The optimal taxation of unskilled labor with job search and social assistance

Jan Boone*and Lans Bovenberg†
Dept. of Economics
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands

26-4-2002

Abstract

In order to explore the optimal taxation of low-skilled labor, we extend the standard model of optimal non-linear income taxation in the presence of quasi-linear preferences in leisure by allowing for involuntary unemployment, job search, an exogenous welfare benefit, and a non-utilitarian social welfare function. In trading off more low-skilled employment against more work effort of higher skilled workers, the government balances distortions on the search margin with those on work effort. Positive marginal tax rates at the bottom may help to encourage job search if this search is taxed on a net basis. Lower welfare benefits and search costs tend to reduce marginal tax rates throughout the skill distribution.

Key words: labor-market search, social assistance, unemployment, low-skilled labor, non-linear income taxation, participation margin, bunching.

1 Introduction

Widening wage dispersion raises the question how public policy should protect the living standards of unskilled workers, as policy makers are increas-

*j.boone@kub.nl
†a.l.bovenberg@kub.nl

ingly concerned about the adverse incentive effects of generous income support. In response to these concerns, many countries have cut taxes on unskilled work in order to combat poverty while at the same time encouraging unskilled workers to look for work. Both the United States and several European countries have already introduced or are considering in-work tax benefits for unskilled work in the form of an Earned Income Tax Credit (EITC). These tax policies are part of active labor-market policies and 'welfare-to-work' programs, where governments fight poverty by raising employment of unskilled workers.

To investigate the optimal response of tax policy to income support provided through the welfare system and to declining relative wages of unskilled labor, we extend the standard model of optimal non-linear income taxation developed by Mirrlees (1971). In particular, we incorporate the labor-market imperfections that induce governments to provide income support, namely search costs and involuntary unemployment.

In the presence of a search margin, the government has to account for not only the standard incentive compatibility constraint on work effort, but also a participation constraint on the willingness of low-skilled agents to look for work. Indeed, from an analytical point of view, our main contribution is to add a participation constraint to the optimal tax problem, including the decision regarding which types should optimally participate in job search. In doing so, we extend both the optimal tax literature, which typically abstracts from participation constraints, and the literature on optimal non-linear monopoly pricing, which generally assumes that the lowest participating type is exogenously given. Within a non-linear pricing framework, Rochet and Stole (2002) recently added an endogenous participation constraint by allowing agents to differ in both outside options and preferences for quality. Our analysis deviates from that of Rochet and Stole (2002) in two respects. First, in the non-linear pricing problem explored in Rochet and Stole (2002), the monopolist cares only about profits earned on the participating agents. In our optimal tax problem, in contrast, also agents who do not participate appear in the objective function because the government is interested in the utilities of both participating and nonparticipating types. The second difference with Rochet and Stole (2002) is that we allow agents to differ in only one dimension; agents feature different skill levels but exhibit the same search costs. Within the context of our labor-market application, this is a reasonable assumption, which is in fact employed by most of the labor-market literature on search (see, for instance, Mortensen and Pissarides (1999)). This assumption implies that the participation constraint is binding only at the bottom of the skill distribution. Rochet and Stole (2002), in contrast, derive a binding participation constraint for each type.

The literature on optimal income taxation has modelled unemployment of unskilled agents as these agents reducing the hours they work in their jobs when they face low gross wages and rapidly rising marginal tax rates. Accordingly, low productivity workers are bunched in low- or zero-production jobs. By introducing a participation margin and positive search costs, we introduce another type of bunching at the bottom of the skill distribution: unskilled agents do not search for work and thus drop out of the labor force. Empirical work, in fact, reveals that unskilled workers adjust their labor supply in response to tax and benefit programs on mainly the extensive margin (i.e. leaving the labor force altogether, for example through early retirement) rather than the intensive margin (i.e. reducing the hours they work in their jobs). This explains the policy concern about welfare programs and high taxes on unskilled work discouraging low-income earners from looking for work. Indeed, our model is consistent with the stylized fact that low-skilled agents feature the highest long-term unemployment rates (see OECD (2001)). The participation margin also gives rise to more intuitive comparative statics. To illustrate, in a traditional model without a participation margin, higher welfare benefits reduce unemployment. In a model with a binding participation margin, in contrast, higher welfare benefits raise unemployment, as it becomes more expensive to draw these types into the labor force.

Also Saez (2000) incorporates the two labor-supply margins of not only hours worked but also labor-force participation in an optimal income tax model.[1] Our approach differs from that of Saez in four important respects. First of all, whereas Saez assumes that all unemployed have voluntarily left the labor force, we account also for involuntary unemployment. Agents thus face two risks: being born with low ability and being involuntarily unemployed. More generally, we are more explicit than Saez (2000) about the labor-market imperfections affecting the costs and effectiveness of labor-market search, including the welfare implications of these imperfections.

A second difference with Saez' approach is that our model attributes the same tastes to the employed and the unemployed. Saez, in contrast, allows agents to differ in not only skills but also preferences for full-time leisure. Hence, unemployment is related to not only skill levels but also preferences. This complicates the welfare analysis. In particular, the arguments for pro-

[1] Diamond (1980) and Choné and Laroque (2002) incorporate a participation margin in an optimal tax framework in which work effort is exogenous so that the intensive margin is absent. Our paper is similar to Saez (2000) in that we model both the intensive and extensive margins of labor supply, allowing us to explore the interaction between these two labor-supply margins.

viding generous benefits to agents without labor incomes would seem considerably weaker in Saez' model (in which a person's employment status is the result of voluntary choices based on individual preferences) than in our setting (in which unemployment is directly related to having had the bad luck of either having been born with few skills or not having been able to find a suitable job). Saez, however, assumes that the welfare weights that the social planner attaches to various agents depend only on disposable income and not on tastes for full-time leisure.

Our analysis differs from Saez (2000) also in that the government takes the welfare benefit as exogenously given when optimizing the tax system. Hence, the government can employ only the non-linear income tax to optimize social welfare. Indeed, in practice, taxes and social assistance are often set by distinct agencies based on rather different interests and considerations. One can view the minimum income floor set by social assistance benefits as being determined by considerations outside our model. Alternatively, one can view our analysis as exploring how the tax system can be employed to address the possibly sub-optimal aspects of social assistance.

A final difference is that Saez (2000) allows for more general preferences that are not necessarily quasi-linear in leisure. Whereas his results are thus more general than ours, our specific assumptions on preferences allow for more analytical results on comparative statics with respect to public spending, labor-market imperfections (such as the costs and effectiveness of search) and institutional features of the welfare system. This sheds additional light on the determinants of the optimal tax schedule. Indeed, a substantial literature (see, e.g., Boadway, Cuff and Marchand (2000), Ebert (1992), Weymark (1986, 1987), and Lollivier and Rochet (1983)) has turned to quasi-linear preferences in leisure in order to obtain more intuition for the determinants of the optimal non-linear income tax, as these preferences allow for closed-form solutions of the standard optimal non-linear income tax problem. Our quasi-linear preferences also imply that a utilitarian government cares about the distribution of consumption rather than the distribution of work effort. Indeed, policy debates typically focus on raising consumption rather than reducing work effort of the poor. We extend the literature on optimal non-linear income taxation with quasi-linear preferences in four ways; we allow for involuntary unemployment, a participation (or search) constraint, an exogenous welfare benefit, and a non-utilitarian social welfare function. These four extensions make this literature more relevant for addressing the timely policy question of how the optimal tax system should treat low-skilled employment in the face of income support.

The rest of this paper is structured as follows. After section 2 introduces the model, section 3 sets out the optimal tax problem. Section 4 investigates the case without a binding participation constraint. This section generalizes results in Ebert (1992) and Boadway, Cuff, and Marchand (2000) to the case of a non-utilitarian government. Since we are especially interested in the optimal taxation of low earnings, we discuss two reasons why the least skilled workers may be bunched in the absence of a participation margin. These two types of bunching provide important benchmarks for bunching on account of a participation margin. We elaborate in particular on the case in which agents stop working on the intensive rather than extensive margin by reducing their work effort to zero. Section 5 explores the consequences of a binding participation margin. Among other things, it discusses how labor-market imperfections and the features of the welfare system impact the optimal income tax. Numerical simulations in section 6 illustrate the quantitative importance of the participation margin. Finally, section 7 concludes.

2 The model

The economy 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 $f(n)$, and $F(n)$ represents the corresponding cumulative distribution function. The support of the distribution of abilities is given by $[n_0, n_1]$,[2] while $f(.)$ is differentiable and satisfies $1 - n_0 f(n_0) > 0$.

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, $v(0) = 0$, $\lim_{x \downarrow 0} v'(x) = \infty$, $\lim_{x \longrightarrow +\infty} v'(x) = 0$ and

[2] The finite upper bound of the skill distribution n_1 implies that the optimal marginal tax rate is zero at the top of the income distribution (see Sadka (1976) and Seade (1977)). Diamond (1998) considers various unbouded distributions, including the Pareto distribution for skills above the mode. We focus on the bottom rather than the top of the skill distribution and therefore do not extensively consider the sensitivity of our results with respect to the distribution of skills at the top of the skill distribution.

$\lim_{x\to+\infty} v(x) - xv'(x) = +\infty$. The concavity of $v(.)$ implies that agents are risk averse and thus want to obtain insurance against the risks of involuntary unemployment and 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 $v(.)$ 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 four important consequences. 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 $x(.)$ and $z(.)$ (see, for instance, Fudenberg and Tirole (1991)). 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. Indeed, Kanbur, Keen and Tuomala (1994) observe that policy debates focus on raising consumption rather than reducing work effort (or increasing leisure) of the poor. However, whereas Kanbur, Keen, and Tuomala (1994) adopt a non-welfarist social welfare function to do justice to this policy concern of fighting poverty, we continue in the welfarist tradition but assume a special, quasi-linear utility function.

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 before-tax income of each individual $z(n)$. We depart from the standard optimal tax literature by incorporating non-verifiable job search: agents have to search for a job and the government cannot verify search intensities. In particular, we allow agents to adjust their labor supply on not only the intensive margin (i.e. by varying hours of work) but also the extensive margin (i.e. by deciding whether or not to look for a job). In particular, by searching with intensity $s \in [0,1]$, agents find a job with probability s. Search costs $\gamma(s)$ are given by

$$\gamma(s) = \begin{cases} \gamma s & \text{if} \quad s \in [0, \bar{s}] \\ +\infty & \text{otherwise,} \end{cases}$$

where $\gamma \geq 0$ is a parameter representing the magnitude of the search costs. $\bar{s} < 1$ captures the idea that agents can fail to find a job, even if they search at full capacity $\bar{s}$. By modelling the costs and effectiveness of search, the parameters γ and $(1-\bar{s})$ represent labor-market imperfections that give rise to unemployment. In contrast to Saez (2000), we allow for not only labor-supply responses on the extensive margin but also involuntary unemployment. Agents thus differ in both ability n and employment status and face two types of risks: being born with low ability n and being involuntarily unemployed.

If an agent does not succeed in finding a job, (s)he receives a welfare (or social assistance) benefit $b \geq 0$.[3] Since the government cannot observe the abilities and search intensities of unemployed agents,[4] the welfare benefit does not depend on n and is exogenously given to the agent. An agent of ability n thus selects search intensity s to maximize expected utility

$$U(n) = \max_{s} \left\{ -\gamma(s) + su(n) + (1-s)\, v\,(b) \right\}.$$

Substituting in the search cost function $\gamma\,(s)$ introduced above, one can easily verify that the optimal choice of s for type n amounts to

$$s\,(n) = \begin{cases} 0 & \text{if} \quad u\,(n) < \gamma + v\,(b) \\ \bar{s} & \text{if} \quad u\,(n) \geq \gamma + v\,(b)\,. \end{cases} \tag{1}$$

The linear specification of the search cost function thus implies that a worker either does not search at all and is voluntarily unemployed or that he searches at the level $\bar{s}$ (and faces a probability of $(1-\bar{s})$ of involuntary unemployment). We refer to the constraint $u\,(n) \geq \gamma + v\,(b)$ as the participation or individual rationality constraint. The government has to respect this participation constraint because it cannot observe search. The search cost parameter γ can be interpreted as the cost for entering the labor market. Indeed, an agent enters the labor market by actively looking for a job only if the additional pay-off from work, $u(n) - v(b)$, exceeds the fixed entry cost γ. Positive entry cost introduce a difference between stopping to work on the intensive and extensive margin of labor supply: working zero hours in a job (i.e. responding on the intensive margin) is different from staying outside the labor force by not looking for a job (i.e. responding on the extensive margin).

[3] An alternative interpretation of b is a categorial unemployment insurance benefit. Indeed, the benefit is paid only to those who have not found a job. In most countries, however, unemployment benefits depend on the previously earned wage incomes and are thus likely to increase with ability n. This is the main reason why we interpret b as a social assistance payment, i.e. the minimum income level provided by the government. Another interpretation of b is an early retirement or disability benefit that is paid if an agent does not have a job.

[4] The government, however, can observe whether or not an agent has found a job. Hence, we do not require that $b = \tilde{T}(0)$, where $\tilde{T}(z)$ is the tax schedule as a function of gross income.

After a worker has found a job, (s)he has to determine her 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 ex-post utility of a type n agent who finds a job is determined by type n's choice of gross income z:

$$u(n) = \max_{z} \left\{ v\left(z - \tilde{T}(z)\right) - \frac{z}{n} \right\}, \tag{2}$$

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

$$u'(n) = \frac{z(n)}{n^2}. \tag{3}$$

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*

$$\begin{aligned} z'(n) &\geq 0, \\ x'(n) &\geq 0, \end{aligned} \tag{4}$$

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:

$$z(n) \geq 0. \tag{5}$$

The government maximizes ex-ante expected utility (i.e. expected utility before ability and labor market status have been revealed):

$$W \equiv \int_{n_0}^{n_1} \left[-\gamma s(n) + s(n)u(n) + (1 - s(n))v(b)\right] f(n)\phi(n)dn.$$

We normalize the rank-order weights $\phi(n)$ such that $\int_{n_0}^{n_1} f(n)\phi(n) = 1$, and assume $\phi'(n) \leq 0$.[5] The government is utilitarian if the rank-order weights

[5] The 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.

are constant, i.e. $\phi(n) = 1$ for all n. 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)[6]). 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

$$\int_{n_0}^{n_1} f(n)\, s(n)\, [b + T(n)]dn = E + b, \tag{6}$$

where E represents exogenously given exhaustive government expenditure, and $T(n) \equiv z(n) - x(n)$ denotes the tax paid by type n. The government can employ only the non-linear income tax to optimize social welfare and takes public spending E and the welfare benefit as given.

3 The optimal tax problem

In optimizing social welfare, the government faces five constraints: the first-order and second-order incentive compatibility constraints (3) and (4), the participation constraint (1), the non-negativity constraint on gross incomes (5), and the government budget constraint (6). Instead of $x(n)$, we employ $u(n)$ as a control variable in order to facilitate the inclusion of first-order incentive compatibility (3) into our optimization problem.[7] 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

$$\max_{\substack{s(.),u(.),z(.),\\ \omega(.)\geq 0}} \int_{n_0}^{n_1} \left\{ \begin{array}{c} [s(n)(u(n)-\gamma) + (1-s(n))\, v(b)]\, \phi(n) f(n) - \lambda_u(n) \left[u'(n) - \frac{z(n)}{n^2}\right] \\ -\lambda_z(n)\, [z'(n) - \omega(n)] - \eta(n)\, (\gamma - u(n) + v(b)) \\ +\lambda_E f(n)\, [s(n)\, (T(n) + b) - b - E] - \delta(n)\, [0 - z(n)] \end{array} \right\} dn,$$

[6]This 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 $n > n_0$.

[7]This is the usual approach in optimal non-linear income tax problems introduced by Mirrlees (1971). If preferences are quasi linear in leisure, however, it is more convenient to employ $x(n)$ and $nu(n)$ rather than $z(n)$ and $u(n)$ as controls, see Boadway, Cuff, and Marchand (2000). We stick to the Mirrlees approach because we gain less by departing from it due the introduction of a participation constraint on $u(n)$.

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, $\eta(n)$ denotes the multiplier of the participation constraint, and $\delta(n)$ is the Lagrange multiplier of the non-negativity constraint on before-tax income.

To further simplify the optimization problem, we observe that, since $z(n) \geq 0$, incentive compatibility (3) implies that utilities do not decline with skill (i.e. $u'(n) \geq 0$). Accordingly, if the participation constraint $u(n) \geq \gamma + v(b)$ is met for skill $\bar{n}$, it is met also for all higher skills $n > \bar{n}$. Defining n_w as the lowest skill looking for work, we thus have $s(n) = 0$ for $n < n_w$ and $s(n) = \bar{s}$ for $n \geq n_w$. Accordingly, our model is consistent with the stylized fact that the extensive margin of labor-supply responses is especially relevant for low-income earners. Empirical work does show that these workers respond to tax and benefit programs mainly on the extensive rather than the intensive margin of labor supply. This explains the policy concern that generous benefit programs and taxes on low earnings discourage low-income earners from looking for work.

The agents with skill $n < n_w$ can be viewed as being voluntarily unemployed. In our model in which all agents feature the same preferences, the lower skills $n < n_w$ are voluntarily unemployed, whereas the higher skills $n > n_w$ all look for work but may be involuntarily unemployed (if $\bar{s} < 1$). This is in contrast to Saez (2000), who assumes that agents differ not only in skill levels but also in preferences for full-time leisure. In his setting, all unemployment is voluntary – with unemployed agents exhibiting a higher preference for full-time leisure than the employed agents of the same skill do. Hence, voluntary unemployment is related to not only skill levels but also preferences. In both Saez (2000) and our model, higher skilled agents may be unemployed, but for different reasons. Whereas in Saez (2000) the higher skilled agents without work value leisure highly and thus do not look for work, in our model these agents actively search for work but have been unfortunate enough not to have found a suitable job.

Lemma 2 *With voluntary unemployment (i.e. $n_w > n_0$), the participation margin is strictly binding (i.e. $u(n_w) = \gamma + v(b)$).*

The reason is that if $u(n_w) > \gamma + v(b)$, also agents with skills just below n_w would like to search for jobs that yield the same gross income as type n_w even though they have to work harder to produce this gross income than skill n_w.[8] Hence, a solution in which $u(n_w) > \gamma + v(b)$ and in which agents

[8]Formally, $u(n_w) > \gamma + v(b)$ can be written as $v(x_w) - \frac{z_w}{n_w} > v(b) + \gamma$, where x_w and z_w

below n_w do not search (so that n_w is the lowest skill that looks for a job) is not incentive compatible.

These observations allow us to reformulate the social planner's problem as

$$\max_{\substack{n_w,u(.),z(.),\\ \omega(.)\geq 0}} \bar{F}(n_w)\, v(b) + \left[1-\bar{F}(n_w)\right]\left(-\gamma\bar{s} + (1-\bar{s})\, v(b)\right) \tag{7}$$
$$+\int_{n_w}^{n_1}\left\{\begin{array}{c} \bar{s}u(n)\phi(n)f(n) - \lambda_u(n)\left[u'(n) - \frac{z(n)}{n^2}\right] \\ -\lambda_z(n)\left[z'(n)-\omega(n)\right] + \lambda_E\left[f(n)\bar{s}T(n)\right] + \delta(n)\, z(n)\end{array}\right\} dn$$
$$-\lambda_E\left\{b\left[F(n_w) + (1-F(n_w))(1-\bar{s})\right] + E\right\}$$
$$-\eta_w\left(\gamma - u(n_w) + v(b)\right),$$

where $\bar{F}(n_w) \equiv \int_{n_0}^{n_w}\phi(n)f(n)dn$. η_w denotes the Lagrange multiplier on the participation constraint for type n_w. It measures the social value of increasing employment by forcing more people to search, and can therefore be interpreted as the value of a work test (and the required information on search intensity) inducing more skills to look for work.

