. Since an increase in α reduces $\frac{\phi_\alpha(t)}{\phi_\alpha(n_0)}$ for $t > n_0$, a more redistributive government makes it more likely that this condition is satisfied. Intuitively, as the government becomes more redistributive, it tries to raise the utility levels of the low types (compared to high types) to such an extent that it is more likely to violate the second-order incentive constraints.

This type of bunching does not affect the marginal utility cost of government revenue λ_E as given by (7): a higher level of government spending is still optimally financed through a uniform increase in $z(n)$ by all agents. The marginal utility cost of government spending therefore continues to correspond to the average utility costs of the associated increase in work effort over the entire population. With the same marginal utility cost of public funds λ_E , the marginal tax rates (8) and the consumption path (9) in the non-bunched intervals are not affected by bunching. Accordingly, with bunching occurring at the bottom of the income distribution, the marginal tax rate faced by the lowest skilled agent who is not bunched, $n_b > n_0$, is positive (i.e. $\tau(n_b) > 0$). This is in contrast to the case without bunching, where the lowest worker n_0 faces a zero marginal tax rate. Intuitively, a positive marginal tax rate for the lowest non-bunched worker generates positive distributional effects only if it redistributes resources towards bunched workers $n < n_b$, who feature the lowest consumption levels.

4.1. comparative statics

Bunching due to violation of monotonicity yields strictly positive marginal tax rates at the bottom of the labor market and features the following comparative statics properties.

Lemma 4 *In case of bunching due to violation of monotonicity, an increase in the public spending requirement E yields the following effects:*

$$\begin{aligned}\frac{d\lambda_E}{dE} &= 0, \\ \frac{dn_b}{dE} &= 0, \\ \frac{dx(\cdot)}{dE} &= 0, \\ \frac{dz(\cdot)}{dE} &= 1.\end{aligned}$$

Higher public spending leaves marginal tax rates, and hence consumption levels, unaffected. Since the bunching interval $[n_0, n_b]$ is completely determined by the skill distribution and the function $\phi(\cdot)$ (see equation (15)), the level of public spending does not impact the size of the bunching interval, either. These comparative static results do not seem to be particularly realistic. We therefore turn to a model in which the non-negativity constraint in hours worked is binding.

5. Bunching with zero work effort

Redistributive tax and transfer systems may discourage those with little skills from working. Indeed, unemployment rates are highest at the bottom of the labor market. Unemployment can be modelled in the context of an optimal tax model as the low skilled optimally choosing to work zero hours. With the non-negativity constraint on hours worked being binding, gross

and net incomes are constant over a range of skills. More precisely, gross incomes $z(n)$ are zero over this range. This, together with (2), implies that also utility is constant over the bunching interval. Moreover, second-order incentive compatibility $z'(n) \geq 0$ implies that this bunching can occur only at the bottom of the income distribution. Accordingly, a skill level n_z exists so that $z(n) = 0$ for $n \in [n_0, n_z]$.

Since $z'(n) \geq 0$ (see lemma 1), $z(n_0) \geq 0$ implies that the non-negativity constraint on gross incomes is satisfied also for all other skills $n > n_0$. We can thus formulate the government's maximization problem (6) as

$$\begin{aligned} \max_{\substack{n_z, x_z, u(\cdot), \\ z(\cdot), \omega(\cdot) \geq 0, \\ z(n_z) \geq 0}} \int_{n_z}^{n_1} \left\{ \begin{array}{l} u(n)\phi(n)f(n) - \lambda_u(n) \left[u'(n) - \frac{z(n)}{n^2} \right] \\ -\lambda_z(n) [z'(n) - \omega(n)] + \lambda_E [f(n)T(n)] \end{array} \right\} dn \quad (17) \\ + \bar{F}(n_z)v(x_z) + \lambda_E E. \end{aligned}$$

Proposition 2 *If the solution in lemma 2 implies $z(n_0) < 0$, then the solution to problem (17) can be characterized as follows. First, a non-empty interval $[n_0, n_z]$ exists such that*

$$\begin{aligned} z(n) &= 0, \\ x(n) &= x_z = x(n_z) \end{aligned}$$

for all $n \in [n_0, n_z]$, where $x(n)$ for $n \geq n_z$ is determined by equation (16). Furthermore, λ_E and n_z are determined by the following two equations⁹

$$\begin{aligned} \lambda_E &= \frac{\frac{\bar{F}(n_z)}{n_z} + G(n_1) - G(n_z)}{\frac{F(n_z)}{n_z v'(x(n_z, \lambda_E))} + 1 - F(n_z)}, \quad (18) \\ &= \frac{(1 - F(n_z)) n_z v(x(n_z, \lambda_E)) + E + F(n_z) x(n_z, \lambda_E)}{\int_{n_z}^{n_1} \{ [nf(n) - [1 - F(n)]] v(x(n, \lambda_E)) - f(n) x(n, \lambda_E) \} dn}. \quad (19) \end{aligned}$$

If $x(n)$ as determined by equation (16) is monotonically increasing in n , then equation (19) is upward sloping, equation (18) is downward sloping and

$$\begin{aligned} \lambda_E &\leq G(n_1), \\ \tau(n_z) &> 0. \end{aligned}$$

For skills $n \geq n_z$, consumption levels and marginal tax rates continue to be determined by equations (9) and (13). Unlike bunching due to the violation of monotonicity, $z = 0$ bunching does impact the marginal utility cost of government revenue λ_E . In particular, if

⁹The first equation has the advantage that it is easy to interpret (see below), but the right-hand side still contains λ_E in the $v'(x(n_z, \lambda_E))$ term. The appendix shows that this equation can be rewritten as $\lambda_E = \frac{\bar{F}(n_z) + (n_z - \frac{F(n_z)}{f(n_z)})(G(n_1) - G(n_z))}{n_z - F(n_z) \frac{1 - F(n_z)}{f(n_z)}}$ so that λ_E is written as an explicit function of n_z .

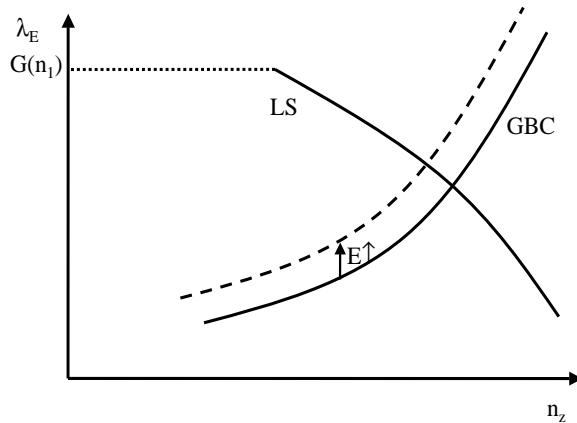


Figure 1: *Equilibrium in n_z, λ_E space: Labor Supply (LS) and Government Budget Constraint (GBC).*

government spending is reduced, this decreases work effort only outside the bunching interval (for skills $n > n_z$). Within the bunching interval, consumption is raised so that utility of the bunched agents increases with the same amount as the marginal worker n_z (see the first term in the numerator at the right-hand side of (18)). Lower government spending thus results in both reduced work effort and higher private consumption. For the constrained households, consumption is valued relatively less (the non-negativity constraint on work effort acts like an implicit subsidy on consumption, i.e. $v'(x(n_0)) < 1/n_0$). This explains why the marginal cost of public funds is lower with $z = 0$ bunching than without it (i.e. $\lambda_E \leq G(n_1)$).

The marginal tax rate facing the least skilled worker n_z is positive (i.e. $\tau(n_z) > 0$). The reason is that a positive tax rate for the lowest skilled worker yields positive distributional effects because it redistributes resources from the productive workers (i.e. the skills $n > n_z$) towards the non-productive workers, who feature the lowest consumption levels. With $z = 0$ bunching, marginal tax rates remain positive in the interior. However, $\lambda_E \leq G(n_1)$ and (13) imply that marginal tax rates at $n > n_z$ are smaller than without $z = 0$ bunching. Intuitively, the benefits of redistribution are smaller if low-skilled agents can use additional resources only to increase consumption (which yields less marginal utility than lower work effort does).

Figure 1 shows the equations (18) and (19) in (n_z, λ_E) space. To obtain more intuition about equation (18), we write it as follows

$$v'(x_z) \int_{n_0}^{n_z} \phi(n) f(n) dn = \lambda_E F(n_z) + n_z v'(x_z) [\lambda_E (1 - F(n_z)) - (G(n_1) - G(n_z))]. \quad (20)$$

Suppose the government considers a small increase in consumption for all unemployed workers $dx_z > 0$,¹⁰ then the left-hand side of this equation represents the increase in social welfare

¹⁰The government raises consumption (i.e. $dx_z > 0$ for given λ_E) through an increase in n_z . Note that the partial effect of n_z (taking x_z as given) on equation (20) is zero so that n_z affects the costs and benefits of redistribution only indirectly through its effect on $x(n_z)$.

corresponding to the higher consumption levels of the bunched least skilled workers who do not supply any labor. The right-hand side measures the costs of the higher consumption level of these workers. First, the higher consumption level of the $F(n_z)$ workers needs to be financed at the marginal cost of public funds λ_E . In addition, in order to maintain incentive compatibility, the gross incomes z of the types above n_z have to be reduced by $v'(x_z)n_z dx_z$.¹¹ This has to be multiplied by the welfare costs of raising gross incomes with one unit for all types above n_z . As argued in explaining equation (13), these costs are given by the term in square brackets on the right-hand side of (20).

