ON THE OPTIMALITY OF A MINIMUM WAGE: NEW INSIGHTS FROM OPTIMAL TAX THEORY *

Mathias HUNGERBÜHLER † Etienne LEHMANN ‡§
University of Namur CREST and IZA

5th February 2008

Abstract

We build a theoretical model to study whether a minimum wage can be welfare-improving if it is implemented in conjunction with an optimized nonlinear income tax. We consider this issue in a framework where search frictions on the labor market generate unemployment. Workers differ in productivity. The government does not observe workers’ productivity but only their wages. Hence, the redistributive policy solves an adverse selection problem. We show that a minimum wage is optimal if the bargaining power of the workers is relatively low. However, if the government controls the bargaining power, then it is preferable to set a sufficiently high bargaining power.

Keywords: Optimal taxation; Minimum wage; search-matching unemployment; Bunching; Wage bargaining

JEL Code: D86 H21 H23 J64 J68

*We thank participants of seminars in ERMES - Paris 2, CREST and CORE and at the PET 2007 meeting in Nashville and the T2M 2008 meeting in Cergy. In particular, we thank Pierre Cahuc, Guy Laroque, Cyril Nouveau, the editor Pierre Pestieau and two anonymous referees for fruitful comments. A part of this research was realized while Hungerbühler was visiting ERMES-Paris 2 and another part while Lehmann was visiting the Department of Economics of the Université Catholique de Louvain.

†E-mail: mathias.hungerbuhler@fundp.ac.be. Department of Economics, FUNDP Namur, Rempart de la Vièrge 8, 5000 Namur, Belgium.

‡Email: etienne.lehmann@ensae.fr, CREST-INSEE, Laboratoire de Macroeconomic, Timbre J360, 15 Boulevard Gabriel Peri, 92 245 Malakoff Cedex, France. Phone: +33 1 41 17 37 38 Fax: +33 1 41 17 76 66

§Corresponding Author
I Introduction

The minimum wage is one of the most controversial economic policies. On the ground of equity considerations, a minimum wage aims to play a redistributive role by increasing income for the least skilled workers. One might however counter-argue that redistributive taxation can achieve this goal in a more efficient way. On the ground of efficiency considerations, the minimum wage is often blamed for its adverse effects on labor demand. This is true as long as labor markets are perfectly competitive. However, the minimum wage can be helpful to correct for noncompetitive wage setting (see e.g. Robinson 1933 or Stigler 1946). In our opinion, it is therefore necessary to include optimal taxation and labor market imperfections when one considers the normative issue of the minimum wage.

In this paper, we propose a theoretical model to study whether a minimum wage can be welfare-improving if it is implemented in conjunction with an optimized nonlinear income tax à la Mirrlees (1971). To integrate explicitly the unemployment effects of a minimum wage, we consider this issue in a framework where search frictions on the labor market à la Mortensen and Pissarides (1999) generate endogenous “involuntary” unemployment (i.e. some workers are willing to work at the equilibrium wage, but fail to find a job).

In our model, workers differ with respect to productivity. They decide whether to search for a job, while firms search for workers to fill their job vacancies. If a worker and a firm are paired, they Nash-bargain the wage. The government observes wages, but not productivity. Hence, it faces an adverse selection problem. Since the productivity of a match is revealed through the wage, and since the negotiated wage is the one that maximizes the Nash Product, incentive constraints depend only on Nash Products. However, and contrary to the standard model in contract theory, workers’ participation constraints depend on a different variable than incentive constraints. In our case, the participation depends on the workers’ expected incomes while searching. We show that in such a context, bunching at the bottom of the wage distribution appears at the (second-best) optimum if the workers’ bargaining power is relatively low. In our model where wages are negotiated, we interpret such a bunching as an argument in favor of a binding minimum wage.

This first result holds under the assumption that the government cannot influence the workers’ bargaining power. One might however argue, that the government can – especially in a long-run perspective – influence the workers’ bargaining power to some degree. We show that, if the government can control the bargaining power, then it is desirable to increase a relatively low bargaining power, in which case our previous argument for the minimum wage disappears. The minimum wage thus acts in our model as an (imperfect) substitute for a rise in workers’ bargaining power.

The impact of a minimum wage in the case of a monopsony in the labor market has been studied among others by Robinson (1933) and Stigler (1946). Firms do not face
competition on the labor market and thus distort wages downwards, thereby reducing labor supply and eventually employment. Therefore, a binding minimum wage can restore efficiency and increase employment (along the labor supply), provided its level is not above the equilibrium wage in a perfectly competitive labor market. Our contribution differs in several ways from Stigler’s. First, we integrate taxation into the framework. As already noted by Stigler, tax measures might achieve the same result as the minimum wage, and possibly even in a more efficient way. We however show that a minimum wage is — under certain conditions — even optimal when tax measures are available. Second, while Stigler only analyses the efficiency problem, we analyze the impact of the minimum wage in a framework where the government wants to redistribute from low-income to high-income individuals, and thus faces an efficiency-equity trade-off. Third, we introduce “involuntary” unemployment. In Stigler’s simple monopsony model, every individual who is willing to work at the (monopsony) market wage is able to find a job. We assume search and matching frictions, which imply that not every individual who is willing to work at the equilibrium wage can find a job. Some people fail to find one and become “involuntarily” unemployed.

Many papers have already investigated whether the minimum wage can be useful in combination with an optimized redistributive tax (see Franklin Allen 1982, Stephen Allen 1987, Guesnerie and Roberts 1987, Drèze and Gollier 1993, Marceau and Boadway 1994, Boadway and Cuff 2001, among others). In particular, Stephen Allen (1987) considers a model with two types of imperfectly substitutable workers and endogenous hours of work à la Stiglitz (1982). He shows that a minimum wage is never optimal in conjunction with the optimized nonlinear income tax, because a rise in the minimum wage strengthens the relevant incentive constraint. Lee and Saez (2007) challenge this result. Their model includes labor supply responses along both the intensive and the extensive margin. Another difference is that the rationing on the low skilled labor demand induced by the minimum wage is “efficient”, in the sense that workers with the lowest surplus at the minimum wage are the first to drop into unemployment. In their model minimum wage can be useful in addition to nonlinear taxation. We abandon this framework of two types of imperfect-substitute labors and we instead base our labor demand margin on matching frictions. In our model, search frictions drive a wedge between (marginal) productivity and the wage, while productivity and hours of work are exogenous. Furthermore, we consider a model with a continuum of productivity which is — as we will show — more relevant for considering the issue of the form of the optimal redistributive allocation.

Finally, we extend the model of optimal redistributive taxation in a search equilibrium framework developed by Hungerbühler, Lehmann, Parmentier and Van der Linden (2006, 1For the case of the (non-)desirability of minimum wages in the context of a monopsony and redistributive taxation, see Cahuc and Laroque (2007).
henceforth HLPV). HLPV assume the Hosios (1990) condition, according to which workers’ bargaining power is equal to the elasticity of the matching function with respect to the mass of unemployed. This condition implies that the economy without government intervention is efficient, and thus constitutes an interesting benchmark. There is however no reason – neither theoretical nor empirical – why this condition should hold in reality. Empirical studies show that a bargaining power that is lower than the elasticity of the matching function is the most plausible case. We show that if this is true, then a wage rise for a given level of the Nash Product increases workers’ expected income. This effect then opens the room for a welfare-improving role of the minimum wage.

The paper is organized as follows. Section II presents the basic model, including incentive and participation constraints. Next, Section III solves the model for a given bargaining power. In particular, we show that a minimum wage is optimal if the bargaining power is sufficiently low. Section IV considers the case when the government can control the bargaining power. Finally, Section V concludes.