The shadow values $\lambda_z(n), \delta(n), \eta_w$ are associated with three types of bunching. $\lambda_z(n) < 0$ (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 (3) with $z(n) = 0$). This is called $z = 0$ *bunching*. The search margin gives rise to an additional type of bunching, namely the case in which $n_w > n_0$ and $\eta_w \neq 0$. Accordingly, types $n_0 \leq n < n_w$ do not search. Hence, as voluntarily unemployed, they are bunched without any labor income and collect social assistance b. This we call $s = 0$ *bunching*. Just as $z = 0$ bunching, $s = 0$ bunching can occur only at the bottom of the skill distribution. It implies that the participation constraint is binding (see Lemma 2).

Lemma 3 *If $\gamma > 0$, a binding participation constraint $u(n_w) = \gamma + v(b)$ implies $z(n) > 0$ for $n \geq n_w$.*

represent net and gross incomes of the lowest skilled agent who searches for work n_w. The latter inequality implies $v(x_w) - \frac{z_w}{n_w - \varepsilon} > v(b) + \gamma$ for $\varepsilon > 0$ small enough. Hence, an agent with skill $(n_w - \varepsilon)$ is better-off looking for a job with before-tax income z_w (corresponding to after-tax income of x_w) than not entering the labor market.

Lemmas 2 and 3 imply that $z = 0$ and $s = 0$ bunching are mutually exclusive if search costs γ are positive. An agent should not search for a job in which he does not produce anything (i.e. $z(n_w) = 0$) if the participation margin is binding (i.e. $u(n_w) = \gamma + v(b)$). Leaving such a worker outside the effective labor force yields no first-order impact on private welfare (as welfare of the marginal worker stays the same)[9] nor on the incentive compatibility constraints of higher skilled agents (since $z(n_w) = 0$ implies $u'(n_w) = 0$).[10] Doing so, however, yields a first-order gain in terms of government revenue. The government saves more on in-work benefits than what it spends on additional welfare benefits (as a binding participation margin with $z(n_w) = 0$ implies that $v(-T(n_w)) = \gamma + v(b)$ and thus $-T(n_w) > b$ if $\gamma > 0$). Indeed, society reduces economy-wide search costs without any losses for the affected worker. Providing search subsidies $-(T(n_w) + b) > 0$ to a non-productive worker (i.e. $z(n_w) = 0$) is thus sub-optimal. Compared to in-work benefits, social assistance is a more efficient instrument to fight the poverty of such agents because this saves search costs.

4 Standard solutions without binding search margin

In order to clearly identify the impact of the search margin on the optimal tax system, we first discuss the case in which the participation constraint is not binding (i.e. $u(n_0) > v(b) + \gamma$ so that $u(n) \geq u(n_0) > v(b) + \gamma$ and $\eta_w = 0$). As all agents look for a job, all unemployment is involuntary. The results in this section generalize Ebert (1992) and Boadway, Cuff, and Marchand (2000) to a non-utilitarian government. Since we are especially interested in the optimal taxation of low earnings, we elaborate on two reasons why the least skilled workers may be bunched in the absence of a participation margin, namely bunching on account of violation of the monotonicity requirement (in sub-section 4.2) and $z = 0$ bunching (in sub-section 4.3). These two types of bunching provide important benchmarks for bunching on account of a participation margin (analyzed in section 5).

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

[9] If the search margin is not binding (as is the case in the next section), in contrast, inducing the worker to search yields a first-order gain in private welfare. This may justify the additional budgetary costs $-T(n_w) - b > 0$.

[10] If $z(n_w) > 0$, leaving one additional skill outside the labor force has consequences for incentive compatibility (since $u'(n_w) = z(n_w)/n_w^2 > 0$) and hence the intensive margin of workers. This may justify $b + T(n_w) < 0$ (see sub-section 5.3).

can thus formulate the government's maximization problem (7) as

$$\max_{\substack{n_z,x_z,u(.),\\ z(.),\omega(.)\geq 0,\\ z(n_0)\geq 0}} \int_{n_0}^{n_1} \left\{ \begin{array}{c} \bar{s}u(n)\phi(n)f(n) - \lambda_u(n)\left[u'(n) - \frac{z(n)}{n^2}\right] \\ -\lambda_z(n)\left[z'(n) - \omega(n)\right] + \lambda_E\left[f(n)\bar{s}T(n)\right] \end{array} \right\} dn \quad (8)$$

$$-\gamma\bar{s} + (1-\bar{s})\,v(b) - \lambda_E\left[(1-\bar{s})\,b + E\right].$$

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

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

4.1 No bunching

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

Lemma 4 *If* $n_w = n_0$, $\eta_w = 0$, $z(n_0) > 0$ *and* $\lambda_z(n) = 0$ *for all* $n \in [n_0, n_1]$, *the solution to maximization problem (8) satisfies*

$$\lambda_E = G(n_1), \quad (9)$$

$$\tau(n) = \frac{\frac{G(n)}{G(n_1)} - F(n)}{nf(n)} \geq 0 \text{ for all } n \in [n_0, n_1], \quad (10)$$

$$v'(x(n)) = \frac{1}{n(1-\tau(n))}, \quad (11)$$

$$u(n) = \frac{1}{n}\left(K - \bar{E} + \int_{n_0}^{n} v(x(t))\,dt\right), \quad (12)$$

$$z(n) = n(v(x(n)) - u(n)) = nv(x(n)) - \int_{n_0}^{n} v(x(t))\,dt + \bar{E} - K, \quad (13)$$

$$W = v(b)(1-\bar{s}) - \bar{s}\gamma + \bar{s}u(n_0)n_0G(n_1) + \bar{s}\int_{n_0}^{n_1}[G(n_1) - G(n)]v(x(n))dn, \quad (14)$$

where

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

$$\bar{E} \equiv b\frac{(1-\bar{s})}{\bar{s}} + \frac{E}{\bar{s}},$$

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. 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 (10) while using $\lambda_E = G(n_1)$, we obtain the following expression for the marginal tax rate:

$$\lambda_E \tau(n) n f(n) = \lambda_E (1 - F(n)) - (G(n_1) - G(n)). \tag{15}$$

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 (15).

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, higher ability agents find it more attractive to mimic type n. To prevent these substitution effects, the government has to decrease gross incomes of all workers who are more skilled than type n by one unit.[11]

[11] The following two steps show that $z(.)$ 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).

The right-hand side of (15) 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 (10) 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 $n_0<n<n_1$).[12] 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 $\lambda_E=\int_{n_0}^{n_1}\frac{\phi(n)}{n}f(n)dn$. 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 (10). The higher this productive capacity, the larger are the efficiency costs associated with a higher marginal tax rate and therefore the lower the marginal tax rate should be.[13]

Consumption of each skill depends only on the marginal wage rate $n(1-\tau(n))$ and not on government spending requirements $\bar{E}$, as consumption depends only on substitution effects and all income effects go into work effort. Expression (13) implies that additional public spending requirements are optimally financed by uniformly increasing gross incomes of all agents, i.e. $dz(n)/d\bar{E}=1$. 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.

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 $t>n$.

[12]This 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 (10) is positive.

[13]For 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.

4.2 Bunching due to violation monotonicity

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. More formally, the restriction $\omega(n) \geq 0$ becomes binding so that $\lambda_z(n) < 0$. 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 top (see Ebert (1992)).

In contrast to the other two types of 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.[14] Moreover, the numerical simulations based on a lognormal distribution (see section 6) suggest that with reasonable distributions featuring most of their probability mass near the median this bunching occurs only at the bottom of the skill distribution, as marginal tax rates increase especially rapidly at these skills.

Lemma 5 *The consumption path implied by equation (11) is non-monotone at the bottom (i.e. $x'(n_0) < 0$) if and only if*

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

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

$$x_b = x(n_b),$$
$$z_b = z(n_b),$$

where n_b is determined by the following equation

$$f(n_b)\left[\frac{\bar{F}(n_b)}{F(n_b)} - \frac{G(n_b)\,n_b}{F(n_b)}\right] = G(n_1)\,F(n_b) - G(n_b), \tag{16}$$

while $\tau(n_b) > 0$. Consumption for types $n \in [n_b, n_1]$ is determined by

$$v'(x(n)) = \frac{f(n)}{\frac{G(n_1)-G(n)}{\lambda_E} + f(n)\,n - [1-F(n)]}, \tag{17}$$

[14] Generalizing 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)).

where

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

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

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

where

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

and $\bar{E} \equiv b\frac{(1-\bar{s})}{\bar{s}} + \frac{E}{\bar{s}}$.

This type of bunching does not affect the marginal utility cost of government revenue λ_E as given by (9): 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 (10) and the consumption path (11) 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, when the lowest non-bunched 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.

4.2.1 comparative statics

Lemma 6 *In case of bunching due to violation of monotonicity, an increase in the public spending requirement $\bar{E}$ yields the following effects:*

$$\frac{d\lambda_E}{d\bar{E}} = 0,$$
$$\frac{dn_b}{d\bar{E}} = 0,$$
$$\frac{dx(.)}{d\bar{E}} = 0,$$
$$\frac{dz(.)}{d\bar{E}} = 1.$$

Higher public spending (due to a fall in $\bar{s}$ or a rise in E or b) 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 $\phi(.)$ (see equation (16)), the level of public spending does not impact the size of the bunching interval, either.

4.3 Bunching with zero work effort

If public spending becomes so low that $z(n_0)$ becomes zero, a further decrease in spending $\bar{E}$ implies that the non-negativity constraint on $z(n_0)$ is binding. With this constraint 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 (3), 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 $n \in [n_0, n_z]$.

This type of bunching provides an important benchmark case for the next section, in which we analyze a binding participation margin with positive search costs and an exogenous welfare benefit. The equilibrium with $z = 0$ bunching is the outcome also with a binding participation margin if the labor market does not suffer from any imperfections (i.e. both search costs and involuntary unemployment are zero), while the government can optimally set transfers to agents who do not produce anything $b = -\left.\tilde{T}(z)\right|_{z=0}$. Indeed, $z = 0$ bunching captures the case in which agents stop working on the intensive rather than the extensive margin.[15]

With positive search costs, a binding non-negativity constraint (i.e. $z(n_z) = 0$ so that $x(n_z) = z(n_z) - T(n_z) = -T(n_z)$) and a non-binding participation constraint (i.e. $u(n_z) = v(x(n_z)) - z(n_z)/n_z = v(x(n_z) > v(b) + \gamma$ so that $x(n_z) > b$) imply that the government provides a search subsidy to types $n \in [n_0, n_z]$ (i.e. $-T(n) - b > 0$). Intuitively, agents engage in costly search for a non-productive job (i.e. a job with $z = 0$) only if the government subsidizes them for doing so.

Lemma 7 *If the solutions in lemma 4 or lemma 5 imply* $z(n_0) < 0$, *then the solution to problem (8) can be characterized as follows. First, a non-empty*

[15] If the government cannot distinguish between working zero hours and having no job (so that $-\tilde{T}(0) = b$ and thus $-T(n_z) = b$), the non-negativity constraint on $z(n)$ can be binding only if search costs are zero (i.e. $\gamma = 0$). In that case, the non-negativity constraint on $z(n)$ is equivalent to the participation constraint; with zero search costs and the government not being able to distinguish between participation at zero hours and not participating, the extensive margin coincides with the intensive margin at $z = 0$.

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 (17). Furthermore,* λ_E *and* n_z *are determined by the following two equations*[16]

$$\lambda_E = \frac{\frac{\bar{F}(n_z)}{n_z} + G(n_1) - G(n_z)}{\frac{F(n_z)}{n_z v'(x(n_z, \lambda_E))} + 1 - F(n_z)}, \quad (18)$$

$$(1 - F(n_z)) n_z v(x(n_z)) + \bar{E} + F(n_z) x(n_z) = \int_{n_z}^{n_1} \{[nf(n) - [1 - F(n)]] v(x(n)) - f(n) x(n)\} dn. \quad (19)$$

If $x(n)$ *as determined by equation (17) is monotonically increasing in* n, *equation (19) is upward sloping. If in addition to the monotonicity of* $x(n)$, *the distribution of skills satisfies the monotone hazard rate property*

$$\frac{d\left(\frac{f(n)}{1-F(n)}\right)}{dn} > 0 \text{ for all } n \in [n_0, n_1], \quad (20)$$

then equation (18) is downward sloping.
Finally,

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

For skills $n \geq n_z$, marginal tax rates and consumption levels continue to be determined by equations (11) and (15). Unlike bunching due to the violation of monotonicity, $z = 0$ bunching does impact the marginal utility cost of government revenue λ_E. In particular, raising government spending increases work effort only outside the bunching interval (for skills $n > n_z$). Within the bunching interval, consumption is reduced so that utility of the bunched agents declines with the same amount as the marginal worker n_z (see the first term in the numerator at the right-hand side of (18)). Higher government spending is thus financed by not only more work effort but also less private consumption (the additional government revenue from lower consumption of

[16] The first equation has the advantage that it is easy to interpret, but the right-hand side also contains λ_E in the $v'(x(n_z))$ term. The appendix shows that this equation can be rewritten as $\lambda_E = \frac{\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right)(G(n_1) - G(n_z))}{n_z - F(n_z)\frac{1-F(n_z)}{f(n_z)}}$.

the bunched individuals is given by the first term in the denominator at the right-hand side of (18)). 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. $\tau(n_z) > 0$). Intuitively, 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 (15) 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. Equation (18) is downward sloping if $x'(n) > 0$ and condition (20) is satisfied. The intuition is the following. As n_z increases, a more skilled worker determines the consumption level of the bunched skills $[n_0, n_z]$. Since $x'(n) > 0$, this implies the least skilled agents enjoy higher consumption levels. At these higher consumption levels of these low-skilled agents, redistributing resources from these skills to the higher skills should not raise social welfare (i.e. $\lambda_u(n)$ should be continuous at n_z). This can be the case only if also the skills above n_z enjoy higher consumption and hence higher utility in response to the rise in n_z. This implies that λ_E must decline (see (17)).[17]

Equation (19), which can be interpreted as the government budget constraint, is upward sloping in (n_z, λ_E) space. At constant λ_E, a more skilled marginal worker n_z raises transfers to all non-productive workers (as $-T(n) = x(n) = x(n_z)$ for $n = n_z$ while the monotonicity requirement implies that $x(n_z)$ rises with n_z) and reduces net taxes on all productive workers (as putting an additional worker out of work requires reducing the net tax level of that type – and thus of all workers above it). To bring the government budget back into balance, a higher level of λ_E must reduce consumption ((17) implies that a higher level of λ_E reduces consumption at fixed n) and raise the work effort of more productive skills. Indeed, the further we move up on this curve, the more we redistribute resources from the rich to the poor

[17]This is true only if a decline in λ_E does not increase consumption of the skills $[n_0, n_z]$ by more than that of the skills $[n_z, n_1]$. Condition (20), which is satisfied for the lognormal distribution employed in the simulations below, is sufficient for a fall in λ_E to benefit the group $[n_z, n_1]$ relatively more than $[n_0, n_z]$.

(the associated positive income effects for the poor induce more low-skilled agents to stop working).

4.3.1 comparative statics

Lemma 8 *In case of $z = 0$ bunching and under the assumption that condition (20) holds, an increase in $\bar{E}$ (due to a fall in $\bar{s}$ or a rise in E or b) yields the following effects*

$$\frac{dn_z}{d\bar{E}} < 0,$$
$$\frac{d\lambda_E}{d\bar{E}} > 0,$$
$$\frac{dx_z}{d\bar{E}} < 0,$$
$$\frac{dT(n)}{d\bar{E}} > 0$$

for n close enough to n_z, and for all $n > n_z$ we find that

$$\frac{d\tau(n)}{d\bar{E}} > 0,$$
$$\frac{dx(n)}{d\bar{E}} < 0,$$
$$\frac{du(n)}{d\bar{E}} < 0.$$

A higher level of public spending shifts the cost of redistribution upwards, thereby reducing the number of agents who are non productive. In contrast to the case without $z = 0$ bunching, a higher level of government spending raises the marginal cost of public funds. Intuitively, at higher levels of public spending, more of the required resources need to come from additional work effort rather than from less private consumption. Unlike the case with bunching due to violation of monotonicity, more public spending reduces the size of the bunching interval. Higher welfare benefits raising $\bar{E}$ thus increase the number of agents who exert positive work effort, as higher spending requirements leave less room for generous in-work benefits. In addition to all bunched agents, also agents $n > n_z$ consume less (in the face of a higher marginal tax burden) and suffer a decline in utility.[18]

[18] The 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 public spending yields not only higher work effort

5 Solutions with binding search margins

This section considers the case where the participation margin is binding (i.e. $u(n_w) = v(b) + \gamma$ and $\eta_w \neq 0$). With a binding participation margin, we can distinguish the cases without ($n_w = n_0$) and with voluntary unemployment ($n_w > n_0$). These two cases are discussed in turn.

5.1 No voluntary unemployment

If the search margin is binding, the solution to the government's optimization problem has the following characteristics.

Proposition 9 *In the case where $u(n_0) = v(b) + \gamma$ and $\eta_0 > 0$, we have*

$$z(n_0) > 0$$

$$f(n_0)\,\bar{s}\lambda_E\,(b + T(n_0)) \geq \eta_0 u'(n_0) > 0 \tag{21}$$

$$\frac{b + T(n_0)}{z(n_0)} \geq \tau(n_0) > 0 \tag{22}$$

$$\lambda_E = \frac{G(n_1)}{1 - \tau(n_0) f(n_0) n_0} > G(n_1) \tag{23}$$

As above, $x(n)$ is determined by equation (17). Utility $u(.)$ and production $z(.)$ are given by the same expressions as in lemma 4, while welfare can be written as

$$W = v(b) + \bar{s}(v(b) + \gamma)[n_0 G(n_1) - 1] + \bar{s}\int_{n_0}^{n_1} (G(n_1) - G(n))\, v(x(n))\, dn$$

The observation that production $z(n_0)$ is strictly positive was proved in lemma 3. Expression (21) is the condition that implies a corner solution $n_w = n_0$. Intuitively, with a binding participation constraint (i.e. $u(n_0) = v(b) + \gamma$), the lowest skill n_0 is indifferent between searching or not. Encouraging the least able type to look for a job therefore yields no direct first-order welfare effects for this type. The first inequality in (21) therefore compares the external benefits and costs of inducing the least skilled to search. The left-hand side represents the external benefits of encouraging these types to look for work; the government budget benefits from low-skilled employment if

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 $-T(n) - b > 0$.

welfare benefits exceed in-work benefits (i.e. $b+T(n_0)>0$). The right-hand side stands for the external costs of employing the least skilled in terms of tightening the incentive compatibility constraint for work effort. With the lowest skill being employed, higher skills find it more attractive to mimic these skills by providing less work effort.