We interpret equation (18) (or equivalently equation (20)) as labor supply because it equates the marginal benefits and marginal costs of reducing labour supply (in persons), i.e. increasing n_z . This labor-supply curve is downward sloping in (n_z, λ_E) space. As n_z goes up (for given λ_E), second-order incentive compatibility (i.e. $x'(n) > 0$) implies that consumption $x(n_z)$ increases for all inactive agents $[n_0, n_z]$. As a result of decreasing marginal utility ($v''(x) < 0$), the marginal utility benefit of more consumption for these workers $v'(x(n_z))$ declines. To ensure that these lower marginal benefits continue to equal the marginal costs of more consumption of these agents in (20), λ_E must fall. $z = 0$ bunching is essential in explaining why the benefits of redistribution, as measured by the left-hand side of (20) fall with n_z ; since $z(n)$ cannot be reduced for types $[n_0, n_z]$, the government can redistribute toward these types only by raising their consumption. This redistribution features decreasing returns because of the concavity of $v(x)$. Indeed, this concavity implies that the benefits of redistribution, as measured by λ_E in (18), decline with n_z .

Equation (19) is derived from the government budget constraint (5), which can be written as

$$-F(n_z)x(n_z) + \int_{n_z}^{n_1} T(n)f(n)dn = E.$$

This curve is upward sloping in (n_z, λ_E) space. The reason is as follows. As the bunching interval widens and n_z goes up, second-order incentive compatibility implies that $x(n_z)$ increases as well (for given λ_E). The higher consumption levels for the least skilled, non-working types have to be financed by the types above n_z . These latter, working types thus must face higher marginal tax rates and enjoy lower consumption levels. This requires an increase in λ_E (see (16)). Indeed, the cost of redistribution, as measured by λ_E implicit in equation (19), rise the number of agents who are unemployed (i.e. n_z).

5.1. comparative statics

This subsection performs comparative statics with respect to government expenditure E and the rank-order weights $\phi(\cdot)$. This illustrates that comparative static exercises remain rather

¹¹To see this, note that for the type t just above n_z , the incentive compatibility constraint can be written as $v(x(t)) - \frac{z(t)}{t} = v(x_z)$. At constant net income $x(t)$ constant, this equation still holds if $\frac{dz}{dx_z} = -n_z v'(x_z)$ for the type t just above n_z . As regards the incentive compatibility constraints for all other types above n_z , a uniform decrease in gross incomes $z(n)$ with one unit for all $t > n_z$ leaves all these constraints $v(x(t)) - z(t)/t \geq v(x(t')) - z(t')/t$ (for $t', t > n_z$) unaffected. Hence, such a uniform decrease in z does not result in any substitution effects for types $t > n_z$.

straightforward in a model of non-linear income taxation with unemployment. First, consider an increase in E . This shifts the government budget constraint (GBC) curve upwards while leaving the labour supply (LS) curve unchanged (see Figure 1). Hence, this figure reveals that the bunching interval narrows (i.e. n_z falls) while the marginal welfare cost of public funds λ_E rises. More precisely, we derive the following.

Lemma 5 *In case of $z = 0$ bunching, an increase in E yields the following effects*

$$\begin{aligned}\frac{dn_z}{dE} &< 0, \\ \frac{d\lambda_E}{dE} &> 0, \\ \frac{dx_z}{dE} &< 0, \\ \frac{dT(n)}{dE} &> 0\end{aligned}$$

for n close enough to n_z . We also find that

$$\begin{aligned}\frac{d\tau(n)}{dE} &> 0, \\ \frac{dx(n)}{dE} &< 0, \\ \frac{du(n)}{dE} &< 0.\end{aligned}$$

for all $n > n_z$

The negative income effect associated with the higher tax level that is required to fund the additional public spending raises labor supply, thereby reducing the number of agents who are non productive. Thus, unlike the case with bunching due to violation of monotonicity, more public spending reduces the size of the bunching interval. Moreover, in contrast to the case without $z = 0$ bunching, a higher level of government spending raises the marginal cost of public funds. Intuitively, with unemployed workers, the required resources to finance additional spending not only come from additional work effort of employed agents but also from lower consumption levels of the unemployed. The utility costs of reducing private consumption become higher if consumption of the unemployed is crowded out further as a result of more public spending. To contain these higher costs of lower consumption levels of the unemployed (as a result of lower tax credit $-T(n_z) = x_z$ for these skills), the government makes the tax system more redistributive at higher levels of public spending so that more of the required resources come from more work effort of skilled agents. In the face of the higher marginal tax burden associated with a more redistributive tax system, also employed workers $n > n_z$ reduce their consumption.¹² Hence, whereas the unemployed cut their consumption on account

¹²The comparative static results illustrate that the case with $z = 0$ bunching resembles the so-called rich economy if skills are observable to the government (see Boone and Bovenberg (2001)). In both cases, higher

of a higher average tax burden, the employed reduce their consumption because of a higher marginal tax burden.

Turning to the effects of a change in the rank-order weights as introduced in definition 1, we recall (see Lemma 3) that without bunching a more redistributive government reduces everyone's consumption level. Moreover, to raise utility at the lower end of the skill distribution, the least skilled agents work less. If these agents are unemployed, however, the government cannot raise their utility by having them work less.

Lemma 6 *If the government becomes more redistributive, consumption behaves as follows*

$$\begin{aligned}\frac{dx(n)}{d\alpha} &> 0 \text{ for } n \in [n_0, \tilde{n}), \\ \frac{dx(n)}{d\alpha} &< 0 \text{ for each } n \in \langle \tilde{n}, n_1 \rangle,\end{aligned}$$

for some $\tilde{n} \in \langle n_0, n_1 \rangle$. Next, there exists $n^* \in \langle \tilde{n}, n_1 \rangle$ such that

$$\begin{aligned}\frac{du(n)}{d\alpha} &> 0 \text{ for } n \in [n_0, n^*), \\ \frac{du(n)}{d\alpha} &< 0 \text{ for each } n \in \langle n^*, n_1 \rangle,\end{aligned}$$

Finally, if $n_z \geq \hat{n}$, where \hat{n} is defined as

$$\frac{d[G_\alpha(n_1) - G_\alpha(\hat{n})]}{d\alpha} = 0,$$

then a more redistributive government raises n_z , i.e.

$$\frac{dn_z}{d\alpha} > 0$$

In contrast to the case without $z = 0$ bunching (see lemma 3), with $z = 0$ bunching we obtain the intuitive result that a higher α produces more redistribution in terms of consumption. As α increases, inactive workers enjoy more consumption because the government cannot raise their utility by reducing their work effort. The types $n \in \langle \tilde{n}, n_1 \rangle$ pay for the higher welfare level of the least skilled and thus face higher marginal tax rates, which reduce their consumption levels.

To find the effect of α on unemployment (and thus n_z), consider the effect of α on the LS and GBC curves in figure 1. $n_z \geq \hat{n}$ is a sufficient condition for a more redistributive government (at a given level of λ_E) reducing consumption $x(n)$ (and raises marginal tax rates) for every type, including n_z . This implies that the GBC curve moves downwards and the labor supply

public spending yields not only higher work effort but also lower consumption for the non-productive individuals. Moreover, it reduces the number of these non-productive workers. Indeed, in both cases, all agents are searching for work (i.e. $n_w = n_0$). Furthermore, there is a skill level below which agents are not productive in their jobs and collect search subsidies $-T(n) - b > 0$.

curve rotates upward.¹³ In particular, the budget surplus resulting from **lower consumption and higher tax revenues** (at fixed n_z) allows for a decline in λ_E so that the GBC curve shifts downwards and to the right. As regards the LS curve (20), a more redistributive government reduces $x_z = x(n_z)$ for given n_z and λ_E . This raises marginal utility $v'(x_z)$. Together with a higher weight attached to the least skill workers, this increases the marginal benefit of higher consumption for the least skilled, non-working types $x_z = x(n_z)$. To re-establish equality between the benefits and costs of additional consumption for the non-working types in (20), the shadow value λ_E must increase so that a more redistributive government shifts the labor supply curve upward.

