Optimal income taxation with endogenous participation and involuntary unemployment.*

Etienne LEHMANN†
CREST, IZA and
Université Catholique de Louvain

Alexis PARMENTIER‡
EPEE - Université d'Evry and
Université Catholique de Louvain

Bruno VAN DER LINDEN§
IRES - Université Catholique de Louvain,
FNRS, ERMES - Université Paris 2 and IZA

December 14, 2007

Abstract

This paper characterizes the optimal redistributive taxation when individuals are heterogeneous in two exogenous dimensions: their skills and their values of non-market activities. Matching frictions on labor markets generate involuntary unemployment. Wages, labor demand and participation are endogenous. When the government has a Maximin objective and observes only wage levels, we show that the distribution of the elasticity of participation is key. If this elasticity decreases with the skill level, at the optimum, the average tax rate is increasing, marginal tax rates are positive everywhere, while wages and participation rates are distorted downwards compared to their laissez faire values.

*We thank for their comments Pierre Cahuc, Guy Laroque, Fabien Postel-Vinay and participants at seminars at Louvain la Neuve and CREST. Any errors are ours. This research has been funded by the Belgian Program on Interuniversity Poles of Attraction (P6/07 Economic Policy and Finance in the Global Economy: Equilibrium Analysis and Social Evaluation) initiated by the Belgian State, Prime Minister's Office, Science Policy Programming.

†Adress: CREST-INSEE, Timbre J360, 15 Boulevard Gabriel Péri, 92245, Malakoff Cedex, France. Email: etienne.lehmann@ensae.fr

‡Adress: EPEE - Université d'Evry Val d'Essone, 4 Boulevard François Mitterand, 91025, Evry Cedex, France. Email: alexis.parmentier@univ-evry.fr

§Adress: IRES - Département d'économie, Université Catholique de Louvain, Place Montesquieu 3, B1348, Louvain-la-Neuve - Ottignies, Belgium. Email: bruno.vanderlinden@uclouvain.be

"Another desire is to have a model in which unemployment can arise and persist for reasons other than a preference for leisure"

Mirrlees (1999).

I Introduction

The uneven distribution of productive talents is since long a matter of concern in public economics. Low productivity workers are however not only low-wage earners. They are also less often employed. In the literature on optimal redistributive taxation initiated by Mirrlees (1971), non-employment, if any, is synonymous with non-participation. Without downplaying the importance of participation decisions, a broader view should recognize that some people do not find a job at the market wage despite they do search for one ("involuntary unemployment"). Non-competitive wage formations provide explanations of involuntary unemployment. Under such imperfections, more progressive taxes on earnings moderate wages. This in turn stimulates labor demand and reduces unemployment (see e.g. Lockwood and Manning 1993). Our paper characterizes optimal redistributive taxation when higher tax levels are detrimental to participation and tax progression influences wages and hence employment.

Our economy is made of a continuum of skill-specific labor markets. On each of them, matching frictions à la Mortensen and Pissarides (1999) generate involuntary unemployment. In this setting, several wage formation mechanisms generate our results. Among them we select the Competitive Search Equilibrium of Moen (1997), henceforth CSE. Then, the wage level maximizes the expected surplus of a job-seeker. We prove that a more progressive tax schedule reduces the equilibrium wage. For, it reduces the marginal benefit an employed worker obtains from a unit increase in her pre-tax wage without changing the marginal loss of chances to become employed. Through this channel, tax progressivity stimulates labor demand and reduces involuntary unemployment. To deal with participation, we assume that whatever their skill, individuals differ by their value of remaining out of the labor force. We show that a higher level of tax reduces the participation rate. In sum, taxes are distorsive via the labor demand margin and via the participation margin.

Our main results relate to the case where taxation of employed workers can only be conditioned on their wage (generating an adverse selection problem with random participation à la Rochet and Stole 2002) and the government has a Maximin (Rawlsian) objective. We show that optimal taxation strongly depends on the whole shape of elasticities of participation along the skill distribution. In the most plausible case (Saez 2002) where these elasticities are decreasing, we prove that optimal marginal tax rates are positive everywhere and optimal average tax rates are increasing. The reason is that a more progressive tax schedule increases the level of tax at the top of the skill distribution where participation decisions are less elastic and decreases the level of tax where participation reacts more strongly to the tax pressure. A more progressive tax schedule in addition distorts wages downwards. At the optimum, the marginal loss generated by the latter distortion equalizes the net gain due to the former adjustment in participation decisions.

In the optimal taxation literature that follows Mirrlees (1971), marginal tax rates have to be positive everywhere, except at the top of the skill distribution when this distribution is bounded.[1] The average tax rate cannot be increasing everywhere, except if the skill distribution is unbounded (Hindricks et alii 2006). In these models where labor supply along the intensive margin is the only source of deadweight loss, positive marginal tax rates distorts gross income downwards. Our results contrast with the Mirrlees literature since we do not need unbounded distribution of skills to find positive marginal tax rates and increasing average tax rates.

The comprehensive survey of Blundell and MacCurdy (1999) suggests that labor supply responses along the intensive margin are empirically very small. There is now a growing evidence that the extensive margin matters more. Diamond (1980) and Choné and Laroque (2005) have studied optimal income taxation when individuals' decisions are limited to a dichotomic choice about whether to work or not. At the optimum, the level of taxes trades off the equity gain of a higher level of tax against the efficiency loss of a lower level of participation. However, there are no distortion of wages in these models which assume a competitive labor market and exogenous productivity levels. Saez (2002) and Boone and Bovenberg (2004) have proposed mixed models of taxation where both extensive and intensive margins of the labor supply are present. They exhibit cases where wages are distorted upwards through negative marginal tax rates at the bottom of the skill distribution. Cahuc and Laroque (2007) have introduced monopsonistic labor markets in the model of Diamond (1980). They explain how the optimal taxation can bypass the distortions induced by the monopsony.

Hungerbühler et alii (2006, henceforth HLPV) have proposed an optimal income tax model with endogenous involuntary unemployment due to matching frictions. HLPV also find that wages have to be distorted downwards, marginal tax rates have to be positive everywhere and the average tax rate increases with the level of wage. The present paper differs from HLPV in three respects. First, the cost of participation takes a unique value in HLPV. Hence, there is an endogenous threshold of skill such that every worker with a higher (lower) skill level participate (does not participate). Conversely, we allow the opportunity cost of participation to vary within and between skill levels. This leads to a much more realistic description of participation decisions. In this sense, HLPV is a particular case of the present paper where the elasticity of participation is infinite at the threshold, and zero above. Second, the wage-setting mechanisms are different. Instead of the CSE, HLPV assumes Nash bargaining under the Hosios (1990) condition. While both mechanisms imply that the laissez faire allocation is efficient (in the Benthamite sense), the CSE is much more flexible and allows to deal with a wider class of matching functions.

[1] See Choné and Laroque (2007) for a counterexample with negative marginal tax rates under a specific objective function. See Diamond (1998) for positive marginal tax rates everywhere under an unbounded Pareto skill distribution.

Finally, following Saez (2001) we here interpret the description of our optima through marginal tax reforms, which we think is more intuitive.

The paper is organized as follows. The environment is presented in the next section. In particular, we show that the wage-moderating and unemployment-reducing effects of tax progressivity which holds under monopoly unions (Hersoug 1984), right-to-manage bargaining (Lockwood and Manning), efficiency wages (Pisauro 1991) or matching models with Nash bargaining (Pissarides 1998) also holds in a CSE framework. Section III describes the Maximin optimum in the case where taxation can be conditioned on skill and wage levels. This case serves has a benchmark to compare to the more realistic case treated in Section IV where the level of taxes only varies with the wage.

II The model

Individuals are risk neutral and characterized by a double heterogeneity. Their exogenous productivity (or skill) is denoted $a \in \mathbb{R}_+$. Their value of non-market activities is denoted $\delta \in \mathbb{R}_+$. Bringing up children or home production are typical examples of these activities. A person looking for a job spends time on searching, and therefore cannot enjoy these activities. For this reason, δ measures an opportunity cost of participation and not a disutility of work.[2] Labor markets are assumed to be segmented by skill.

The size of the population is normalized to 1. The unconditional density function of productivity is denoted $f(a)$ whose support is $[a_0, a_1]$, where $0 \leq a_0 < a_1 \leq +\infty$. Conditional on a, the value of non-market activities δ has a c.d.f:

$$G(a, \Sigma) = \Pr[\delta \leq \Sigma | a]$$

and a density $g(a, \Sigma) = \partial G(a, \Sigma) / \partial \Sigma$. For all a in $[a_0, a_1]$, the support of $g(a, .)$ is an interval that includes 0 as a lower bound. Note that this formulation includes the possibility that a and δ are correlated. Both $f(.)$ and $g(., .)$ are positive and continuous functions on their support.

To introduce involuntary unemployment, we assume the existence of matching frictions *à la* Mortensen and Pissarides (1999) and Pissarides (2000). Different wage-setting mechanisms have been proposed in this framework. As the literature dealing with optimal redistribution in a competitive framework (Mirrlees 1971), the role of taxation is not to restore efficiency but to redistribute income. Therefore, we need a wage-setting mechanism that maximizes the sum of utility levels in the absence of taxes. The CSE introduced by Moen (1997) and Shimer (1996) induces this property. Nash bargaining under the Hosios (1990) condition is a standard alternative[3]. The Hosios condition is however very unlikely to hold, in particular when matching functions are not iso-elastic. A wider class of matching technologies can be considered with the CSE approach.

[2] However, our model can easily be extended to include a skill-specific disutility of work.

[3] Under this condition, worker's bargaining power is equal to the elasticity of the matching function with respect to the number of unemployed workers.

We now describe the CSE environment. Firms post skill-specific vacancies and associate to each of them a take-it-or-leave-it (gross or pre-tax) wage offer w. A worker of type a can only search for a job paid at a single wage level. Therefore, within skill-specific labor markets, there are different wage-specific submarkets. On each submarket (a, w), the number of filled jobs is a function[4] $H(V_{a,w}, U_{a,w})$ of the numbers $V_{a,w}$ of vacancies and $U_{a,w}$ of job-seekers. The matching function captures unmodeled heterogeneities and information imperfections that remain on each submarket.