The government thus faces a trade-off between obtaining revenues from either inducing all agents to search so as to maximize overall employment or encourage a smaller group of employed agents to work harder. One can state this dilemma also as one between raising production through more employment and raising it through a higher labor productivity level (interpreting y as work effort rather than hours worked), or, alternatively, as one between increasing labor supply on the extensive margin and raising it on the intensive margin. Indeed, we can rewrite equation (21) as (22), that is in terms of a trade-off between the distortion of the participation (or extensive) margin and the effort (or intensive) margin.[19] In particular, all workers should search if the average tax on work at the bottom (i.e. the left-hand side of (22)) exceeds the marginal tax rate on additional work effort at the bottom (i.e. the right-hand side of (22)). Intuitively, if the distortion on the extensive margin is more serious than the distortion on the intensive margin of labor supply, then all agents should search because combatting distortions on the extensive margin by inducing agents to search is more important than fighting distortions on the intensive margin by encouraging agents to work harder.

The marginal tax rate on the lowest skill is positive (i.e. $\tau(n_0)>0$). This result contrasts with the familiar result from the optimal tax literature that, in the absence of bunching at the bottom and a participation constraint, the lowest skill should face a zero marginal tax rate (see Seade (1977)). Intuitively, with a binding search constraint, the government is forced to raise utility $u(n_0)=v(b)+\gamma$ of the least skilled type n_0 above the level (associated with a zero marginal tax rate $\tau(n_0)=0$) chosen in the absence of this constraint. To prevent higher skills from mimicing the more attractive

[19] With a binding participation contraint, violation of the monotonicity constraint is less likely. The reason is that a binding participation contraint limits the increase in the marginal tax rate at the bottom of the skill distribution by lifting the marginal tax rate at the bottom $\tau(n_0)$ above zero. In mathematical terms, with a larger value of λ_E as a result of a binding participation contraint, $v'(x(n,\lambda_E))=\frac{f(n)}{\frac{G(n_1)-G(n)}{\lambda_E}+f(n)n-[1-F(n)]}$ is more likely to be decreasing in n (i.e. as λ_E increases, the derivative of the right-hand side with respect to n becomes smaller). Lemma 20 in the appendix characterizes the case in which both the participation and monotonicity constraint $x'(n)\geq 0$ are binding at n_0. Unlike the case without a binding participation constraint, this type of bunching affects the marginal utility cost of government revenue λ_E.

income bundle of the lowest type n_0, the government distorts the work effort of this latter type. In particular, by reducing (net and gross) incomes $x(n_0)$ and $z(n_0)$, a strictly positive marginal tax rate $\tau(n_0) > 0$ makes the income bundle $(x(n_0), z(n_0))$ less attractive for higher types. This result parallels that of a monopolist engaging in second-order price discrimination by offering a menu of goods with different qualities and prices. As shown by Mussa and Rosen (1978), such a monopolist offers an inefficiently low quality level to the lowest type in order to be able to extract more rents from buyers featuring higher reservation prices. Also in that application, the participation constraint of the lowest type gives rise to a distortion at the bottom. In our application, the government optimally adjusts the taxation of low-skilled agents in order to induce these agents to continue to look for work, while at the same time minimizing the tax revenues that have to be given up to higher types. This result of a positive marginal tax rate at the bottom illustrates the importance of examining the tax system in conjunction with social assistance. Indeed, the government reduces the tax burden on low-skilled workers as an instrument to contain welfare spending rather than to alleviate poverty of these workers.

Without the binding participation margin, the marginal cost of public funds depends only on the distribution of skills (i.e. $\lambda_E = G(n_1)$, see (9)). If the participation margin is binding, λ_E needs to be larger (i.e. $\lambda_E > G(n_1)$).[20] The intuition behind the higher marginal cost of public funds is that the additional behavioral margin of labor-market search makes labor supply behavior more sensitive to tax distortions. In particular, agents can respond to taxation by not only adjusting work effort but also search behavior. Hence, in addition to marginal tax rates, also average tax rates impact labor supply. Indeed, a uniform increase in the tax burden is no longer equivalent to a lump-sum tax, since such a tax increase induces low-skilled agents to stop searching for a job.

[20] This contrasts with the $z = 0$ bunching case, where λ_E is smaller rather than larger than $G(n_1)$.

5.1.1 comparative statics with respect to public spending

Lemma 10 *If* $u(n_0) = v(b) + \gamma$ *and* $\eta_0 > 0$, *marginal changes in the public spending requirement* $\bar{E}$ *yield the following comparative static effects:*

$$\begin{aligned}
\frac{dn_w}{d\bar{E}} &= 0, \\
\frac{d\lambda_E}{d\bar{E}} &> 0, \\
\frac{dT(n_0)}{d\bar{E}} &< 0, \\
\frac{du(n_0)}{d\bar{E}} &= 0, \\
\frac{dz(n_0)}{d\bar{E}} &< 0, \\
\frac{du(n_1)}{d\bar{E}} &< 0,
\end{aligned}$$

and for all $n \geq n_0$

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

Rather than uniformly increasing the tax burden throughout the skill distribution (as in the case without a binding participation margin (see section 4)), the government optimally protects the utility level of the least able workers so as to prevent these workers from leaving the labor market. Instead, it gradually increases the tax burden with skill so as to minimize the distortions in work effort. Compared to the case without a binding participation margin, the rich thus finance a larger part of the additional government spending, as this participation margin renders the labor-supply response of low-skilled workers more elastic. The higher marginal tax rates associated with a more progressive tax system induce all workers to cut their consumption. This erodes the tax base, thereby increasing the marginal cost of public funds λ_E (see also equation (23)).

5.1.2 comparative statics with respect to the welfare system

To explore the interaction between social assistance and the labor tax, we consider the impact of a higher welfare benefit b. These effects are similar to the comparative statics in lemma 10, except that $\frac{du(n_0)}{db} > 0$ as an increase in b raises $u(n_0) = v(b) + \gamma$. Indeed, a higher welfare benefit induces the government to cut the average tax burden on low-skilled workers in order to encourage them to continue to look for a job. The least skilled workers thus indirectly benefit from more generous social assistance, even though only the involuntarily unemployed collect these benefits. Both the lower tax burden on low earnings and the higher welfare benefits for the involuntary unemployed are financed by skilled workers. In this way, the increased importance of the participation margin raises marginal tax rates throughout the skill distribution, thereby worsening distortions on the intensive margin of labor supply. Reflecting these distortionary substitution effects, all workers cut their consumption. Despite their lower consumption level, the least able workers enjoy a higher level of utility arising from a lower average tax burden resulting in less work effort.[21]

Lemma 11 *If $u(n_0) = v(b) + \gamma$ and $\eta_0 > 0$, marginal changes in search costs γ yield the following comparative static effects:*

$$\frac{dn_w}{d\gamma} = 0,$$
$$\frac{d\lambda_E}{d\gamma} > 0,$$

$$\frac{dT(n_0)}{d\gamma} < 0,$$
$$\frac{du(n_0)}{d\gamma} > 0,$$
$$\frac{dz(n_0)}{d\gamma} < 0,$$
$$\frac{du(n_1)}{d\gamma} < 0,$$

and for all $n > n_0$

$$\frac{d\tau(n)}{d\gamma} > 0,$$
$$\frac{dx(n)}{d\gamma} < 0.$$

[21] These impacts on low-skilled workers are similar to those found by Weymark (1987) for an increase in the welfare weight for the least skilled agent. Indeed, a binding participation constraint for the least able types n_0 implies that the government attaches a higher weight to transfers to these types than in the absence of the participation constraint. A higher welfare benefit increases this weight further.

Lower search costs γ indicate more efficient labor market matching associated with a more flexible labor market. Alternatively, a decline in γ can be viewed as resulting from more active labor-market policies helping agents to find work. Search costs are reduced, for example, by offering agents jobs in the public sector in the context of welfare-to-work programs. Lower search costs substitute for in-work benefits as an instrument to induce low-skilled workers to search. The higher tax revenues from low-skilled workers allow the government to cut marginal taxes throughout the skill distribution. In this way, welfare-to-work programs help to alleviate the distortions on the intensive margin of labor supply. *Ceteris paribus* the level of welfare benefits, these programs redistribute resources from the poor to the rich.

5.2 With voluntary unemployment

With a binding participation margin, the first inequality in (21) does not necessarily hold. In that case, the government finds it optimal not to encourage the least able agents to search for a job; these agents thus remain voluntarily unemployed. Types $n_0 \leq n < n_w$ are bunched without any labor income and instead collect the welfare benefit b. Consequently, the government employs passive welfare benefits instead of active labor-market policies to fight the poverty of these low-skilled agents. The lower bound on the observed wage distribution n_w can be viewed as the effective minimum wage (or wage floor) implied by the welfare benefit b and the taxes on labor income.

For the marginal worker n_w, the equivalent of the first inequality in (21) is binding:[22]

$$b + T\left(n_w\right) = \tau(n_w)z(n_w). \tag{24}$$

This expression states that for the marginal worker the distortion on the extensive margin (i.e. the left-hand side of (24)) should equal the distortion on the intensive margin (i.e. the right-hand side of (24)).[23] To understand this equation, one should first note that the marginal type is indifferent between

[22] As noted above, we ignore violations of monotonicity in the main text. With a binding participation constraint and voluntary unemployment, violation of the monotonicity requirement is less likely (see also the numerical simulations below). The reason is that the search margin takes those low-skilled agents out of the labor market who, in the absence of the search margin, would be bunched (with positive work effort) on account of violation of monotonicity.

[23] This expression can be stated as $\frac{b+T(n_w)}{z(n_w)} = \tau(n_w)$, where the right-hand side is closely related to the replacement rate (in after-tax terms). $\frac{b+T(n_w)}{z(n_w)}$ indicates which part of gross labor income in a marginal job $z(n_w)$ is taxed away through withdrawn welfare benefits b and additional labor taxes $T(n_w)$.

participating and not participating, hence this agent's utility does not enter the equation determining n_w. Effects on the government budget and other agents thus determine the optimal employment level. The left-hand side of (24) represents the direct budgetary implications of raising employment by reducing n_w: by bringing a marginal worker into work, the government saves a welfare benefit b and collects additional tax revenue $T(n_w)$. The indirect implications, namely the effects on other workers, are captured by the right-hand side of (24)). Bringing a marginal type n_w into work encourages workers who are marginally more skilled to work less hard – as they can now mimic type n_w.[24] An optimal tax system balances the welfare implications of this latter behavioral response on the intensive margin of the more productive workers (represented by the right-hand side of (24)[25]) with the budgetary implications of the behavioral response on the extensive margin of the marginal workers. The government thus faces a trade-off between obtaining revenues from either inducing more agents to search or encouraging a smaller group of agents to work harder. We now characterize the solution.

Proposition 12 *With voluntary unemployment (i.e. $n_w > n_0$), the solution can be characterized as follows*

$$v'(x_w) = \frac{v(x_w) - v(b) - \gamma}{x_w - b}, \tag{25}$$

$$\lambda_E = \frac{G(n_1) - G(n_w)}{1 - F(n_w) - n_w f(n_w) + \frac{f(n_w)}{v'(x_w)}}, \tag{26}$$

$$\begin{aligned} & n_w (1 - F(n_w)) (v(b) + \gamma) + F(n_w) b + \bar{E} \\ = & \int_{n_w}^{n_1} \{[tf(t) - [1 - F(t)]] v(x(t)) - x(t) f(t)\} dt \end{aligned} \tag{27}$$

in the unknowns x_w, n_w and λ_E, where $x(n)$ is determined by equation (17)

[24] If the government could observe skills, it would eliminate the search distortion of the marginal worker because it does not need to worry about the implications of doing so for the intensive margin of more efficient workers (see Boone and Bovenberg (2001)).

[25] To see this, note that a higher employment level implies that marginal workers must enjoy more utility from work in order to convince them to search. This forces the government to increase also the utilities of all types above n_w so as to maintain incentive compatibility of work effort. Since $u'(n_w) = \frac{z(n_w)}{n_w^2}$, the type just above n_w enjoys an increase of utility $\frac{z(n_w)}{n_w^2}$ as n_w starts participating and hence decreases his production by $dz = \frac{z(n_w)}{n_w}$. Such a cut in gross incomes for all types above n_w leaves incentive compatibility intact. Using (15), the net welfare effect of reducing gross output in this way, $\left[\lambda_E (1 - F(n_w)) - \int_{n_w}^{n_1} \frac{\phi(t) f(t)}{t} dt\right] \frac{z(n_w)}{n_w}$, can be written (in terms of government revenues per additional marginal worker employed) as the right-hand side of (24).

for $n \geq n_w$ and $x_w = x(n_w)$. The solution satisfies the properties $x_w > b$, $n_w \in \langle n_0, n_1 \rangle$, $\lambda_E > 0$, and (if $\gamma > 0$)

$$\left.\frac{d\lambda_E}{dn_w}\right|_{(n_w,\lambda_E)\ satisfying\ (26)} > 0 > \left.\frac{d\lambda_E}{dn_w}\right|_{(n_w,\lambda_E)\ statisfying\ (27)}. \tag{28}$$

The solution for the marginal tax rate can be written as

$$\tau(n) = 1 - \frac{\frac{G(n_1)-G(n)}{\lambda_E} + f(n)n - [1 - F(n)]}{nf(n)}, \tag{29}$$

for $n \geq n_w$. The (in work) utility of a type $n \geq n_w$ is determined by

$$u(n) = \frac{1}{n}\left(K + \int_{n_w}^{n} v(x(t))\,dt\right),$$

where K is defined as

$$K \equiv \frac{1}{1 - F(n_w)}\left[\begin{array}{c} \int_{n_w}^{n_1} \{[tf(t) - [1 - F(t)]]\, v(x(t)) - x(t) f(t)\}\, dt \\ -F(n_w)\, b - \bar{E} \end{array}\right].$$

Gross income of a type $n \geq n_w$ is determined by

$$z(n) = n(v(x(n)) - u(n)).$$

Overall welfare amounts to

$$\begin{aligned} W &= v(b) + \bar{s}(v(b) + \gamma)[n_w(G(n_1) - G(n_w)) - (1 - F(n_w))] + \\ &\quad + \bar{s}\int_{n_w}^{n_1} (G(n_1) - G(n))\, v(x(n))\, dn. \end{aligned}$$

The system is recursive in that equation (25) determining the consumption level of the marginal worker, $x_w = x(n_w)$, depends on neither λ_E nor n_w. As shown in the appendix, this expression is equivalent to (24). It determines the minimum production level (and given the binding participation constraint also the minimum consumption level) that makes it worthwhile for the government to encourage an agent to search for work and thus employ active labor-market policies rather than welfare benefits as an instrument to alleviate poverty. The production in a marginal job needs to be sufficient to offset the search costs and the decline in production of more efficient workers. The consumption level x_w can thus be viewed as the reservation income for the government to have agents search.

To see what determines the consumption gap between the lowest working type and the unemployed, $x_w - b$, we substitute the Taylor expansion $v(b) =$

$v(x_w) + v'(x_w)(b - x_w) + 1/2v''(\xi)(b - x_w)^2$ (where $b < \xi < x_w$) into (25). This yields $\gamma = -1/2v''(\xi)(b - x_w)^2$, so that reservation consumption is a mark-up on the welfare benefit

$$x_w = b + \sqrt{2\gamma/(-v''(\xi))} \tag{30}$$

The optimal gap between the consumption level of marginal workers, x_w, and that of the unemployed, b, thus depends on search costs γ and the concavity of utility from consumption, $v(.)$. Large search costs imply that marginal workers need to be rather productive (and thus enjoy relatively high consumption levels) in order to make it worthwhile to have them search for a job. A concave utility function, in contrast, implies that unequal consumption levels of x and b become rather costly. Hence, the government optimally limits the gap between consumption of workers and the unemployed.

Just as with the bunching cases in section 4, we find the optimum by deriving two relationships between λ_E and the marginal skill (n_w in this case). In (n_w, λ_E) space (as illustrated in figure 2), (26) is upward sloping and the government budget constraint (27) is downward sloping in equilibrium.[26] (26) models the optimal trade-off between the incentive constraints (i.e. first-order incentive compatibility and the participation constraint) at different employment levels. This relationship captures the marginal benefits of using active labor-market policies (rather than passive welfare benefits) as an instrument to combat poverty. These benefits rise with the level of unemployment. More generous welfare benefits and larger search costs reduce these benefits, thereby shifting the upward-sloping relationship (26) downward in (n_w, λ_E) space. The government budget constraint (27) represents the costs of these active labor-market policies.

As shown in figure 2, the curve (26) features an asymptote at $n_w = \tilde{n}_w$ in (λ_E, n_w) space. In order to interpret $\tilde{n}_w$, we define $g(n_w, \lambda_E)$ as government expenditure that can be financed (with a marginal utility cost of public funds λ_E and a marginal skill level n_w)

$$\begin{aligned} g(n_w, \lambda_E) \equiv\ & \bar{s} \int_{n_w}^{n_1} \{[tf(t) - [1 - F(t)]]\, v(x(t)) - f(t)\, x(t)\}\, dt \\ & -n_w (1 - F(n_w))\, \bar{s}\, (v(b) + \gamma) + b\, [1 - (1 - F(n_w))\, \bar{s}]\,, \end{aligned}$$

where $x(t)$ is determined by equation (17).

[26] If the two equations do not intersect because equation (26) lies everywhere above (27), then $n_w = n_0$. Hence, we are in the situation considered in section 5.1 without any voluntary unemployment. If equation (26) lies everywhere below (27), then the incentive constraints do not allow the government to finance its spending requirement.

Lemma 13 *Type $\tilde{n}_w$ provides a zero net contribution to the budget in the sense that*

$$\left.\frac{\partial g(n_w,\lambda_E)}{\partial n_w}\right|_{(n_w,\lambda_E)\ \textit{satisfying (26)}} = \begin{cases} >0 & \textit{if} \quad n_w \in [n_0,\tilde{n}_w\rangle \\ =0 & \textit{if} \quad n_w = \tilde{n}_w \\ <0 & \textit{if} \quad n_w \in \langle \tilde{n}_w, n_1] \end{cases}.$$

Furthermore,

$$T(\tilde{n}_w) + b > 0.$$

At $n_w = \tilde{n}_w$, the government collects the largest possible tax revenues (net of spending on welfare benefits). Accordingly, $g(\tilde{n}_w, \lambda_E)$ is the maximum amount of government spending E that can financed. A Leviathan government, which maximizes tax revenues $g(n_w, \lambda_E)$ and thus sets $n_w = \tilde{n}_w$, distorts the extensive margin (i.e. $T(\tilde{n}_w) + b > 0$) because this allows it to collect more tax revenues from types $n_w > \tilde{n}_w$. Indeed, by leaving type n'_w for which $T(n'_w) + b = 0$ out of the labor force, types $n_w > n'_w$ can no longer mimic n'_w so that the government can raise more tax revenues from these types $n_w > n'_w$. To offset the adverse effects of additional employment on tax revenues from higher types, the government budget must benefit directly from the employment of a marginal type (i.e. $T(\tilde{n}_w) + b > 0$).