The shifts of the LS and GBC curves imply that a more redistributive government raises unemployment. Intuitively, a more redistributive government wants to increase utility of the least skilled. Since these agents are not working, this can be done only through raising consumption x_z of the marginal worker n_z . Under condition $n_z \geq \hat{n}$, however, consumption of the marginal worker would decline at fixed λ_E and n_z while lower consumption levels produce a fiscal surplus. Both to help the poor and to establish government balance, consumption of the least skilled is raised by (given that $x'(n) > 0$) widening the bunching interval.¹⁴

Our results contrast to Stähler (2002), who finds that higher government expenditure raises unemployment. The reason for our different results is that we allow for income effects in labor supply and endogenously determine the minimum standard of living. In particular, Stähler adopts a utility function that is linear in consumption. This implies that a utilitarian government does not exhibit a redistributive motive. To introduce a rationale for redistribution, Stähler introduces an exogenously given minimum standard of living M , below which the utility level of individuals is not allowed to fall. To encourage high-ability agents to work positive hours, the government must give them positive rents. As E increases, the government can no longer afford these rents for high ability types. Employment thus falls to reduce these rents. In our approach, in contrast, the minimum consumption level of unemployed agents is endogenously determined (our utility function is concave in consumption giving a direct rationale for redistribution). Hence, scarcer public funds result not only in a higher tax burden on higher skilled agents but also in a lower consumption level for unemployed agents (which indeed it does as $dx_z/dE < 0$). The lower minimum standard of living mitigates the effects of a higher level of government spending on incentives to work. In fact, with our utility function, the adverse income effects of the higher tax burden cause agents to reduce their work effort, thereby raising voluntary unemployment. In Stähler's framework with a utility function that is linear in consumption, in contrast, income effects on labor supply are absent.

¹³Since $v'(x(n_1)) = \frac{1}{n_1}$, equation (20) implies that α does not affect the labor-supply curve at $n_z = n_1$. Hence, the LS curve rotates round the point $(n_z, \lambda_E) = (n_1, \frac{1}{n_1})$.

¹⁴If $n_z < \hat{n}$, we cannot exclude that a higher α increases $x(n_z)$ for given n_z and λ_E to such an extent that n_z has to fall to satisfy the government budget constraint.

6. Conclusion

Our paper has contributed to the literature on bunching in mechanism design problems. Bunching is typically seen as technically rather difficult and lacking realism. Our paper suggests, in contrast, that bunching is relatively straightforward to deal with. Indeed, this type of bunching can be characterized in a two-dimensional diagram in which comparative statics can be performed rather easily. Moreover, bunching is realistic in some settings. More specifically, in optimal tax problems, bunching due to violation of the non-negativity constraint on hours worked is relevant because it provides a more realistic description of the bottom of the labor market than bunching on account of violation of second-order incentive compatibility. In particular, the least skilled are unemployed as their productivity level is not high enough to offset welfare benefits and search costs. Another advantage of this description of the labor market (compared to bunching as a result of second-order incentive compatibility considerations) is that public revenue requirements affect marginal tax rates, private consumption and the marginal cost of public funds. Moreover, a more redistributive government raises reduces the consumption levels of the poorest agents. This in contrast to a model without bunching.

7. References

- Atkinson, A.B., 1995, *Public Economics in Action. The Basic Income/Flat Tax Proposal* (Clarendon Press, Oxford).
- Boadway, R., K. Cuff, and M. Marchand, 2000, Optimal Income Taxation with Quasi-Linear Preferences Revisited, *Journal of Public Economic Theory*, Vol. 2, pp. 435-460.
- Boone, J., and A.L. Bovenberg, 2001, 'Unemployment vs. In-work Benefits with Search Unemployment and Observable Abilities,' Center Discussion Paper No. 77, Tilburg University, the Netherlands.
- Diamond, P.A., 1998, 'Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rate,' *American Economic Review*, Vol. 88, pp. 83-95.
- Ebert, U., 1992, 'A Reexamination of the Optimal Nonlinear Income Tax,' *Journal of Public Economics*, Vol. 49, pp. 47-73.
- Fudenberg, D., and J. Tirole, 1991, *Game Theory* (MIT Press, Cambridge).
- 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*, Vol. 25, pp 329-369.
- Kamien, M.I., and N.L. Schwartz, 1981, *Dynamic optimization: The calculus of variations and optimal control in economics and management* (North-Holland, Amsterdam).
- Lollivier, S., and J. Rochet, 1983, Bunching and Second-Order Conditions: A Note on Optimal Tax Theory, *Journal of Economic Theory*, Vol. 31, pp. 392-400.
- Mirrlees, J.A., 1971, An Exploration in the Theory of Optimal Income Taxation,' *Review of Economic Studies*, Vol. 38, pp. 175-208.

Stähler, F., 2002, Budget cuts, social assistance and voluntary unemployment, *Journal of Public Economic Theory*, Vol. 4 (4), pp. 573-579.

Weymark, J.A., 1986, A Reduced-Form Optimal Nonlinear Income Tax Problem, *Journal of Public Economics*, Vol. 30, pp. 199-217.

Weymark, J.A., 1987, Comparative Static Properties of Optimal Nonlinear Taxes, *Econometrica*, Vol. 55, pp. 1165-1185.

8. Appendix

Proof of lemma 1

We write

$$\begin{aligned} u(x(n), z(n), n) &= v(x(n)) - \frac{z(n)}{n}, \\ u(x(m), z(m), n) &= v(x(m)) - \frac{z(m)}{n}, \end{aligned}$$

and note that incentive compatibility implies

$$u(x(n), z(n), n) - u(x(m), z(m), n) \geq 0$$

for each $n \in [n_0, n_1]$ and each $m \in [n_0, n_1]$.

We look at the difference $u(x(n), z(n), n) - u(x(m), z(m), n)$ in two different ways. First, we fix n and let m vary and define

$$g(m) = \{u(x(n), z(n), n) - u(x(m), z(m), n)\}.$$

The consumption and gross income schedules $x(\cdot)$ and $z(\cdot)$ (and thereby the tax schedule $T(z(n))$) satisfy incentive compatibility, if an agent of type n finds it optimal to report $m = n$. That is, the function $g(\cdot)$ has a minimum at $m = n$. The first order condition for this minimum ($g'(m)|_{m=n} = 0$) can be written as

$$[u'_x x'(m) + u'_z z'(m)]_{m=n} = 0, \tag{21}$$

where $u'_x = v'(x(m)) > 0$ and $u'_z = -\frac{1}{n} < 0$. This equation implies that $z'(n) = 0$ if and only if $x'(n) = 0$.

Now fix m and let n vary and define the function $\tilde{g}(\cdot)$ as

$$\tilde{g}(n) = \{u(x(n), z(n), n) - u(x(m), z(m), n)\}.$$

This function achieves a minimum at $n = m$. The first-order condition for this ($\tilde{g}'(n)|_{n=m} = 0$) can be written as follows (using equation (21))

$$[u'_n - u'_n(x(m), z(m), n)]_{n=m} = 0,$$

or equivalently

$$\frac{z(n)}{n^2} - \frac{z(m)}{m^2} = 0$$

at $n = m$. The second-order condition for the minimization of $\tilde{g}(\cdot)$ evaluated at $n = m$ amounts to

$$\frac{1}{n^2} z'(n) \geq 0.$$

It follows from this condition that $z'(n) \geq 0$. Using equation (21), we find that also $x'(n) \geq 0$.

In order to prove that the conditions in the lemma also guarantee that the second-order condition holds globally, we use a proof by contradiction.¹⁵ So suppose this is not the case. In particular, assume that there exist two types n and n' such that

$$u(x(n'), z(n'), n) > u(x(n), z(n), n).$$

This can be written as

$$\int_n^{n'} \frac{\partial u(x(t), z(t), n)}{\partial t} dt > 0,$$

or equivalently

$$\int_n^{n'} \left[v'(x(t))x'(t) - \frac{1}{n} z'(t) \right] dt > 0.$$

Assume that $n' > n$,¹⁶ then $\frac{1}{t} < \frac{1}{n}$ for each $t > n$ implies that

$$\int_n^{n'} \left[v'(x(t))x'(t) - \frac{1}{t} z'(t) \right] dt > \int_n^{n'} \left[v'(x(t))x'(t) - \frac{1}{n} z'(t) \right] dt.$$

Using equation (21), we find

$$0 > \int_n^{n'} \left[v'(x(t))x'(t) - \frac{1}{n} z'(t) \right] dt.$$

However, this contradicts the inequality with which we started this proof. Hence, there cannot be two types n and n' such that $u(x(n'), z(n'), n) > u(x(n), z(n), n)$. Q.E.D.

Proof of lemma 2

The first-order conditions (Euler equations) for optimizing (6) with respect to $\omega(\cdot)$, $u(\cdot)$ and $z(\cdot)$ amount to (if $z(n_0) > 0$)

$$\omega(n) = \arg \max_{\omega \geq 0} \lambda_z(n) \omega, \quad (22)$$

$$\lambda'_u(n) = f(n) \left(\frac{\lambda_E}{v'(x(n))} - \phi(n) \right), \quad (23)$$

$$\lambda'_z(n) = -\frac{\lambda_u(n)}{n^2} + \lambda_E f(n) \left(\frac{1}{nv'(x(n))} - 1 \right), \quad (24)$$

¹⁵This proof follows closely the argument by Guesnerie and Laffont (1984).

¹⁶The proof for the case where $n' < n$ is similar to the one given here.

together with the transversality conditions

$$\begin{aligned}\lambda_u(n_0) &= \lambda_u(n_1) = 0, \\ \lambda_z(n_0) &= \lambda_z(n_1) = 0,\end{aligned}$$

and the government budget constraint (5).