As usual in the optimal tax literature that follows Mirrlees (1971), we consider a static model. The timing of events is the following:[5]

1. The government commits to a tax and benefits policy.

2. Firms decide how many vacancies of each type they create and which wage levels they attach to each of them. This sets the values of $V_{a,w}$.

3. Individuals of type (a, δ) decide whether they participate to the labor market of type a. If they participate, they renounce δ and choose the submarket (a, w) on which they search for a job. This determines the values of $U_{a,w}$.

4. On each labor market, the matching process determines the number of filled jobs on each submarket. Each employed worker of type (a, δ) produces a units of goods. Wages are paid. Taxes are collected and redistributed. Agents consume.

We solve the model by backward induction. The next subsection describes the policy instruments of the government. Subsection II.2 solves the CSE described at stages 2 and 3 on each skill-specific labor market, while Subsection II.3 studies the fiscal incidence on this equilibrium. Subsection II.4 states the government's budget constraint and objectives. Finally, subsection II.5 is devoted to the *laissez-faire* equilibrium.

II.1 The government's instruments

Taxes and benefits depend on what the government observes. The government does not observe individuals' value of non-market activities δ. In addition, the government cannot distinguish unemployed people from non-participants.[6] Following Diamond (1980), the skill level a cannot be observed for non-employed individuals. Therefore, the government is constrained to give the same level of benefit b to non-participant and unemployed individuals irrespectively of their types (a, δ).

[4] Assuming a skill-specific matching functions $H(., ., a)$ does not change the results, but burdens the notations.

[5] In this timing, stages two and three can be inverted or can occur simultaneously (see Moen 1997 and Shimer Rogerson and Wright 2005).

[6] In reality, search behavior can to some extent be monitored. The joint optimization of taxation and job-search monitoring is beyond the scope of the present paper. For tractability reasons, we stick on a limiting case of unobservable search that sounds the most realistic.

We distinguish two alternative informational assumptions about employed workers. Since firms can direct their search according to the level of skill, it may sound theoretically reasonable that the government infers the skill of a worker from the observation of the matching process. Then, it can condition the tax schedule on wages and skill levels $T(.) = T(a, w)$. Section III considers this case. In reality, one hardly observes that tax schemes are skill-dependent. One reason is that inferring individuals' skills, while theoretically feasible, is actually very complex and thereby costly for the government. Moreover, two individuals would face different tax schedules only because of their innate characteristics. This is a form of discrimination, which might be forbidden by law. So, in Section IV, we consider the case where the government can only condition the tax schedule on the wage levels $T(.) = T(w)$.

In sum, a worker of type (a, δ) can be in three positions. She can be employed with a post-tax income $w - T(.)$. She can be unemployed and then receives a benefit b. Finally, if she does not participate, she gets the benefit plus the value of her non-market activities $b + \delta$.

II.2 Participation decisions and the Competitive Search Equilibrium

Before solving stages 2 and 3, we describe the matching technology. On each submarket (a, w), the matching function determines the number of filled jobs $H(V_{a,w}, U_{a,w})$ as a function of the number $V_{a,w}$ of vacancies and of the number $U_{a,w}$ of job-seekers. Following Petrongolo and Pissarides (2001), we assume:

Assumption AS 1 *The matching function $H(.,.)$ is twice-continuously differentiable on $\mathbb{R}^2_+$, is increasing in both arguments and exhibits constant returns to scale. Moreover, $H(V, 0) = H(0, U) = 0$, and for any $(V, U) \in \mathbb{R}^2_{+*}$, one has $H(V, U) < \min(V, U)$, $\lim_{V \mapsto +\infty} H(V, U) \leq U$ and $\lim_{U \mapsto +\infty} H(V, U) \leq V$.*

A typical specification that verifies these assumptions is the CES function with a low elasticity of substitution:

$$H(V, U) = \frac{U \cdot V}{[U^\rho + V^\rho]^{\frac{1}{\rho}}} = \left[U^{-\rho} + V^{-\rho}\right]^{-\frac{1}{\rho}} \qquad \text{with} \quad \rho > 0 \tag{1}$$

Define tightness θ as the ratio V/U. The probability that a vacancy meets an applicant is $q(\theta) \overset{\text{def}}{\equiv} H\left(1, \frac{1}{\theta}\right) = H(V, U)/V$. Due to congestion externalities, the job-filling probability decreases with the number of vacancies and increases with the number of job-seekers. Because of constant returns to scale, only tightness matters and $q(\theta)$ is a decreasing function of θ. Symmetrically, the probability that a job-seeker finds a job is an increasing function $p(\theta) \overset{\text{def}}{\equiv} H(\theta, 1) = H(V, U)/U$ of tightness. Firms and individuals being atomistic, they take as given tightness $\theta_{a,w}$ on each submarket and hence the probabilities $p(\theta_{a,w})$ and $q(\theta_{a,w})$.

At stage 3, an individual of type (a, δ) gets $\delta + b$ if she does not participate and expects $p(\theta_{a,w})(w - T(.)) + (1 - p(\theta_{a,w}))\, b$ if she searches for a job paid w. Let $\Sigma_a \overset{\text{def}}{\equiv} \sup_w \; p(\theta_{a,w})(w - T(.) - b)$

be the highest expected surplus an individual of productivity a can expect from participation. At this stage, Σ_a is conditional on $V_{a,w}$. An individual of type (a, δ) participates if and only if:

$$\Sigma_a \geq \delta$$

Once an individual has chosen to participate, what only matters is her productivity a and no longer her value of non-market activities δ. We henceforth call her a participant of type a.

Consider a submarket (a, w). If this submarket offers too low an expected surplus to participants, i.e. if $p(\theta_{a,w})(w - T(.) - b) < \Sigma_a$, then less participants choose to search on this submarket, $U_{a,w}$ decreases, thereby increasing tightness $\theta_{a,w}$. By this mechanism, the decisions of participants induces an equalization of the expected surpluses across submarkets, so the equality

$$p(\theta_{a,w})(w - T(.) - b) = \Sigma_a \tag{2}$$

holds for all w. Equation (2) defines a relation between a, w, $\theta_{a,w}$ and Σ_a. Along this relation, participants of type a are ready to search for a lower post-tax wage $w - T(.)$ if this is compensated by a higher probability of finding a job $p(\theta_{a,w})$. Given the distributions of vacancies $V_{a,w}$ and wages across submarkets, Equation (2) determines the number of job-seekers $U_{a,w}$ on each of them.

At stage 2, when a firm creates a vacancy of type a and offers a wage w, her expected profit equals $q(\theta_{a,w})(a - w) - \kappa_a$, where $\kappa_a > 0$ denotes the cost of creating this vacancy. This cost includes the screening of applicants and the investment in equipment. Under free-entry, the expected profit in any submarket is nonpositive at the subgame perfect equilibrium. Otherwise, creating an additional vacancy on a submarket (a, w') where $q(\theta_{a,w'})(a - w') > \kappa_a$ would be a profitable deviation. Hence, at equilibrium, firms create a positive number of vacancies on submarket (a, w) only if the zero-profit condition $q(\theta_{a,w})(a - w) = \kappa_a$ holds. Let $L(a, w)$ be the employment probability under the zero profit condition. One has

$$L(a, w) \stackrel{\text{def}}{\equiv} p\left(q^{-1}\left(\frac{\kappa_a}{a - w}\right)\right) \tag{3}$$

Due to constant returns to scale in the matching technology, the probability of being employed is independent of the number of participants and depends only on the skill and the wage levels. Assumption AS 1 implies that $L(.,.)$ is twice-differentiable and verifies

$$\partial L(a, w) / \partial w < 0$$

As the wage increases, tightness has to decline to rise the probability of filling a vacancy. Function $L(.,.)$ summarizes labor demand behavior in our economy.

At the subgame perfect equilibrium, the wage has to solve:[7]

$$w_a = \arg\max_w \quad L(a,w) \cdot (w - T(.) - b) \tag{4}$$

This condition pins down the participants' expected surplus:

$$\Sigma_a = L(a, w_a) \cdot (w_a - T(.) - b) \tag{5}$$

To show how subgame perfection implies (4) and (5), assume by contradiction that there exists a wage w' such that $L(a,w')(w' - T(.) - b) > \Sigma_a$. Consider then one deviating firm that offers w' instead of w_a. This deviation shifts the distribution of vacancies across submarkets. The distribution of job-seekers shifts then in order to keep the indifference condition (2). Hence, one must have[8] $(a - w')\, q\left(\theta_{a,w'}\right) - \kappa_a > 0$, so the deviation is profitable. In sum,

Definition 1 *At the CSE on labor market of skill a, the wage w_a is given by (4), the employment probability equals $L(a,w_a)$, the expected surplus Σ_a is given by (5), the participation rate is $G(a,\Sigma_a)$ and the employment rate is $L(a,w_a) \cdot G(a,\Sigma_a)$.*

Although the government does not observe whether each non-employed individual is unemployed or out-of-the labor force, the government knows functions $L(.,.)$ and $G(.,.)$ and the wage schedule $a \mapsto w_a$. Therefore, it has the ability to decompose each skill-specific employment rate as the product of a labor demand term (the employment probability[9]) and a labor supply term (the participation rate).

To deal with optimal taxation, we only need the $L(.,.)$ function, the distribution of types given by $f(.)$ and $G(.)$ and the wage-setting condition (4) or equivalently (5). There are alternative micro-foundations for (4) within a matching environment. When the matching function is of a Cobb-Douglas form, Nash bargaining under the Hosios (1990) condition leads to (4).[10] Another possibility is to assume that a skill-specific utilitarian monopoly union selects the wage w_a after individuals' participation decisions, but before firms' decisions about vacancy creation. Due to the free-entry condition, the monopoly union takes into account the consequence of its wage decision on the employment probability as described by function $L(a,w)$. A Utilitarian monopoly union therefore maximizes $L(a,w) \cdot (w - T(.)) + (1 - L(a,w)) \cdot b$, which is equivalent to (4) (see Mortensen and Pissarides 1999).

[7] A *priori*, different wage levels can solve (4) depending on the curvature of $T(.)$. However, we focus only on situations where the optimal $T(.)$ is such that the solution of (4) is unique. This assumption is also implicitly made in the optimal taxation literature *à la* Mirrlees (1971) where different levels of labor supply can *a priori* solve a worker's program.