In the solution to the planner's problem, we have $n_w < \tilde{n}_w$. Accordingly, employment is subsidized in the sense that a reduction in employment would yield additional revenues (i.e. $\frac{\partial g(n_w,\lambda_E)}{\partial n_w} > 0$). The government budget constraint (27) is thus downward sloping over the relevant range (i.e. around the equilibrium value determined by (26) and (27)):[27] a higher level of employment (i.e. a decrease in n_w) harms the government budget and thus requires an increase in the marginal cost of public funds λ_E to bring the budget back into balance. This contrasts with the case in which the non-negativity constraint on gross income is binding. In that case, the government budget constraint is upward sloping: a reduction in employment raises the marginal costs of public funds.

5.2.1 comparative statics with respect to public spending

Lemma 14 *With $u(n_w) = v(b) + \gamma$ and $n_w > n_0$, we have*

$$\frac{dn_w}{dE} > 0,$$
$$\frac{d\lambda_E}{dE} > 0,$$

[27] This assumes that search costs γ are positive. If $\gamma = 0$, the government budget constraint is horizontal at the equilibrium. Intuitively, in the absence of search costs, inducing agents to search through active labor-market policies does not involve any budgetary costs.

$$\frac{dT(n_1)}{dE} > 0,$$
$$\frac{dz(n_1)}{dE} > 0.$$

For agents $n > n_w$, we find that

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

More public spending raises the marginal costs of active labor-market policies, which as indicated above are a net drain on the government budget (since $n_w < \tilde{n}_w$). As a direct consequence, marginal skills n_w leave the labor market. In contrast to the case in which both $z = 0$ bunching and a binding participation margin are absent (see sub-sections 4.1 and 4.2), more public spending raises marginal tax rates and the marginal cost of public funds. Intuitively, in the presence of the additional behavioral margin of search, government spending can no longer be financed in a non-distortionary fashion. Indeed, the binding participation margin implies that a uniform increase in labor taxes (i.e. $dT(n) = dT > 0$ for $n_w \geq n_0$) no longer acts as a lump-sum tax but harms labor supply of the marginal worker. To alleviate distortions on the extensive labor-supply margin of low-skilled workers, the government raises a relatively large share of the required additional tax revenues from higher skills. The implied increase in marginal tax rates induces workers to substitute away from consumption towards leisure. As public spending increases, the implied distortions on the intensive labor-supply margin worsen, thereby further increasing the marginal costs of public spending. Equation (65) in the appendix writes marginal cost of public funds (26) as

$$\lambda_E = \left[\frac{G(n_1) - G(n_w)}{1 - F(n_w)}\right] \frac{1}{1-\chi}, \tag{31}$$

where $\chi \equiv \frac{n_w f(n_w)}{1-F(n_w)} \frac{b+T(n_w)}{z(n_w)} = \frac{n_w f(n_w)}{1-F(n_w)} \tau(n_w)$. Whereas the first term (between square brackets) stands for the marginal cost in the absence of behavioral responses, the second term $\frac{1}{1-\chi}$ represents additional costs as a consequence of behavioral responses.

We can compare the response to higher public spending also with the corresponding response in an economy in which the non-negativity constraint on gross income is binding (see sub-section 4.3). In both economies, part of the additional public spending is financed through lower private consumption rather than additional work effort. Moreover, higher public spending raises marginal tax rates. A major difference, however, is that the number of productive workers rises rather than falls in an economy in which workers stop working on the intensive margin (i.e. an economy in which the non-negativity constraint on gross income is binding).[28] The reason for this different employment response in the latter economy is that more productive workers help the government budget by reducing spending on generous in-work benefits. If the participation constraint is binding and the welfare benefit b is exogenous, in contrast, more employment increases budgetary pressures as it requires more spending on costly active labor market policies to induce workers to enter the labor market. This illustrates how a binding participation margin fundamentally alters the structure of an economy and hence the constraints faced by a government setting its tax policy.

5.2.2 comparative statics with respect to the welfare system

Lemma 15 *With $u(n_w) = v(b) + \gamma$ and $n_w > n_0$, we have*

$$\frac{dn_w}{db} > 0, \frac{dn_w}{d\gamma} > 0,$$

and values $\bar{E} < g(\tilde{n}_w, +\infty)$ and $\bar{\gamma} > 0$ exist such that

$$\frac{d\lambda_E}{db} > 0, \frac{d\lambda_E}{d\gamma} > 0$$

for all $E \in \langle \bar{E}, g(\tilde{n}_w, \lambda_E) \rangle$ and $\gamma \in \langle 0, \bar{\gamma} \rangle$. In this case, we also find

$$\begin{aligned} \frac{d\tau(n)}{db} &> 0, \frac{d\tau(n)}{d\gamma} > 0, \\ \frac{dx(n)}{db} &< 0, \frac{dx(n)}{d\gamma} < 0 \end{aligned}$$

for all $n \in \langle n_w, n_1 \rangle$.
Although a set of rather low values of $E \in \langle E_*, E^* \rangle$ exists such that

$$\frac{d\lambda_E}{db} < 0, \frac{d\lambda_E}{d\gamma} < 0$$

[28] These differences in employment responses resemble the different responses with observable types (see Boone and Bovenberg (2001)) in a so-called normal economy (in which additional spending reduces productive employment) versus a so-called rich economy (in which additional spending raises productive employment).

for all $E \in \langle E_*, E^* \rangle$*, we can exclude this possibility if* $\tau(n_w) > 0, \frac{d\tau(n)}{dn} \leq 0$ *around* n_w *and a rise in* b *is not Pareto improving. In this latter case, in addition to the effects on* $\tau(n)$ *and* $x(n)$ *above, we have*

$$\frac{dT(n)}{db} < 0, \frac{dT(n)}{d\gamma} < 0$$

for n *close to* n_w*.*

Larger unemployment benefits raise unemployment through two channels. First, the marginal cost of active labor-market policies increase as higher in-work benefits have to be paid to low-skilled agents in order to induce them to enter the labor market. This effect shifts the government budget locus (27) upward. Second, the marginal benefits of active labor-market policies decrease as the welfare system takes on a larger role in combatting poverty. As a direct consequence, less redistribution has to be carried out through the tax system. Indeed, a higher welfare benefit raises the productivity requirements for marginal workers, x_w, so that the benefit locus of active labor market policies (26) shifts down. With the marginal costs of these policies rising and the associated benefits falling, voluntary unemployment unambiguously rises. With higher welfare benefits, society relies more on these passive benefits rather than on in-work tax benefits to redistribute resources towards low-skilled agents.

The impact of larger welfare benefits on the marginal costs of public funds (and hence on marginal tax rates and consumption levels of workers) depends on which effect is stronger: the fall in benefits of active labor market policies or the increase in costs of these policies. The cost effect dominates if in (n_w, λ_E) space the downward-sloping government budget locus (27) is relatively flat compared to the upward-sloping benefit locus (26). This is in fact the normal case, implying that a higher welfare benefit typically increases the marginal costs of public funds and hence marginal tax rates.[29] In fact, one can show that this is indeed the case if search costs γ are relatively small (so that the budget locus is relatively flat) or public spending E is large (so that the benefit curve is close to its asymptote $\tilde{n}_w$ so that this curve is steep). Furthermore, the marginal cost of public funds (and hence

[29] The benefit effect dominates the cost effect (and the marginal cost of public funds λ_E thus declines) only if the benefit curve (26) is flat enough. The benefit curve is flat if agents are almost bunched in the sense that consumption levels do not rise quickly with n. That is, the equilibrium value of n_w is close to $\underline{n}$ in figure 2. In that case, the higher productivity requirements associated with a higher welfare benefit induces a substantial number of agents to stop searching, thereby saving substantial search costs (especially because the steep government budget curve signals that search costs γ are high).

marginal tax rates) increases with b if marginal tax rates are positive and declining with skill close to the marginal worker n_w and if a rise in b is not Pareto improving.[30] In that case, in addition to the unemployed, also marginal workers (i.e. workers with skills close to n_w) benefit from higher welfare benefits. The result that λ_E and hence marginal tax rates rise as a result of larger welfare benefits is analogous to the unambiguous impacts of more generous social assistance on these variables if all workers search (see section 5.1.2). Also in that case, higher welfare benefits raise in-work benefits for marginal workers, which are financed through higher tax rates on skilled workers, implying higher marginal tax rates and a more progressive tax system.

We can compare the comparative static results on higher welfare benefits for an economy in which the non-negativity constraint on gross incomes rather than the participation constraint is binding. At very low welfare benefits (and relatively low spending requirements E), increasing these benefits raises productive employment, as higher welfare benefits reduce the budgetary room to pay generous in-work benefits to low-skilled workers in unproductive jobs. More agents have to work in order to help finance the larger benefits to the involuntarily unemployed. If welfare benefits are raised, the participation margin eventually becomes binding, and further increasing the welfare benefit reduces employment. Hence, the relationship between welfare benefits and productive employment is U-shaped. At low welfare benefits, the adverse income effects associated with higher taxation on account of higher welfare discourage agents from exiting the labor market through the intensive margin. At high welfare benefits, in contrast, larger welfare benefits encourage agents to exit the labor market through the extensive margin.

As a result of these different employment impacts of welfare benefits, also the relationship between these benefits and the level of in-work benefits appears to be U-shaped. At low welfare benefits, the participation margin is not binding. As welfare benefits are raised from a low initial level, these benefits absorb the budgetary room for generous in-work benefits as an instrument to fight poverty. At low welfare benefits, social assistance and in-work benefits are thus substitutes in fighting poverty. As welfare benefits are increased further, however, the participation constraint for marginal workers becomes binding and the government needs to raise in-work benefits to encourage agents to look for work. At high levels of social assistance, therefore, in-work benefits and welfare benefits become complements: in-work benefits help to offset the impact of more generous social assistance on the partici-

[30] Although we do not analyze the case in which b is set to maximize welfare, it seems reasonable to exclude the case in which a rise in b yields a Pareto improvement.

pation constraint. This U-shaped relationship between in-work benefits and social assistance reveals that generous in-work benefits are called for in both countries with low and high welfare benefits, but for different reasons. In countries with low welfare benefits (such as the United States), in-work benefits are aimed at poverty alleviation. In countries with more generous social assistance (such as most European countries), in contrast, in-work benefits protect the incentives to participate in the labor market.

The effects of smaller search costs have the same sign as the effects of smaller unemployment benefits. Hence, in the normal case, government policies aimed at reducing search costs help to decrease marginal tax rates throughout the skill distribution.

5.3 Excessive employment

This section shows that in-work benefits may be so large that search is actually subsidized, i.e. $b + T(n_w) < 0$.

Proposition 16 *If $n_w > n_0$ it is possible that*

$$\begin{aligned} \eta_w &< 0, \\ b + T(n_w) &< 0, \\ \tau(n_w) &< 0. \end{aligned}$$

In this case we have

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

for all types $n > n_w$ where $\tau(n) < 0$.

With positive search costs and a binding non-negativity constraint on gross income, sub-section 4.3 showed that the government provides a search subsidy to marginal workers n_z. The proposition shows that with voluntary unemployment (implying that the participation constraint rather than the non-negativity constraint on gross income is binding), the government may also optimally provide a search subsidy to marginal workers n_w. The condition for the optimal level of voluntary unemployment $f(n_w)\bar{s}\lambda_E(b + T(n_w)) = \eta_w u'(n_w)$ implies that a search subsidy for the marginal worker $b + T(n_w) < 0$ is associated with $\eta_w < 0$ so that a further increase in employment harms welfare. Moreover, the negative shadow value η_w implies that participation restriction $u(n_w) = v(b) + \gamma$ is binding from below: $u(n) \leq v(b) + \gamma$ for $n < n_w$ is the relevant constraint. Instead of attracting types $n \geq n_w$ into work, the government wants to keep types $n < n_w$ out of the labor force. This provides a rationale for hiring and firing costs. Employment could also be discouraged by imposing a minimum earnings requirement on jobs. Hence, by prohibiting jobs that earn less than $z_{\min} > 0$, the government facilitates generous tax benefits to low-skilled workers.

In the presence of search subsidies $b + T(n_w) < 0$, the government faces a trade-off between redistributing to agents with low productivity and containing search costs γ.[31] With exogenous welfare benefits, it can redistribute to low productivity agents only by providing rather generous in-work benefits. Additional redistribution, however, results in more entry into the labor force. With search subsidies $b + T(n_w) < 0$, the government budget bears part of the search costs associated with this additional entry: an increase in employment implies a direct burden on the government budget, as the additional in-work benefits $-T(n_w)$ exceed the welfare benefit b the government saves. These additional budgetary costs of redistribution stop the government from redistributing more to the bottom of the skill distribution and explain why the shadow price η_w is negative. Search subsidies are likely to be optimal if not only b but also E are rather low (so that the government has the budgetary means to provide generous in-work benefits to unskilled workers), while at the same time search costs γ are high (so that the government wants to contain entry costs). The simulations in section 6 confirm this.

We overturn the standard result from the literature on optimal non-linear income taxation that marginal tax rates should be positive in the interior of the skill distribution. The negative marginal tax rate $\tau(n_w)$[32] can be explained in two ways. First, the government likes to redistribute to less productive agents but at the same time wants to avoid excessive search. To reconcile these two objectives, it grants the largest in-work subsidies to types above n_w rather than to n_w (as this would draw types $n < n_w$ into the labor force). Hence, the marginal tax rate is negative at n_w. Second, in order to separate the types at the bottom of the labor force (and thus make it

[31] In the presence of search subsidies, voluntary unemployment can exist only with positive search costs. Indeed, agents refrain from search only if they have to incur search costs in order to take advantage of the higher public benefits in employment (i.e. $-T(n_w) > b$).

[32] Crucial for our result that marginal tax rates can be negative is that the government can not freely set b to optimize social welfare. If the government could freely manipulate the welfare benefit b, it would increase it. In this way, it could redistribute resources to low skilled agents without having to rely on in-work benefits that require low skilled agents to engage in costly search.With inward migration, however, the government may find it hard to affect the outside option of migrants. Indeed, a government facing inward migration of low-skilled labor may want to make the tax system less progressive compared to the case in which it does not face the participation constraint of migrants. Moreover, the welfare benefit b may be set by another government agency than the tax authorities determining labor taxes. The tax authorities may thus have to treat b as being exogenous.

unattractive for the types below n_w to mimic type n_w by searching for a job), the government induces marginal workers n_w to supply excessive labor on the intensive margin (which is cheaper for type n_w than for types below n_w), implying that excessive labor supply on the extensive margin spreads to the intensive margin. Indeed, in contrast to traditional optimal tax problems, the self selection constraints are binding from below in the sense that they prevent low types from mimicing high types (rather than preventing high types from mimicing low types).

As noted in lemma 14, more public spending reduces employment. In the presence of search subsidies, this helps to alleviate excessive employment. This side benefit of higher public spending contains the marginal cost of public funds. Indeed, expression (31) implies that the marginal costs of public funds are below $G(n_1)$, its level in a model without a binding participation constraint and without $z = 0$ bunching (see sub-sections 4.1 and 4.2).[33]

6 Numerical simulations

The simulations assume that the government is utilitarian and that utility from consumption is given by $v(x) = 2x^{\frac{1}{2}}$. Log productivity is distributed according to a truncated normal distribution with a mean of 3 and a standard deviation of 0.5. The distribution is truncated at a maximum skill n_1 of 100 and, in all simulations except the first one, at a minimum skill n_0 of 4. The various simulations assume different values for the government spending requirement $\bar{E}$, the welfare benefit b, and search costs γ.

In order to exclude bunching due to violation of monotonicity, we increase the minimum skill n_0 to 10, which is about half the median skill of 20.[34] In this case (see the first column of Table 1), the marginal tax rate rises rapidly from zero at $n_0 = 10$ to its maximum of about 19% close to the median skill, after which it gradually declines to about 15% at the 95th percentile of the skill distribution (at about $n = 45$). These marginal tax rates do not depend on the spending requirement $\bar{E} \geq 0$, as $z(n_0)$ stays positive even at $\bar{E} = 0$.

[33] The first term at the right-hand side of (31), $\frac{G(n_1)-G(n_w)}{1-F(n_w)}$, is declining in n_w and thus smaller than $G(n_1)$ if $n_w > n_0$. The second term is smaller than one if $T(n_w) + b < 0$. Hence, $\lambda_E \leq G(n_1)$. This inequity holds also for the case that the non-negativity constraint on $z(n)$ is binding.

[34] Such a high minimum skill can be justified by assuming that the government sets its gross minimum wage equal to this minimum skill n_0 and provides disability benefits to all agents featuring less skills than $n_0 = 10$. This interpretation assumes that the government can observe which workers have a skill level below $n_0 = 10$. For the case that the government can observe individual skills levels exactly, see Boone and Bovenberg (2001).

At zero public spending, $u(n_0) = 18.1$ so that the participation margin does not play a role as long as $v(b) + \gamma = 2b^{\frac{1}{2}} + \gamma < 18.1$.

If we adopt the benchmark minimum skill of $n_0 = 4$, the monotonicity requirement gives rise to bunching between the minimum skill of 4 and $n_b = 5.4$ (see the second column of Table 1). The few agents with these few skills feature a consumption level x of 10.8 and, with a public spending requirement $\bar{E}$ of 120, gross incomes z of 1.8. In this case, the marginal tax rate starts at a relatively high level close to 40% at $n_b = 5.4$, after which it declines to about 24% at the medium skill and 16% at the 95th percentile of the skill distribution. The participation margin is not relevant as long as $2b^{\frac{1}{2}} + \gamma < u(n_0) = 6.1$.

Bunching on account of a binding monotonicity requirement is replaced by $z = 0$ bunching if public spending is reduced to $\bar{E} = 25$ while $b = 10$ and $\gamma = 1$ (see the third column of Table 1). The resulting "American" economy features low government spending and a low welfare benefit. Skills below $n_z = 10.5$ are not productive and thus feature zero gross incomes. Hence, 10% of the population that is not involuntarily unemployed does not provide any work effort. The marginal tax rate at n_z amounts to 22%, after which it rises somewhat to a maximum of close to 25% at $n = 15$, before it declines again to some 16% at the 95th percentile. The in-work benefit at the marginal skill level n_z of $-T(n_z) = 68$ compares to a tax payment $T(n_{95\%}) = 255$ corresponding to an average tax rate $T(n_{95\%})/z(n_{95\%})$ of 15% at the 95th percentile. A somewhat higher spending level $\bar{E} = 50$ requires workers to exert more work effort, thereby reducing the non-productive part of the (non-involuntarily unemployed) population to close to 6% as the marginal skill level declines to $n_z = 9.4$ (see the fourth column of Table 1).

By increasing not only the spending level to $\bar{E} = 50$ but also the welfare benefit to $b = 40$, we arrive at our benchmark case (see the first column of Table 2). This case features a binding participation margin with voluntary unemployment. In particular, skills $n < n_w = 11.1$ do not search for a job but rather collect the welfare benefit, implying that 12 % of the (non-involuntarily unemployed) population does not enter the labor force. The marginal tax rate at the wage floor n_w amounts to 15%, reaching a maximum of 21% near the median, before gradually declining to 16% at the 95th percentile. The average tax rate increases from just below zero at the marginal worker n_w to 15% at the 95th percentile.