II The model

Our model follows the framework built in HLPV to deal with the optimal tax problem of Mirrlees (1971) within the equilibrium unemployment theory of Mortensen and Pissarides (1999) and Pissarides (2000). To keep things as simple as possible, we consider a static setting which has become standard in the models of search equilibrium with taxation (see also Boone and Bovenberg, 2002). There is a mass 1 of risk-neutral individuals. They can be either employed, unemployed or out of the labor force. Their preferences on consumption and leisure are assumed quasi-linear in consumption. Individuals differ with respect to their productivity \( a \in [a_0, a_1] \) with \( 0 \leq a_0 < a_1 < \infty \) according to the positive and continuous density function \( f(a) \) and the cumulative density function \( F(a) \). These functions are common knowledge, while the productivity is private information to the worker. The timing of the model is as follows:

1. First, the government commits to its policy. The policy consists in a tax schedule \( T(\cdot) \), a welfare benefit \( b \) and a minimum wage \( w \). Since the government cannot observe the individuals’ productivities but only their wages, this tax schedule is conditional on the gross wage \( w \) only. The government cannot observe whether a non-employed individual has searched for a job but not found one (hence, she is unemployed) or simply not searched for a job at all (hence, she is out of the labor force). Consequently, all non-employed individuals get the same welfare benefit \( b \) (whatever their productivity and their participation decision). In section IV only,

\[2\text{Our results still hold if } a_1 \to \infty \text{ as long as } \int_{a_0}^{a_1} adF(a) \text{ is finite.}\]
we assume that the government has an additional policy instrument: It can control the workers’ bargaining power $\beta$.

2. In a second step, individuals and firms decide to participate in the labor market. The individuals have the binary choice to invest all their leisure time in search for a job or to stay out of the labor force and enjoy utility from leisure, $d > 0$, and the welfare benefit $b$. They do decide to participate in the market if their expected income is above the utility of staying out of the labor force. Firms can decide to open job vacancies. To do so, they first have to invest capital to build up the workstation. This capital investment is assumed irreversible and productivity-specific, i.e. to be able to hire a worker of ability $a$, the firm has to buy $\kappa_a$ units of capital. Conversely, an individual of ability $a$ can only work at a firm that has made the appropriate investment in equipment, i.e. that has invested the exact amount $\kappa_a$ of capital. We assume that

$$\frac{\dot{\kappa}_a}{\kappa_a} \leq \frac{1}{a}$$  \hspace{1cm} (1)

Firms decide to enter into the labor market as long as their expected profit is positive.

3. Next, individuals and firms match on their skill-specific labor markets. To model the search frictions, we use a standard matching function. The number of matches is a function of the number of individuals searching for a job and of firms searching for a worker. Capital investment and labor markets are productivity-specific. In this sense, we assume directed search. The outcome of this matching process determines the unemployment rate in our model. During this matching process, the firm observes the productivity of the worker. Unmatched firms loose their invested capital.

4. The worker and the firm bargain the gross wage that is paid to the worker.

5. Production and transfers occur.

---

3 A firm might consist of a large number of different jobs. The government observes in a firm’s account only the sum of the total investment costs $\kappa_a$ and total profits $a - w$, but cannot disentangle the costs and profits of each job. For this reason, we assume that no tax schemes based on investment costs $\kappa_a$ or profits $a - w_a$ are available to the government.

4 This assumption ensures that in an economy without government intervention, the unemployment rate among the high-skilled individuals is not higher than the unemployment among the low-skilled individuals.

5 While the government does not observe a worker’s productivity, we assume that firms observe it after a match. Hence, if a type-$a$ worker searches on a type-$t$ labor market with $t \neq a$ and finds a job, the match becomes unproductive and the worker is fired. Since search is costly for workers, a type-$a$ worker has no incentive to search on another labor market.
II.1 The matching process, participation decisions and employment

The matching function $H(U_a, V_a)$ in labor market $a$ gives the number (density) $H_a$ of employed individuals of type $a$ as a function of the number $U_a$ of workers searching for a job, and the number $V_a$ of job vacancies. This matching function is assumed to represent heterogeneities and frictions that we do not model explicitly. It is usually assumed that the matching function $H(.,.,.)$ is increasing in both its arguments, concave and homogeneous of degree 1. Empirical studies have found that a Cobb-Douglas approximation of the matching function fits the data well (Petrongolo and Pissarides, 2001). We therefore assume that the matching function is given by

$$H(U_a, V_a) = A(U_a)^\gamma (V_a)^{1-\gamma}$$

with $\gamma \in (0, 1)$ (2)

where $A > 0$ is a scale parameter of the matching function. If individuals of type $a$ search for a job, then $U_a = f(a)$, otherwise $U_a = 0$. The probability for a firm to hire a worker $H_a/V_a$ is a decreasing function of the number of vacancies, while the probability $L_a = H_a/U_a$ for a searching worker to find a job increases in the number of vacancies.

A firm has to invest $\kappa_a$ units of capital to open a type-$a$ vacancy. When the firm finds a suitable worker for this vacancy, this match produces $a$ units of goods. Note that the investment takes place before the matching to the worker. Since there are matching frictions on the labor market, some firms do not find a worker. In that case, the loss of the firm is equal to the investment cost $\kappa_a$. If the firm finds a worker for its vacancy, then they have to bargain on the gross wage $w_a$ and the firm’s profit writes $a - w_a - \kappa_a$. Since the probability that a firm finds a worker of type $a$ is equal to $H_a/V_a$, the expected profit from posting a vacancy can be written as $(H_a/V_a)(a - w_a) - \kappa_a$. Firms enter the market as long as this expected profit is positive. Hence, it is nil at equilibrium, and therefore

$$\frac{H_a}{V_a} = \kappa_a$$

(3)

Given (2), one has $H_a/V_a = A \cdot (V_a/U_a)^{-\gamma}$. The free-entry condition (3) therefore determines the ratio $V_a/U_a$, and thereby the probability $L_a = H_a/U_a = A \cdot (V_a/U_a)^{1-\gamma}$ for a type-$a$ searching worker to find a job, which is given by

$$L_a = A^{\frac{1}{\gamma}} \left( \frac{a - w_a}{\kappa_a} \right)^{\frac{1-\gamma}{\gamma}}$$

(4)

If the wage $w_a$ decreases, then the firm’s surplus $a - w_a$ increases relative to the vacancy cost $\kappa_a$, and hence, firms create more vacancies and the probability of finding a job increases for type-$a$ searching workers. An additional firm that enters the market increases employment and therefore gross output. But it also increases the resources spent for
capital investments. The impact of an additional vacancy on net output (net of investment costs) is then ambiguous and depends on the number of vacant jobs that are already on the labor market. If the wage is sufficiently low, the firm has incentives to enter the market, even though this might not be optimal from a social point of view, because too many resources might then be spent for capital investments. The gross output generated by workers of type \( a \) is equal to \( aH_a = aL_a f(a) \). Let

\[
Y_a = aL_a - \frac{V_a \kappa_a}{f(a)}
\]

Output net of investment costs on the type-\( a \) labor market can then be written as \( aH_a - V_a \kappa_a = Y_a f(a) \). Multiplying Equation (3) by \( V_a \), one gets

\[
Y_a(w_a) \equiv w_aL_a = A^\frac{1}{\gamma} \left( \frac{a - w_a}{\kappa_a} \right)^{\frac{1-\gamma}{\gamma}} w_a
\] (5)

Because of free entry, firms’ expected profits equal 0, so their total surplus \((a - w_a) L_a f(a)\) equals total investment costs \(V_a \kappa_a\). As a consequence, net output consists only in the total gross wages of workers, \( w_aL_a f(a) \). Net output is an inverse-U shaped function of gross wage.