[8] Equation (2) implies $\Sigma_a = p(\theta_{a,w'})(w' - T(.) - b)$. So, $\Sigma_a < L(a,w')(w' - T(.) - b)$ induces successively that $p(\theta_{a,w'}) < L(a,w')$, $\theta_{a,w'} < q^{-1}\left(\frac{\kappa_a}{a-w'}\right)$ and $q(\theta_{a,w'}) > \kappa_a/(a-w')$, since $p(.)$ is increasing and $q(.)$ is decreasing.

[9] Note that $1 - L(a,w_a)$ corresponds to the skill-specific unemployment rate.

[10] See Hungerbühler et alii (20006). Further restrictions on κ_a are needed to keep $L(a,w)$ below 1.

II.3 The fiscal incidence

In this subsection, we explain how the type a CSE is influenced by the tax/benefit system $(b, T(.))$. When $T(.)$ is differentiable, the first-order condition[11] associated to (4) writes:

$$0 = \frac{\partial \log L}{\partial w}(a, w) + \frac{\eta(.)}{w} \tag{6}$$

where[12]

$$\eta(.) = \frac{1 - \frac{\partial T}{\partial w}(.)}{1 - \frac{T(.)+b}{w}} = \frac{\partial \log (w - T(.) - b)}{\partial \log w} \tag{7}$$

When the wage increases by one unit, the term $\partial \log L/\partial w (a, w)$ measures the relative decrease in the employment probability, while $\left(1 - \frac{\partial T}{\partial w}(.)\right)/(w - T(.) - b)$ measures the relative increase in the ex-post surplus. At the equilibrium, Equations (6) requires that these two effects sum to zero. So, $\eta(.)$ has to be positive. As the expected surplus is positive, this implies that the marginal tax rate has to be lower than 1.

The effect of a tax reform on the equilibrium wage is entirely summarized by the change in the elasticity $\eta(.)$. This term is the wage elasticity of the ex-post surplus of an employed worker $x = w - T(w) - b$. A given relative rise in the wage increases this ex-post surplus less when $\eta(.)$ is lower. The elasticity $\eta(.)$ is reminiscent of the Coefficient of Residual Income Progression which measures the wage elasticity of net earnings (Musgrave and Musgrave 1976). $\eta(.)$ is actually the Coefficient of Residual Income Progression divided by one minus the net replacement ratio $b/(w - T(.))$. We henceforth refer to $\eta(.)$ at the CSE as the *Extended Coefficient of Residual Income Progression* (ECRIP for short). The *higher* the ECRIP for some wage level w, the *less* progressive is the tax schedule around w.

To understand how a tax reform influences the equilibrium wage, consider a tax change in the neighborhood of w_a such that the ECRIP decreases by $\Delta\eta < 0$. Then, according to (6), the equilibrium wage w_a decreases (See Appendix A). As the tax schedule becomes more progressive, a given rise in the wage leads to a smaller gain in the ex-post surplus $w - T(.) - b$, while the loss of employment probability remains unchanged. So the expected surplus can increase if the wage is reduced. Since the CSE maximizes this expected surplus Σ_a, the wage declines and the employment probability rises.

This wage-moderating effect of tax progressivity is a key mechanism in our model. It is well-known in the equilibrium unemployment literature and holds also under Nash bargaining (see Lockwood Manning 1993, Pissarides 1998), monopoly union (Hersoug 1984) or efficiency wages (Pisauro 1991). A similar effect of tax progressivity on total income is also present in models with competitive labor markets. In the latter, a rise in tax progressivity typically

[11] The solution to (4), if any, necessarily lies in $(-\infty, a - \kappa_a]$. Since $L(a, a - \kappa_a) = 0$, $w = a - \kappa_a$ does not solve (4). From a theoretical viewpoint, the wage can be negative whenever $T(.)$ is negative enough to keep some agents of type a participating to the labor market (i.e. $w - T(.) > b$).

[12] When the government can observe the skill level of an employed worker as in Section III, then η is a function of the skill, the wage and the benefit level $\eta(.) = \eta(a, w)$. When the government does not observe the skill level, as in Section IV, then η depends only on the wage and on the benefit level $\eta(.) = \eta(w)$.

decreases labor supply through a substitution effect. The wage moderating effect occurs in all these models because a more progressive tax-schedule reduces the marginal gain for the worker from obtaining a higher gross wage, while the cost remains unchanged. The models only differ by the economic mechanisms that explain why a reduction in gross wage induces a gain for the worker. In our model, a lower gross wage increases the employment probability, thereby reducing unemployment. Sørensen (1997) and Røed and Strøm (2002) give some empirical evidence in favor of the wage-moderating and the unemployment-reducing effects of tax progressivity.

In addition to its effect on wage and unemployment through the ECRIP, taxation also influences participation decisions. To isolate this effect, consider a tax reform that decreases the tax level $T(.)$, without changing the ECRIP.[13] Such tax reform does neither change the wage level, nor the employment probability. However, the ex-post surplus $w_a - T(.) - b$ is raised, so the surplus Σ_a an agent of type a can expect from participation increases. Therefore, such a reform increases the participation rate $G(a, \Sigma_a)$, thereby the employment rate $L(a, w_a) \cdot G(a, \Sigma_a)$.

II.4 The government

In this subsection, we first present the government's budget constraint and then its various objectives. The government's budget constraint writes:

$$b\left(\int_{a_0}^{a_1} \{1 - L(a, w_a) \cdot G(a, \Sigma_a)\} f(a)\, da\right) = \int_{a_0}^{a_1} T(.) \cdot L(a, w_a) \cdot G(a, \Sigma_a) \cdot f(a)\, da$$

An additional employed worker of type a generates additional tax revenues $T(.)$ and saves welfare benefits b. Therefore, on average, an additional participant of type a increases public resources by $(T(.) + b) \cdot L(a, w_a)$. We define

$$Y(a, w) \stackrel{\text{def}}{\equiv} w \cdot L(a, w) \tag{8}$$

$Y(a, w)$ is the average taxable income generated by an additional participant. So, given (5) and (8), the additional tax revenue per participant equals $Y(a, w_a) - \Sigma_a$. The government's budget constraint can then be written as:

$$b = \int_{a_0}^{a_1} [Y(a, w_a) - \Sigma_a] \cdot G(a, \Sigma_a) \cdot f(a)\, da \tag{9}$$

The aggregation of gross wages over skill levels is split between the financing of the benefit and the surpluses obtained by employed workers.

We will consider different normative criteria. The first is the Benthamite (utilitarian) objective which maximizes the sum of utilities. There are $G(a, \Sigma_a) f(a)$ participants of type a whose net income is $w - T(.)$ if they are employed and b otherwise, while non participants obtain $b + \delta$.

[13] That is, consider a rise in $\log(w - T(.) - b)$ by a constant amount, so that $\eta(.) = \partial \log(w - T(.) - b) / \partial w$ remains unchanged.

So, the Benthamite objective writes:

$$\begin{aligned}
\mathcal{U} &= \int_{a_0}^{a_1} \left\{ (L(a,w_a)(w_a - T(.)) + (1 - L(a,w_a))b) \cdot G(a,\Sigma_a) + \int_{\Sigma_a}^{+\infty} (b+\delta) \cdot g(a,\delta) \cdot d\delta \right\} f(a) \cdot da \\
&= \int_{a_0}^{a_1} \left\{ (\Sigma_a + b) \cdot G(a,\Sigma_a) + \int_{\Sigma_a}^{+\infty} (b+\delta) \cdot g(a,\delta) \cdot d\delta \right\} f(a) \cdot da
\end{aligned}$$

where the second equality uses (5). Given the government's budget constraint (9), this objective can be rewritten as:

$$\mathcal{U} = \int_{a_0}^{a_1} \left\{ Y(a,w_a) \cdot G(a,\Sigma_a) + \int_{\Sigma_a}^{+\infty} \delta \cdot g(a,\delta) \cdot d\delta \right\} f(a) \cdot da \tag{10}$$

The benthamite objective aggregates average earnings plus the value of non-market activities over the whole population, no matter how they are distributed. In this sense, the Benthamite criterion is an extreme case. At the other extreme, the Maximin (Rawlsian) criterion only values the utility of the least well-off. Unemployed individuals get b, which is always lower than the employed workers' and non participants' income, respectively $(w - T(.))$ and $(b+\delta)$. Therefore, a Maximin government aims at maximizing b.

II.5 The *laissez faire* equilibrium

Before characterizing the laissez faire equilibrium, we make two additional assumptions on respectively functions $Y(.,.)$ and $L(.,.)$. Following Equations (3) and (8), these assumptions actually amount to restricting the matching technology.

According to (4), in the absence of taxes and benefits, the wage maximizes $Y(a,.)$. A unique solution is guaranteed if:

Assumption AS 2 *For any* (a,w) :

$$\frac{\partial^2 \log Y}{\partial w^2}(a,w) < 0$$

A wage increase has a direct positive effect on $Y(a,.)$ and a negative effect through the employment probability. As the wage tends to 0, so does $Y(a,.)$. As the wage tends to $a - \kappa_a$, $L(a,w)$ tends to 0 according to Assumption 1 and Equation (3). As a consequence, $Y(a,.)$ admits an interior maximum within $(0, a-\kappa_a)$. Using superscript LF for the *laissez faire* allocation, Assumption 2 implies that there is a unique *laissez faire* value of the wage which verifies $\partial \log L / \partial w \left(a, w_a^{\mathrm{LF}}\right) + 1/w_a^{\mathrm{LF}} = 0$. Furthermore, w_a^{LF} increases with the level of skill if:

Assumption AS 3 *For any* (a,w) :

$$\frac{\partial^2 \log L}{\partial a \partial w}(a,w) > 0$$

The wage-elasticity of the labor demand equals $w \cdot (\partial \log L / \partial w)$. So, the derivative of this elasticity with respect to the skill a is $w \cdot \left(\partial^2 \log L / \partial a \partial w\right)$. Therefore, Assumption 3 implies that, in absolute value, the wage-elasticity of the labor demand decreases with the skill level. This feature is largely confirmed by the empirical literature (see Hamermesh 1993).

Appendix B verifies that the CES specification (1) satisfies these two assumptions if the cost of creating a vacancy decreases or increases at most proportionately with respect to the skill level a, that is[14] $a\ \dot{\kappa}_a \leq \kappa_a$. There is a lack of clear-cut empirical evidence on the link between κ_a and a. However, most of the literature assumes that the cost of creating a vacancy is either constant or proportional to productivity (Pissarides, 2000, page 10).