Since by assumption $\lambda_z(n) = 0$ and thus $\lambda'_z(n) = 0$ for all n , (24) can be written as

$$\frac{1}{v'(x(n))} = n + \frac{1}{n} \frac{\lambda_u(n)}{\lambda_E f(n)}. \quad (25)$$

The first-order condition for maximizing individual utility with respect to $z(n)$ in equation (1) amounts to

$$v'(z(n) - \tilde{T}(z(n))) \left(1 - \tilde{T}'(z(n))\right) - \frac{1}{n} = 0,$$

or equivalently,

$$v'(x(n)) = \frac{1}{n(1 - \tau(n))}.$$

Using this in equation (25) to eliminate $v'(x(n))$, we find

$$\tau(n) = \frac{-\lambda_u(n)}{\lambda_E n^2 f(n)}. \quad (26)$$

Substituting equation (25) into (23) to eliminate $v'(x(n))$, we arrive at

$$\lambda'_u(n) = \frac{1}{n} \lambda_u(n) + \lambda_E f(n) n - f(n) \phi(n). \quad (27)$$

This is a linear differential equation that can be solved analytically (using the method of the varying constant):

$$\lambda_u(n) = n \left[c_0 + \lambda_E \int_{n_0}^n f(t) dt - \int_{n_0}^n \frac{f(t) \phi(t)}{t} dt \right] \quad (28)$$

for some constant c_0 . The transversality condition $\lambda_u(n_0) = 0$ yields $c_0 = 0$ so that

$$\lambda_u(n) = n [\lambda_E F(n) - G(n)]. \quad (29)$$

The transversality condition $\lambda_u(n_1) = 0$ implies

$$\lambda_E = G(n_1).$$

Substitution of this and (29) into equation (26) to eliminate $\lambda_u(n)$ and λ_E yields

$$\tau(n) = \frac{\frac{G(n)}{G(n_1)} - F(n)}{nf(n)}.$$

The definition of $u(n)$ allows us to write

$$z(n) = n(v(x(n)) - u(n)). \quad (30)$$

To determine what $u(n)$ looks like, we substitute this expression for $z(n)$ into the incentive compatibility constraint $u'(n) = \frac{z(n)}{n^2}$ to arrive at the following differential equation:

$$u'(n) = -\frac{1}{n}u(n) + \frac{1}{n}v(x(n)).$$

This linear differential equation can be solved analytically (with method of varying constant):

$$u(n) = \frac{1}{n} \left(K - E + \int_{n_0}^n v(x(t)) dt \right) \quad (31)$$

for some constant K . To determine K , we substitute (31) into (30) and the result into the government budget constraint (5) to eliminate $z(t)$. This yields

$$K = \int_{n_0}^{n_1} \{ [tv(x(t)) - x(t)] f(t) - [1 - F(t)] v(x(t)) \} dt.$$

Finally, the expression for W in the lemma can be derived by writing individual utility (from equation (31)) as $u(n) = \frac{1}{n} \left(n_0 u(n_0) + \int_{n_0}^n v(x(t)) dt \right)$, substitute this expression into welfare $W = \int_{n_0}^{n_1} f(n) \phi(n) u(n) dn$ to eliminate $u(n)$, and employ partial integration. Q.E.D.

Proof of Lemma 3

Note first that $v'(x(n))$ can be written as (using (8) to eliminate τ from (9))

$$v'(x(n)) = \frac{f(n)}{F(n) + nf(n) - \frac{G_\alpha(n)}{G_\alpha(n_1)}},$$

where

$$G_\alpha(n) = \int_{n_0}^n \frac{\phi_\alpha(t)}{t} f(t) dt.$$

The effect of α on $x(n)$ is found by differentiating $\frac{G_\alpha(n)}{G_\alpha(n_1)}$ with respect to α .

$$\text{sign} \left(\frac{d \left(\frac{G_\alpha(n)}{G_\alpha(n_1)} \right)}{d\alpha} \right) = \text{sign} \left(\frac{\int_{n_0}^n \frac{d(\ln \phi_\alpha(t))}{d\alpha} \frac{\phi_\alpha(t) f(t)}{t} dt}{\int_{n_0}^n \frac{\phi_\alpha(t) f(t)}{t} dt} - \frac{\int_{n_0}^{n_1} \frac{d(\ln \phi_\alpha(t))}{d\alpha} \frac{\phi_\alpha(t) f(t)}{t} dt}{\int_{n_0}^{n_1} \frac{\phi_\alpha(t) f(t)}{t} dt} \right),$$

which is positive since $\frac{d(\ln \phi_\alpha(t))}{d\alpha}$ is decreasing in t by definition 1. $\frac{d \frac{G_\alpha(n)}{G_\alpha(n_1)}}{d\alpha} > 0$ implies that $\frac{dx(n)}{d\alpha} < 0$ and (from (8)) $\frac{d\tau(n)}{d\alpha} > 0$ for all n .

To find the effect of α on $u(n)$, we differentiate equation (31) with respect to α .

$$\frac{du(n)}{d\alpha} = \frac{1}{n} \left(\frac{dK}{d\alpha} + \int_{n_0}^n v'(x(t)) \frac{dx(t)}{d\alpha} dt \right)$$

with $\frac{dx(t)}{d\alpha} < 0$.

The following argument shows that $\frac{dK}{d\alpha} > 0$. Suppose not, i.e. $\frac{dK}{d\alpha} < 0$. Then $\frac{du(n)}{d\alpha} < 0$ for all n . Hence, the allocation of $x(n)$ and $z(n)$ before α increased yields higher welfare than the new allocation after α went up. This contradicts the optimality of this (new) allocation because the old allocation is still feasible. This implies that $\frac{dK}{d\alpha} > 0$ and $\frac{du(n)}{d\alpha} > 0$ for some n . However, it is not possible that $\frac{du(n)}{d\alpha} > 0$ for all n because this would contradict the optimality of the allocation before α was increased.

Since $\int_{n_0}^n v'(x(t)) \frac{dx(t)}{d\alpha} dt$ is decreasing in n , the previous paragraph implies that there exists $\tilde{n} \in \langle n_0, n_1 \rangle$ such that all agents with $n < \tilde{n}$ enjoy a rise in utility while types $n > \tilde{n}$ suffer a reduction in utility on account of a higher α . Q.E.D.

Proof of Proposition 1

Combining equations (25) and (29), we find

$$v'(x(n)) = \frac{f(n)}{\frac{G(n_1) - G(n)}{\lambda_E} + f(n)n - [1 - F(n)]} \quad (32)$$

where $\lambda_E = G(n_1)$. This solution yields a path for consumption x which is not monotone at the bottom if and only it implies $x'(n_0) < 0$ or equivalently

$$\left. \frac{d \left[n - \frac{\frac{G(n)}{G(n_1)} - F(n)}{f(n)} \right]}{dn} \right|_{n=n_0} < 0.$$

This inequality can be written as

$$\frac{2n_0}{\phi(n_0)} < \frac{1}{\int_{n_0}^{n_1} \frac{f(t)\phi(t)}{t} dt}.$$

What does the optimal solution look like if $x'(n_0) < 0$? The main departure from the proof of lemma 2 is that $\lambda_z(n) < 0$ for n close to n_0 so that the optimal $\omega(n)$ determined by equation (22) equals $\omega(n) = z'(n) = 0$ for these types. Lemma 1 then implies that also $x'(n) = 0$ for these types. Accordingly, types $[n_0, n_b]$ are bunched together.

Equations (22), (23), (24) and the transversality conditions still apply. Moreover, the analysis in the proof of lemma 2 is correct for non-bunched types $n \geq n_b$. To determine the size of the bunching interval and the marginal cost of public funds, we derive two equations in n_b and λ_E . The first equation is found by integrating (23) and using the transversality conditions $\lambda_u(n_0) = \lambda_u(n_1) = 0$ to arrive at

$$0 = -1 + \lambda_E \int_{n_0}^{n_1} \frac{f(n)}{v'(x(n))} dn.$$

Since $x(n) = x_b$ for all $n \in [n_0, n_b]$, we can rewrite this as

$$\lambda_E = \frac{1}{F(n_b) \frac{1}{v'(x_b)} + \int_{n_b}^{n_1} \frac{f(n)}{v'(x(n))} dn}. \quad (33)$$

Using the three steps labelled A, B and C below, we write this as

$$\lambda_E = \frac{\bar{F}(n_b) + \left(n_b - \frac{F(n_b)}{f(n_b)}\right) (G(n_1) - G(n_b))}{n_b - F(n_b) \frac{1-F(n_b)}{f(n_b)}} \quad (34)$$

Step A. Eliminating $v'(x(n))$ from $\int_{n_b}^{n_1} \frac{f(n)}{v'(x(n))} dn$ by using equation (32) for $n \geq n_b$ and employing partial integration, we find $\int_{n_b}^{n_1} \frac{f(n)}{v'(x(n))} dn = -n_b \frac{G(n_1)}{\lambda_E} + \frac{1}{\lambda_E} [n_b G(n_b) + [1 - \bar{F}(n_b)]] - n_b F(n_b) + n_b$.