A type-\( a \) participating worker finds a job with probability \( L_a \). In this case, she gets the wage \( w_a \) and has to pay income taxes \( T(w_a) \). If she doesn’t find a job, her income consists of the welfare benefit \( b \). The (ex-ante) expected income of a searching individual equals \( L_a [w_a - T(w_a)] + (1 - L_a) b \). If the individual decides to stay out of the labor force, then she gets the welfare benefit \( b \) and enjoys her leisure time which gives her utility \( d \). Hence, the individual participates in the labor market as long as \( L_a [w_a - T(w_a)] + (1 - L_a) b \geq b + d \). By defining worker’s ex-post surplus as \( x_a \equiv w_a - T(w_a) - b \) and worker’s expected surplus from participation as \( \Sigma_a \equiv x_a L_a \), one gets

\[
\Sigma_a = A^\frac{1}{\gamma} \left( \frac{a - w_a}{\kappa_a} \right)^{\frac{1-\gamma}{\gamma}} (w_a - T(w_a) - b)
\] (6)

Then, the participation constraint for type-\( a \) workers simplifies to

\[
\Sigma_a \geq d
\] (7)

II.2 The wage bargain

At this stage of the game, the entry costs are sunk. If there is an agreement between the firm and the worker on the wage \( w \), the output is produced and the firm pays the worker the negotiated wage. In the absence of an agreement, nothing is produced, and the worker only gets the welfare benefit \( b \). The ex-post surplus of the worker is therefore equal to \( x = w - T(w) - b \), whereas the surplus of the firm equals \( a - w \). As it is standard in the
literature, we assume that the wage negotiation amounts to maximize the Nash Product defined by
\[
\mathcal{N}(w, x, a) \equiv (a - w)^{1 - \beta} x
\] (8)
where \(\beta \in (0, \gamma]\) denotes the worker’s relative bargaining power. In this paper, we are only interested by the case where workers’ bargaining power \(\beta\) is lower than the elasticity of the matching function \(\gamma\), that is
\[
\beta \leq \gamma
\] (9)

Economic theory does not say anything about the relation between the workers’ bargaining power \(\beta\) and the elasticity of the matching function, \(\gamma\). Hence, there is no reason why the often-assumed Hosios condition (according to which \(\beta = \gamma\)) should hold. One therefore has to rely on empirical estimates to get a realistic ranking between \(\beta\) and \(\gamma\). In sum, those empirical studies\(^6\) suggest a value of the elasticity of the matching function \(\gamma\) between 0.5 and 0.7 together with a value for the workers’ bargaining power \(\beta\) around or below 0.4. Hence, our assumption that \(\beta < \gamma\) seems the most realistic one.

As we will show later, the minimum wage appears at the optimum as soon as \(\beta < \gamma\). If the wage that maximizes the Nash product \(\mathcal{N}(w, x, a)\), lies below the minimum wage \(\underline{w}\), then the minimum wage must be paid to the worker. This leads to
\[
\underline{w}_a \equiv \arg \max_{w > \underline{w}} \mathcal{N}(w, x, a) = \max \left[ \frac{w \beta [1 - T'(w_a)] a + (1 - \beta) (T(w_a) + b)}{1 - \beta \cdot T'(w_a)} \right]
\] (10)
The second equality holds only for values of \(w_a\) where the function \(w \mapsto T(.)\) admits a derivative. To simplify notations in what follows, we define the \textit{maximized} Nash product \(\mathcal{N}_a\) as
\[
\mathcal{N}_a = (a - w_a)^{1 - \beta} (w_a - T(w_a) - b)
\] (11)
so \(\mathcal{N}_a = \mathcal{N}(w_a, x_a, a)\). Given this definition, we can rewrite the worker’s expected surplus from participation as
\[
\Sigma_a = A^{\frac{1}{\gamma}} (\kappa_a)^{\frac{\gamma - 1}{\gamma}} \mathcal{N}_a \ (a - w_a)^{1 - \gamma} - \frac{1 - \beta}{\gamma}
\] (12)

\section{II.3 Incentive constraints}

In this section, we characterize the set of allocations that can be decentralized with the tax function \(T(.)\), the welfare benefit \(b\) and the minimum wage \(\underline{w}\). Since the government

---

\(^6\)The comprehensive survey of Petrongolo and Pissarides (2001) suggests a value of \(\gamma\) between 0.5 and 0.7. While 0.5 is a typical value in calibration exercises (See e.g. Calvani and Zylberberg 2004), Shimer (2005) takes \(\gamma = 0.72\). Estimating the bargaining power requires a structural econometrics of the full model. Flinn (2006) finds a value of \(\beta\) around 0.4 on US data. His estimate of \(\gamma\) is around 0.2 which is far below estimates in the literature. Flinn does not use data on vacancies which may explain this discrepancy. Fève and Langot (1996) find \(\gamma = 0.43\) and \(\beta = 0.36\) on French data. Cahuc et alii (2006) find a value of \(\beta\) between 0 and 0.2 for unskilled workers and between 0.2 and 0.4 for high skilled workers in France. Yashiv (2003) finds estimates of \(\beta\) between 0.1 and 0.3 on Israeli data. These two last studies do not estimate the matching function.
does not observe the productivity level \( a \) of employed individuals, the tax can only be conditioned on the wage level \( w \). Therefore, the government faces an adverse selection problem and can only implement allocations that lead agents to reveal their ability. By choosing the wage \( w_a \), agents implicitly send a “message” about their productivity. The particularity of this problem in our context is that the message is jointly determined by the worker and the firm of a match. However, since the wage maximizes the Nash product \( \mathcal{N}(w, x, a) \), we treat this problem as if a single agent chooses the wage that maximizes \( \mathcal{N}(w, x, a) \). Therefore we can apply standard techniques (see e.g. Salanié 2002) and express incentive constraints in terms of Nash products. Using the taxation principle, it is equivalent to design a tax function \( T(w) \) or to let the firm-worker pair choose among the menu of proposed bundles \((w_a, x_a)\). To be optimal, the allocations must induce the individual matches to truthfully reveal their type, which is the case for a firm-worker pair of type \( a \) if and only if

\[
\text{for all } a' \neq a \quad \mathcal{N}(w_a, x_a, a) \geq \mathcal{N}(w_{a'}, x_{a'}, a) \tag{13}
\]

In other words, a worker-firm pair of type \( a \) prefers the wage \( w_a \) designed for it (which induces the workers’ ex-post surplus to be \( x_a = w_a - T(w_a) - b \)), instead of wage \( w_{a'} \) designed for any other type \( a' \) (which induces the workers’ ex-post surplus to be equal to \( x_{a'} = w_{a'} - T(w_{a'}) - b \)). Since at a constant value for the Nash product, we have that the marginal rate of substitution between the wage \( w \) and the worker’s ex-post surplus \( x \)

\[
\frac{\partial x}{\partial w} \bigg|_{\mathcal{N}(\ldots, a)} = \frac{1 - \beta}{\beta} \frac{x}{a - w} \tag{14}
\]

is decreasing in \( a \), the single-crossing property is fulfilled. This allows a full description of incentive compatible allocations.