We now show that the CSE maximizes the Benthamite objective in a *laissez faire* economy. We get the following Proposition:

Proposition 1 *The laissez faire allocation maximizes the Benthamite objective.*

Proof. For each a and Y, the function $\Sigma \mapsto Y \cdot G(a, \Sigma) + \int_{\Sigma}^{+\infty} \delta \cdot g(a, \delta) \cdot d\delta$ reaches a unique maximum for $\Sigma = Y$. Therefore, when we compare any allocation $a \mapsto (w_a, \Sigma_a)$ to the laissez faire one, we get using (10):

$$\begin{aligned}
\mathcal{U}^{\mathrm{LF}} &= \int_{a_0}^{a_1} \left\{ Y\left(a, w_a^{\mathrm{LF}}\right) \cdot G\left(a, \Sigma_a^{\mathrm{LF}}\right) + \int_{\Sigma_a^{\mathrm{LF}}}^{+\infty} \delta \cdot g(a, \delta) \cdot d\delta \right\} f(a) \cdot da \\
&\geq \int_{a_0}^{a_1} \left\{ Y\left(a, w_a^{\mathrm{LF}}\right) \cdot G(a, \Sigma_a) + \int_{\Sigma_a}^{+\infty} \delta \cdot g(a, \delta) \cdot d\delta \right\} f(a) \cdot da \\
&\geq \int_{a_0}^{a_1} \left\{ Y(a, w_a) \cdot G(a, \Sigma_a) + \int_{\Sigma_a}^{+\infty} \delta \cdot g(a, \delta) \cdot d\delta \right\} f(a) \cdot da = \mathcal{U}
\end{aligned}$$

The first inequality holds because $\Sigma_a^{\mathrm{LF}} = L\left(a, w_a^{\mathrm{LF}}\right) \cdot w_a^{\mathrm{LF}} = Y\left(a, w_a^{\mathrm{LF}}\right)$ at the *laissez faire*, according to (5) and (8). The second inequality holds because w_a^{LF} maximizes $Y(a, .)$. ■

Because of our wage-setting mechanism (4), wages maximize "efficiency" (defined according to the Benthamite criterion) in the absence of taxes. This proposition implies that taxation is used only for redistributive purposes and not to restore efficiency.

III Optimal Taxation when workers' productivity is observable

In this section, we consider the case where the government can condition the taxation of employed workers on their wage and skill levels. In this case, the government can freely select the level of tax $T(a, w_a)$ and the degree of tax progression as measured by the ECRIP $\eta(a, w_a)$ at the CSE of each skill-specific labor market. Through the wage-setting mechanism (6), the government can therefore decentralize any wage level w_a by choosing an appropriate ECRIP. Then, any level

[14] Throughout the paper, a dot over a variable denotes its total derivative with respect to the skill.

of expected surplus Σ_a can be decentralized trough the choice of the level of taxes (see (5)). We use superscript $*$ to denote the optimal allocation in this case. From (9), the Maximin optimum thus solves:

$$b^* = \max_{w_a,\Sigma_a} \qquad \int_{a_0}^{a_1} \left\{ \left[Y\left(a, w_a\right) - \Sigma_a \right] \cdot G\left(a, \Sigma_a\right) \cdot f\left(a\right) \right\} da \tag{11}$$

This problem is solved in two steps. First, the wage is adjusted to maximize the average taxable income $Y(a,.)$ per participant. The optimal level of wage w_a^* equals the *laissez faire* value w_a^{LF} and thus the wage at the Benthamite optimum, according to Proposition 1. Hence, the redistributive motive of the government does not distort the labor demand. Second, for each a, the optimal level of expected surplus solves:

$$G\left(a, \Sigma_a^*\right) = \left(Y\left(a, w_a^*\right) - \Sigma_a^*\right) g\left(a, \Sigma_a^*\right) \tag{12}$$

At the optimum, a mechanical effect and a behavioral effect of a rise in Σ_a^* on the maximin objective have to balance. According to Equations (5) and (8), the term $Y(a, w_a^*) - \Sigma_a^*$ measures the expected level of public resources per participant at the optimum. The left-hand side of (12) measures the *mechanical* effect. Since $\Sigma_a^* = L(a, w_a^*)(w_a^* - T(a, w_a^*) - b^*)$, increasing participants' expected surplus requires a decline in public resources $T(w_a^*) + b^*$ per additional employed. For a given number of participants, the government therefore collects less resources to finance the benefit b, which is detrimental to the Maximin objective.

The right-hand side of (12) measures the *behavioral* effect. A rise in Σ_a^* induces more agents of skill a to participate. For a given level of tax, this increase in participation raises the resources available to finance the benefit b, whenever $Y(a, w_a^*) - \Sigma_a^* > 0$.

Let

$$\pi\left(a, \Sigma\right) \stackrel{\mathrm{def}}{\equiv} \frac{\Sigma \cdot g\left(a, \Sigma\right)}{G\left(a, \Sigma\right)} \tag{13}$$

be the (endogenous) elasticity of the participation rate with respect to the participants' expected surplus Σ. Conditional on the (gross) wage level w_a, $\pi(a,.)$ is also the elasticity of the participation rate with respect to the ex-post surplus of employed workers $w_a - T(.) - b$. Empirical studies about participation (mentioned by e.g. Saez 2002) estimate the latter elasticity. With this notation, the optimal Σ_a^* solves:

$$\Sigma_a^* = Y\left(a, w_a^*\right) \frac{\pi\left(a, \Sigma_a^*\right)}{1 + \pi\left(a, \Sigma_a^*\right)} \tag{14}$$

When participation is more elastic along the optimum, the participants' expected surplus increases and the level of public resources decreases. Furthermore, participants' surplus and thereby participation rates are lower at the Maximin optimum compared to the laissez faire allocation where $\Sigma_a^{\mathrm{LF}} = Y(a, w_a^*)$. From the budget constraint (9), decreasing the expected surplus Σ_a below $Y(a, w_a^*)$ is needed to finance the benefit.

From (5), (8) and (14), the optimal allocation can be decentralized by the tax function[15]

$$T(a,w) = \frac{1}{1+\pi(a,\Sigma_a^*)} w - b^* \tag{15}$$

Formula (15) is similar to the one obtained by Diamond (1980), Saez (2002) and Choné and Laroque (2005) in models of optimal income taxation with extensive labor supply responses, but without involuntary unemployment. This similarity is due to the skill-specificity of the tax schedule. This property enables to optimize independently along the demand and the participation margins.

IV Optimal Taxation when workers' productivity is not observable

In this section, we assume that the government can condition the taxation of employed workers on their wage but not on their skill level. Since the second heterogeneity about the value of non-market activities is only relevant for non-participants, the government faces an adverse selection problem with random participation. According to the *taxation principle* (see e.g. Guesnerie 1995), the set of allocations generated by non-linear tax schedules is the set of allocations that can be implemented by direct truthful-revealing mechanisms. Following Mirrlees (1971), it is much more convenient to solve the government's problem in terms of truthful-revealing allocations. However, we follow Saez (2001) in interpreting the optimality conditions in terms of marginal reforms of tax instruments.

For each skill a, the government designs a wage level w_a and a level of surplus $x_a = w_a - T(w_a) - b$. This allocation has to be incentive compatible. At the subgame perfect equilibrium, considering a market a, the allocation (w_a, x_a) should be superior to any other allocation $(w_{a'}, x_{a'})$. Given the wage-setting mechanism (4), this leads to the following incentive constraints:

$$\forall \left(a, a'\right) \in [a_0, a_1]^2 \qquad L(a,w)\ x_a \geq L(a, w_{a'})\ x_{a'} \tag{16}$$

To verify that the single-crossing condition holds, let us compute the marginal rate of substitution between the wage and the employed workers' surplus:

$$\left.\frac{\partial x}{\partial w}\right|_{L(a,w)\ x=cst} = -x\ \frac{\partial \log L}{\partial w}(a,w)$$

Assumption 3 ensures that this rate decreases in type a, so we can apply the techniques used in adverse selection models. Following Guesnerie and Laffont (1984) we consider only allocations that are continuous in skill levels and piecewise-differentiable. Appendix C shows that an

[15] With this skill specific tax schedule, the wage at the CSE maximizes $(w - T(a,w) - b)\, L(a,w)$, which equals

$$\frac{\pi(a,\Sigma_a^*)}{1+\pi(a,\Sigma_a^*)} w \cdot L(a,w) = \frac{\pi(a,\Sigma_a^*)}{1+\pi(a,\Sigma_a^*)} Y(a,w)$$

The equilibrium wage therefore equals its laissez faire value $w_a^{\mathrm{LF}} = w_a^*$.

allocation $a \mapsto (w_a, x_a, \Sigma_a = x_a L(a, w_a))$ verifies (16) if and only if,

$$\forall a \in [a_0, a_1] \qquad \frac{\dot{\Sigma}_a}{\Sigma_a} = \frac{\partial \log L}{\partial a}(a, w_a) \tag{17}$$

and the monotonicity constraint according to which $a \mapsto w_a$ is non-decreasing holds. Equation (17) is derived by applying the envelope theorem to Equation (5). It is key since it specifies how the profiles of wages and of expected surplus are connected when taxation cannot be conditioned on the skill level.

The government maximizes the same objective as in section III but under (17) and the monotonicity constraint. It is therefore restricted to choose a non-decreasing function $a \mapsto w_a$ and an initial expected surplus Σ_{a_0}. Then, $a \mapsto \Sigma_a$ is deduced thanks to (17), and the level of the benefit b is finally given by (9). When the monotonicity constraint is not binding, there is no bunching and the optimal Maximin allocation satisfies (see Appendix D):[16]

$$\frac{\partial Y}{\partial w}(a, w_a) \cdot G(a, \Sigma_a) \cdot f(a) = \frac{\partial^2 \log L}{\partial a \partial w}(a, w_a) \cdot Z_a \tag{18}$$

$$\text{where} : Z_a = \int_a^{a_1} \{\Sigma_t - [Y(t, w_t) - \Sigma_t] \cdot \pi(t, \Sigma_t)\} \cdot G(t, \Sigma_t) \cdot f(t)\, dt \tag{19}$$

This optimum solves a trade-off between a direct effect and an indirect one that we now explain. A rise Δw in the wage designed for workers of skills in the infinitesimal interval $[a, a + \delta a]$ affects the expected taxable income per participant by $\Delta Y = \partial Y / \partial w (a, w_a) \cdot \Delta w$. If $w_a < w_a^*$, (respectively $>$) the wage increases more (less) than the employment probability decreases. So the net effect is positive (negative). Multiplying ΔY by the participation rate $G(a, \Sigma_a)$ and the measure $f(a)\, \delta a$ of the corresponding agents gives the left-hand side of (18) times $\delta a \cdot \Delta w$. This is the direct wage distorsive effect.