Raising public spending or the welfare benefit further, we produce a "European" economy featuring high spending and welfare levels. To illustrate, if we increase the welfare benefit to $b = 70$, about half of the population does not enter the labor force (see the second column of Table 2). In this case, the marginal tax rate at the marginal skill level n_w amounts to 43%. At the 95th percentile, the marginal rate is 21%. With a more heavily distorted participation margin and lower employment, a larger part of the spending requirement must be financed by the higher skills. Indeed, the 95th percentile skill level now faces an average tax rate close to 25%. Raising public spending to $\bar{E} = 100$ while leaving the welfare benefit b at 40, we find that marginal tax rates approach 50% at the effective minimum wage $n_w = 17.6$ (see the third column of Table 2)

Search is subsidized if both public spending and the welfare benefit are quite low at respectively $\bar{E} = 25$ and $b = 10$, while search costs are large at $\gamma = 10$ (see the fourth column of Table 2).[35] The substantial search costs yield a high wage floor of $n_w = 15.4$. At the same time, low public spending and welfare benefits imply that the government subsidizes search at a rate of about 2% (i.e. $(T(n_w) + b)/z(n_w) = -.02$). The marginal tax rate rises from this level to a maximum of 13% around a skill level of 35.

The case with a binding participation margin but without any involuntary unemployment emerges at high spending levels $\bar{E} = 150$ and low search costs and welfare benefits. In the absence of a participation margin, the least skill agents would feature negative utility levels $u(4) = -1.36$. Hence, even without search costs and a welfare benefit (i.e. $b = \gamma = 0$), the participation margin is binding (see the last column of Table 2).[36] Hence, the marginal costs of public funds increases beyond $G(n_1) = 0.056$. This eliminates bunching due to a binding monotonicity requirement and lifts the marginal tax rate at the bottom to over 60%. In this case, therefore, just as in the first simulation in which we raised n_0 to 10, bunching is eliminated altogether and the agents featuring the least skills thus exert positive work effort.

A final simulation considers the benchmark case with a non-utilitarian government who attaches a welfare weight of $\phi(n) = \frac{1/\sqrt{n}}{\int_{n_0}^{n_1} f(t)/\sqrt{t}dt}$ to skill n. This simulation (see the last column of Table 1) shows that, *ceteris paribus* the welfare benefit, a more egalitarian government lowers unemployment. Intuitively, such a government is willing to spend more resources on active

[35] This case is similar in all respects to the U.S. economy discussed above, except that search costs are considerably larger. These larger search costs cause agents to stop working on the extensive rather than the intensive margin. In both cases, however, the government provides search subsidies to marginal workers.

[36] Another way to arrive at this case without dramatically raising public spending and reducing b and γ is to increase n_0. To illustrate, we arrived at this case by combining $n_0 = 7.5$ and $\gamma = 1$ with European spending levels $\bar{E} = 100$ and American welfare levels $b = 10$. In this case, the increase in λ_E was not sufficient to eliminate bunching on account of violation of monotonicity. For the relevant equations in this case, see Lemma 20 in the appendix.

labor-market policies for the low skilled. In fact, in-work benefits are so generous that all agents search, even though the least productive 16% of the population does not produce anything in their jobs.

7 Conclusions

This paper has explored how the income tax system should optimally respond to an exogenously given welfare system in the presence of costly labor-market search and non-verifiable skills. We showed that optimal unemployment is determined by the requirement that distortions on the external margin balance those on the intensive margin. On the one hand, generous in-work benefits help to alleviate distortions on the participation margin by encouraging more low-skilled workers to actively look for work. On the other hand, such benefits make it more attractive for high-ability agents to mimic lower ability agents, thereby distorting work effort. The government thus faces a trade-off between boosting (low productive) employment and raising the work effort of higher skilled workers. When the government lacks information on individual skills, a distorted participation margin is therefore the price for combatting poverty while at the same time protecting the labor supply of higher skilled workers. A similar trade-off appears in determining an optimal retirement scheme. In particular, rather than linking public retirement benefits to the retirement age in an actuarially fair way, the government may want to favor early retirement to aid low productivity individuals suffering from poor health.

In the presence of low search costs, low welfare benefits, a concentrated skill distribution and large public spending, the government may find it optimal to employ a progressive tax system providing generous in-work benefits to low-skilled workers in order to induce all agents to search. Such a progressive tax system features a positive marginal tax rate at the bottom. This contrasts with the familiar result from the optimal tax literature that, in the absence of bunching at the bottom, those with the lowest skills should face a zero marginal tax rate (see Seade (1977)). This new result shows that the welfare system and the participation margin may importantly affect the optimal tax system.

Social assistance and positive search costs may overturn also the well-known result from the optimal tax literature that marginal taxes should be positive in the interior of the income distribution. We showed that the government may optimally increase in-work benefits with gross income (implying a negative marginal tax on work effort) in order to limit excessive entry into the labor market. In particular, whereas positive marginal tax rates at the bottom help to encourage search if labor-force participation of low-skilled workers is taxed on a net basis, negative marginal taxes rates for low-skilled workers help to discourage excessive entry of low-skilled workers if this entry is subsidized. This latter case provides a rationale also for minimum wages and hiring and firing costs.

The incorporation of labor-market imperfections and the welfare system into a model of optimal non-linear income taxation enabled us to investigate how these new elements impact the optimal income tax. We showed, for example, that lower search costs allow the government to cut marginal taxes throughout the skill distribution. In this way, active labor-market policies that enhance the flexibility of the labor market alleviate distortions on the intensive margin of labor supply. More generous welfare benefits, in contrast, tend to raise marginal tax rates, as the government cuts the average tax burden on low-skilled workers in order to encourage these workers to continue to look for a job. Accordingly, skilled workers finance not only more generous social assistance, but also a lower tax burden of low-skilled workers.

As a benchmark case for the main contribution of this paper, namely the introduction of a binding participation margin in a model of optimal income taxation, we elaborated on the case in which agents stop working on the intensive margin by reducing their work effort to zero. The comparative static results for this case help to show how a binding participation margin alters the constraints faced by a government setting its tax policy. To illustrate, the employment response to higher welfare benefits depends crucially on whether or not the participation margin is binding. If it is not, the adverse income effects associated with higher taxation on account of higher welfare discourage more agents from exiting the labor market through the intensive margin. With a binding participation margin, in contrast, higher welfare benefits induce additional agents to leave the labor market through the extensive margin.

In future research, we would like to investigate optimal tax policy if the government can simultaneously set welfare benefits, search obligations, and other categorical social insurance benefits (such as disability benefits based on a signal of skill type). A study of these issues would need to allow for imperfect information on search behavior, household structure, and skill types. In exploring the optimal trade-off between passive welfare benefits and active labor-market policies, we also would like to account for negative external effects of unemployment.

8 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.

Choné, P., and G. Laroque, 'Optimal Incentives for Labor Force Participation,' INSEE Working Paper 2001-25, Paris.

Diamond, P.A., 1980, 'Income Taxation with Fixed Hours of Work,' Journal of Public Economics, Vol. 13, pp. 101-110.

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).

Kanbur, R., M. Keen, and M. Tuomala, 'Optimal Non-Linear Income Taxation for the Alleviation of Income-Poverty,' *European Economic Review*, Vol. 38, pp. 1613-1632.

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.

Mortensen, D.T. and C.A. Pissarides, 1999, New developments in models of search in the labor market, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3 (North-Holland, Amsterdam).

Mussa, M. and S. Rosen, 1978, Monopoly and product quality, *Journal of Economic Theory*, Vol. 18, p. 301-317.

OECD, 2001, *Education at a Glance*, Paris.

Rochet, J. and L. Stole, 2002, Nonlinear Pricing with Random Participation, *Review of Economic Studies*, Vol. 69, p. 277-311,

Sadka, E., 1976, On Income Distribution, Incentive Effects and Optimal Income Taxation, *Review of Economic Studies*, Vol. 43, p. 261-267.

Saez, E., 2000, Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses, NBER Discussion Paper No. 7708.

Seade, J.K., 1977, On the Shape of Optimal Tax Schedules, *Journal of Public Economics*, Vol. 7, pp. 203-235.

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.

9 Appendix: proofs of results

9.1 Preliminary results

The dynamic optimization problems in this paper are not standard. In particular, section 5 includes the constraint $u(n_w) \geq v(b) + \gamma$, where both the function $u(.)$ and the type n_w are endogenous. Similarly, section 4.3 features the constraint $z(n_z) \geq 0$, where both $z(.)$ and n_z are endogenous. Since standard references as Kamien and Schwartz (1981) do not provide the necessary conditions for these problems, we derive the required lemma ourselves. The lemma below provides the necessary conditions for the optimality of $u(.)$, which can be readily adapted to obtain the optimality of $z(.)$.

The maximization with respect to $u(.)$ and n_w exhibits the following structure

$$\max_{u(.),n_w \geq n_0} \int_{n_w}^{n_1} L(n, u(n), \dot{u}(n))\, dn + \Psi(n_w, u_w),$$

where $u_w \equiv u(n_w)$ is the starting value for $u(.)$ at n_w, and

$$\begin{aligned} L(n, u(n), \dot{u}(n)) &\equiv \bar{s}u(n)f(n)\phi(n) - \lambda_u(n)\left[u'(n) - \frac{z(n)}{n^2}\right] \\ &\quad +\lambda_E f(n)\bar{s}\underbrace{\left[z(n) - v^{-1}\left(\frac{z(n)}{n} + u(n)\right)\right]}_{=T(n)}, \\ \Psi(n_w, u_w) &\equiv \left\{ \begin{array}{c} \bar{F}(n_w)\, v(b) + \left[1 - \bar{F}(n_w)\right](-\gamma\bar{s} + (1-\bar{s})\, v(b)) + \\ -\lambda_E \left[F(n_w) + (1 - F(n_w))(1-\bar{s})\right] b - \eta_w(\gamma - u_w + v(b)) \end{array} \right\}. \end{aligned}$$

Hence, we find for this particular optimization problem

$$\begin{aligned}
L_u &= -\bar{s} f(n)\left(\frac{\lambda_E}{v'(x(n))} - \phi(n)\right), \\
L_{\dot{u}} &= -\lambda_u(n), \\
\Psi_{u_w} &= \eta_w, \\
\Psi_{n_w} &= f(n_w)\,\bar{s}\,[\gamma + v(b) - \lambda_E b], \\
L(n_w) &= \bar{s}u(n_w)f(n_w)\phi(n_w) + \lambda_E f(n_w)\bar{s}T(n_w).
\end{aligned}$$

Substituting these expressions into the following lemma, we obtain the necessary conditions in this appendix.

Lemma 17 *Necessary conditions for $u^*(n), u^*(n_1), u^*(n_w^*) = u_w^*$ and $n_w^* \geq n_0$ to solve the following maximization problem*

$$\max_{u(.), n_w \geq n_0} \int_{n_w}^{n_1} L(n, u(n), \dot{u}(n))\, dn + \Psi(n_w, u_w),$$

where $u(n_1)$ and u_w are free, n_1 is exogenously fixed and n_w can be chosen freely as long as $n_w \geq n_0$, are

$$\begin{aligned}
L_u(n) - \frac{dL_{\dot{u}}(n)}{dn} &= 0, \\
L_{\dot{u}}(n_1) &= 0, \\
\Psi_{u_w} - L_{\dot{u}}(n_w) &= 0, \\
\Psi_{n_w} + L_{\dot{u}}(n_w)\,\dot{u}(n_w) - L(n_w) &= 0 \text{ if } n_w > n_0, \\
\left[\Psi_{n_w} + L_{\dot{u}}(n_w)\,\dot{u}(n_w) - L(n_w)\right]_{n_w = n_0} &\leq 0 \text{ if } n_w = n_0,
\end{aligned}$$

where L_u ($L_{\dot{u}}$) denotes the derivative of the function $L(n, u, \dot{u})$ with respect to u ($\dot{u}$) and $L_u(n)$ ($L_{\dot{u}}(n)$) denotes this derivative evaluated at the point n.

Proof. Define the function $g : \Re \rightarrow \Re$ as

$$g(a) = \int_{n_w^* + a\delta_n}^{n_1} L\left(n, u^*(n) + ah(n), \dot{u}^*(n) + a\dot{h}(n)\right) dn + \Psi(n_w^* + a\delta_n, u_w^* + a\delta_u)$$

for some function $h(.)$ and scalars δ_n and δ_u. Since, by definition, $u^*(n), u^*(n_1), u^*(n_w^*) = u_w^*$ and $n_w^* \geq n_0$ solve the maximization problem above, the function $g(.)$ reaches a maximum at $a = 0$ for each admissable combination of $h(.), \delta_n$ and δ_u. Taking the derivative of $g(.)$ with respect to a and evaluating it at $a = 0$, we write the necessary condition for this maximum $g'(0) = 0$ as

$$-\delta_n L(n_w^*) + \int_{n_w^*}^{n_1} \left[L_u(n)\, h(n) + L_{\dot{u}}(n)\, \dot{h}(n)\right] dn + \Psi'_{n_w}(n_w^*, u_w^*)\,\delta_n + \Psi'_{u_w}(n_w^*, u_w^*)\,\delta_u = 0.$$

Through integration by parts, we find

$$\left\{ \begin{array}{c} L_{\dot{u}}(n_1)\,h(n_1) - L_{\dot{u}}(n_w^*)\,h(n_w^*) + \int_{n_w^*}^{n_1}\left[L_u(n) - \frac{dL_{\dot{u}}(n)}{dn}\right]h(n)\,dn + \\ \left(\Psi'_{n_w}(n_w^*, u_w^*) - L(n_w^*)\right)\delta_n + \Psi'_{u_w}(n_w^*, u_w^*)\,\delta_u \end{array} \right\} = 0.$$

$\delta_n = \delta_u = h(n_1) = h(n_w) = 0$ yields an admissable alternative function. Hence to satisfy the equation above, we must have

$$\int_{n_w^*}^{n_1}\left[L_u(n) - \frac{dL_{\dot{u}}(n)}{dn}\right]h(n)\,dn = 0$$

for all admissable $h(.)$. This implies

$$L_u(n) - \frac{dL_{\dot{u}}(n)}{dn} = 0$$

for each n. Substituting this into the equation above, we obtain

$$L_{\dot{u}}(n_1)\,h(n_1) - L_{\dot{u}}(n_w^*)\,h(n_w^*) + \left(\Psi'_{n_w}(n_w^*, u_w^*) - L(n_w^*)\right)\delta_n + \Psi'_{u_w}(n_w^*, u_w^*)\,\delta_u = 0.$$

Also $\delta_n = \delta_u = h(n_w) = 0$ yields an admissable alternative function. Since $u(n_1)$ is free so that $h(n_1)$ can be both positive and negative, it must be the case that

$$L_{\dot{u}}(n_1) = 0.$$

We are thus left with

$$-L_{\dot{u}}(n_w^*)\,h(n_w^*) + \left(\Psi'_{n_w}(n_w^*, u_w^*) - L(n_w^*)\right)\delta_n + \Psi'_{u_w}(n_w^*, u_w^*)\,\delta_u = 0.$$

Note that the relationship between δ_u, δ_n and $h(.)$ is

$$\delta_u = h(n_w^*) + \dot{u}(n_w^*)\,\delta_n.$$

This can be seen as follows

$$\begin{aligned} a\delta_u &= u(n_w^* + a\delta_n) - u_w^* \\ &\approx u(n_w^*) - u^*(n_w^*) + a\dot{u}(n_w^*)\,\delta_n \\ &= ah(n_w^*) + a\dot{u}(n_w^*)\,\delta_n, \end{aligned}$$

where a first-order Taylor expansion yields the second line, while the third line follows from the definition of $h(.)$. Substitution of this relation into the equation above to eliminate $h(n_w^*)$ yields

$$\left(\Psi'_{u_w}(n_w^*, u_w^*) - L_{\dot{u}}(n_w^*)\right)\delta_u + \left(\Psi'_{n_w}(n_w^*, u_w^*) + L_{\dot{u}}(n_w^*)\,\dot{u}(n_w^*) - L(n_w^*)\right)\delta_n = 0.$$

$\delta_n = 0$ yields an admissable alternative. Since $u(n_w)$ is free and δ_u can therefore be both positive and negative, we must have

$$\Psi'_{u_w}(n_w^*, u_w^*) - L_{\dot{u}}(n_w^*) = 0.$$

Finally, we are left with

$$\left(\Psi'_{n_w}(n_w^*, u_w^*) + L_{\dot{u}}(n_w^*)\,\dot{u}(n_w^*) - L(n_w^*)\right)\delta_n = 0.$$

Here we face two possibilities. First, if $n_w^* > n_0$, δ_n can be both positive and negative so that

$$\Psi'_{n_w}(n_w^*, u_w^*) + L_{\dot{u}}(n_w^*)\,\dot{u}(n_w^*) - L(n_w^*) = 0.$$

Second, if $n_w^* = n_0$, the only admissable values for δ_n are nonnegative (for $a \downarrow 0$). Hence,

$$\left[\Psi'_{n_w}(n_w^*, u_w^*) + L_{\dot{u}}(n_w^*)\,\dot{u}(n_w^*) - L(n_w^*)\right]_{n_w = n_0} \leq 0$$

because otherwise one would optimally select $n_w > n_0$, contradicting the optimality of $n_w^* = n_0$. ■

9.2 Proofs of results in section 3

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 $x(.)$ and $z(.)$ (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 $g(.)$ has a minimum at $m = n$. The first order condition for this minimum ($g'(m)|_{m=n} = 0$) can be written as

$$[u'_x x'(m) + u'_z z'(m)]_{m=n} = 0, \tag{32}$$

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 $\tilde{g}(.)$ 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 (32))

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

or equivalently

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

at $n = m$. The second-order condition for the minimization of $\tilde{g}(.)$ 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 (32), 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.[37] 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 $n' > n$,[38] then $\frac{1}{t} < \frac{1}{n}$ for each $t > n$ implies that

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

[37] This proof follows closely the argument by Guesnerie and Laffont (1984).

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

Using equation (32), 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 3

The statement in the lemma can be rephrased as follows. If $\gamma > 0$ then $z(n_w) = 0$ implies $u(n_w) > v(b) + \gamma$, that is the participation margin is not binding. The first-order condition for n_w in maximization problem (7) can be written as

$$\eta_w u'(n_w) \leq f(n_w) \bar{s} \left[(-\gamma - v(b) + u(n_w)) \phi(n_w) + \lambda_E (b + T(n_w)) \right],$$

where the inequality can be strict only if $n_w = n_0$. Now assume (by contradiction) that not only $z(n_w) = 0$ but also $u(n_w) = v(b) + \gamma$. The first-order condition for n_w can then be written as (using $u'(n_w) = \frac{z(n_w)}{n_w^2} = 0$)

$$0 \leq f(n_w) \bar{s} \lambda_E (b + T(n_w))$$

so that $-T(n_w) \leq b$. With $z(n_w) = 0$, $-T(n_w) = x(n_w)$ and $v(x(n_w) = u(n_w)$ so that $-T(n_w) \leq b$ implies $u(n_w) \leq v(b)$. This contradicts the binding participation margin $v(x(n_w)) = u(n_w) = v(b) + \gamma$ if $\gamma > 0$. Q.E.D.