**Lemma 1** An allocation is incentive compatible if and only if there exists a type \( a_d \) such that all types with \( a \geq a_d \), and only them, do participate and

\[
\text{either } \Sigma_{a_0} \geq d \text{ and } a_d = a_0 \quad \text{or} \quad a_d > a_0 \text{ and } \Sigma_{a_d} \geq d \tag{15}
\]

for any \( a \in [a_d, a_1] \)

\[
N_a = N_{a_d} \cdot \exp \left[ \frac{1 - \beta}{\beta} \int_{a_d}^{a} \frac{dt}{t - w_a} \right] \tag{16}
\]

\[
a \mapsto w_a \text{ is a non-decreasing function over } [a_d, a_1] \tag{17}
\]

Because the single-crossing property is fulfilled, the incentive constraints (13) can be reduced to a first-order incentive constraint (16) and a second-order (or monotonicity)

\footnote{For further discussions, see HLPV. In particular, since the government observes only wages, and since we rule out side-payments or tax evasion, the firm and the worker of a match cannot send separate messages.}
constraint (17). Note that according to this formulation, $N_a$ has to be a continuous function of $a$, although $w_a$ has not to be.

It will turn convenient to rewrite (16) in terms of expected surplus $\Sigma_a$. Using Equation (12) at $a_d$ and at $a$ to express $N_a$ and $N_{a_d}$ as a function of $\Sigma_a$ and $\Sigma_{a_d}$, the first-order incentive constraint is equivalent to

$$\Sigma_a = \Sigma_{a_d} \cdot \left( \frac{a - w_a}{a_d - w_{a_d}} \right)^{\frac{1-\gamma}{\gamma}} \cdot \left( \frac{\kappa_a}{\kappa_{a_d}} \right)^{\frac{\beta}{\gamma}} \cdot \exp \left[ \frac{1 - \beta}{\beta} \int_{a_d}^{a} \frac{dt}{t - w_t} \right]$$  \hfill(18)

Given that the Nash Product is continuous in $a$ while the wage $w_a$ may have points of discontinuity, the expected surplus $\Sigma_a$ can also be discontinuous. In particular, the first-order incentive constraint alone does not imply that $\Sigma_a$ is increasing in $a$. This further implies that the participation pattern might not be monotonic (that is, there is one threshold type such that all types with higher ability participate while those with lower ability do not participate). This is unlike a standard adverse selection model, where the first-order incentive constraint leads to a monotonic participation pattern. The difference between our model and the standard one stems from the fact that the participation and incentive constraint depend on different (though related) variables: Participation depends on the expected surplus $\Sigma_a$ while the incentive constraints depend on the Nash Products $N_a$. However, we obtain a monotonic participation pattern as soon as we take account of the second-order incentive constraint (17). In the standard adverse selection model, this monotonicity constraint is necessary to ensure that any allocation satisfying the first-order incentive constraint verifies all the incentive constraints (13). In our model, the monotonicity constraint plays thus an additional role for the treatment of participation constraints (7). They depend on expected surplus $\Sigma_a$, which weight wages $w_a$ and ex-post surplus $x_a$ differently than Nash Products $N_a$. Appendix A shows that

**Lemma 2** If (16) holds and $a_s < a_d$, then either $\Sigma_{a_s} < \Sigma_{a_d}$ or $a \mapsto w_a$ has to decrease somewhere between $a_s$ and $a_d$.

Therefore, imposing (17) in addition to (16) ensures that only the participation constraint at the lowest participating skill level can be relevant.\footnote{From this point of view, we are thus back at a standard adverse selection model. However, to obtain this participation monotonicity, we used the second-order incentive constraint. Hence, we can for the further resolution of our problem not rely on a traditional first-order approach.} Equation (15) distinguishes two cases. Either participation constraints (7) are never binding, in which case one must have $\Sigma_{a_0} \geq d$. Or, there exists a skill level $a_d$, such that all individuals that are more (less) productive than this threshold choose to participate (not to participate) to the
labor market. The participation constraint then only binds at \( a_d \) and the threshold verifies \( \Sigma_{a_d} = d \).\(^{10}\)

Equation (10) explains how to decentralize the incentive compatible allocations. A minimum wage *per se* does not change the set of incentive-compatible allocations, for minimum wage policy is based on the same information (the level of gross wage) as taxation. However, as we will show, there are cases where the second-order incentive constraint is binding at the bottom of the wage distribution. This bunching can be decentralized with the tax function. However, it seems more realistic to us that the government uses a binding minimum wage to have a bunching of skills at the bottom of the wage distribution, i.e. it sets a minimum wage to avoid that worker-firm matches choose a wage below some threshold wage. Clearly, the government could also simply put a very high tax for all wages below the threshold wage, and it would achieve the same result. But the most natural way to achieve it, seems still the minimum wage. In particular, a high tax at any wage that is below the threshold wage can be interpreted as a fine for the violation of the minimum wage law (see also Cahuc and Laroque 2007 for a discussion). We thus henceforth consider that when (17) is binding at the bottom of the skill distribution, a minimum wage is implemented to achieve this bunching.

II.4 The government’s objective and budget constraint

As in HLPV, we assume that the government cares only about the distribution of expected utilities, i.e. \( \Sigma_a + b \) for the participating types and \( b + d \) for the non participating ones. We assume the following objective to the government:

\[
\Omega = F(a_d) \Phi(b + d) + \int_{a_d}^{a_1} \Phi[L_a(w_a - T(w_a)) + (1 - L_a)b] f(a) \, da
\]

where \( \Phi(.) \) is a twice continuously differentiable, increasing and concave function. This formulation implies that the government compensates individuals for their innate productivity \( a \), but not for their labor market status.\(^{11}\) It admits as a limiting case the *maximin* criterion. Using the definition of the worker’s expected surplus (6), one can

\(^{10}\)Since \( \Sigma_a \) is not continuous, it may be the case that

\[
\lim_{a \to a_d, a < a_d} \Sigma_a \leq d \leq \lim_{a \to a_d, a > a_d} \Sigma_a
\]

\(^{11}\)Adding this latter motive leads to an objective function of the form

\[
F(a_d) \Phi(b + d) + \int_{a_d}^{a_1} [L_a \Phi(w_a - T(w_a)) + (1 - L_a) \Phi(b)] f(a) \, da
\]

This alternative formulation does not change the mechanisms that lead to our main results, but makes the model less tractable.
rewrite this objective as
\[
\Omega = F(a_d) \Phi(b + d) + \int_{a_d}^{a_1} \Phi(b + \Sigma_a) f(a) \, da
\] (19)

The budget constraint can be written as
\[
\int_{a_d}^{a_1} T(w_a) L_a f(a) \, da = \left[ F(a_d) + \int_{a_d}^{a_1} (1 - L_a) f(a) \, da \right] b + E
\]
where \( E \geq 0 \) denotes exogenous public expenditures. Since firms' profits net of vacancy costs are nil and only labor incomes are taxed, the government's budget constraint can be replaced by the resource constraint
\[
\int_{a_d}^{a_1} Y_a(w_a) f(a) \, da = \int_{a_d}^{a_1} \Sigma_a f(a) \, da + E + b
\] (20)
The left-hand side of (20) denotes total production net of vacancy costs and the right-hand side the distribution of these available resources to all individuals.

III The optimal policy for a given bargaining power

In this section, we characterize the optimal allocation when the bargaining power is exogenously set a lower level than the one prescribed by the Hosios condition. The government maximizes its objective (19), subject to the budget constraint (20), the first-order incentive constraint (18), the second-order incentive constraint (17) and the participation constraint (7). Following Lemma 1, the government’s problem is reduced to choosing a threshold value \( a_d \) of productivity and an initial value \( \Sigma_{a_d} \) subject to (15) and a non-decreasing function \( a \mapsto w_a \) defined over \([a_d,a_1]\). We henceforth call this strategy the second-order approach since the second-order incentive constraint (17) is included in the government’s program.