The indirect effect is the consequence of the decentralization of the wage increase Δw. According to (6), only agents whose productivity belongs to $[a, a + \delta a]$ should be confronted with a less progressive tax, that is a local rise in the ECRIP $\Delta \eta$. Let $\delta w = w_{a+\delta a} - w_a$ so that the interval $[a, a + \delta a]$ of the skill distribution maps into the interval $[w_a, w_a + \delta w]$ of the wage distribution. The rise Δw of the wage designed over $[a, a + \delta a]$ is decentralized by a rise $\Delta \eta$ of the ECRIP in the wage interval $[w, w + \delta w]$. According to Appendix A which exploits (6), one has:

$$\Delta \eta \cdot \frac{\delta w}{w} = \frac{\partial^2 \log L}{\partial a \partial w}(a, w_a) \cdot \Delta w \cdot \delta a$$

Figure 1 uses log scales to represent the ex-post surplus as a function of the gross wage. In particular, the slope of the tax schedule equals the ECRIP (see 7). A rise $\Delta \eta$ in the ECRIP corresponds to a locally steeper tax schedule. Consequently, for those employed of types t above $a + \delta a$, the surplus increases by $\Delta x_t / x_t = \Delta \eta \cdot (\delta w / w)$. Since for them, the ECRIP is unaffected,

[16] If otherwise, there is bunching over an interval $[\underline{a}, \overline{a}]$, then the necessary condition integrates (18) over $[\underline{a}, \overline{a}]$, as described by (28) in the Appendix.

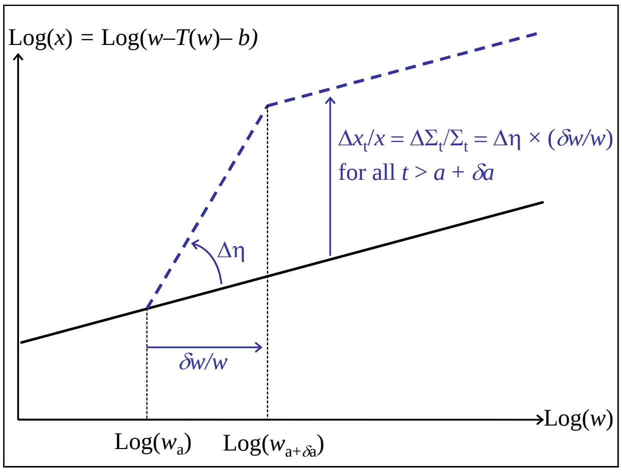

Figure 1: Tax reform required to decentralize a wage increase Δw over the interval $[a, a + \delta a]$.

their wage and employment remain unchanged and thereby their expected surplus increases by a common *relative* amount that equals

$$\frac{\Delta \Sigma_t}{\Sigma_t} = \frac{\Delta x_t}{x_t} = \frac{\partial^2 \log L}{\partial a \partial w}(a, w_a) \cdot \Delta w \cdot \delta a \qquad \text{for } t > a + \delta a \tag{21}$$

Let Z_a be the total loss of resources available to finance b due to a unit relative rise in the expected surplus for all types t above a. Multiplying Z_a by the relative increase in the expected surpluses given by (21) equals the right-hand side of (18) times $\delta a \cdot \Delta w$.

To understand better the indirect effect, it is necessary to detail the impact of a rise in Σ_t on Z_a. As in Section III, a rise in Σ_t has a *mechanical* negative effect and a positive *behavioral* effect on the Maximin objective. From (9), these effects are captured by changes in $(Y(t, w_t) - \Sigma_t) \cdot G(t, \Sigma_t) \cdot f(t)$ all along the distribution of skills. Through the mechanical effect, conditional on the participation rate, a unit *relative* rise in Σ_t reduces the resources available to pay the benefit b by $\Sigma_t \cdot G(t, \Sigma_t) \cdot f(t)$. Through the behavioral effect, a unit *relative* increase in Σ_t pushes up the participation rate by $\pi(t, \Sigma_t) \cdot G(t, \Sigma_t) \cdot f(t)$. This has a beneficial effect on public resources equal to $(Y(t, w_t) - \Sigma_t) \cdot \pi(t, \Sigma_t) \cdot G(t, \Sigma_t) \cdot f(t)$. In sum, Z_a cumulates the mechanical and behavioral effects on public resources of a unit relative increase in expected surplus for all types t above a.

The optimal condition on the initial level of expected surplus writes

$$Z_{a_0} = 0 \tag{22}$$

(see Appendix D). Increasing the initial level of the expected surplus Σ_{a_0}, while keeping wages unchanged, induces a proportional rise in all expected surpluses (see 17). At the optimum, the sum of all mechanical effects has to compensate the sum of all behavioral effects, which is granted by (22).

We now examine whether the optimal Maximin allocation of Section III can be decentralized when taxation cannot be conditioned on skill levels.

Proposition 2 *If $a \mapsto \pi(a, \Sigma_a^*)$ is constant, the allocation $a \mapsto (w_a^*, \Sigma_a^*)$ solves the Maximin problem when taxation cannot be conditioned on skill levels. This allocation is decentralized by*

$$T(w) = \frac{1}{1 + \pi(a, \Sigma_a^*)} w - b^*$$

where b^* is given by (11)

Proof. When $a \mapsto \pi(a, \Sigma_a^*)$ is constant, the tax function (15) that decentralizes $a \mapsto (w_a^*, \Sigma_a^*)$ depends only on the wage and no longer on the level of skill. ■

When the elasticity of participation is constant, the respective magnitudes of the mechanical and behavioral effects are identical across skill-specific labor markets. Therefore, from Equations (14) and (19), one has $Z_a = 0$ for all a. Hence, there is no incentive for the government to distort wages in order to redistribute expected surpluses between participants of different skills (see 18).

The assumption of a constant elasticity of participation is convenient but not plausible. As a first reason, this elasticity is a function of the expected surplus. Hence, it is actually endogenous except if[17]

$$G(a, \Sigma) = A_a \cdot \Sigma^{\pi_a} \qquad \text{where } A_a > 0 \text{ and } \pi_a > 0 \tag{23}$$

With this specification of the c.d.f. G, Proposition 2 requires in addition that $\pi_a = \pi$ for all a.

In the absence of structural estimation of the model, the best approximations of the elasticities of participation along the optimum are arguably the corresponding elasticities in actual economies. Note that under the particular specification (23) these elasticities are identical. Empirical studies suggest that participation decisions are more elastic at the bottom of the skill distribution (see the discussion in Saez 2002). The next proposition characterizes the Maximin optimum in the plausible case where the elasticity of participation is decreasing in skill levels along the optimum.

Proposition 3 *If along the Maximin optimum $a \mapsto \pi(a, \Sigma_a)$ is decreasing in a, then*

i) $w_a \leq w_a^{LF}$ and $w_a < w_a^{LF}$ almost everywhere in (a_0, a_1), while $w_{a_0} = w_{a_0}^{LF}$ and $w_{a_1} = w_{a_1}^{LF}$

ii) $\Sigma_a < \Sigma_a^{LF}$ everywhere in $[a_0, a_1]$, meaning that the participation rate is distorted downwards.

iii) The average tax rate $T(w)/w$ is an increasing function of the wage and the marginal tax rates $T'(w)$ are positive everywhere. The in-work benefit (if any) at the bottom-end of the distribution is lower than the benefit $-T(w_{a_0}) < b$.

[17] When we adopt this specification, we implicitly assume that A_a is such that one always has $\Sigma_a \leq (A_a)^{-1/\pi_a}$. Otherwise, the participation rate equals one and becomes inelastic.

This proposition is proved in Appendix D.1. An intuition behind *i*) is the following. A reduction in the pre-tax wage level below its *laissez faire* value is distorsive. Remembering (9), this is detrimental to the Maximin objective. To see why it is nevertheless optimal, recall that a reduction in the wage is decentralized thanks to a locally more progressive tax schedule. Consequently, the expected surpluses Σ_t for types t above a decrease (see Figure 1). As we saw above, this reduction has a mechanical positive effect and a behavioral negative effect on the Maximin objective. Whether the mechanical effect dominates or not depends on the elasticity of participation. When the elasticity of participation is decreasing in the skill level, lowering the expected surplus for high skilled workers is beneficial since for them the mechanical effect dominates. For each skill a, the wage is distorted downwards until at the margin the negative direct effect on the average taxable income equals the positive indirect effect on the expected surplus of participants endowed with a skill higher than a. Distorting wages downwards is thus optimal inside (a_0, a_1). However, at the top of the skill distribution, there are no higher skilled workers whose surplus could be reduced. Consequently, there is no distortion at the top. Reducing the wage at the bottom end of the skill distribution reduces the expected surplus for all participants. Since this has no beneficial effect (see Equation 22), there is no distortion at the bottom.[18]

An alternative intuition is the following. In the case where taxation can be conditioned on the skill level, from (15), the optimal tax schedule is such that $(T(a,w)+b)/w$ decreases with the elasticity of participation. When this elasticity decreases with the skill level, the government wishes that the ratio $(T(a,w)+b)/w$ be increasing in the skill level. When taxation can no more be conditioned on the skill level, leaving aside the situation where $\pi(a,\Sigma_a^*)$ is constant,the government can no longer implement (15). Then, the government chooses a tax scheme such that $(T(w)+b)/w$ is still increasing, yet flatter. By definition (7), the ECRIP is therefore below 1 and the tax schedule is progressive. Hence, the wage is distorted downwards (see Equation (6)).