9.3 Proofs of results in section 4

Proof of lemma 4

The first-order conditions (Euler equations) for optimizing (7) with respect to $\omega(.), u(.)$ and $z(.)$ amount to (if $z(n_0) > 0$, $n_0 = n_w$ and $\eta_0 = 0$)

$$\omega(n) = \arg\max_{\omega \geq 0} \lambda_z(n) \omega, \quad (33)$$

$$\lambda_u'(n) = \bar{s} f(n) \left(\frac{\lambda_E}{v'(x(n))} - \phi(n) \right), \quad (34)$$

$$\lambda_z'(n) = -\frac{\lambda_u(n)}{n^2} + \lambda_E f(n) \bar{s} \left(\frac{1}{n v'(x(n))} - 1 \right), \quad (35)$$

together with the transversality conditions

$$\lambda_u(n_0) = \lambda_u(n_1) = 0,$$
$$\lambda_z(n_0) = \lambda_z(n_1) = 0,$$

and the government budget constraint (6).

Since by assumption $\lambda_z(n) = 0$ and thus $\lambda_z'(n) = 0$ for all n, (35) can be written as

$$\frac{1}{v'(x(n))} = n + \frac{1}{n}\frac{\lambda_u(n)}{\lambda_E f(n)\bar{s}}. \tag{36}$$

The first-order condition for maximizing individual utility with respect to $z(n)$ in equation (2) amounts to

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

or equivalently,

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

Using this in equation (36) to eliminate $v'(x(n))$, we find

$$\tau(n) = \frac{-\lambda_u(n)}{\lambda_E n^2 f(n)\bar{s}}. \tag{37}$$

Substituting equation (36) into (34) to eliminate $v'(x(n))$, we arrive at

$$\lambda_u'(n) = \frac{1}{n}\lambda_u(n) + \lambda_E \bar{s} f(n) n - \bar{s} f(n)\phi(n) \tag{38}$$

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

$$\lambda_u(n) = n\left[c_0 + \bar{s}\left(\lambda_E \int_{n_0}^{n} f(t)\,dt - \int_{n_0}^{n} \frac{f(t)\phi(t)}{t} dt\right)\right] \tag{39}$$

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

$$\lambda_u(n) = n\bar{s}\left[\lambda_E F(n) - G(n)\right]. \tag{40}$$

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

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

Substitution of this and (40) into equation (37) 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

$$z(n) = n(v(x(n)) - u(n)). \tag{41}$$

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):

$$u(n) = \frac{1}{n}\left(K - \bar{E} + \int_{n_0}^{n} v(x(t))\,dt\right) \tag{42}$$

for some constant K. To determine K, we substitute (42) into (41) and the result into the government budget constraint (6) 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 (42)) 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 $W = -\gamma\bar{s} + (1-\bar{s})\,v(b) + \int_{n_0}^{n_1} \bar{s}f(n)\,\phi(n)\,u(n)dn$ to eliminate $u(n)$, and employ partial integration. Q.E.D.

Proof of Lemma 5

Combining equations (36) and (40), we find

$$v'(x(n)) = \frac{f(n)}{\frac{G(n_1)-G(n)}{\lambda_E} + f(n)\,n - [1 - F(n)]} \tag{43}$$

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 4 is that $\lambda_z(n) < 0$ for n close to n_0 so that the optimal $\omega(n)$ determined by equation (33) 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 (33), (34), (35) and the transversality conditions still apply. Moreover, the analysis in the proof of lemma 4 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 (34) and using the transversality conditions $\lambda_u(n_0) = \lambda_u(n_1) = 0$ to arrive at

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

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

$$\lambda_E = \frac{1}{F(n_b)\frac{1}{v'(x_b)} + \int_{n_b}^{n_1} \frac{f(n)}{v'(x(n))} dn}. \tag{44}$$

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

$$\lambda_E = \frac{\bar{F}(n_b) + \left(n_b - \frac{F(n_b)}{f(n_b)}\right)(G(n_1) - G(n_b))}{n_b - F(n_b)\frac{1-F(n_b)}{f(n_b)}} \tag{45}$$

Step A. Eliminating $v'(x(n))$ from $\int_{n_b}^{n_1} \frac{f(n)}{v'(x(n))} dn$ by using equation (43) for $n \geq n_b$ and employing partial integration, we find $\int_{n_b}^{n_1} \frac{f(n)}{v'(x(n))} dn = -n_b \frac{G(n_1)}{\lambda_E} + \frac{1}{\lambda_E}\left[n_b G(n_b) + \left[1 - \bar{F}(n_b)\right]\right] - n_b F(n_b) + n_b$.

Step B. Combining this with the observation that $v'(x_b) = v'(x(n_b))$ and using (43) for $n = n_b$, we establish

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

Step C. Substituting this into the denominator of equation (44) and solving for λ_E, we arrive at (45).

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 (35))

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

We solve $\lambda_u(n)$ for $n < n_b$ by integrating (34) and employing the transversality condition $\lambda_u(n_0) = 0$:

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

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

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

Using (43) 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

$$\lambda_E = \frac{G(n_1) - G(n_b) + \frac{G(n_b) f(n_b) n_b}{F(n_b)} - \frac{f(n_b)}{F(n_b)} \bar{F}(n_b)}{1 - F(n_b)}. \tag{46}$$

Equating the two expressions ((45) and (46)) for λ_E, we find (16) determining n_b in the lemma. Substituting equation (16) into either (45) or (46), 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 4, taking into account that $x(n) = x_b$ for $n \in [n_0, n_b]$. Q.E.D.

Proof of Lemma 6

$\frac{d\lambda_E}{d\bar{E}} = 0$ follows immediately from the result that $\lambda_E = G(n_1)$. Since $\bar{E}$ does not impact λ_E, consumption $x(.)$ (which is determined by (17)) is not affected by $\bar{E}$. (16) shows that n_b is determined completely by the distribution of skills and the rank order weights $\phi(.)$ and is thus not affected by $\bar{E}$ so that $\frac{dn_b}{d\bar{E}} = 0$. Finally, the equation for $z(.)$ in lemma 5 (taking the previous results into account) implies that $\frac{dz(n)}{d\bar{E}} = 1$ for all $n \in [n_0, n_1]$. Q.E.D.

Proof of Lemma 7

The main departure from the proof of lemma 4 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 (35), $\lambda_z(n_0) < 0$ implies that there exists $n_z > n_0$ such that $\lambda_z(n) < 0$ for all $n \in [n_0, n_z\rangle$. Hence, the optimal $\omega(n)$ determined by equation (33) equals $\omega(n) = z'(n) = 0$ for $n \in [n_0, n_z\rangle$. 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 4. In particular, equations (33), (34) and (35) 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 4 is correct for types $n \geq n_z$. The main difference with the proof of lemma 4 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 (38) starting from the endpoint n_1 and using $\lambda_u(n_1) = 0$

$$\lambda_{u+}(n) = -n\bar{s}\left[\lambda_E(1 - F(n)) - (G(n_1) - G(n))\right] \tag{47}$$

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 (34) 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

$$\lambda_{u-}(n) = \bar{s}\left[F(n)\frac{\lambda_E}{v'(x_z)} - \bar{F}(n)\right] \tag{48}$$

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

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

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

As in the proof of lemma 5, consumption $x(n)$ for $n \geq n_z$ is determined by equation (43). By thus solving the path for $x(.)$ 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 (43) for $n = n_z$ to eliminate $v'(x_z)$ from (18) and solving for λ_E, we arrive at the first relation between n_z and λ_E[39]

$$\lambda_E = \frac{\bar{F}(n_z) + \left(n_z - \frac{F(n_z)}{f(n_z)}\right)(G(n_1) - G(n_z))}{n_z - F(n_z)\frac{1-F(n_z)}{f(n_z)}}. \tag{49}$$

[39] Note the similarity with equation (45) above, which also follows from the condition that $\lambda_u(n)$ is continuous at the end of the bunching interval.

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 $u(n) = \frac{1}{n}\left(K_z - \bar{E} + \int_{n_z}^{n} v(x(t))\,dt\right)$ (from (42)) into (41) to eliminate $u(n)$, we obtain

$$z(n) = nv(x(n)) - K_z + \bar{E} - \int_{n_z}^{n} v(x(t))\,dt. \tag{50}$$

To solve for K_z, we employ (6), 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 $n \geq n_z$)

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

By integrating by parts, we rewrite this equation as

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

Substituting this expression for $\left(K_z - \bar{E}\right)$ into equation (50) and using $z(n_z) = 0$, we arrive at the second relation between $n_z$and λ_E, namely equation (19) in the lemma.

The next two lemmas derive some properties of the two equations (18) (or equivalently (49)) and (19).

Lemma 18 *Equation (18) in (n_z, λ_E) space has the following properties:*
♦ *it goes through the points $(n_0, G(n_1))$ and $\left(n_1, \frac{1}{n_1}\right)$,*
♦ *it is downward sloping for all $n_z \in [n_0, n_1]$ if $x(n)$ as determined by equation (17) is increasing in n and the distribution of skills satisfies the monotone hazard rate property (20),*
♦ *a point (n_z, λ_E) with $\lambda_E > G(n_1)$ cannot be part of a solution to the optimization problem (8) because it violates monotonicity for $n \geq n_z$.*

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

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

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 (47)) $\lambda_{u+}(n_0) = n_0 (G(n_1) - \lambda_E)$. Hence, $\lambda_{u+}(.)$ can be written as (from (34))

$$\lambda_{u+}(n, \lambda_E) = n_0 (G(n_1) - \lambda_E) + \int_{n_0}^{n} f(t) \left(\frac{\lambda_E}{v'(x(t))} - \phi(t) \right) dt, \qquad (51)$$

where $x(t)$ is determined by (43).

We can write $\lambda_{u-}(.)$ as (from (34))

$$\lambda_{u-}(n, n_z, \lambda_E) = \int_{n_0}^{n} f(t) \left(\frac{\lambda_E}{v'(x(n_z))} - \phi(t) \right) dt. \qquad (52)$$

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 (17) 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

$$\frac{d\lambda_E}{dn_z} = -\frac{F(n_z) \lambda_E \frac{-v''(n(n_z))}{(v'(x(n_z)))^2} x'(n_z)}{n_z - F(n_z) \frac{1-F(n_z)}{f(n_z)}}.$$

Hence, $\frac{d\lambda_E}{dn_z} < 0$ if $x'(n) > 0$ for all n and if the denominator is positive for all n, that is $n - F(n)\frac{1-F(n)}{f(n)} > 0$ for all n. To see that the latter condition follows from condition (20) note that

$$n_0 - F(n_0)\frac{1-F(n_0)}{f(n_0)} = n_0 > 0$$

Further, it is routine to verify that the monotone hazard rate property (20) implies

$$\frac{d\left(n - F(n)\frac{1-F(n)}{f(n)}\right)}{dn} > 0$$

for all n. Hence it follows that $n - F(n)\frac{1-F(n)}{f(n)} > 0$ for all $n \in [n_0, n_1]$.
♦ Consider a point (n_z, λ_E) with $\lambda_E > G(n_1)$. Then continuity of (49) implies that a point (n_z', λ_E') exists such that $n_z' > n_z$ and $\lambda_E' > G(n_1)$, which implies (from (47)) 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 (51) and (52)

$$\lim_{n\uparrow n_z'} x(n) > x(n_z').$$

By choosing n_z' close enough to n_z, this inequality implies $x'(n_z) < 0$ where $x(n_z)$ is evaluated at λ_E'. Since λ_E' is close to λ_E and $x(n)$ (as determined by equation (17)) 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 $x(.)$ is violated for $n \geq n_z$ and thus cannot be part of a solution to (8). ■

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

Proof. Define

$$\psi(n_z, \lambda_E) \equiv \int_{n_z}^{n_1} \{f(n)[nv(x(n)) - x(n)] - [1 - F(n)]v(x(n))\} dn$$
$$- \left[(1 - F(n_z)) n_z v(x(n_z)) + \bar{E} + F(n_z) x(n_z)\right].$$

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

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

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

where both inequalities hold if $x'(n) \geq 0$. Hence,

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

■

We thus have two curves in (n_z, λ_E) space. Assuming that condition (20) holds, 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 4 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 $(\tilde{n}_z, \lambda_E)$ 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 $(\tilde{n}_z, \lambda_E)$ 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 $(\tilde{n}_z, \lambda_E)$ 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 $\tau(n_z) > 0$. (37) 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,

$$\frac{\lambda_E}{v'(x(n_z))} - \phi(n_z) < 0 \tag{53}$$

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 (53). 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 8

As established in the proof of lemma 7, 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 $\bar{E}$. The proof of lemma 19 implies that (19) shifts upward (and to the left) as $\bar{E}$ increases. Hence, n_z falls and λ_E rises with $\bar{E}$.

Since λ_E reduces $x(n_z)$ and we have $x'(n_z) \geq 0$ and $dn_z/d\bar{E} < 0$, we find that $x_z = x(n_z)$ falls with $\bar{E}$. Furthermore, (17) 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 $\bar{E}$ 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, $\frac{dx_z}{d\bar{E}} < 0$ implies that a value $n^* > n_z$ exists such that

$$\frac{dT(n)}{d\bar{E}} > 0$$

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

9.4 Proofs of results in section 5

The main text focuses on one type of bunching at the time. The next lemma considers possible combinations of bunching. In particular, it excludes several bunching combinations and provides the necessary conditions for an optimum in the other cases not considered in the main text.

The second-order condition $z'(n) \geq 0$ implies that $z(n) \geq 0$ is satisfied for all $n > n_w$ if it is met for n_w. We therefore have to add only the constraint $z(n_w) \geq 0$ (with the corresponding Lagrange multiplier δ_w) to the optimization problem (7):

$$\begin{aligned} &\max_{\substack{n_w, u(.), z(.), \\ \omega(.) \geq 0}} \bar{F}(n_w) v(b) + \left[1 - \bar{F}(n_w)\right] (-\gamma \bar{s} + (1 - \bar{s}) v(b)) \\ &+ \int_{n_w}^{n_1} \left\{ \begin{array}{c} \bar{s} u(n) \phi(n) f(n) - \lambda_u(n) \left[u'(n) - \frac{z(n)}{n^2}\right] \\ -\lambda_z(n) \left[z'(n) - \omega(n)\right] + \lambda_E \left[f(n) \bar{s} T(n)\right] \end{array} \right\} dn \\ &- \lambda_E \left\{ b \left[F(n_w) + (1 - F(n_w))(1 - \bar{s})\right] + E \right\} \\ &- \eta_w (\gamma - u(n_w) + v(b)) + \delta_w z(n_w) \end{aligned}$$

Lemma 20 *Based on the values for the Lagrange multipliers $\eta_w, \delta_w = \delta(n_w)$ and $\lambda_z(n_w)$, we can distinguish the following cases:*

	$\eta_w < 0$			$\eta_w = 0$			$\eta_w > 0$	
	$\delta_w > 0$	$\delta_w = 0$		$\delta_w > 0$	$\delta_w = 0$		$\delta_w > 0$	$\delta_w = 0$
$\lambda_z(n) = 0$	$\emptyset$	A	$\lambda_z = 0$	$\emptyset$	C	$\lambda_z = 0$	$\emptyset$	F
$\lambda_z(n) < 0^*$	$\emptyset$	B	$\lambda_z < 0$	D	E	$\lambda_z < 0$	$\emptyset$	G

* *In particular we focus on the case where $\lambda_z(n) < 0$ for n close to n_w*

The cases denoted by $\emptyset$ can be excluded, while the cases B and G are not considered in the main text. The equations characterizing the solution in these *latter two cases are as follows. As in propositions 9 and 12,* n_w *is the lowest type with positive search effort. In addition to that, a type* n_b *exists such that*

$$x(n) = x(n_b)$$

for all $n \in [n_w, n_b]$*, where* $x(.)$ *is determined by equation (17) for all* $n \geq n_b$*. The equations determining* $u(.)$ *and* $z(.)$ *are the same as those in proposition 12, with* $x(.)$ *being constant over the range* $[n_w, n_b]$*.*
If $n_w = n_0$*, the endogenous variables* n_b, λ_E *and* η_0 *are determined by*

$$\begin{aligned}
\int_{n_0}^{n_b} \left\{ \frac{-\frac{\eta_0 + \int_{n_0}^{n} \bar{s} f(t) \left(\frac{\lambda_E}{v'(x_b)} - \phi(t) \right) dt}{n^2} +}{\lambda_E f(n) \bar{s} \left(\frac{1}{n v'(x_b)} - 1 \right)} dn \right\} &= 0, \\
\eta_0 + \int_{n_0}^{n} \bar{s} f(t) \left(\frac{\lambda_E}{v'(x_b)} - \phi(t) \right) dt &= \lambda_u(n_b), \\
\left\{ \begin{array}{c} \int_{n_0}^{n_b} \{ [t v(x_b) - x_b] f(t) - [1 - F(t)] v(x_b) \} dt + \\ \int_{n_b}^{n_1} \{ [t v(x(t)) - x(t)] f(t) - [1 - F(t)] v(x(t)) \} dt \end{array} \right\} &= n_0 (v(b) + \gamma) + \bar{E},
\end{aligned}$$

where $\lambda_u(n)$ *is determined by equation (54).*
If $n_w > n_0$*,* x_w *is determined by (25) and the endogenous variables* n_b, n_w, λ_E *and* η_w *are determined by*

$$\begin{aligned}
v'(x(n_b)) &= v'(x_w), \\
\int_{n_w}^{n_b} \left\{ \frac{-\frac{\eta_w + \int_{n_w}^{n} \bar{s} f(t) \left(\frac{\lambda_E}{v'(x_w)} - \phi(t) \right) dt}{n^2} +}{\lambda_E f(n) \bar{s} \left(\frac{1}{n v'(x_w)} - 1 \right)} dn \right\} &= 0, \\
\eta_w + \int_{n_w}^{n} \bar{s} f(t) \left(\frac{\lambda_E}{v'(x_w)} - \phi(t) \right) dt &= \lambda_u(n_b),
\end{aligned}$$

$$\begin{aligned}
& n_w (1 - F(n_w)) (v(b) + \gamma) + b \left[\frac{1}{\bar{s}} - (1 - F(n_w)) \right] + \frac{E}{\bar{s}} \\
= & \int_{n_w}^{n_b} \{ [t v(x_w) - x_w] f(t) - [1 - F(t)] v(x_w) \} dt + \\
& \int_{n_b}^{n_1} \{ [t v(x(t)) - x(t)] f(t) - [1 - F(t)] v(x(t)) \} dt,
\end{aligned}$$

where $\lambda_u(n)$ *is determined by equation (54).*

Proof

Lemma 3 implies that $\eta_w \neq 0$ cannot go together with $\delta_w > 0$ (i.e. $z(n_w) \geq 0$ is a binding constraint). Furthermore, using lemma 17 in the appendix, the transversality condition for $z(.)$ at n_w (satisfying $z(n_w) \geq 0$) can be written as

$$\lambda_z(n_w) + \delta_w = 0.$$

$\delta_w > 0$ thus excludes $\lambda_z(n_w) = 0$.