The usual way to solve such an adverse selection model with a continuum of types is to use optimal control techniques. In our case, this would imply that one takes the Nash Product \( N_a \) as state variable and the wage \( w_a \) as control variable. However, this strategy does not work in our case. In particular, the participation constraint for the lowest participating type \( a_d \) writes
\[
N_{a,d} = d A^{-\gamma} \left( \kappa_{a_d} \right)^{\frac{1-\gamma}{\gamma}} (a_d - w_{a_d})^{\frac{1-\gamma}{\gamma}} - \frac{1-\gamma}{\gamma}
\] (21)

Hence, the initial value of the state variable \( N_{a,d} \) depends on the initial value \( w_{a,d} \) of the control. To our knowledge, optimal control (see e.g. the textbook by Kamien and Schwartz

\[12\text{Given these choices, Equation (18) gives } \Sigma_a. \text{ The budget constraint (20) yields the value of the welfare benefit } b. \text{ Finally we get the workers' ex-post surplus } x_a = w_a - T(w_a) - b, \text{ thereby the level of tax from (6).} \]
1991) cannot handle problems with such a specific initial condition. Therefore, rather than using optimal control, we derive the necessary conditions by considering the consequences of infinitesimal variations in all variables.\textsuperscript{13} Note also that we cannot use a traditional first-order approach\textsuperscript{14}: As explained in Section III.3, a monotonic participation pattern only emerges when the second-order incentive constraint is integrated. We thus need to impose at least partially the second-order constraint to obtain a monotonic participation pattern. Hence, to show that the optimal solution involves a constant wage at the bottom of the skill distribution, we follow what we call a “local”\textsuperscript{15} first-order approach: We optimize wages on the interval \((a_d, a_1]\) (instead of \([a_d, a_1]\)), while imposing that the monotonicity constraint holds on this interval. We then look at the wage of the least skilled participating type \(a_d\) (without imposing the monotonicity constraint for this type). If his wage \(w_{a_d}\) is below or equal to the limiting wage\textsuperscript{16} \(w_{a_d}^+\), then the problem is solved and there is no violation of the second-order incentive constraint and thus no bunching at the bottom of the skill distribution. We however show that this is not the case: One necessarily obtains \(w_{a_d} > w_{a_d}^+\) and hence bunching at the bottom of the wage distribution (which, as discussed before, can be interpreted as a binding minimum wage). Note that the violation of the second-order incentive constraint is necessarily only local, simply because we imposed this constraint to hold everywhere else. This however does not mean that bunching at the optimum is only a local phenomenon.

For the remainder of this section, we first need to describe the equity-efficiency tradeoff for wages above the lowest one. We thus compute the shadow cost of public funds (III.1), and then the optimality condition for the wage in the absence of bunching (III.2). We then derive the optimal initial value of expected surplus \(\Sigma_{a_d}\) (III.3). The results in these three subsections are valid under both the local first-order and second-order approaches. We then come to the main Proposition of this paper (III.4), that is the optimality condition for the minimum wage level. We do this first by showing that the local first-order approach leads to a solution that violates the monotonicity constraint (17). Then, we show that starting from a situation with no bunching at the bottom, implementing a binding minimum wage \(\underline{w}\) increases strictly the government’s objective. We finally turn to the condition with respect to the participation threshold (III.5). The mathematical derivations of the optimality conditions are in Appendix B.

\textsuperscript{13}It can however easily be shown that most of the first-order conditions (all except the condition on the minimum wage) would be the same when using optimal control techniques (see also HLPV, 2006).

\textsuperscript{14}i.e. solve the problem without imposing the second-order incentive constraint and then show that this monotonicity constraint is violated at the bottom of the distribution.

\textsuperscript{15}Usually, the first-order approach relaxes the monotonicity constraint everywhere. We relax it only locally.

\textsuperscript{16}For any function \(y_a\) of \(a\) where \(y = w, x, N, \Sigma\), we denote \(y_{a_d}^+\) the limiting value \(\lim_{a = a_d, a > a_d} y_a\).
III.1 The shadow cost of public funds

Consider a rise in the level of public expenditures $E$, holding the threshold $a_d$ and the expected surplus $\Sigma_a d$ for this threshold type and the wage distribution constant. Then, neither $Y_a(w_a)$ nor $\Sigma_a$ is affected by this change. Therefore, the only change in Equation (20) is that $b$ decreases one-to-one when $E$ rises. The social utility of non-participating individuals decreases by $\Phi'(b + d)$ while the social expected utility of participating individuals of type $t$ decreases by $\Phi'(b + \Sigma_t)$. Hence, the shadow cost of public funds $\lambda$ is given by

$$\lambda = \Phi'(b + d) F(a_d) + \int_{a_d}^{a_1} \Phi'(\Sigma_t + b) f(t) dt$$  \hspace{1cm} (22)

III.2 Optimal wages $w_a$ in the absence of bunching

We now describe the optimality condition for the wage $w_a$ at a point of the skill distribution where there is no bunching (and at a point where $w_a$ is continuous in $a$). We consider a marginal translation $\Delta w$ of the function $a \mapsto w_a$ on an infinitesimal interval $[a, a + \delta a]$. Appendix B.1 shows that the optimality condition can be written as

$$0 = \lambda \frac{\partial Y_a(w_a)}{\partial w_a} f(a) - \frac{1 - \beta}{\beta (a - w_a)^2} Z_a + \left(\frac{1 - \beta}{\beta} - \frac{1 - \gamma}{\gamma}\right) \frac{1}{a - w_a} [\Phi'(\Sigma_a + b) - \lambda] \Sigma_a f(a)$$  \hspace{1cm} (23)

where

$$Z_a = \int_a^{a_1} (\lambda - \Phi'(\Sigma_t + b)) \Sigma_t f(t) dt$$  \hspace{1cm} (24)

To see the intuition behind this optimality condition, consider the optimization problem for agents of type $a$. For given levels of $a_d$ and of $\Sigma_a$, the first-order incentive constraint (16) implies that the Nash product for type $a$ is predetermined and not affected by the change in the wage $w_a$.

The first term in equation (23) stands for the efficiency part of the trade-off. An increase in the wage rate $w_a$ implies that less vacancies are created, which has two consequences. First, it decreases employment and therefore gross output. But it also decreases the resources used for investments in capital to build workstations. The effect on
output net of investment costs equals $\partial Y_a/\partial w_a$ and is therefore ambiguous (see equation (5)). Multiplying this by the shadow cost $\lambda$ of public funds and the density $f(a)$ of type-$a$ workers gives the efficiency term in (23).

The second term in equation (23) represents the impact on informational rents of a higher gross wage for type-$a$ workers. When firm-worker matches endowed with productivity $a$ have a higher gross wages (while keeping the Nash product $N_a$ fixed), more productive firm-worker matches find it more attractive to mimic them. In other words, a type-$t$ above $a$ worker-firm pair finds it profitable (in terms of Nash Product 8) to choose the wage $w_a$ designed for type-$a$ jobs instead of the wage $w_t$ designed for them.

![Figure 1: The informational rent effect](image)

Figure 1 illustrates this point. It displays the iso-Nash curves for types $a$ and $t$ with $t > a$. Because productivity is higher, the elasticity of the firm’s surplus with respect to the gross wage is smaller for worker-firm pairs of type $t$ than for pairs of type $a$. Consequently, the iso-Nash curves corresponding to type $a$ are steeper than iso-Nash curves corresponding to type $t$ (see Equation 14). For a given level of Nash product accruing to type $a$, the government can propose different bundles, for instance a low wage combined with a low ex-post surplus $(w_a, x_a)$ or a higher wage combined with a higher ex-post surplus $(w'_a, x'_a)$. If the government proposes $(w_a, x_a)$, it has to give a Nash product at least equal to $N_t$ to prevent type-$t$ worker-firm pairs to mimic type-$a$ pairs. If the government however proposes the allocation $(w'_a, x'_a)$, then, it has to give a Nash product $N'_t$, which is strictly higher than $N_t$. Hence, the higher the wage $w_a$, the higher the Nash product for types $t$ above $a$, which means that a higher wage increases informational rents. In the “informational rent” term of equation (23), the term in front of $Z_a$ measures the rate at which the growth rate of the worker’s maximized Nash product has to increase to prevent more productive matches from mimicking the type $a$ match.