Step B. Combining this with the observation that $v'(x_b) = v'(x(n_b))$ and using (32) for $n = n_b$, we establish

$$F(n_b) \frac{1}{v'(x_b)} + \int_{n_b}^{n_1} \frac{f(n)}{v'(x(n))} dn = \frac{1}{\lambda_E} \left[(G(n_1) - G(n_b)) \left(\frac{F(n_b)}{f(n_b)} - n_b \right) + [1 - \bar{F}(n_b)] \right] + n_b - F(n_b) \frac{1 - F(n_b)}{f(n_b)}.$$

Step C. Substituting this into the denominator of equation (33) and solving for λ_E , we arrive at (34).

The second relationship between λ_E and n_b follows from the transversality condition $\lambda_z(n_0) = 0$ and the definition of n_b as the end of the bunching interval: $\lambda_z(n) = 0$ for all $n \geq n_b$ (while $\lambda_z(n) < 0$ for $n \in \langle n_0, n_b \rangle$). $\lambda_z(n_0) = \lambda_z(n_b) = 0$ implies $\int_{n_0}^{n_b} \lambda'_z(n) dn = 0$, or equivalently (using equation (24))

$$- \int_{n_0}^{n_b} \frac{\lambda_u(n)}{n^2} dn + \lambda_E \int_{n_0}^{n_b} f(n) \left(\frac{1}{nv'(x(n))} - 1 \right) dn = 0.$$

We solve $\lambda_u(n)$ for $n < n_b$ by integrating (23) and employing the transversality condition $\lambda_u(n_0) = 0$:

$$\lambda_u(n) = -\bar{F}(n) + \frac{\lambda_E}{v'(x_b)} F(n).$$

Substituting this expression and $x_b = x(n_b) = x(n)$ for $n < n_b$ into the previous equation to eliminate $\lambda_u(n)$ and $x(n)$ and using integration by parts, we find

$$\left\{ \begin{aligned} & -\frac{1}{n_b} \bar{F}(n_b) + G(n_b) - \frac{\lambda_E}{v'(x_b)} \left[-\frac{F(n_b)}{n_b} + \int_{n_0}^{n_b} \frac{f(n)}{n} dn \right] + \\ & + \frac{\lambda_E}{v'(x_b)} \int_{n_0}^{n_b} \frac{f(n)}{n} dn - \lambda_E F(n_b) \end{aligned} \right\} = 0.$$

Using (32) for $n = n_b$ to eliminate $v'(x_b)$ and solving for λ_E , we arrive at the second relation between λ_E and n_b

$$\lambda_E = \frac{G(n_1) - G(n_b) + \frac{G(n_b)f(n_b)n_b}{F(n_b)} - \frac{f(n_b)}{F(n_b)} \bar{F}(n_b)}{1 - F(n_b)}. \quad (35)$$

Equating the two expressions ((34) and (35)) for λ_E , we find (15) determining n_b in the lemma. Substituting equation (15) into either (34) or (35), we arrive at $\lambda_E = G(n_1)$. We find the equation determining $z(n)$ in the same way as in the proof of lemma 2, taking into account that $x(n) = x_b$ for $n \in [n_0, n_b]$. Q.E.D.

Proof of Lemma 4

$\frac{d\lambda_E}{dE} = 0$ follows immediately from the result that $\lambda_E = G(n_1)$. Since E does not impact λ_E , consumption $x(\cdot)$ (which is determined by (16)) is not affected by E . (15) shows that n_b is determined completely by the distribution of skills and the rank order weights $\phi(\cdot)$ and is thus not affected by E so that $\frac{dn_b}{dE} = 0$. Finally, the equation for $z(\cdot)$ in proposition 1 (taking the previous results into account) implies that $\frac{dz(n)}{dE} = 1$ for all $n \in [n_0, n_1]$. Q.E.D.

Proof of Proposition 2

The main departure from the proof of lemma 2 is that the transversality condition for $z(n_0)$ is now changed to $\lambda_z(n_0) < 0$. The reason is that (by assumption) the restriction $z(n_0) \geq 0$ is binding. In other words, one would like to reduce $z(n_0)$ in order to raise welfare (which is exactly what $\lambda_z(n_0) < 0$ implies) but is prevented from doing so by the restriction $z(n_0) \geq 0$. Together with equation (24), $\lambda_z(n_0) < 0$ implies that there exists $n_z > n_0$ such that $\lambda_z(n) < 0$ for all $n \in [n_0, n_z]$. Hence, the optimal $\omega(n)$ determined by equation (22) equals $\omega(n) = z'(n) = 0$ for $n \in [n_0, n_z]$. Lemma 1 then implies that also $x'(n) = 0$ for these types so that types $n \in [n_0, n_z]$ are bunched together with $z(n) = 0$.

The other conditions for optimality are similar to the ones in the proof of lemma 2. In particular, equations (22), (23) and (24) together with the transversality conditions $\lambda_z(n_1) = \lambda_u(n_0) = \lambda_u(n_1) = 0$ continue to apply. Indeed, the analysis in the proof of lemma 2 is correct for types $n \geq n_z$. The main difference with the proof of lemma 2 is that we derive two equations in n_z and λ_E to determine the size of the bunching interval and the marginal cost of public funds.

The first relationship is found by solving equation (27) starting from the endpoint n_1 and using $\lambda_u(n_1) = 0$

$$\lambda_{u+}(n) = -n[\lambda_E(1 - F(n)) - (G(n_1) - G(n))] \quad (36)$$

for all $n \in [n_z, n_1]$. The + in λ_{u+} indicates that it is the solution for λ_u from above n_z . Solving equation (23) starting from n_0 using $\lambda_u(n_0) = 0$ and taking into account that $x(n) = x_z$ for $n \leq n_z$, we find

$$\lambda_{u-}(n) = \left[F(n) \frac{\lambda_E}{v'(x_z)} - \bar{F}(n) \right] \quad (37)$$

for all $n \in [n_0, n_z]$. Setting $\lambda_{u+}(n_z) = \lambda_{u-}(n_z)$, we obtain

$$F(n_z) \frac{\lambda_E}{v'(x(n_z))} - \bar{F}(n_z) = -n_z \lambda_E (1 - F(n_z)) + n_z (G(n_1) - G(n_z)), \quad (38)$$

which can be rewritten as equation (18) in the proposition.

As in the proof of proposition 1, consumption $x(n)$ for $n \geq n_z$ is determined by equation (32). By thus solving the path for $x(\cdot)$ from above, we observe that $x_z = x(n_z)$ because

$\omega(n) = 0$ for $n \leq n_z$ implies that consumption does not fall further. Substituting (32) for $n = n_z$ to eliminate $v'(x_z)$ from (38) and solving for λ_E , we arrive at the first relation between n_z and λ_E ¹⁷

$$\lambda_E = \frac{\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) (G(n_1) - G(n_z))}{n_z - F(n_z) \frac{1-F(n_z)}{f(n_z)}}. \quad (39)$$

The second relation between n_z and λ_E follows from the government budget constraint and the condition $z(n_z) = 0$. In particular, substituting $u(n) = \frac{1}{n} \left(K_z - E + \int_{n_z}^n v(x(t)) dt\right)$ (from (31)) into (30) to eliminate $u(n)$, we obtain

$$z(n) = nv(x(n)) - K_z + E - \int_{n_z}^n v(x(t)) dt. \quad (40)$$

To solve for K_z , we employ (5), which can here be written as (using $T(n) = -x(n_z)$ for $n \leq n_z$ and $T(n) = z(n) - x(n)$ for $n \geq n_z$)

$$-F(n_z) x(n_z) + \int_{n_z}^{n_1} f(n) \left[nv(x(n)) - K_z + E - \int_{n_z}^n v(x(t)) dt - x(n) \right] dn = E.$$

By integrating by parts, we rewrite this equation as

$$(1 - F(n_z))(K_z - E) = -E - F(n_z) x(n_z) + \int_{n_z}^{n_1} \left\{ \begin{array}{l} f(n) [nv(x(n)) - x(n)] \\ -(1 - F(n)) v(x(n)) \end{array} \right\} dn.$$

Substituting this expression for $(K_z - E)$ into equation (40) and using $z(n_z) = 0$, we arrive at the second relation between n_z and λ_E , namely equation (19) in the proposition.

The next two lemmas derive some properties of the two equations (18) (or equivalently (39)) and (19).

Lemma 7 *Equation (18) in (n_z, λ_E) space has the following properties:*

- ◇ *it goes through the points $(n_0, G(n_1))$ and $(n_1, \frac{1}{n_1})$,*
- ◇ *it is downward sloping for all $n_z \in [n_0, n_1]$ if $x(n)$ as determined by equation (16) is increasing in n ,*
- ◇ *a point (n_z, λ_E) with $\lambda_E > G(n_1)$ cannot be part of a solution to the optimization problem (17) because it violates monotonicity for $n \geq n_z$.¹⁸*

Proof. ◇ Substitution of $n_z = n_0$ in (18) yields

$$\lambda_E = \frac{0 + (n_0 - 0)(G(n_1) - 0)}{n_0 - 0} = G(n_1).$$

¹⁷Note the similarity with equation (34) above, which also follows from the condition that $\lambda_u(n)$ is continuous at the end of the bunching interval.