Let us turn to the distortion along the participation margin claimed in *ii*). According to Proposition 1, the laissez faire maximizes the Benthamite objective (10). The mechanical effect of a rise in Σ_t is absent under the Benthamite objective, while it is present under the Maximin objective. Furthermore, the behavioral effect $g(a,\Sigma_a)(Y(a,w_a)-\Sigma_a)$ is equivalent under both objectives. Therefore, one has $\Sigma_a^{\mathrm{LF}} = Y(a,w_a^*) \geq Y(a,w_a) > \Sigma_a$. Moreover, $T(w)+b$ has to be positive everywhere to distort the participation rate downwards. These results are consistent with Choné and Laroque (2005). Combining points *i*) and *ii*), one observes that employment rates may be higher or lower at the Maximin optimum than at the laissez faire. This is because the employment probability is almost everywhere distorted upwards, while participation rate is distorted downwards, so the total effect is ambiguous.

Point *iii*) is explained as follows. For wages to be distorted downwards, the tax schedule has

[18] This last result only holds in the absence of bunching

to be progressive, that is the wage elasticity $\eta(.)$ of employed workers surplus $w - T(w) - b$ has to be lower than 1. This has two consequences. First, $w - T(w) - b$ has to increase in w less than proportionally. Thus, $(T(w) + b)/w$ has to be an increasing function of the wage and so is the average tax rate $T(w)/w$. Second, marginal tax rates have to be higher than $(T(w) + b)/w$, the latter being positive to distort participation downwards. Therefore, marginal tax rates $T'(w)$ have to be positive everywhere, including at the top of the income distribution.

To complete the analysis, we characterize the optimum under the less plausible assumption that the elasticity of participation is increasing in the skill level.

Proposition 4 *If along the Maximin optimum $a \mapsto \pi(a, \Sigma_a)$ is increasing in a, then*

i) $w_a \geq w_a^{LF}$ and $w_a > w_a^{LF}$ almost everywhere in (a_0, a_1), while $w_{a_0} = w_{a_0}^{LF}$ and $w_{a_1} = w_{a_1}^{LF}$

ii) $\Sigma_a < \Sigma_a^{LF}$ that is, the participation rate is distorted downwards.

iii) The in-work benefit (if any) at the bottom-end of the distribution is lower than the unemployment benefit $-T(w_{a_0}) < b$ and the marginal tax rate is positive at the top.

The formal proof is in Appendix D.2. Since the elasticity of participation is increasing in a, the optimal pre-tax wages are now distorted upwards almost everywhere. The participation decisions are distorted downwards for the same reason as above. Since the employment probability is distorted downwards almost everywhere, the employment rate is unambiguously distorted downwards. Distorting wages upwards requires a regressive taxation ($\eta > 1$). Hence, $w - T(w) - b$ has now to increase more than proportionally with respect to the wage, so $(T(w) + b)/w$ is now decreasing in the wage. Consequently, we cannot sign marginal tax rates, except at the top. This is because the no-distorsion at the top implies that the marginal tax rate equals to $(T(w) + b)/w$, which is positive. Finally, since marginal tax rates cannot be signed, it is not possible to conclude whether average tax rates are increasing or not.

Appendices

A Decentralization

From equation (6), we define

$$\mathcal{W}(w, a, \Delta\eta) \stackrel{\text{def}}{\equiv} \frac{\partial \log L}{\partial w} + \frac{\eta(w) + \Delta\eta}{w}$$

The second-order condition of (4) writes $\mathcal{W}'_w(w_a, a, 0) \leq 0$. Whenever this condition holds with a strict inequality, we can apply the implicit function theorem on Equation (6) to obtain:

$$\frac{\partial w}{\partial \Delta\eta} = -\frac{1}{\mathcal{W}'_w} \cdot \frac{1}{w} \qquad \frac{\partial w}{\partial a} = -\frac{1}{\mathcal{W}'_w} \cdot \frac{\partial^2 \log L}{\partial a \partial w}$$

Therefore, to an infinitesimal interval of the skill distribution $[a, a + \delta a]$ corresponds an interval of the wage distribution $[w, w + \delta w]$ where $w + \delta w = w_{a+\delta a}$, with

$$\delta w = -\frac{1}{\mathcal{W}'_w} \cdot \frac{\partial^2 \log L}{\partial a \partial w} \cdot \delta a$$

Moreover, a rise in the wage Δw requires a rise in the ECRIP

$$\Delta\eta = -\Delta w \cdot w \cdot \mathcal{W}'_w$$

Hence, a rise Δw over $[a, a+\delta a]$ is decentralized by a rise $\Delta\eta$ of the ECRIP over the wage interval $[w_a, w_a+\delta w]$ such that (21) holds.

B Microfoundation of $L(.,.)$

Under the CES specification (1) of the matching function one has

$$p(\theta) = \frac{V}{[U^\rho + V^\rho]^{\frac{1}{\rho}}} = \left(\frac{\theta^\rho}{1+\theta^\rho}\right)^{\frac{1}{\rho}} \qquad q(\theta) = \frac{U}{[U^\rho + V^\rho]^{\frac{1}{\rho}}} = (1+\theta^\rho)^{-\frac{1}{\rho}}$$

Hence, $q^{-1}(x) = (x^{-\rho}-1)^{1/\rho}$ and $p\left(q^{-1}(x)\right) = (1-x^\rho)^{1/\rho}$. Hence, following (3), we get:

$$L(a,w) = \left[1 - \left(\frac{a-w}{\kappa_a}\right)^{-\rho}\right]^{\frac{1}{\rho}} \tag{24}$$

Therefore

$$\frac{\partial \log L}{\partial w}(a,w) = -\frac{1}{a-w}\frac{\left(\frac{a-w}{\kappa_a}\right)^{-\rho}}{1-\left(\frac{a-w}{\kappa_a}\right)^{-\rho}} = -\frac{1}{a-w}\cdot\frac{1}{\left(\frac{a-w}{\kappa_a}\right)^{\rho}-1}$$

which is negative. Moreover,

$$\frac{\partial^2 \log L}{\partial w^2}(a,w) = -\frac{1}{(a-w)^2}\left(\frac{1}{\left(\frac{a-w}{\kappa_a}\right)^{\rho}-1} + \frac{\rho\left(\frac{a-w}{\kappa_a}\right)^{\rho}}{\left(\left(\frac{a-w}{\kappa_a}\right)^{\rho}-1\right)^2}\right)$$

which is also negative. Since $\log Y(a,w) = \log L(a,w) + \log w$, one has $\partial^2 \log Y/\partial w^2(a,w) = \partial^2 \log L/\partial w^2(a,w) - 1/w^2$. Hence $\rho > 0$ ensures that Assumption 2 is verified. Moreover,

$$\frac{\partial^2 \log L}{\partial a \partial w}(a,w) = \left(\frac{1}{a-w}\right)^2 \cdot \frac{1}{\left(\frac{a-w}{\kappa_a}\right)^{\rho}-1} + \frac{1}{a-w}\cdot\frac{\rho\left(\frac{a-w}{\kappa_a}\right)^{\rho}\left(\frac{1}{a-w}-\frac{\dot\kappa_a}{\kappa_a}\right)}{\left(\left(\frac{a-w}{\kappa_a}\right)^{\rho}-1\right)^2}$$

Therefore, $\rho \geq 0$ and $\frac{1}{a-w} \geq \frac{\dot\kappa_a}{\kappa_a}$ are sufficient conditions for Assumption 3.

C Incentive Compatible allocations

Let $a \mapsto (w_a, x_a, \Sigma_a)$ be an allocation such that for all a, $\Sigma_a = L(a, w_a)\cdot x_a$ and for all a and a' (16) is verified. Condition (16) can be rewritten as:

$$\log\Sigma_{a'} - \log\Sigma_a \leq \log L\left(a', w_{a'}\right) - \log L\left(a, w_{a'}\right)$$

Using the symmetric inequality where a and a' are inverted gives:

$$\log L\left(a', w_a\right) - \log L\left(a, w_a\right) \leq \log\Sigma_{a'} - \log\Sigma_a \leq \log L\left(a', w_{a'}\right) - \log L\left(a, w_{a'}\right) \tag{25}$$

Assume $a' > a$ and consider the two extreme parts of (25). They implies that

$$0 \leq \int_a^{a'} \left\{ \frac{\partial \log L}{\partial a}(t, w_{a'}) - \frac{\partial \log L}{\partial a}(t, w_a) \right\} dt$$

Since $a' > a$, and $\partial^2 \log L(a, w) / \partial a \partial w > 0$, this last inequality requires $w_{a'} \geq w_a$. Take $a' > a$. Then from (25) we get

$$\frac{\log L(a', w_a) - \log(a, w_a)}{a' - a} \leq \frac{\log \Sigma_{a'} - \log \Sigma_a}{a' - a} \leq \frac{\log L(a', w_{a'}) - \log(a, w_{a'})}{a' - a}$$

As a' tends to a, the left-hand side of this condition tends to $\partial \log L(a, w_a) / \partial a$. Since $a' \to w_{a'}$ is continuous, the right-hand side tends to $\partial \log L(a, w_a) / \partial a$ as well. Hence, $t \mapsto \Sigma_t$ admits a right-derivative for $t = a$, which equals to $\partial \log L(a, w_a) / \partial a$. Redoing the same reasoning for $a' < a$ insures that (17) holds for all a.

To show the reciprocal, let $a \mapsto (w_a, x_a, \Sigma_a)$ be an allocation such that for all a, $\Sigma_a = L(a, w_a) \cdot x_a$, $a \mapsto w_a$ is non-decreasing and (17) holds. We have to show that (16) holds for all $a' \neq a$. Assume that $a' < a$ (respectively $a' > a$). Then we have for all $t \in [a', a]$ (resp. for all $t \in [a, a']$), that $w_t \geq w_{a'}$ (respectively $w_t \leq w_{a'}$). Since $\partial^2 \log L(a, w) / \partial a \partial w > 0$ this implies that:

$$\int_{a'}^{a} \left\{ \frac{\partial \log L}{\partial a}(t, w_t) - \frac{\partial \log L}{\partial a}(t, w_{a'}) \right\} dt \geq 0$$

which induces

$$\int_{a'}^{a} \frac{\partial \log L}{\partial a}(t, w_t)\, dt \geq \log L(a, w_{a'}) - \log L\left(a', w_{a'}\right)$$

Integrating (17) between a' and a, we see that the left-hand side of the last inequality equals to $\log \Sigma_a - \log \Sigma_{a'}$. Therefore, one has

$$\log \Sigma_a \geq \log \Sigma_{a'} + \log L(a, w_{a'}) - \log L\left(a', w_{a'}\right)$$

which is equivalent to (16).