The derivation for the necessary conditions for optimality follows largely the derivations in the proofs of propositions 9 and 12. More precisely, the first-order conditions (or Euler equations) for $n_w, u(.), z(.)$ and $\omega(.)$ coincide. The difference is that in cases B and G, $\lambda_z(n) < 0$ around n_w so that $\omega(n) = 0$ around n_w. In particular, there exists $n_b > n_w$ such that $\omega(n) = 0$ for all $n \in [n_0, n_b]$. n_b is determined as follows. Since the transversality condition implies that $\lambda_z(n_w) = 0$ and since (by definition) n_b is the end of the bunching interval (so we have $\lambda_z(n_b) = 0$), it must be the case that

$$\int_{n_w}^{n_b} \lambda_z'(n)\, dn = 0.$$

Using the expression for $\lambda_z(n)$ in equation (35), we can write this as

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

Solving for $\lambda_u(n)$ by integrating equation (34) and taking into account that $x(n)$ is constant over the range $[n_w, n_b]$ and the transversality condition $\lambda_u(n_w) = \eta_w$, we obtain

$$\int_{n_w}^{n_b} \left\{ \begin{array}{c} -\frac{\eta_w + \int_{n_w}^{n} \bar{s} f(t) \left(\frac{\lambda_E}{v'(x_w)} - \phi(t) \right) dt}{n^2} + \\ \lambda_E f(n)\, \bar{s} \left(\frac{1}{nv'(x_w)} - 1 \right) \end{array} dn \right\} = 0,$$

where $\eta_w = \eta_0$ and $x_w = x_b$ if $n_w = n_0$.

$\lambda_u(n)$ can be solved also from n_1, which yields the solution given by (54). To ensure that $\lambda_u(n)$ is continuous at n_b, we need that

$$\eta_w + \int_{n_w}^{n} \bar{s} f(t) \left(\frac{\lambda_E}{v'(x_w)} - \phi(t) \right) dt = \lambda_u(n_b),$$

where $\lambda_u(n_b)$ is determined by (54) and $\eta_w = \eta_0, x_w = x_b$ if $n_w = n_0$.

Finally, the government budget constraint needs to be satisfied, which can be written as

$$
\begin{aligned}
& n_w\left(1-F\left(n_w\right)\right)\left(v\left(b\right)+\gamma\right)+b\left[\frac{1}{\bar{s}}-\left(1-F\left(n_w\right)\right)\right]+\frac{E}{\bar{s}} \\
= & \int_{n_w}^{n_b}\left\{\left[tv\left(x_w\right)-x_w\right]f\left(t\right)-\left[1-F\left(t\right)\right]v\left(x_w\right)\right\}dt+ \\
& \int_{n_b}^{n_1}\left\{\left[tv\left(x\left(t\right)\right)-x\left(t\right)\right]f\left(t\right)-\left[1-F\left(t\right)\right]v\left(x\left(t\right)\right)\right\}dt,
\end{aligned}
$$

where $x_w = x_b$ if $n_w = n_0$. Q.E.D.

Proof of Proposition 9

Since the optimal value for n_w is the corner solution $n_w = n_0$, the first-order condition for optimizing (7) with respect to n_w (evaluated at $n_w = n_0$) amounts to

$$\eta_0 u'\left(n_0\right) \leq f\left(n_0\right)\bar{s}\lambda_E\left(b+T\left(n_0\right)\right),$$

which yields the first inequality in (21). The optimality conditions for $\omega\left(.\right), u\left(.\right)$ and $z\left(.\right)$ are the same as equations (33), (34) and (35). We know from Lemma 2 that $z\left(n_0\right) > 0$ if the participation constraint is binding (and $\gamma > 0$), which establishes the first inequity in the proposition. Furthermore, $z\left(n_0\right) > 0$ implies that the transversality conditions for $z\left(.\right)$ are

$$\lambda_z\left(n_0\right) = \lambda_z\left(n_1\right) = 0.$$

Using the adjusted transversality condition derived in lemma 17 for $u\left(.\right)$ at $n_w = n_0$, we obtain

$$
\begin{aligned}
\lambda_u\left(n_0\right)+\eta_0 &= 0, \\
\lambda_u\left(n_1\right) &= 0.
\end{aligned}
$$

The government would like to reduce $u\left(n_0\right)$, but is prevented from doing so because of the constraint $u\left(n_0\right) \geq v\left(b\right)+\gamma$. Hence, $\lambda_u\left(n_0\right) < 0$ and $\eta_0 > 0$. The latter inequity and $u'\left(n_0\right) = \frac{z(n_0)}{n_0^2} > 0$ imply the second inequality in (21).

Also here, the differential equation determining $\lambda_u\left(n\right)$ is given by equation (38) (since $\lambda_z\left(n\right) = 0$ and thus $\lambda_z'\left(n\right) = 0$ for all n). By using $\lambda_u(n_1) = 0$, we can solve this differential equation as

$$\lambda_u\left(n\right) = -n\bar{s}\left[\lambda_E\left(1-F\left(n\right)\right)-\left(G\left(n_1\right)-G\left(n\right)\right)\right]. \tag{54}$$

Substituting this solution into (36), we arrive at (17).

$\lambda_u(n_0) = -\eta_0 < 0$ and (54) (for $n = n_0$) imply that

$$n_0 \bar{s} [\lambda_E - G(n_1)] = \eta_0 > 0, \tag{55}$$

so that $\lambda_E > G(n_1)$.

As above, $\tau(n)$ is determined by (37), hence (employing the transversality condition $\lambda_u(n_0) = -\eta_0$ to eliminate $\lambda_u(n_0)$ in (37) for $n = n_0$)

$$\tau(n_0) = \frac{\eta_0}{\lambda_E n_0^2 f(n_0) \bar{s}} > 0,$$

which establishes the second inequality in (22). Using equation (21), this can be rewritten as the first inequality in (22):

$$\tau(n_0) = \frac{\eta_0}{\lambda_E n_0^2 f(n_0) \bar{s}} \leq \frac{b + T(n_0)}{u'(n_0) n_0^2} = \frac{b + T(n_0)}{z(n_0)},$$

where we have used incentive compatibility $u'(n_0) = \frac{z(n_0)}{n_0^2}$ to eliminate $u'(n_0)$ from the inequality.

Substituting $\eta_0 = \tau(n_0) \lambda_E n_0^2 f(n_0) \bar{s}$ (see the previous equation) to eliminate η_0 from (55), we arrive at

$$\lambda_E (1 - n_0 f(n_0) \tau(n_0)) = G(n_1),$$

which yields the equality in (23).

Finally, the expression for welfare W follows from substituting $u(n_0) = v(b) + \gamma$ into (14). Q.E.D.

Proof of lemma 10 and lemma 11

Since $n_w = n_0$ is a corner solution, marginal changes in $\bar{E}$ and γ do not affect n_w. We prove $\frac{d\lambda_E}{d\bar{E}}, \frac{d\lambda_E}{d\gamma} > 0$ by contradiction. Suppose that $\frac{d\lambda_E}{d\bar{E}}, \frac{d\lambda_E}{d\gamma} < 0$. $u(n_0) = v(b) + \gamma$ implies that n_0 does not lose (if E increases) and gains (if b or γ increase). (17) and $\frac{d\lambda_E}{d\bar{E}}, \frac{d\lambda_E}{d\gamma} < 0$ imply $\frac{dx(n)}{d\bar{E}}, \frac{dx(n)}{d\gamma} > 0$ for all $n > n_0$. Writing utility as (from (42))

$$u(n) = \frac{n_0}{n} u(n_0) + \frac{1}{n} \int_{n_0}^{n} v(x(t))\, dt,$$

we find that all types $n > n_0$ gain if $\bar{E}$, b, or γ rise. This Pareto improvement contradicts the optimality of the original paths for $u(.)$, $z(.)$ and $x(.)$. Hence, $\frac{d\lambda_E}{d\bar{E}}, \frac{d\lambda_E}{d\gamma} > 0$.

$\frac{d\lambda_E}{d\bar{E}} > 0$ implies (from (17)) $\frac{dx(n)}{d\bar{E}} < 0$ and thus (from $v'(x(n)) = \frac{1}{n(1-\tau(n))}$) $\frac{d\tau(n)}{d\bar{E}} > 0$ for all $n \geq n_0$.

Since $u(n_0) = v(x(n_0) - z(n_0)/n_0$, $\frac{dx(n_0)}{d\bar{E}} < 0$ and $\frac{du(n_0)}{d\bar{E}} \geq 0$ yield $\frac{dz(n_0)}{d\bar{E}} < 0$. Writing $\frac{du(n_0)}{d\bar{E}} \geq 0$ as

$$v'(x(n_0))\frac{dx(n_0)}{d\bar{E}} - \frac{1}{n_0}\frac{dz(n_0)}{d\bar{E}} \geq 0,$$

we find that $v'(n_0) > \frac{1}{n_0}$ (because of $\tau(n_0) > 0$) implies

$$\frac{dz(n_0)}{d\bar{E}} < \frac{dx(n_0)}{d\bar{E}},$$

and thus (using $T(n_0) = z(n_0) - x(n_0)$)

$$\frac{dT(n_0)}{d\bar{E}} < 0.$$

Finally, in order to derive the sign of $\frac{du(n_1)}{d\bar{E}}$ we write $\frac{du(n)}{d\bar{E}}$ as (using (42))

$$\frac{du(n)}{d\bar{E}} = \frac{n_0}{n}\underbrace{\frac{du(n_0)}{d\bar{E}}}_{\geq 0} + \frac{1}{n}\int_{n_0}^{n} v'(x(t))\underbrace{\frac{dx(t)}{d\bar{E}}}_{<0}dt.$$

$\frac{du(n_1)}{d\bar{E}} > 0$ would imply $\frac{du(n)}{d\bar{E}} > 0$ for all $n \geq n_0$, thereby contradicting the optimality of the original solution paths. Q.E.D.

Proof of Proposition 12

With an interior solution $n_w > n_0$, the first-order condition for maximizing (7) with respect to n_w amounts to

$$\eta_w u'(n_w) = f(n_w)\{\bar{s}[-\gamma - v(b) + u(n_w) + \lambda_E(b + T(n_w))]\}. \tag{56}$$

The transversality conditions are

$$\lambda_u(n_w) + \eta_w = 0, \tag{57}$$
$$\lambda_u(n_1) = 0, \tag{58}$$
$$\lambda_z(n_w) = 0, \tag{59}$$
$$\lambda_z(n_1) = 0. \tag{60}$$

The first transversality condition follows from lemma 17 while $z(n_w) > 0$ (see lemma 3) yields the third transversality condition. The second transversality condition implies that differential equation (38) for $\lambda_u(n)$ can be solved as (54), which implies (together with (36)) that (17) holds. Solving this expression for λ_E, we arrive at (26).

Using (57) to eliminate $\lambda_u(n_w)$, we can write the solution (54) for $n = n_w$ as

$$-\frac{\eta_w}{n_w} + \bar{s}\left[\lambda_E\left(1 - F\left(n_w\right)\right) - \left(G\left(n_1\right) - G\left(n_w\right)\right)\right] = 0. \qquad (61)$$

We know from 2 that an interior solution $n_w > n_0$ implies $u(n_w) = v(b) + \gamma$, so that we can write (56) as

$$\eta_w = \frac{f\left(n_w\right)\bar{s}\lambda_E\left(b + T\left(n_w\right)\right)}{u'\left(n_w\right)}. \qquad (62)$$

Substituting (57) into (37) to eliminate $\lambda_u(n_w)$ and using (62) to eliminate η_w from the resulting expression, we obtain

$$\begin{aligned}\tau\left(n_w\right) &= \frac{b + T\left(n_w\right)}{u'\left(n_w\right)n_w^2} \\ &= \frac{b + T\left(n_w\right)}{z\left(n_w\right)}. \qquad (63)\end{aligned}$$

where we have used the incentive compatibility $u'(n_w) = \frac{z(n_w)}{n_w^2}$ to derive the second expression. Since $v'(x_w) = \frac{1}{n_w(1-\tau(n_w))}$, we find (using (63) to eliminate $\tau(n_w)$)

$$\begin{aligned}v'\left(x_w\right) &= \frac{1}{n_w}\frac{z\left(n_w\right)}{z\left(n_w\right) - b - T\left(n_w\right)} \\ &= \frac{v\left(x_w\right) - \gamma - v\left(b\right)}{x_w - b}, \qquad (64)\end{aligned}$$

where we have used (from the binding participation constraint) $z(n_w) = n_w(v(x_w) - u(n_w)) = n_w(v(x_w) - v(b) - \gamma)$ and $T(n_w) = z(n_w) - x_w$ to eliminate $z(n_w)$ and $T(n_w)$ from the first equality in (64). To see that x_w is uniquely defined, we write (25) as

$$-v'\left(x_w\right)b + v\left(b\right) + \gamma = v\left(x_w\right) - x_w v'\left(x_w\right).$$

At $x_w = b$, the left-hand side exceeds the right-hand side. Furthermore, as $x \to +\infty$, the right-hand side exceeds the left-hand side because $\lim_{x\to+\infty} v(x) - xv'(x) = +\infty$ and $v''(x) < 0$. Hence, a value $\hat{x}_w > b$ exists so that (25) holds. This value is unique because for $x_w > b$ the slope of the right-hand side $(-v''(x_w)x_w > 0)$ exceeds that of the left-hand side $(-v''(x_w)b > 0)$.

Substituting (61) into (62) to eliminate η_w, we find

$$\begin{aligned}\lambda_E &= \frac{\frac{G(n_1)-G(n_w)}{1-F(n_w)}}{1 - \frac{f(n_w)}{1-F(n_w)}\frac{1}{u'(n_w)}\frac{T(n_w)+b}{n_w}} \\ &= \frac{G\left(n_1\right) - G\left(n_w\right)}{1 - F\left(n_w\right)}\frac{1}{1-\chi}, \qquad (65)\end{aligned}$$

where $\chi \equiv \frac{f(n_w)n_w}{1-F(n_w)}\frac{T(n_w)+b}{z(n_w)}$ and we have used incentive compatibility $u'(n_w) = \frac{z(n_w)}{n_w^2}$ to eliminate $u'(n_w)$.

The government budget constraint is given by

$$\int_{n_w}^{n_1} f(n)\bar{s}\left[z(n) - x(n)\right]dn = E + \left[F(n_w) + (1 - F(n_w))(1 - \bar{s})\right]b. \quad (66)$$

In analogy of (42) and (50), we can write

$$u(n) = \frac{1}{n}\left(K_w + \int_{n_w}^{n} v(x(t))\,dt\right), \quad (67)$$

$$z(n) = nv(x(n)) - K_w - \int_{n_w}^{n} v(x(t))\,dt.$$

Substituting this into (66) to eliminate $z(n)$ and solving for K_w, we arrive at

$$K_w = \frac{1}{1 - F(n_w)}\left[\begin{array}{c} \int_{n_w}^{n_1}\left\{\left[tv(x(t)) - x(t)\right]f(t) - \left[1 - F(t)\right]v(x(t))\right\}dt \\ -bF(n_w) - \bar{E}\end{array}\right].$$

Using (67), $u(n_w) = v(b) + \gamma$ can be written as $\frac{1}{n_w}K_w = v(b) + \gamma$. Substituting the expression for K_w above, we establish (27).

We now prove that the solution satisfies (28). First, consider (26). The numerator is positive and decreasing in n_w. Denote the denominator by

$$D(n_w) \equiv 1 - F(n_w) - n_w f(n_w) + \frac{f(n_w)}{v'(x_w)}. \quad (68)$$

The assumptions $1 - n_0 f(n_0) > 0$ and $v'(\hat{x}_w) > \frac{1}{n_1}$ imply

$$\begin{array}{rcl} D(n_0) & > & 0, \\ D(n_1) & < & 0. \end{array}$$

Let $\tilde{n}_w$ denote the smallest value such that $D(\tilde{n}_w) = 0$ (by continuity of $f(.)$ such a value exists).[40] Hence, in (n_w, λ_E) space, (26) features an asymptote at $n_w = \tilde{n}_w$ (see figure 2). Since the numerator is positive, $\lim_{n_w \uparrow \tilde{n}_w} \lambda_E(n_w) = +\infty$. Furthermore, (26) is monotonically increasing in (n_w, λ_E) space for $n_w \in [n_0, \tilde{n}_w]$ if $x'(n) > 0$ for all n (as we assume here). To see this, suppose that $\left.\frac{\partial \lambda_E}{\partial n_w}\right|_{(26)} > 0$ does not hold. Since equation (26) is identical to equation (17) (where the relation between $v'(.)$ and λ_E is a positive one), this implies

[40] If $f(.)$ satisfies the monotone hazard rate property, $\tilde{n}_w$ is the unique solution to $D(\tilde{n}_w) = 0$.

that $v'(x(n))$ is increasing in n for some $n > n_w$ so that $x'(n) < 0$ for these values. This contradicts the monotonicity requirement $x'(n) \geq 0$ in lemma 1. Accordingly, a non-binding monotonicity requirement implies $\left.\frac{\partial\lambda_E}{\partial n_w}\right|_{(26)} > 0$.

To determine the slope of (27), we define

$$\begin{aligned}\phi(n_w, \lambda_E) \equiv & \int_{n_w}^{n_1} \{[tv(x(t,\lambda_E)) - x(t,\lambda_E)] f(t) - [1 - F(t)] v(x(t,\lambda_E))\} dt \\ & - \left\{n_w (1 - F(n_w))(v(b) + \gamma) + bF(n_w) + \bar{E}\right\}.\end{aligned}$$

To find the slope of (27) (which is $\phi(n_w, \lambda_E) = 0$) in (n_w, λ_E) space, we determine $\phi_{n_w} = \frac{\partial\phi}{\partial n_w}$ and $\phi_{\lambda_E} = \frac{\partial\phi}{\partial\lambda_E}$

$$\begin{aligned}\phi_{n_w} &= -[n_w v(x(n_w,\lambda_E)) - x(n_w,\lambda_E)] f(n_w) + [1 - F(n_w)] v(x(n_w,\lambda_E)) \\ &\quad - \{[1 - F(n_w)](v(b) + \gamma) - n_w f(n_w)(v(b) + \gamma) + f(n_w) b\} \\ &= -(v(x_w) - v(b) - \gamma)(n_w f(n_w) - [1 - F(n_w)]) + f(n_w)(x_w - b) \\ &= (x_w - b) v'(x_w) D(n_w) > 0, \qquad (69)\end{aligned}$$

where $D(n_w)$ is defined in (68) and the last equality holds only for the values of λ_E and n_w satisfying (26) (since we used (25) to eliminate $(v(x_w) - v(b) - \gamma)$).

$$\begin{aligned}\phi_{\lambda_E} &= \int_{n_w}^{n_1} \{[tv'(x(t,\lambda_E)) - 1] f(t) - [1 - F(t)] v'(x(t,\lambda_E))\} \frac{dx(t,\lambda_E)}{d\lambda_E} dt \\ &= \int_{n_w}^{n_1} \{v'(x(t,\lambda_E))(tf(t) - [1 - F(t)]) - f(t)\} \frac{dx(t,\lambda_E)}{d\lambda_E} dt \\ &= \int_{n_w}^{n_1} f(t) \frac{\frac{G(n_1)-G(t)}{\lambda_E}}{\frac{G(n_1)-G(t)}{\lambda_E} + f(t)t - [1 - F(t)]} \left[-\frac{dx(t,\lambda_E)}{d\lambda_E}\right] dt > 0, \qquad (70)\end{aligned}$$

where we have used equation (17) to derive the third equality by eliminating $v'(x(t,\lambda_E))$. The sign of ϕ_{λ_E} follows because the concavity of $v(x)$ together with (17) imply $\frac{dx(t,\lambda_E)}{d\lambda_E} < 0$. With $\phi_{n_w} > 0$ and $\phi_{\lambda_E} > 0$, (27) is downward sloping in (n_w, λ_E) space at the point where (27) and (26) intersect.