¹⁸See below for a sketch of the analysis if there is both $z = 0$ bunching and bunching due to violation of monotonicity.

Similarly, we obtain for $n_z = n_1$

$$\lambda_E = \frac{1 + \left(n_1 - \frac{1}{f(n_1)}\right) 0}{n_1 - 0} = \frac{1}{n_1}.$$

Since $G(n_1) > \frac{1}{n_1}$, this curve must be downward sloping over parts of the range $[n_0, n_1]$.

◇ Note that (from (36)) $\lambda_{u+}(n_0) = n_0(G(n_1) - \lambda_E)$. Hence, $\lambda_{u+}(\cdot)$ can be written as (from (23))

$$\lambda_{u+}(n, \lambda_E) = n_0(G(n_1) - \lambda_E) + \int_{n_0}^n f(t) \left(\frac{\lambda_E}{v'(x(t))} - \phi(t) \right) dt, \quad (41)$$

where $x(t)$ is determined by (32).

We can write $\lambda_{u-}(\cdot)$ as (from (23))

$$\lambda_{u-}(n, n_z, \lambda_E) = \int_{n_0}^n f(t) \left(\frac{\lambda_E}{v'(x(n_z))} - \phi(t) \right) dt. \quad (42)$$

A point (n_z, λ_E) on curve (18) satisfies

$$\lambda_{u+}(n_z, \lambda_E) = \lambda_{u-}(n_z, n_z, \lambda_E),$$

which can be written as

$$n_0(G(n_1) - \lambda_E) = \int_{n_0}^{n_z} f(t) \left(\frac{\lambda_E}{v'(x(n_z))} - \frac{\lambda_E}{v'(x(t))} \right) dt.$$

Differentiating this expression with respect to λ_E and n_z , we find

$$\left[-n_0 \frac{G(n_1)}{\lambda_E} + \int_{n_0}^{n_z} f(t) \lambda_E \underbrace{\left\{ \frac{-v''(n(n_z))}{(v'(x(n_z)))^2} \frac{dx(n_z)}{d\lambda_E} - \frac{-v''(n(t))}{(v'(x(t)))^2} \frac{dx(t)}{d\lambda_E} \right\}}_{(*)} dt \right] d\lambda_E =$$

$$\int_{n_0}^{n_z} f(t) \lambda_E \frac{-v''(n(n_z))}{(v'(x(n_z)))^2} \frac{dx(n_z)}{dn_z} dt dn_z.$$

Differentiating equation (16) with respect to λ_E , we derive

$$v''(x(t)) \frac{dx(t)}{d\lambda_E} = \frac{1}{f(t)} (v'(x(t)))^2 \frac{G(n_1) - G(t)}{\lambda_E^2}.$$

Substituting this into the part labelled (*) in the equation above, we obtain

$$(*) = \frac{1}{f(n_z)} \frac{G(n_1) - G(n_z)}{\lambda_E^2} - \frac{1}{f(t)} \frac{G(n_1) - G(t)}{\lambda_E^2}.$$

By integrating by parts, we can write $\frac{d\lambda_E}{dn_z}$ as

$$\begin{aligned}\frac{d\lambda_E}{dn_z} &= -\frac{F(n_z) \lambda_E \frac{-v''(n(n_z))}{(v'(x(n_z)))^2} x'(n_z)}{n_z - F(n_z) \frac{1-F(n_z)}{f(n_z)}} \\ &= -\frac{\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) (G(n_1) - G(n_z))}{\left(n_z - F(n_z) \frac{1-F(n_z)}{f(n_z)}\right)^2} F(n_z) \frac{-v''(n(n_z))}{(v'(x(n_z)))^2} x'(n_z).\end{aligned}$$

where we have used the expression for λ_E in equation (39). To prove that $\frac{d\lambda_E}{dn_z} < 0$ if $x'(n) > 0$ for all n , we need the following inequality to hold for all n_z

$$\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) (G(n_1) - G(n_z)) \geq 0$$

Note that for $n_z = n_0$ and for $n_z = n_1$ this inequality holds strictly. Furthermore, it is clear from equation (18) that $\lambda_E > 0$ for all $n_z \in [n_0, n_1]$. Using $\lambda_E > 0$ and (39), we note that

$$\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) (G(n_1) - G(n_z)) < 0$$

if and only if

$$n_z - \frac{F(n_z)}{f(n_z)} (1 - F(n_z)) < 0.$$

Assume (by contradiction) that such a point n_z exists where the last inequality holds. Since $n_0 - \frac{F(n_0)}{f(n_0)} (1 - F(n_0)) > 0$ there must exist a point n' where $n' - \frac{F(n')}{f(n')} (1 - F(n')) = 0$. Then for n_z just above n' we find that $n_z - \frac{F(n_z)}{f(n_z)} (1 - F(n_z)) < 0$ and hence it must be the case that

$\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) (G(n_1) - G(n_z)) < 0$ for n_z just above n' . Substituting $n' - \frac{F(n')}{f(n')} (1 - F(n')) = 0$ into this inequality yields

$$\begin{aligned}&\bar{F}(n') + \left(n' - \frac{n'}{1 - F(n')}\right) (G(n_1) - G(n')) \\ &= F(n') \left[\frac{\int_{n_0}^{n'} \phi(t) f(t) dt}{F(n')} - \frac{\int_{n'}^{n_1} \frac{n'}{t} \phi(t) f(t) dt}{1 - F(n')} \right] > 0\end{aligned}$$

because $\phi(n)$ is decreasing in n and $\frac{n'}{t} < 1$ for $t \in \langle n', n_1 \rangle$. Hence, by continuity, for n_z just above n' we find that $\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) (G(n_1) - G(n_z)) > 0$. This together with $n_z - \frac{F(n_z)}{f(n_z)} (1 - F(n_z)) < 0$ for n_z just above n' leads to $\lambda_E < 0$. This contradicts $\lambda_E > 0$ for all $n_z \in [n_0, n_1]$. This implies that there is no $n_z \in [n_0, n_1]$ such that $n_z - \frac{F(n_z)}{f(n_z)} (1 - F(n_z)) < 0$. Therefore we find that

$$n_z - \frac{F(n_z)}{f(n_z)} (1 - F(n_z)) \geq 0 \quad (43)$$

$$\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) (G(n_1) - G(n_z)) \geq 0 \quad (44)$$

for all $n_z \in [n_0, n_1]$.

◇ Consider a point (n_z, λ_E) on the curve with $\lambda_E > G(n_1)$. Then continuity of (39) implies that another point (n'_z, λ'_E) on the curve exists such that $n'_z > n_z$ and $\lambda'_E > G(n_1)$, which implies (from (36)) that $\lambda_{u+}(n_0, \lambda'_E) < 0$. Since $\lambda_{u-}(n_0, n_z, \lambda_E) = 0$, λ_{u+} crosses λ_{u-} from below at $n = n'_z$, i.e.

$$\lim_{n \uparrow n'_z} \frac{\partial \lambda_{u+}(n, \lambda'_E)}{\partial n} > \lim_{n \uparrow n'_z} \frac{\partial \lambda_{u-}(n, n'_z, \lambda'_E)}{\partial n},$$

or equivalently (from (41) and (42))

$$\lim_{n \uparrow n'_z} x(n) > x(n'_z),$$

where $x(n'_z)$ is evaluated at λ'_E . Since λ'_E can be chosen close to λ_E and $x(n)$ (as determined by equation (16)) is continuous in λ_E , we must have

$$x'(n'_z) < 0$$

evaluated at λ_E . Therefore, a point (n_z, λ_E) with $\lambda_E > G(n_1)$ implies that monotonicity of $x(\cdot)$ is violated for $n \geq n_z$ and thus cannot be part of a solution to (17). ■

Lemma 8 Equation (19) is upward sloping in (n_z, λ_E) space if $x'(n) \geq 0$.