D The government's problem

Let $\sigma_a = \log \Sigma_a$ and $c_a = \dot{w}_a$. We use optimal control by considering σ_a and w_a as the state variables and c_a as the control.

$$\max_{c_a, w_a, \sigma_a} \int_{a_0}^{a_1} [Y(a, w_a) - \exp \sigma_a] \cdot G(a, \exp \sigma_a) \cdot f(a)\, da$$

$$s.t \quad : \quad \dot{\sigma}_a = \frac{\partial \log L}{\partial a}(a, w_a) \qquad \dot{w}_a = c_a \qquad c_a \geq 0$$

Following Guesnerie and Laffont (1984), the monotonicity constraint on w_a is captured by the positivity constraint on the control. Let q_a and μ_a be the multipliers associated to the equations of motion of respectively σ_a and w_a so $Z_a = -q_a$. The Hamiltonian writes

$$\mathcal{H}(c, w, \sigma, q, \mu, a) \stackrel{\text{def}}{\equiv} [Y(a, w) - \exp \sigma] \cdot G(a, \exp \sigma) \cdot f(a) + q \cdot \frac{\partial \log L}{\partial a}(a, w) + \mu \cdot c$$

Except on the finite number of points where we have allowed $a \mapsto w_a$ to be non-differentiable, one has:

$$-\dot{\mu}_a = \frac{\partial \mathcal{H}}{\partial w} = \frac{\partial Y}{\partial w}(a, w_a) \cdot G(a, \Sigma_a) \cdot f(a) + q_a \cdot \frac{\partial \log L}{\partial a \partial w}(a, w_a) \tag{26}$$

$$-\dot{q}_a = \frac{\partial \mathcal{H}}{\partial \sigma} = -\left\{ G(a, \Sigma_a) - [Y(a, w_a) - \Sigma_a] \cdot g(a, \Sigma_a) \right\} \cdot \Sigma_a \cdot f(a) \tag{27}$$

together with the transversality conditions $q_{a_0} = q_{a_1} = \mu_{a_0} = \mu_{a_0} = 0$.

Using $q_{a_1} = 0$, $Z_a = -q_a$, one has $Z_a = \int_a^{a_1} \dot{q}_t \cdot dt$. Hence, (27) gives (19). The transversality condition $q_{a_0} = 0$ gives (22). Maximizing the Hamiltonian with respect to the control c_a implies either that $\mu_a = 0$ and the monotonicity constraint $\dot{w}_a \geq 0$ is not binding, or $c_a = 0$ and $\mu_a \leq 0$. As a co-state variable μ_a is continuous and differentiable for any a where $a \mapsto w_a$ is differentiable.

- If $c_a > 0$, then $\mu_a = 0$. Since c_a is piecewise continuous, it remains positive around a so μ_a remains nil and $\dot{\mu}_a = 0$. Therefore (26) and $Z_a = -q_a$ give (18).
- If otherwise $c_a = 0$ and $\mu_a < 0$, then the wage is constrained to be constant over an interval and there is *bunching*. Let $(\underline{a}, \overline{a})$ be a maximized interval of bunching, such that there is neither bunching in the neighborhood on the left of $\underline{a}$, nor in the neighborhood on the right of $\overline{a}$. By continuity of $a \mapsto \mu_a$ one has $\mu_{\underline{a}} = \mu_{\overline{a}}$, so $0 = \int_{\underline{a}}^{\overline{a}} \dot{\mu}_a \cdot da$. Therefore (26) implies

$$\int_{\underline{a}}^{\overline{a}} \frac{\partial Y}{\partial w}(a, w_a) \cdot G(a, \Sigma_a) \cdot f(a) \cdot da = \int_{\underline{a}}^{\overline{a}} Z_a \cdot \frac{\partial \log L}{\partial a \partial w}(a, w_a) \cdot da \tag{28}$$

The following Lemma is useful

Lemma 1 *If $Z_a \geq 0$ (resp. $\leq$) over an interval $I \subset [a_0, a_1]$ then $a \mapsto Y(a, w_a)/\Sigma_a$ is non-decreasing (resp. non-increasing) over I.*

Proof. Equations (8) implies:

$$\frac{d \log Y(a, w_a)}{da} = \frac{\partial \log Y}{\partial w}(a, w_a) \cdot \dot{w}_a + \frac{\partial \log L}{\partial a}(a, w_a) = \frac{\partial Y}{\partial w}(a, w_a) \cdot \frac{\dot{w}_a}{Y(a, w_a)} + \frac{\dot{\Sigma}_a}{\Sigma_a}$$

the last equality coming from (17). Therefore, we get:

$$\frac{d \log \frac{Y(a, w_a)}{\Sigma_a}}{da} = \frac{\partial Y}{\partial w}(a, w_a) \cdot \frac{\dot{w}_a}{Y(a, w_a)} \tag{29}$$

Either there is bunching, in which case $\dot{w}_a = 0$, or Equation (18) applies which implies that $\partial Y(a, w_a)/\partial w$ and hence the right-hand-side of (29) have the sign of Z_a. ■

D.1 Proof of Proposition 3

First, we show that $Z_a \geq 0$ for all a and $Z_a > 0$ almost everywhere. From (19), one has:

$$\dot{Z}_a = \left\{ \left(\frac{Y(a, w_a)}{\Sigma_a} - 1 \right) \pi(a, \Sigma_a) - 1 \right\} \cdot g(a, \Sigma_a) \cdot (\Sigma_a)^2 \cdot f(a) \tag{30}$$

Assume by contradiction that Z is negative at some point. Since $a \mapsto Z_a$ is continuous, there exists an interval where Z remains negative. Given the transversality conditions $Z_{a_0} = Z_{a_1} = 0$, this implies the existence of an interval $[\underline{a}, \overline{a}]$ such that $Z_{\underline{a}} = Z_{\overline{a}} = 0$ and such that $Z_a \leq 0$ for all $a \in [\underline{a}, \overline{a}]$. Since $Z_{\underline{a}} = 0$ and Z_a is negative in the neighborhood on the right of $\underline{a}$, one has $\dot{Z}_{\underline{a}} \leq 0$. Given (30) this implies that:

$$\frac{Y(\underline{a}, \Sigma_{\underline{a}})}{\Sigma_{\underline{a}}} - 1 \leq \frac{1}{\pi(\underline{a}, \Sigma_{\underline{a}})}$$

From Lemma 1, one has that $a \mapsto (Y(a, w_a)/\Sigma_a)$ is weakly decreasing on $[\underline{a}, \overline{a}]$ so:

$$\frac{Y(\overline{a}, \Sigma_{\overline{a}})}{\Sigma_{\overline{a}}} - 1 \leq \frac{Y(\underline{a}, \Sigma_{\underline{a}})}{\Sigma_{\underline{a}}} - 1$$

Finally, since $Z_{\overline{a}} = 0$ and Z_a is negative in the neighborhood on the left of $\overline{a}$, one has $\dot{Z}_{\overline{a}} \geq 0$. Given (30) this implies that:

$$\frac{1}{\pi(\overline{a}, \Sigma_{\overline{a}})} \leq \frac{Y(\overline{a}, \Sigma_{\overline{a}})}{\Sigma_{\overline{a}}} - 1$$

These three inequalities imply that $\pi(\overline{a}, \Sigma_{\overline{a}}) \geq \pi(\underline{a}, \Sigma_{\underline{a}})$, which contradicts the assumption the $a \to \pi(a, \Sigma_a)$ is decreasing. Hence, one has $Z_a \geq 0$ and there is no interval where $Z_a = 0$. So, Z_a is positive almost everywhere.

In the absence of bunching, Equation (18) implies that for all a, $w_a \leq w_a^*$ and that one has almost everywhere $w_a < w_a^*$. In the case of bunching at $\tilde{w}$ over an interval $[\underline{a}, \overline{a}]$, one has from (18) and by continuity of $a \mapsto w_a$

$$\frac{\partial Y}{\partial w}(\underline{a}, w_{\underline{a}}) = \frac{\partial Y}{\partial w}(\underline{a}, \tilde{w}) \geq 0$$

So $\tilde{w} \leq w_{\underline{a}}^* < w_a$ for any $a \in (\underline{a}, \overline{a}]$. $Z_{a_1} = 0$ implies that $w_{a_1} = w_{a_1}^*$. Either there is bunching at the bottom of the skill distribution or $Z_{a_1} = 0$ implies that $w_{a_0} = w_{a_0}^*$.

We now show that average tax rates are increasing. Let $w < w'$ be two wage levels and let a and a' be such that $w = w_a$ and $w' = w_{a'}$. Integrating (29) for t between a and a' and using the fact that $\partial Y/\partial w(t, w_t)$ is positive almost everywhere, we obtain that $\log(Y_{a'}/\Sigma_{a'}) > \log(Y_a/\Sigma_a)$. Given (8) and (5), this leads to $(x_{a'}/w_{a'}) < (x_a/w_a)$ and to $(T(w) + b)/w < (T(w') + b)/w'$. Therefore, the mapping $w \mapsto (T(w) + b)/w$ is increasing, and so are average tax rates $T(w)/w$ are increasing too.

Since $Z_a \geq 0$ and $Z_{a_0} = 0$, one has $\dot{Z}_{a_0} \geq 0$. From (30), this implies that $Y(a_0, w_{a_0})/\Sigma_{a_0} \geq 1 + (1/\pi(a_0, \Sigma_{a_0})) > 1$. Since the mapping $a \mapsto Y(a, w_a)/\Sigma_a$ is non-decreasing, it reaches its minimum value for $a = a_0$. So for all a, one has $Y(a, w_a) > \Sigma_a$, which has two consequences. First, participation is always distorted downwards since $\Sigma_a < Y(a, w_a) \leq Y(a, w_a^*) = \Sigma_a^{\mathrm{LF}}$. Second, for all wages, $T(w) + b > 0$. Hence if there is an in-work benefit (that is, if for some wage $T(w) < 0$), then this in-work benefit is lower than unemployment benefits $-T(w) < b$.