From the first-order incentive constraint (16), the Nash product designed for all types
t above $a + \delta a$ has to increase by the same proportion to prevent mimicking. Since wages designed for all these types above $a + \delta a$ are not changed, we get that the expected surplus for these types increases in the same proportion, so $\Delta \Sigma_t / \Sigma_t = \Delta N_t / N_t = \Delta N_{a+\delta a} / N_{a+\delta a}$. The increase in the expected surplus obtained by type-$t$ worker-firm pairs increases the social welfare by $\Phi' (\Sigma_t)$, but implies a budgetary cost equal to $\lambda$. Integrating these two terms over all types $t$ above $a + \delta a$ gives the shadow cost $Z_a$ of a relative increase of the type-$a$ Nash product (see equation (24)).

Taking the limit when $\delta a$ tends to 0 leads to the informational rent effect in (23). We get the following Lemma, which is proved in Appendix B.2:

**Lemma 3** The shadow cost $Z_a$ of a relative increase in the Nash product is positive for all $a < a_1$

In other words, the government desires to avoid informational rents.

A third effect appears in the present model due to the fact that workers’ bargaining power $\beta$ is below the elasticity of the matching function, $\gamma$. In our model, the expected surplus $\Sigma_a$ that the government focuses on does not coincide with the Nash product $N_a$ that the firm and the worker maximize when they negotiate the wage. For a given maximized Nash product $N_a$ (that is predetermined by the incentive constraints), a change in the wage $w_a$ has also an impact on the expected surplus $\Sigma_a$ as described in equation (12).

The intuition is depicted in Figure 2. Both Nash product $N_a$ and workers’ expected surplus $\Sigma_a$ are increasing functions of workers’ ex-post surplus $x_a$ and of firms’ surplus $a - w_a$ (thereby a decreasing function of the gross wage $w_a$). However, if the Hosios condition $\beta = \gamma$ is not fulfilled, they put different relative weights on these two components. Hence, the marginal rate of substitution between gross wages $w_a$ and workers’ ex-post surplus $x_a$ that keep workers’ expected surplus $\Sigma_a$ unchanged differ from the one that keep Nash product $N_a$ unchanged. Since $\beta < \gamma$, the Nash product criterion puts a higher relative weight on gross wages $w$. Hence, increasing the wage $w_a$ for a given Nash product $N_a$ increases workers’ expected surplus $\Sigma_a$, while increasing the wage $w_a$ for a given workers’ expected surplus $\Sigma_a$ decreases the Nash product $N_a$. This implies that the iso-Nash curves are steeper than the corresponding iso-expected-surplus curves. When the employment level decreases by an increase in gross wages from $w_1$ to $w_2$, the net income has to rise from $x_1$ to $x_2$ to give the same Nash product as before to this firm-worker match. In terms of expected surplus, this increase in net income more than compensates for the employment loss due to the relatively small increase in gross wages. The expected surplus therefore increases.

---

19 Formally, $-Z_a$ is the welfare effect of a relative unit change in the Nash product $N_a$, keeping the function $t \mapsto w_t$ unchanged for $t \in [a, a_1]$. 

16
The increase in $\Sigma_a$ is valued at the marginal social utility of type $a$, namely $\Phi'_a$. Since the government thus gives more resources to agents of type $a$, less resources can be affected to redistribution toward other individuals. This decrease in budgetary funds is valued at the marginal cost of public funds $\lambda$. An increase in the wage rate is therefore desirable if type $a$ is a low-productivity type that gets a lower expected surplus than the high-productivity type because of the first-order incentive constraint. Giving resources to this low-productivity type is socially valuable because it increases equity. The contrary is true if type $a$ is a high-productivity type from whom the government wants to take resources in order to redistribute.

The sum of these three terms gives (23).

### III.3 The optimal initial value of expected surplus $\Sigma_{a,d}$

According to Lemma 1, the participation constraint (7) may never bind, in which case one has $a_{d} = a_{0}$ and $\Sigma_{a,d} > d$. This can however never be optimal, because a decline in the initial value of expected surplus $\Sigma_{a,d}$ spills over the whole distribution. Hence, the informational rents of all types decrease. Switching to an allocation with $\Sigma_{a,d} = d$, while $a_{d}$ and the wages $w_{a}$ are kept unchanged thus improves the government's objective. The following Lemma is summarizes this statement. It is formally proved in Appendix B.3

**Lemma 4** The participation constraint binds, so $\Sigma_{a,d} = d$.

### III.4 The optimal minimum wage $w$

In this section, we consider the optimal wage level at the bottom of the skill distribution. We first show intuitively that without imposing the monotonicity constraint at the bottom of the wage distribution, the monotonicity constraint is violated. More precisely, we
impose the wage \( w_a \) to be nondecreasing in \( a \) on the interval \((a_d, a_1]\). We then discuss the mechanisms that determine the wage \( w_{a_d} \) for the lowest participating type and show that this wage \( w_{a_d} \) must be set above \( w_{a_d}^* \), which clearly indicates a violation of the monotonicity constraint. Therefore, this constraint must be binding at the second-best optimum under the second-order approach (that is, imposing the monotonicity constraint for all types). By construction, the violation we show is very local and only concerns type \( a_d \). This does however not imply that the bunching is limited to type \( a_d \). To show this, we consider in a second step the effect of a rise of the minimum wage \( w \). Starting from a situation with no minimum wage, such a rise is welfare improving. Therefore, any allocation without a minimum wage is dominated by an allocation with bunching of wages.

We thus start by discussing the optimal wage \( w_{a_d} \) for the lowest participating type when the monotonicity constraint is not imposed to hold for this marginal type for which the participation constraint is binding. Hence, we search the optimal wage \( w_{a_d} \) for given (optimized) nondecreasing wages on the interval \((a_d, a_1]\). According to Lemma 4, the participation constraint (7) binds and the expected surplus \( \Sigma_{a_d} \) of the marginal type \( a_d \) is equal to his outside option \( d \). Hence, if the wage \( w_{a_d} \) rises (thereby decreasing the employment probability), then the ex-post surplus \( x_{a_d} \) has to increase to keep the expected surplus at \( d \). This change will however have an impact on the Nash Product (see(21)): Since for a low bargaining power \( \beta < \gamma \), the Nash product \( N(\ldots, \ldots) \) puts a higher relative weight on the gross wage \( w \) than the expected surplus \( \Sigma_{a_d} \), an increase in the wage \( w_{a_d} \) for a given expected surplus leads to a decrease in the Nash Product \( N_{a_d} \) (see also Figure 2). However, the first-order incentive constraint (16) indicates that this decrease in the Nash Product \( N_{a_d} \) also decreases Nash Products for all types \( a > a_d \). This in turn leads to a reduction in informational rents for all types \( a > a_d \) (see also (18)). Such a reduction is however beneficial, according to Lemma 3. Hence, the government has an incentive to set this wage \( w_{a_d} \) at a high level to decrease informational rents. Note that this mechanism only holds for types for which the expected surplus is given by the participation constraint. However, according to Lemma 1 the participation constraint is binding only for type \( a_d \). Hence, the government has no incentive to set the wage \( w_{a_d} \) as high as the wage \( w_{a_d}^* \) (the limiting wage \( w_{a_d}^* \)) and one clearly gets a discontinuity in the wage around \( a_d \). Thus, the monotonicity constraint is violated. Therefore we must follow a second-order approach. We now impose the monotonicity constraint to hold for all types, including \( a_d \), and consider the optimal condition for the lowest wage \( w \).