Proof. Define

$$\begin{aligned} \psi(n_z, \lambda_E) \equiv & \int_{n_z}^{n_1} \left\{ \begin{array}{l} f(n) [nv(x(n, \lambda_E)) - x(n, \lambda_E)] \\ - [1 - F(n)] v(x(n, \lambda_E)) \end{array} \right\} dn \\ & - [(1 - F(n_z)) n_z v(x(n_z, \lambda_E)) + E + F(n_z) x(n_z, \lambda_E)]. \end{aligned} \quad (45)$$

(19) can be written as $\psi(n_z, \lambda_E) = 0$. It is routine to verify

$$\psi_{n_z}(n_z, \lambda_E) = -x'(n_z, \lambda_E) \frac{n_z f(n_z) + F(n_z) \left[\frac{G(n_1) - G(n_z)}{\lambda_E} - (1 - F(n_z)) \right]}{\frac{G(n_1) - G(n_z)}{\lambda_E} + n_z f(n_z) - (1 - F(n_z))} \leq 0,$$

because $x'(n_z, \lambda_E) \geq 0$ by assumption, and

$$\begin{aligned} \psi_{\lambda_E}(n_z, \lambda_E) &= \left[\int_{n_z}^{n_1} \left\{ [tv'(x(t, \lambda_E)) - 1] f(t) - [1 - F(t)] v'(x(t, \lambda_E)) \right\} \frac{dx(t, \lambda_E)}{d\lambda_E} dt \right. \\ &\quad \left. - ((1 - F(n_z)) n_z v'(x(n_z, \lambda_E)) + F(n_z)) \frac{dx(n_z, \lambda_E)}{d\lambda_E} \right] \\ &= \left[\int_{n_z}^{n_1} \left\{ v'(x(t, \lambda_E)) (tf(t) - [1 - F(t)]) - f(t) \right\} \frac{dx(t, \lambda_E)}{d\lambda_E} dt \right. \\ &\quad \left. + ((1 - F(n_z)) n_z v'(x(n_z, \lambda_E)) + F(n_z)) \left(-\frac{dx(n_z, \lambda_E)}{d\lambda_E} \right) \right] \\ &= \left[\int_{n_z}^{n_1} f(t) \frac{G(n_1) - G(t)}{\lambda_E} \left[-\frac{dx(t, \lambda_E)}{d\lambda_E} \right] dt \right. \\ &\quad \left. + ((1 - F(n_z)) n_z v'(x(n_z, \lambda_E)) + F(n_z)) \left[-\frac{dx(n_z, \lambda_E)}{d\lambda_E} \right] \right] > 0. \end{aligned}$$

Hence,

$$\frac{d\lambda_E}{dn_z} = -\frac{\psi_{n_z}(n_z, \lambda_E)}{\psi_{\lambda_E}(n_z, \lambda_E)} \geq 0.$$

■

We thus have two curves in (n_z, λ_E) space. We can identify four possible cases:

1. The curve (19) lies everywhere below the curve (18) in (n_z, λ_E) space,
2. the curve (19) lies everywhere above the curve (18) in (n_z, λ_E) space,
3. the curve (19) crosses the curve (18) at a point where (19) is upward sloping and (18) downward sloping and $\lambda_E \leq G(n_1)$.
4. the curve (19) crosses the curve (18) at a point (n_z, λ_E) where $x'(n_z) < 0$.

In case 1, E is so low (probably negative) that no one needs to work in this economy.

In case 2, the solution in lemma 2 implies $z(n_0) > 0$. This can be seen as follows. If instead of deriving equation (19) with $z(n_z) = 0$, we derive an equation with $z(n_z) = \bar{z} > 0$, the curve (19) shifts downwards and hence we find a point of intersection between (19) and (18) where $z(n_0) = \bar{z} > 0$.

In case 3, the intersection point determines the equilibrium values of n_z and λ_E .

In case 4, the point (n_z, λ_E) is not an equilibrium point. $x'(n_z) < 0$ implies that $\lambda_z(n_z) < 0$ and hence types slightly above n_z should be bunched together with type n_z with the same consumption and production ($\omega(n) = 0$ for these types). Since $z(n_z) = 0$, the $z = 0$ bunching interval then extends to types $n > n_z$ beyond n_z . In this case, the procedure to find an equilibrium is as follows. Extend the bunching interval to the smallest value $\tilde{n}_z > n_z$ such that $x'(\tilde{n}_z) \geq 0$. If this point (\tilde{n}_z, λ_E) satisfies the government budget constraint (19), it is the solution to the maximization problem. If it does not satisfy the government budget constraint, there are two possibilities. First, the solution (\tilde{n}_z, λ_E) may be too expensive to be an equilibrium. Then the solution will feature $z(n) = \bar{z} > 0$ for $n \in [n_0, n_z]$ so that $z = 0$ bunching does not occur. Second, the solution (\tilde{n}_z, λ_E) may leave government money on the table. In that case, the bunching interval should be extended beyond \tilde{n}_z .

Finally, we need to prove that $\tau(n_z) > 0$. (26) implies that $\tau(n_z) > 0$ if and only if $\lambda_u(n_z) < 0$. Hence, the proof boils down to showing that $\lambda_{u+}(n_z, \lambda_E) < 0$ if $\lambda_{u+}(n_z, \lambda_E) = \lambda_{u-}(n_z, n_z, \lambda_E)$. Suppose (by contradiction) that $\lambda_{u+}(n_z, \lambda_E) > 0$. Then $\lambda_{u+}(n, \lambda_E)$ is decreasing in n at n_z because $\lambda_{u+}(n_1, \lambda_E) = 0$ and the expression $\lambda'_{u+}(n) = \left(\frac{\lambda_E}{v'(x(n))} - \phi(n) \right)$ changes sign only once (from negative to positive) as a function of n (since $x'(n) \geq 0$ and $\phi'(n) \leq 0$). Hence,

$$\frac{\lambda_E}{v'(x(n_z))} - \phi(n_z) < 0 \tag{46}$$

Next observe that $\lambda_{u-}(n_0, n_z, \lambda_E) = 0$ together with $\lambda_{u-}(n_z, n_z, \lambda_E) > 0$ implies that $\frac{\lambda_E}{v'(x(n_z))} - \phi(n) > 0$ for some $n \leq n_z$. $\phi'(n) \leq 0$ implies in fact that $\frac{\lambda_E}{v'(x(n_z))} - \phi(n_z) > 0$, which contradicts inequality (46). Hence, $\lambda_{u+}(n_z, \lambda_E) = \lambda_{u-}(n_z, n_z, \lambda_E) < 0$ and thus $\tau(n_z) > 0$. Q.E.D.

Proof of lemma 5

As established in the proof of proposition 2, we must have that $x'(n_z) > 0$ so that n_z and λ_E are determined by the intersection of the downward sloping curve (18) and the upward sloping curve (19). Clearly, equation (18) is not affected by a change in E . The proof of lemma 8 implies that (19) shifts upward (and to the left) as E increases. Hence, n_z falls and λ_E rises with E .

Since λ_E reduces $x(n_z)$ and we have $x'(n_z) \geq 0$ and $dn_z/dE < 0$, we find that $x_z = x(n_z)$ falls with E . Furthermore, (16) implies that the rise in λ_E raises the marginal tax rate and reduces consumption for all types $n > n_z$. Writing utility for type $n > n_z$ as

$$u(n) = \frac{n_z}{n} v(x_z) + \frac{1}{n} \int_{n_z}^n v(x(t)) dt,$$

we find that utility declines with E for all $n > n_z$. Finally, the tax paid by type n can be written as

$$\begin{aligned} T(n) &= T(n_0) + \int_{n_0}^n T'(t) dt \\ &= -x_z + \int_{n_z}^n T'(t) dt, \end{aligned}$$

since $z(n) = 0$ for all $n \in [n_0, n_z]$. Hence, $\frac{dx_z}{dE} < 0$ implies that a value $n^* > n_z$ exists such that

$$\frac{dT(n)}{dE} > 0$$

for all $n \in [n_z, n^*]$. Q.E.D.

Proof of Lemma 6

First, we prove the effect of α on n_z by analyzing the effect of α on the LS and GBC curves. To find the effects of α on the GBC curve, write this curve as

$$\psi(n_z, \lambda_E, \alpha) = 0,$$

where the function $\psi(n_z, \lambda_E)$ is defined in equation (45). In this equation, both λ_E and α work through their effects on $x(n)$ only.

To consider the impact of α on $x(n)$ for $n \geq n_z$, we differentiate (16) with respect to α and note that the sign of this effect depends on the sign of

$$h(n) \equiv \frac{d(G_\alpha(n_1) - G_\alpha(n))}{d\alpha} = \int_n^{n_1} \frac{d \ln \phi_\alpha(t)}{d\alpha} \frac{1}{t} \phi_\alpha(t) f(t) dt.$$

The properties of $h(n)$ are as follows (see Figure 2). The normalization $\int_{n_0}^{n_1} \frac{d \ln \phi_\alpha(t)}{d\alpha} \phi_\alpha(t) f(t) dt = 0$ implies that $h(n_0) > 0$ since $\frac{d \ln \phi_\alpha(t)}{d\alpha}$ is declining in skill t so that in $h(n_0)$ the larger values of $\frac{d \ln \phi_\alpha(t)}{d\alpha}$ are weighted more heavily. $\int_{n_0}^{n_1} \frac{d \ln \phi_\alpha(t)}{d\alpha} \phi_\alpha(t) f(t) dt = 0$ and $\frac{d \ln \phi_\alpha(t)}{d\alpha}$ is declining in skill t implies also that there exists a skill level \bar{n} such that $\frac{d \ln \phi_\alpha(n)}{d\alpha} \geq 0$ for $n \leq \bar{n}$ and $\frac{d \ln \phi_\alpha(n)}{d\alpha} < 0$

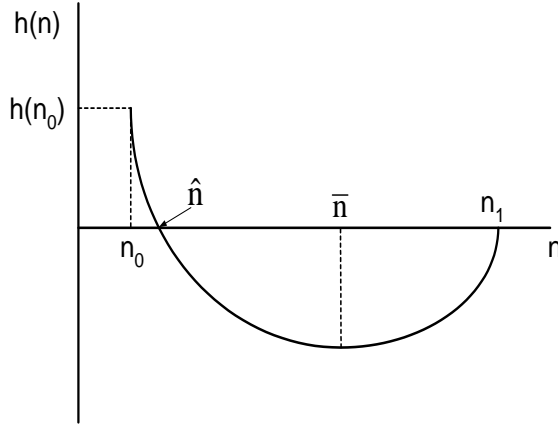


Figure 2: *The function $h(n)$.*

for $n > \bar{n}$. Since $\text{sign}[h'(n)] = -\text{sign}[\frac{d \ln \phi_\alpha(n)}{d\alpha}]$, this implies that $h(n)$ is first decreasing until skill \bar{n} and then increases for $n > \bar{n}$. Since $h(n_1) = 0$, we know that $h(\bar{n}) < 0$. Hence, $h(n_0) > 0$ implies that there exists a skill level \hat{n} between n_0 and \bar{n} where $h(\hat{n}) = 0$. For all $n > \hat{n}$, we have that $h(n) \leq 0$. Hence, the condition $n_z > \hat{n}$ implies that $h(n)$ is negative so that $x(n)$ declines with α for $n \geq n_z$.