Finally, to derive the expression for W, we employ (14) with $F(n_w)$ agents voluntarily unemployed

$$\begin{aligned}W &= v(b)(1 - \bar{s}) + \bar{s}F(n_w)(v(b) + \gamma) + \bar{s}u(n_w) n_w (G(n_1) - G(n_w)) + \\ &\quad + \bar{s} \int_{n_w}^{n_1} (G(n_1) - G(n)) v(x(n)) dn.\end{aligned}$$

Substitution of $u(n_w) = v(b) + \gamma$ yields the expression in the proposition. Q.E.D.

Proof of lemma 13

From the definition of the asymptote $\tilde{n}_w$ above, we have

$$D(n_w) \begin{cases} >0 & \text{if} \quad n_w < \tilde{n}_w \\ =0 & \text{if} \quad n_w = \tilde{n}_w \\ <0 & \text{if} \quad n_w > \tilde{n}_w \end{cases}.$$

The result in the lemma on $\frac{\partial q}{\partial n_w}$ follows from (69), $x_w > b$ and $v'(x_w) > 0$.

From the definition of $\tilde{n}_w$ ($D(\tilde{n}_w) = 0$, i.e. (68)), we observe

$$f(\tilde{n}_w)\left(\frac{1}{v'(x_w)} - \tilde{n}_w\right) = -(1 - F(\tilde{n}_w)) < 0,$$

so that $v'(x(\tilde{n}_w)) = v'(x_w) > \frac{1}{\tilde{n}_w}$. This implies $\tau(\tilde{n}_w) > 0$ and thus (from (63)) $T(\tilde{n}_w) + b > 0$. Q.E.D.

Proof of Lemma 14

Differentiating equations (26) and (27) with respect to n_w, λ_E, E, b and γ, we find

$$\begin{pmatrix} \phi_{n_w} & \phi_{\lambda_E} \\ -\left.\frac{d\lambda_E}{dn_w}\right|_{(26)} & 1 \end{pmatrix}\begin{pmatrix} dn_w \\ d\lambda_E \end{pmatrix} = \begin{pmatrix} \frac{1}{\bar{s}}dE + \begin{bmatrix} n_w(1-F(n_w))v'(b) \\ +\frac{1}{\bar{s}} - (1-F(n_w)) \end{bmatrix} db + n_w[1-F(n_w)]\,d\gamma \\ \frac{d\lambda_E}{dv'(x_w)}\frac{dv'(x_w)}{db}db + \frac{d\lambda_E}{dv'(x_w)}\frac{dv'(x_w)}{d\gamma}d\gamma \end{pmatrix}$$

where ϕ_{n_w} and ϕ_{λ_E} are defined in equations (69) and (70). The determinant of the matrix on the left-hand side equals

$$\det = \begin{vmatrix} \phi_{n_w} & \phi_{\lambda_E} \\ -\left.\frac{d\lambda_E}{dn_w}\right|_{(26)} & 1 \end{vmatrix} = \phi_{n_w} + \phi_{\lambda_E}\left.\frac{d\lambda_E}{dn_w}\right|_{(26)} > 0.$$

Hence,

$$\begin{pmatrix} dn_w \\ d\lambda_E \end{pmatrix} = \frac{1}{\det}\begin{pmatrix} 1 & -\phi_{\lambda_E} \\ \left.\frac{d\lambda_E}{dn_w}\right|_{(26)} & \phi_{n_w} \end{pmatrix} \times \tag{71}$$

$$\begin{pmatrix} \frac{1}{\bar{s}}dE + \begin{bmatrix} n_w(1-F(n_w))v'(b) \\ +\frac{1}{\bar{s}} - (1-F(n_w)) \end{bmatrix} db + [1-F(n_w)]\,d\gamma \\ \frac{d\lambda_E}{dv'(x_w)}\frac{dv'(x_w)}{db}db + \frac{d\lambda_E}{dv'(x_w)}\frac{dv'(x_w)}{d\gamma}d\gamma \end{pmatrix}.$$

We thus find

$$\begin{aligned} \frac{dn_w}{dE} &= \frac{1}{\det}\frac{1}{\bar{s}} > 0 \\ \frac{d\lambda_E}{dE} &= \frac{1}{\det}\left.\frac{d\lambda_E}{dn_w}\right|_{(26)}\frac{1}{\bar{s}} > 0. \end{aligned}$$

From (29) and (17), the effect of E on λ_E implies that

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

Writing $u(n)$ as

$$u(n) = \frac{n_w}{n}\left(v(b)+\gamma\right) + \frac{1}{n}\int_{n_w}^{n} v(x(t))\,dt,$$

we find that

$$\begin{aligned}\frac{du(n)}{dE} &= \frac{1}{n}\underbrace{\left[v(b)+\gamma - v(x(n_w))\right]}_{=-\frac{z(n_w)}{n_w}<0}\underbrace{\frac{dn_w}{dE}}_{>0} + \frac{1}{n}\int_{n_w}^{n} v'(x(t))\underbrace{\frac{dx(t)}{dE}}_{<0}dt\\ &< 0.\end{aligned}$$

This implies $\frac{du(n_1)}{dE} < 0$ while $x(n_1)$ is determined by $v'(x(n_1)) = \frac{1}{n_1}$ and thus independent of E. Hence (as $u(n_1) = v(x(n_1) - z(n_1)/n_1)$, $\frac{dz(n_1)}{dE} > 0$.

Writing $T(n) = z(n) - x(n) = n\left(v(x(n)) - u(n)\right) - x(n)$, we find

$$\frac{dT(n)}{dE} = \underbrace{-n\frac{du(n)}{dE}}_{>0} + \underbrace{\left[nv'(x(n)) - 1\right]}_{=\frac{\tau(n)}{1-\tau(n)}\geq 0}\underbrace{\frac{dx(n)}{dE}}_{<0}. \tag{72}$$

Since $\tau(n_1) = 0$, we find $\frac{dT(n_1)}{dE} > 0$. Q.E.D.

Proof of Lemma 15

Since b and γ appear in similar ways in equation (71), we derive results only for b. Note that (from (17)) $\frac{d\lambda_E}{dv'(x_w)} > 0$ and (from differentiating (25)) $\frac{dv'(x_w)}{db} < 0$ so that (71) yields

$$\frac{dn_w}{db} > 0.$$

We also have

$$\frac{d\lambda_E}{db} = \frac{1}{\det}\left[\begin{array}{c} \left.\frac{d\lambda_E}{dn_w}\right|_{(26)}\underbrace{\left[n_w\left(1-F(n_w)\right)v'(b) + \frac{1}{\bar{s}} - \left(1-F(n_w)\right)\right]}_{>0} + \\ +\phi_{n_w}\underbrace{\frac{d\lambda_E}{dv'(x_w)}\frac{dv'(x_w)}{db}}_{<0}\end{array}\right].$$

We thus obtain $\frac{d\lambda_E}{db} > 0$ if ϕ_{n_w} is close to 0 and $\frac{d\lambda_E}{db} < 0$ if $\left.\frac{d\lambda_E}{dn_w}\right|_{(26)}$ is close to 0. We consider these cases in turn. First, (69) implies that ϕ_{n_w} is close to 0 if either x_w is close to b or if $D(n_w)$ is close to 0. (30) reveals that $x_w - b$ is close to 0 if γ is close to zero. Hence, $\bar{\gamma} > 0$ exists such that for all $\gamma \in \langle 0, \bar{\gamma}\rangle$ ϕ_{n_w} is small enough to ensure $\frac{d\lambda_E}{db} > 0$. Moreover, by continuity, $D(n_w)$ is close to 0 if n_w is close to the asymptote $\tilde{n}_w$ in figure 2. By raising E, the downward sloping government budget constraint in figure 2 shifts upward so that we can obtain an equilibrium value arbitrarily close to $\tilde{n}_w$. Therefore, by choosing E sufficiently large (but below maximum government revenues $g(\tilde{n}_w, +\infty)$), ϕ_{n_w} is close enough to 0 so that $\frac{d\lambda_E}{db} > 0$.

If (26) is downward sloping around n_0 (as depicted in figure 2), a value for n_w exists (denoted by $\underline{n}$ in figure 2) where $\left.\frac{d\lambda_E}{dn_w}\right|_{(26)} = 0$. In this case, we can find values for E such that the intersection of (27) and (26) is at $n_w > \underline{n}$ but is close enough to $\underline{n}$ so that $\frac{d\lambda_E}{db} < 0$. More precisely, let E_* denote the value of E such that the intersection is at $\underline{n}$. Then a value $E^* > E_*$ exists such that $\frac{d\lambda_E}{db} < 0$ for all $E \in \langle E_*, E^*\rangle$.

The effects of b on $x(n)$ and $\tau(n)$ follow immediately from $\frac{d\lambda_E}{db}$ by using (17) and (29), respectively.

Now we argue that $\frac{d\lambda_E}{db} < 0$ can be excluded if $\frac{d\tau(n)}{dn} < 0$ around n_w and if a rise in b does not lead to a Pareto improvement. More precisely, we show that $\frac{d\lambda_E}{db} < 0$ and $\frac{d\tau(n)}{dn} < 0$ around n_w imply that raising b does yield a Pareto improvement.

Utility can be written (from (67)) as

$$u(n) = \frac{1}{n}\left(u(n_w) + \int_{n_w}^{n} v(x(t))\,dt\right) = \frac{1}{n}\left(v(b) + \gamma + \int_{n_w}^{n} v(x(t))\,dt\right)$$

Taking the derivative with respect to b, we arrive at

$$\begin{aligned} \frac{du(n)}{db} &= \frac{1}{n}\left(v(b) + \gamma - v(x_w)\right)\frac{dn_w}{db} + \frac{n_w}{n}v'(b) + \frac{1}{n}\int_{n_w}^{n} v'(x(t))\frac{dx(t)}{db}dt \qquad (73)\\ &= \frac{n_w}{n}\left[v'(b) - \frac{v'(x_w)}{n_w}(x_w - b)\frac{dn_w}{db}\right] + \frac{1}{n}\int_{n_w}^{n} v'(x(t))\frac{dx(t)}{db}dt, \end{aligned}$$

where we have used (25) to eliminate $(v(b) + \gamma - v(x_w))$. Differentiation of $v'(x_w) = 1/[n_w(1 - \tau(n_w))]$ yields $dn_w/dx_w \leq n_w^2(1 - \tau(n_w))(-v''(x_w))$ if $d\tau(n)/dn \leq 0$ around $n = n_w$. Similarly, differentiating (25), we find

$$\frac{dx_w}{db} = \frac{v'(b) - v'(x_w)}{(x_w - b)(-v''(x_w))}.$$

Accordingly, $\frac{dn_w}{db} = \frac{dn_w}{dx_w}\frac{dx_w}{db} \leq n_w^2 \frac{v'(b)-v'(x_w)}{x_w-b}(1-\tau(n_w))$. Substitution of this inequality in (73) to eliminate $\frac{dn_w}{db}$, we obtain

$$\frac{du(n)}{db} \geq \frac{n_w}{n}\left[v'(b) - v'(x_w)n_w(1-\tau(n_w))(v'(b)-v'(x_w))\right] + \frac{1}{n}\int_{n_w}^{n} v'(x(t))\frac{dx(t)}{d\lambda_E}\frac{d\lambda_E}{db}dt. \tag{74}$$

Substituting $v'(x_w) = 1/[n_w(1-\tau(n_w)]$ to eliminate $v'(x_w)$, we find that the first term at the right-hand side of this equation is positive. Also the second term at the right-hand side is positive if $d\lambda_E/db < 0$ (since (17) implies that $\frac{dx(t)}{d\lambda_E} < 0$). Hence, $d\lambda_E/db < 0$ together with $\frac{d\tau(n)}{dn} < 0$ around $n = n_w$ implies that all agents gain from a higher welfare benefit.

(74) implies that a value n^* close to n_w exists such that $\frac{du(n)}{db} > 0$ for all $n \in \langle n_w, n^* \rangle$. Since $\frac{d\lambda_E}{db} > 0$ implies that $\frac{dx(n)}{db} < 0$ for all $n > n_w$, $\frac{du(n)}{db} > 0$ implies that $\frac{dz(n)}{db} < 0$ for all $n \in \langle n_w, n^* \rangle$. Finally, rewriting (72) for the effect of b on $T(n)$, we obtain

$$\frac{dT(n)}{db} = \underbrace{-n\frac{du(n)}{db}}_{<0} + \underbrace{[nv'(x(n)) - 1]}_{=\frac{\tau(n)}{1-\tau(n)}\geq 0}\underbrace{\frac{dx(n)}{db}}_{<0} < 0$$

for all $n \in \langle n_w, n^* \rangle$, where we have used that $\tau(n_w) > 0$ and hence $\tau(n) > 0$ for n close to n_w. Q.E.D.

Proof of proposition 16

Section 6 shows that parameter values exist under which the optimal tax schedule implies $b + T(n_w) < 0$ and $\tau(n_w) < 0$. The first-order condition for n_w (see (56) with $u(n_w) = v(b) + \gamma$) then implies that $\eta_w < 0$.

Employing (72), we find for types n with $\tau(n) < 0$ that

$$\frac{dT(n)}{dE} = \underbrace{-n\frac{du(n)}{dE}}_{>0} + \underbrace{[nv'(x(n)) - 1]}_{=\frac{\tau(n)}{1-\tau(n)}<0}\underbrace{\frac{dx(n)}{dE}}_{<0} > 0,$$

where the effects of E on $u(n)$ and $x(n)$ are derived in lemma 14. Q.E.D.

10 Tables

Table 1. Numerical results without binding participation margin

parameter values	(1)	(2)	(3)	(4)	(5)
n_0	10^*	4^{**}	4^{**}	4^{**}	4^{**}
b	10	5	10	10	40
γ	1	1	1	1	1
$\bar{E}$	50	120	25	50	50
$\phi(n)$	1	1	1	1	$\frac{1/\sqrt{n}}{\int_{n_0}^{n_1} f(t)/\sqrt{t}dt}$
results					
n_b	10	5.4	n.a.	n.a.	n.a.
$F(n_b)$	0	0.0	n.a.	n.a.	n.a.
n_z	10	4	10.5	9.4	12.3
$F(n_z)$	0	0	0.10	0.06	0.16
$\tau(n_b)$	0	0.39	n.a.	n.a.	n.a.
$\tau(n_z)$	0	n.a.	0.22	0.24	0.34
$\tau(n_{md})$	0.19	0.24	0.22	0.23	0.31
$\tau(n_{95\%})$	0.15	0.16	0.16	0.16	0.19
$max\ \tau(n)$	0.19	0.39	0.24	0.26	0.35
$x(n_0)$	100	10.8	67.9	51.0	65.2
$z(n_0)$	69.1	1.8	0	0	0
$u(n_0)$	13.1	6.1	16.5	14.3	16.2
$v(b)+\gamma$	7.3	5.5	7.3	7.3	13.6
$T(n_b)/z(n_b)$	-0.45	-5.0	$-67.9^{\S}$	$-51.0^{\S}$	$-65.2^{\S}$
$T(n_{md})/z(n_{md})$	0.03	0.26	$-.06$	0.04	-0.01
$T(n_{95\%})/z(n_{95\%})$	0.15	0.20	0.15	0.16	0.20
λ_E	0.050	0.056	0.054	0.055	0.057

baseline values: $ln(n) \sim N(3, 0.5)$, $n_1 = 100$ and $v(x) = 2\sqrt{x}$.

* The median ability $n_{md} = 21.1$, while the 95th percentile ability $n_{95\%} = 46.0$.

** The median ability $n_{md} = 20.1$, while the 95th percentile ability $n_{95\%} = 45.6$.

§ This is $T(n_z)$ since the average tax rate is not defined with $z(n_z) = 0$.

Table 2. Numerical results with binding participation margin

parameter values	(1)	(2)	(3)	(4)	(5)
b	40	70	40	10	0
γ	1	1	1	10	0
$\bar{E}$	50	50	100	25	150
results					
n_w	11.1	20.8	17.6	15.4	4
$F(n_w)$	0.12	0.53	0.40	0.30	0
$\tau(n_w)$	0.15	0.43	0.47	-0.02	0.62
$\frac{T(n_w)+b}{z(n_w)}$	0.15	0.43	0.47	-0.02	0.81
$\tau(n_{md})$	0.20	n.a.	0.40	0.08	0.23
$\tau(n_{95\%})$	0.16	0.21	0.19	0.13	0.16
$\max\ \tau(n)$	0.21	0.43	0.47	0.13	0.62
$x(n_w)$	88.1	139	88.1	246	2.4
$z(n_w)$	56.9	122	90.4	231	12.3
$u(n_w) = v(b) + \gamma$	13.6	17.7	13.6	16.3	0
$T(n_w)/z(n_w)$	$-.55$	-0.14	0.03	-0.07	0.81
$T(n_{md})/z(n_{md})$	0.04	n.a.	0.24	-0.03	0.32
$T(n_{95\%})/z(n_{95\%})$	0.15	0.24	0.25	0.09	0.21
λ_E	0.052	0.127	0.095	0.040	0.057

baseline values: $ln(n) \sim N(3, 0.5), n_0 = 4$, $n_1 = 100^*, v(x) = 2\sqrt{x}$ and $\phi(.) = 1$; these parameters imply $n_{md} = 20.1$, $n_{95\%} = 45.6$.

Figure 1: equilibrium in n_z, λ_E space: equations (18) and (19) in Lemma 7

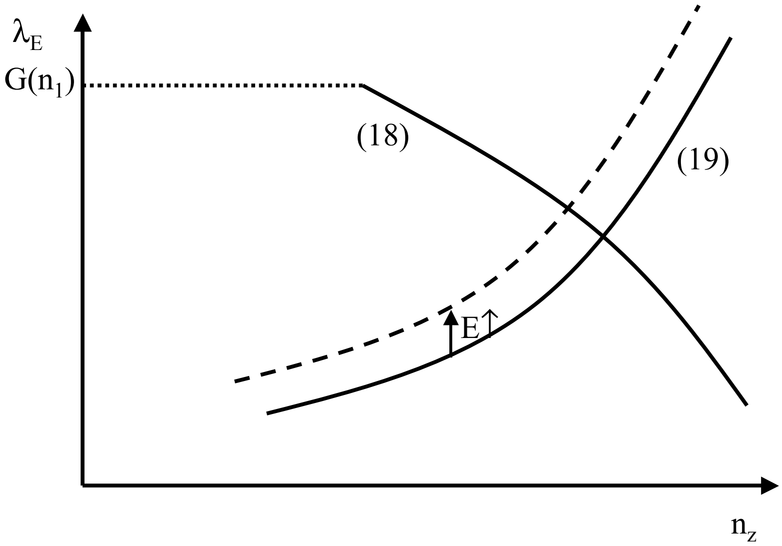

Figure 2: equilibrium in n_w, λ_E space: equations (26) and (27) in proposition 12