If there is bunching at the bottom of the wage distribution, then a positive measure of workers are paid at the same wage \( w \) that we interpret as a minimum wage. Let \( a_m \) denote the highest skill level that is paid at the minimum wage (see Figure 3). We now consider the welfare effect of an increase in the minimum wage \( w \), that is of an increase of
wages over \([a_d, a_m]\) up to \(w_{a_m}\). In this policy change, we keep \(a_d\) and \(\Sigma_{a_d}\) unchanged. The tax function \(T(\cdot)\) is adjusted such that the participation decisions remain unchanged. We show in Appendix B.4 that

\[
\frac{\partial \Omega}{\partial w} = \lambda \int_{a_d}^{a_m} \frac{\partial Y_a(w)}{\partial w_a} f(a) da + \int_{a_d}^{a_m} (\Phi'(\Sigma_a + b) - \lambda) \Sigma_a \frac{1 - \gamma}{\gamma} \left( \frac{1}{a_d - w} - \frac{1}{a - w} \right) f(a) da \\
+ \left( \frac{1 - \beta}{\beta} \frac{1}{a_m - w} - \frac{1 - \gamma}{\gamma} \frac{1}{a_d - w} \right) Z_{a_m}
\]

An increase in the minimum wage has three effects. The first term captures the efficiency effect. In accordance with common wisdom, a higher minimum wage decreases employment (and hence increases unemployment). Even though this leads to a lower gross output, the higher wage also implies that less resources are spent by firms investing in new workstations. Therefore, the effect on net output \(\partial Y_a/\partial w_a(w)\) is ambiguous.

The second term captures the effect on the expected surplus of the workers of types \(a\) in \([a_d, a_m]\) whose wages are given by the minimum wage. The effect of a marginal increase \(\Delta w\) of the minimum wage on the expected surplus of these workers is twofold. First, the rise in the minimum wage reduces the labor demand. Equation (4) implies that a rise \(\Delta w\) of the minimum wage decreases the probability for a type-\(a\) searching worker to find a job by

\[
\frac{\Delta L_a}{L_a} = \frac{1 - \gamma}{\gamma} \frac{\Delta w}{a - w}
\]

Second, the government has to increase \(w - T(w)\) to keep type-\(a_d\) workers in the labor
force. Hence, one has

\[
\frac{\Delta (w - T(w) - b)}{w - T(w) - b} = -\frac{\Delta L_{ad}}{L_{ad}} = \frac{1 - \gamma}{\gamma} \frac{\Delta w}{a_d - w}
\]

to keep type \(a_d\) workers indifferent between participating or not. The net effect on workers’ expected surplus \(\Sigma_a\) is therefore

\[
\frac{\Delta \Sigma_a}{\Sigma_a} = \frac{\Delta L_a}{L_a} + \frac{\Delta (w - T(w) - b)}{w - T(w) - b} = \frac{1 - \gamma}{\gamma} \left( \frac{\Delta w}{a_d - w} - \frac{\Delta w}{a - w} \right)
\]

At a given wage level, the probability of finding a job is relatively less sensitive to wage for workers of a higher productivity. Hence, for \(a > a_d\), the net effect of a rise in minimum wage on workers’ expected surplus is positive. The induced increase in \(\Sigma_a\) has a direct consequence on the government’s objective that is valued at a rate \(\Phi'(\Sigma_a)\), and a negative effect on public funds that is valued at rate \(\lambda\). Summing this effect for all types between \(a_d\) and \(a_m\) gives the second term in the right-hand side of (25).

Finally, a rise in the minimum wage changes the level of the Nash product for the workers of the limiting type \(a_m\). Firms’ ex-post surplus \(a_m - w_{am}\) decreases by \(-\Delta w\) while workers’ ex-post surplus increases by \(\Delta (w - T(w) - b)\), so the net effect equals to

\[
\frac{\Delta N_{am}}{N_{am}} = \frac{1 - \beta \Delta (a_m - w_{am})}{\beta a_m - w_{am}} + \frac{\Delta (w - T(w) - b)}{w - T(w) - b}
\]

\[
= \left( \frac{1 - \gamma}{\gamma} \frac{1}{a_d - w} - \frac{1 - \beta}{\beta} \frac{1}{a_m - w} \right) \Delta w
\]  

(26)

However, this relative change in the Nash product of these workers spills over the whole distribution of productivity to prevent the mimicking off worker-firm pairs of all type \(t\) above \(a_m\) (see (16)). Since \(Z_a\) is the shadow cost of a relative increase in the Nash product, the last term of equation (25) represents the effect of the minimum wage on the individuals whose wage is not constrained by the minimum wage. We can now prove that

**Proposition 1** If \(\beta < \gamma\), a binding minimum wage is optimal.

**Proof.** Assume by contradiction that bunching is not optimal. For this, we consider a function \(a \mapsto w_a\) that is increasing in the neighborhood of \(a_d\). Hence, one has \(a_m = a_d\), so Equation (25) simplifies to

\[
\frac{\partial \Omega}{\partial w} = \left( \frac{1 - \beta}{\beta} - \frac{1 - \gamma}{\gamma} \right) \frac{Z_{ad}}{a_d - w_{ad}}
\]

Since \(\beta < \gamma\), and \(Z_{ad} > 0\) from Lemma 3, any allocation where \(a \mapsto w_a\) is locally increasing around \(a_d\) is dominated by an allocation with a binding minimum wage. ■
Technically speaking, the new feature of our model is that while the worker and the firm maximize the Nash product, the participation constraint depends on another variable, the expected surplus. In the traditional adverse selection model in contract theory, the variables concerning the agent’s maximization problem and the participation constraint are the same. By the participation constraint, the maximized utility of the lowest participating type has to equal the (exogenous) outside option of the individual. Therefore, the principal cannot affect the maximized utility of the lowest participating type. From this utility of the lowest participating type, the incentive constraints then determine the evolution of the maximized utility for all other types, and hence the informational rent given to these types. In our model, things are different. Even though the expected surplus of the lowest participating type must equal the outside option (i.e. the utility of leisure in our case), the government can decrease the level of the maximized Nash product by imposing a very high wage on the lowest participating individual (see equation (21)). This implies through the incentive constraints that the Nash products of all types decrease, and by equation (12), expected surplus of all individuals above the lowest type decreases (See the effect of $w_{a_d}$ in Equation (18)). Hence, this allows the government to decrease informational rents of the agents and use these resources for redistributive purposes, which in turn increases social welfare. However, choosing a very high wage for the lowest participating individual inevitably violates the second-order incentive constraint. As a consequence, there is bunching at the bottom of the wage distribution in the second-order approach, and since this constraint is on the wage, we can interpret this bunching as a minimum wage.

III.5 The optimal participation $a_d$

For completeness, we finally consider a marginal change in the threshold $a_d$, keeping the function $a \mapsto w_a$ unchanged. The optimal participation decision sets the optimal value of the threshold type $a_d$ and writes (see Appendix B.5)

$$
\lambda (Y_{a_d} (w_{a_d}) - d) f (a_d) = \frac{1 - \gamma}{\gamma} \left( \frac{1}{a_d - \bar{w}} - \frac{k_{a_d}}{K_{a_d}} \right) Z_{a_d}
$$

(27)

The left-hand side of (27) corresponds to the efficiency part of the trade-off. When $a_d$ decreases, participation increases by an amount that is proportional to the density $f (a_d)$. The participation of these workers increases total net output by $Y_{a_d}$ but requires that they receive an expected surplus at least equal to $d$ to participate. This net budgetary gain is socially valued at the shadow cost of public funds, $\lambda$. The right-hand side of (27) is the equity part of the trade-off. When $a_d$ decreases, worker-firm pairs with productivity $a_d$ have the possibility to mimic the additional participants. To avoid this mimicking, the government has to give an additional informational rent to type-$a_d$ matches. The term
in front of $Z_{ad}$ is equal to the relative increase in the Nash product that should be given to type-$a_d$ matches to prevent them from mimicking the new entrants. The equity cost multiplies this increase by the shadow cost $Z_{ad}$ of a relative increase in the Nash product $N_{ad}$.