Since an increase in not only λ_E but also α thus reduces $x(n)$ for $n \geq n_z \geq \hat{n}$, both variables exert the same effect on $\psi(\cdot)$ for given $n_z \geq \hat{n}$. Hence, for given $n_z \geq \hat{n}$, an increase in α must be accompanied by a fall in λ_E to establish budget balance ($\psi = 0$) again. In other words, an increase in α shifts the GBC curve downward in figure 1.

Consider now the labour supply LS curve in the form of equation (39). We find that

$$\frac{d\lambda_E}{d\alpha} = \frac{\frac{d\bar{F}(n_z)}{d\alpha} + \left(n_z - \frac{F(n_z)}{f(n_z)}\right) \frac{d(G_\alpha(n_1) - G_\alpha(n_z))}{d\alpha}}{n_z - F(n_z) \frac{1 - F(n_z)}{f(n_z)}}$$

We have established above that $h(n_z) = \frac{d(G_\alpha(n_1) - G_\alpha(n_z))}{d\alpha} < 0$ since $n_z > \hat{n}$. Hence, a sufficient condition for $\frac{d\lambda_E}{d\alpha} > 0$ is

$$\begin{aligned} & \frac{d\bar{F}(n_z)}{d\alpha} + n_z \frac{d(G_\alpha(n_1) - G_\alpha(n_z))}{d\alpha} = \\ & \int_{n_0}^{n_z} \frac{d \ln \phi_\alpha(n)}{d\alpha} \phi_\alpha(n) f(n) dn + \int_{n_z}^{n_1} \frac{d \ln \phi_\alpha(n)}{d\alpha} \frac{n_z}{n} \phi_\alpha(n) f(n) dn \geq 0 \end{aligned}$$

We denote this expression as a function of n_z by $\gamma(n_z)$ and prove that $\gamma(n_z) \geq 0$ for all $n_z \in [n_0, n_1]$. First note that $\gamma'(n_z) = \int_{n_z}^{n_1} \frac{d \ln \phi_\alpha(n)}{d\alpha} \frac{1}{n} \phi_\alpha(n) f(n) dn = h(n_z)$. The function $\gamma(\cdot)$ is thus increasing on the interval $[n_0, \hat{n}]$ and decreasing for $n \in [\hat{n}, n_1]$. Since $\gamma(n_0) = n_0 h(n_0) > 0$ (since $h(n_0) > 0$, see above) and $\gamma(n_1) = 0$ (since $\int_{n_0}^{n_1} \frac{d \ln \phi_\alpha(t)}{d\alpha} \phi_\alpha(t) f(t) dt = 0$ from the definition of α), we must have that $\gamma(n_z) \geq 0$ for all $n_z \in [n_0, n_1]$.

Hence, for $n_z \geq \hat{n}$, an increase in α shifts the GBC downward in figure 1 and the LS curve upward.

Now we turn to effect of α on $x(n)$. Equation (32) implies that

$$\frac{dx(n)}{d\alpha} \begin{matrix} \leq \\ \geq \end{matrix} 0 \text{ if and only if } \frac{d \left[\frac{G(n_1) - G(n)}{\lambda_E} \right]}{d\alpha} \begin{matrix} \leq \\ \geq \end{matrix} 0$$

for $n \in [n_z, n_1]$ so that

$$\begin{aligned} \text{sign} \left(\frac{dx(n)}{d\alpha} \right) &= \text{sign} \left(\frac{\frac{d[G(n_1) - G(n)]}{d\alpha}}{G(n_1) - G(n)} - \frac{\frac{d\lambda_E}{d\alpha}}{\lambda_E} \right) \\ &= \text{sign} \left(\frac{h(n)}{G(n_1) - G(n)} - \frac{\frac{d\lambda_E}{d\alpha}}{\lambda_E} \right). \end{aligned} \quad (47)$$

We will first prove that $\frac{h(n)}{G(n_1) - G(n)}$ is decreasing in n . $\frac{d \left[\frac{h(n)}{G(n_1) - G(n)} \right]}{dn} < 0$ if and only if

$$h'(n) (G(n_1) - G(n)) + G'(n) h(n) < 0,$$

or equivalently

$$\frac{\int_n^{n_1} \frac{d \ln \phi_\alpha(t)}{d\alpha} \frac{\phi_\alpha(t) f(t)}{t} dt}{\int_n^{n_1} \frac{\phi_\alpha(t) f(t)}{t} dt} - \frac{d \ln \phi_\alpha(n)}{d\alpha} < 0.$$

This last inequality holds because $\frac{d \ln \phi_\alpha(t)}{d\alpha}$ is decreasing in t by definition 1 and the first expression is a weighted average of terms $\frac{d \ln \phi_\alpha(t)}{d\alpha}$ for $t \geq n$ which is smaller than the second term $\frac{d \ln \phi_\alpha(n)}{d\alpha}$.

Returning to equation (47) and noting that $\frac{\frac{d\lambda_E}{d\alpha}}{\lambda_E}$ does not vary with n while $\frac{h(n)}{G(n_1) - G(n)}$ is decreasing in n , we have the following three possibilities:

- (i) $\frac{dx(n)}{d\alpha} > 0$ for each $n \in [n_0, n_1]$,
- (ii) $\frac{dx(n)}{d\alpha} < 0$ for each $n \in [n_0, n_1]$ and
- (iii) there exists $\tilde{n} \in \langle n_0, n_1 \rangle$ such that

$$\begin{aligned} \frac{dx(n)}{d\alpha} &> 0 \text{ for each } n \in [n_0, \tilde{n}] \\ \frac{dx(n)}{d\alpha} &< 0 \text{ for each } n \in \langle \tilde{n}, n_1 \rangle. \end{aligned}$$

Note that $v'(x(n_1)) = \frac{1}{n_1}$ and hence $x(n_1)$ is not affected by α . Also note that $x(n) = x(n_z)$ for all $n \in [n_0, n_z]$.

We show that cases (i) and (ii) contradict the optimality of the consumption schedule $x(n)$ by writing welfare in terms of consumption only, as in equation (12). We write welfare as

$$W_\alpha = \int_{n_0}^{n_1} u(n) \phi_\alpha(n) f(n) dn,$$

where

$$\begin{aligned} u(n) &= v(x(n_z)) \text{ for all } n \in [n_0, n_z] \\ u(n) &= \frac{1}{n} \left[n_z v(x(n_z)) + \int_{n_z}^n v(x(t)) dt \right] \text{ for all } n \in [n_z, n_1]. \end{aligned} \quad (48)$$

Substituting these expressions for $u(n)$ into the expression for W_α and using integration by parts, we find

$$W_\alpha = [\bar{F}(n_z) + n_z (G(n_1) - G(n_z))] v(x(n_z)) + \int_{n_z}^{n_1} v(x(n)) (G(n_1) - G(n)) dn.$$

Now consider two values of α denoted by $\alpha' < \alpha''$. Then case (i) can be ruled out because $x_{\alpha''}(n)$ would lead to higher welfare $W_{\alpha''}$ contradicting the optimality of $x_{\alpha'}(n)$ (as changing α impacts only the objective function of the government and does not affect the feasibility and incentive compatibility constraints of a menu $(x(n), z(n))$). In a similar fashion, case (ii) can be ruled out. That leaves only case (iii), which is the one in the lemma.

Since $\frac{dx(n)}{d\alpha} > 0$ for each $n \in [n_0, \tilde{n}]$, equation (48) implies that utility goes up for a group of agents $n \in [n_0, n^*]$ where $n^* > \tilde{n}$. However, it cannot be the case that utility goes up for all agents $n \in [n_0, n_1]$ as this would contradict the optimality of the initial allocation. Since $\frac{dx(n)}{d\alpha} < 0$ for each $n \in \langle \tilde{n}, n_1 \rangle$, equation (48) implies that $n^* < n_1$ and $\frac{du(n)}{d\alpha} < 0$ for $n \in \langle n^*, n_1 \rangle$. Q.E.D.