Whenever the Tax schedule is differentiable, since $w_a \leq w_a^*$ (so $\partial Y/\partial w(a, w_a) \geq 0$, thereby $-w_a(\partial \log L/\partial w(a, w_a)) \geq 1$), Equations (6) and (7) imply that $T'(w_a) \geq (T(w_a) + b)/w_a$. Given that $(T(w_a) + b) > 0$ we obtain that marginal tax rates are always positive.

D.2 Proof of Proposition 4

First, we show that $Z_a \leq 0$ and that almost everywhere, $Z_a < 0$. Assume by contradiction that Z is positive at some point. Since $a \mapsto Z_a$ is continuous, there exist an interval where Z remains positive. Given the transversality conditions $Z_{a_0} = Z_{a_1} = 0$, this implies the existence of an interval $[\underline{a}, \overline{a}]$ such that $Z_{\underline{a}} = Z_{\overline{a}} = 0$ and such that $Z_a \leq 0$ for all $a \in [\underline{a}, \overline{a}]$.

Since $Z_{\underline{a}} = 0$ and Z_a is positive in the neighborhood on the right of $\underline{a}$, one has $\dot{Z}_{\underline{a}} \geq 0$. Given (30) this implies that:

$$\frac{Y(\underline{a}, \Sigma_{\underline{a}})}{\Sigma_{\underline{a}}} - 1 \geq \frac{1}{\pi(\underline{a}, \Sigma_{\underline{a}})}$$

From Lemma 1, one has that $a \mapsto (Y(a, w_a)/\Sigma_a)$ is weakly increasing on $[\underline{a}, \overline{a}]$ so:

$$\frac{Y(\overline{a}, \Sigma_{\overline{a}})}{\Sigma_{\overline{a}}} - 1 \geq \frac{Y(\underline{a}, \Sigma_{\underline{a}})}{\Sigma_{\underline{a}}} - 1$$

Finally, since $Z_{\overline{a}} = 0$ and Z_a is positive in the neighborhood on the left of $\overline{a}$, one has $\dot{Z}_{\overline{a}} \leq 0$. Given (30) this implies that:

$$\frac{1}{\pi\left(\overline{a}, \Sigma_{\overline{a}}\right)} \geq \frac{Y\left(\overline{a}, \Sigma_{\overline{a}}\right)}{\Sigma_{\overline{a}}} - 1$$

These three inequalities imply that $\pi\left(\overline{a}, \Sigma_{\overline{a}}\right) \leq \pi\left(\underline{a}, \Sigma_{\underline{a}}\right)$, which contradicts the assumption the $a \to \pi\left(a, \Sigma_a\right)$ is decreasing.

Hence, one has $Z_a \leq 0$ and there is no interval where $Z_a = 0$. So, Z_a is negative almost everywhere. In the absence of bunching Equation (18) implies that for all a, $w_a \leq w_a^*$, and that almost everywhere, one has $w_a < w_a^*$. In the case of bunching of wage at $\tilde{w}$ over an interval $[\underline{a}, \overline{a}]$, one has from (18) and by continuity of $a \mapsto w_a$

$$\frac{\partial Y}{\partial w}\left(\overline{a}, w_{\overline{a}}\right) = \frac{\partial Y}{\partial w}\left(\overline{a}, \tilde{w}\right) \leq 0$$

So $\tilde{w} \geq w_{\overline{a}}^* > w_a$ for any $a \in [\underline{a}, \overline{a})$. $Z_{a_1} = 0$ implies that $w_{a_1} = w_{a_1}^*$. Either there is bunching at the bottom of the skill distribution or $Z_{a_1} = 0$ implies that $w_{a_0} = w_{a_0}^*$.

Let $w < w'$ be two wage levels and let a and a' be such that $w = w_a$ and $w' = w_{a'}$. Integrating (29) for t between a and a' and using the fact that $\partial Y / \partial w\left(t, w_t\right)$ is negative almost everywhere, we obtain that $\log\left(Y_{a'}/\Sigma_{a'}\right) < \log\left(Y_a/\Sigma_a\right)$. Given (8) and (5), this leads to $\left(x_{a'}/w_{a'}\right) > \left(x_a/w_a\right)$ and to $\left(T\left(w\right) + b\right)/w > \left(T\left(w'\right) + b\right)/w'$. Therefore, the mapping $w \mapsto \left(T\left(w\right) + b\right)/w$ is decreasing.

Since $Z_a \leq 0$ and $Z_{a_1} = 0$, one has $\dot{Z}_{a_1} \geq 0$. From (30), this implies that $Y\left(a_1, w_{a_1}\right)/\Sigma_{a_1} \geq 1 + \left(1/\pi\left(a_1, \Sigma_{a_1}\right)\right) > 1$. Since the mapping $a \mapsto Y\left(a, w_a\right)/\Sigma_a$ is non-increasing, it re+aches its minimum value for $a = a_1$. So for all a, one has $Y\left(a, w_a\right) > \Sigma_a$, which has two consequences. First, participation is always distorted downwards since $\Sigma_a < Y\left(a, w_a\right) \leq Y\left(a, w_a^*\right) = \Sigma_a^{\mathrm{LF}}$. Second, for all wage w, $T\left(w\right) + b > 0$. Hence if there is an in-work benefit (that is if for some wage $T\left(w\right) < 0$) then this in-work benefit is lower than unemployment benefits $-T\left(w\right) < b$. Finally, since $w_{a_1} = w_{a_1}^*$, one has from (6) and (7) that $T'\left(w_{a_1}\right) = \left(T\left(w_{a_1}\right) + b\right)/w_{a_1}$.

References

[1] Blundell, R. and MaCurdy, T. , 1999, Labor Supply: A Review of alternative approaches, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, North-Holland.

[2] Boadway, R. and L. Jacquet, 2006, Optimal marginal and average income taxation under Maxi-min, Queen's University *working paper* 1073.

[3] Boone, J. and L. Bovenberg, 2004, The Optimal Taxation of Unskilled Labor with Job Search and Social Assistance, *Journal of Public Economics*, **88**(11), 2227-2258.

[4] Cahuc, P. and G. Laroque, 2007, Optimal Taxation and Monopsonistic Labor Market: Does Monopsony Justify the Minimum Wage?, *IZA discussion paper* 2955.

[5] Choné, P. and G. Laroque, 2005, Optimal Incentives for Labor force participation, *Journal of Public Economics*, **89**, 395-425.

[6] Choné, P. and G. Laroque, 2007, Negative marginal tax rates and heterogeneity, *mimeo* CREST.

[7] Diamond, P., 1980, Income Taxation with fixed hours of work, *Journal of Public Economics*, **13**, 101-110.

[8] Diamond, P., 1998, Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates, *American Economic Review*, **88**, 83-95.

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

[10] Guesnerie, R. and Laffont, J-J, 1984, A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm, *Journal of Public Economics*, **25**, 329-369.

[11] Hamermesh, D., 1993, *Labor demand*, Princeton University Press.

[12] Hersoug, T., 1984, Union Wage Responses to Tax Changes, *Oxford Economic Papers*, **36**, 37-51.

[13] Hindricks, J., E. Lehmann and A. Parmentier, 2006, Optimal Income Taxation and the shape of average tax rates, *Economics Bulletin*, **8**(2), pp. 1-6.

[14] Hosios, A.,1990, On the Efficiency of Matching and Related Models of Search and Unemployment, *Review of Economic Studies*, **57**, 279-298.

[15] Hungerbühler, M., E. Lehmann, A. Parmentier, B. Van der Linden, 2006, Optimal Redistributive Taxation in a Search Equilibrium Model, *Review of Economic Studies*, **73**, 743-767.

[16] Lockwood B. and Manning A., 1993, Wage Setting and the Tax System: Theory and evidences for the United Kingdom, *Journal of Public Economics*, **52**, 1-29.

[17] Mirrlees, J., 1971, An Exploration in the Theory of Optimum Income Taxation, *Review of Economic Studies*, **38**(2), 175-208.

[18] Mirrlees, J., 1976, Optimal Tax Theory: A Synthesis, *Journal of Public Economics*, **6**(3), 327-358.

[19] Mirrlees, J., 1999, Optimal marginal tax rates at low incomes, *mimeo*, Faculty of Economics, Cambridge, England.

[20] Moen, E., 1997, Competitive Search Equilibrium, *Journal of Political Economy*, **105**(2), 385-411.

[21] Mortensen, D. and Pissarides, C., 1999, New developments in models of search in the Labor Market, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, vol 3 B, North-Holland, Amsterdam.

[22] Musgrave, R. A. and. Musgrave P.B, 1976, *Public Finance in Theory and Practice*, Second Edition, McGraw-Hill, New-York.

[23] Petrongolo, B. and Pissarides, C. A., 2001, Looking into the Black Box: A Survey of the Matching Function, *Journal of Economic Literature*, **38**, 390-431.

[24] Pisauro, G., 1991, The effect of taxes on labour in efficiency wage models, *Journal of Public Economics*, **46**(3), 329-345.

[25] Pissarides, C. A., 1998, The impact of employment tax cuts on unemployment and wages: the role of unemployment benefits and tax structure, *European Economic Review*, **42**, 155-183.

[26] Pissarides, C. A., 2000, *Equilibrium Unemployment Theory*, Second Edition, MIT Press, Cambridge, USA.

[27] Rochet, J-C and L. Stole, 2002, NonLinear Pricing with Random Participation, *Review of Economic Studies*, **69**, 277-311.

[28] Røed, K., S. Strøm, 2002, Progressive Taxes and the Labour Market: Is the Tradeoff Between Equality and Efficiency Inevitable? , *Journal of Economic Surveys*, **16**(1), 77-110.

[29] Rogerson, R., R. Shimer and R. Wright, 2005, Search-Theoretic Models of the Labor Market: A Survey, *Journal of Economic Literature*, **43**, 959-988

[30] Saez, E, 2001, Using Elasticities to Derive Optimal Income Tax Rates, *Review of Economics Studies*, **68**, 205-229.

[31] Saez, E., 2002, Optimal Income Transfer Programs:Intensive Versus Extensive Labor Supply Responses, *Quarterly Journal of Economics*, **117**, 1039-1073.

[32] Shimer, R., 1996, Contracts in frictional labor markets, MIT mimeo.

[33] Sørensen, P. B., 1997, Public finance solutions to the European unemployment problem?, *Economic Policy*, **25**, 223-264