**IV Varying bargaining power**

The previous section considered the optimal policy if the bargaining power is exogenous. One might however argue that the government has some influence on the bargaining power, especially in the long run. There might be different ways how the government can affect the bargaining power. It can change the bargaining procedures by law, or a law might change the way how unions are financed and regulated, etc. Cahuc and Zylberberg (2004) mention as examples the abolition of closed shop by the Thatcher government in the UK and the 1991 employment contract acts in New Zealand. Moreover, the work by Abowd et alii (2005) suggest that workers’ bargaining power is higher in France than in the US, which is consistent with the presumption that institutions influence workers’ bargaining power. Therefore, one might argue that bargaining power can be influenced by some deep institutional reforms in the labor market in a long-run perspective. While the exact modelling of these institutional reforms is clearly beyond the scope of the present paper, this section simply assumes that the government can change directly the workers’ bargaining power $\beta$.

The envelope theorem implies that $\frac{d\Omega}{d\beta} = \frac{\partial \Omega}{\partial \beta}$. Analyzing this last expression, we can derive the following proposition, which is proved in Appendix C:

**Proposition 2** Increasing workers’ bargaining power is optimal as long as worker’s bargaining power $\beta$ is not higher than the elasticity of the matching function $\gamma$.  

To understand this result, one might identify the distortions that are present in the (second-best) optimum. In the absence of redistribution and under the Hosios condition (i.e. $\beta = \gamma$), negotiated wages maximize net output. Since the government wants to redistribute from high- to low-income individuals, it wants to install a high marginal tax rate. According to equation (10), this distorts the wage downwards. However, the bargaining power also distorts the wage levels. A higher bargaining power increases the wage, again according to equation (10). Therefore, a rise in the bargaining power induces a distortion on the wage that (partly) offsets the one induced by the redistributive taxation. The equity-efficiency trade-off becomes less severe the higher the worker’s bargaining power $\beta$. An increase in $\beta$ is thus always desirable, at least up to the point where $\gamma = \beta$.

We are not able to find analytical results for the case where $\beta > \gamma$. In fact, Appendix C shows that increasing the workers’s bargaining power is welfare-improving as long as
$\Sigma_a$ is increasing in $a$, that is, as long as the desired redistribution goes from high- to low-productive workers. Lemma 2 only proves that this is the case if $\beta \leq \gamma$. If $\beta > \gamma$, then the worker’s expected surplus at the optimal solution might not be monotonically increasing in $a$ anymore.

To understand why we need workers’ expected surplus $\Sigma_a$ to be increasing in types to obtain our result that a rise in workers’ bargaining power is welfare-improving, we have to analyze whether a rise in the bargaining power $\beta$ relaxes or strengthens the relevant incentive constraints. When workers’ expected utility $\Sigma_a$ increases with their productivity $a$, the relevant incentive constraint is that a type-$a$ match does not want to mimic slightly less productive matches. In other words, when they negotiate the wage, the relevant incentive constraint induces the worker and the firm of a type-$a$ job to choose the wage $w_a$ designed for them, and not the wage $w_{a-da}$ designed for slightly less productive jobs of type $a-da$. Obviously, the higher the worker’s bargaining power $\beta$, the harder it becomes for the firm to obtain $w_{a-da}$ as the bargaining outcome instead of $w_a$. Therefore, a rise in the bargaining power $\beta$ relaxes the incentive constraints that prevent worker-firm pairs from mimicking less productive worker-firm pairs, which explains why in such a context, the government can achieve a better outcome.

V Conclusion

We have given a sufficient condition for the minimum wage to be a part of the optimal redistributive policy: If the bargaining power is lower than the elasticity of the matching function, then the introduction of a binding minimum wage is welfare-improving. However, if the government can also control the workers’ bargaining power, it should increase it, at least up to a point where our argument in favor of minimum wage does no longer apply. Hence, our argument in favor of the introduction of a binding minimum wage only holds if the government cannot control the bargaining power, and if this parameter is relatively low. In other words, the minimum wage is an imperfect substitute for a rise in workers’ bargaining power. Hence, the government should prefer the latter, in combination with redistributive taxation.

Whether and how the government can affect the bargaining power is still an open question. It would also be interesting to determine the optimal level of the bargaining power when income taxation is simultaneously optimized. We have shown that as long as the workers’ expected surplus remains increasing in types, increasing the bargaining power is welfare-improving because it relaxes the relevant incentive constraint. Hence, since the workers’ expected surplus is increasing whenever the bargaining power is not higher than the elasticity of the matching function, this suggest that the optimal bargaining power in our redistributive context is higher than the one prescribed by Hosios (1990) in a pure
efficiency context. We leave the further characterization of the optimal bargaining power for future research.

References


A Proof of Lemmas 1 and 2

A.1 Necessary conditions

Let $T(.)$ be a function of wages, $b$ a welfare benefit and $w$ a minimum wage level. These policies induced an allocation defined by the wage bargaining solution (10). This subsection gives some properties that the induced allocation has to verify.

Let $I_a$ be a function defined for $a \in [a_d, a_1]$ by

$$I_a \equiv \sup_{w \in [w, a-a]} (a - w) \cdot (w - T(w) - b)^{\frac{\beta}{\sigma}}$$

From (11), one has

$$N_a = (I_a)^{\frac{1-\beta}{\sigma}}$$

(28) Then $w_a$ is a maximizer of $(a - w) \cdot (w - T(w) - b)^{\frac{\beta}{\sigma}}$ and $x_a = w_a - T(w_a) - b$. It is worth noting that $w_a$ and $x_a$ are defined even if workers of type $a$ choose not to participate. Since for all $a$ the bundle $(w_a, x_a)$ maximizes $N(w, w - T(w) - b, a)$, the allocation $a \mapsto (w_a, x_a)$ has to verify the set of incentive constraints (13).

Without any mathematical restrictions on the tax function, for any $w$, the function $a \mapsto (a - w)(w - T(w) - b)^{\frac{\beta}{\sigma}}$ is linear in $a$. Therefore, the function $a \mapsto I_a$ is the convex envelope of linear and increasing functions of $a$. This ensures that function $a \mapsto I_a$ is convex and increasing in $a$. Therefore, $a \mapsto I_a$ is continuous on $[a_d, a_1]$ and admits a derivative everywhere except on a countable set (we use henceforth the acronym $a.e.$ for “almost everywhere”). Therefore $a \mapsto N_a$ is continuous on $[a_d, a_1]$. By the envelope theorem, we get that whenever $a \mapsto I_a$ admits a derivative $\dot{I}_a$, this derivative verifies

$$\dot{I}_a \equiv (x_a)^{\frac{\beta}{1-\sigma}} \quad \iff \quad \dot{I}_a \equiv \frac{1}{I_a} \frac{1}{a - w_a}$$

(29)

Since $a \mapsto I_a$ is convex, one has that $a \mapsto x_a$ is non-decreasing. To show that $a \mapsto w_a$ is non-decreasing too, we assume by contradiction it is not. Let then assume there exists $a' > a$ such that $w_{a'} < w_a$. Recall that according to the definition of the Nash product in (8), $N(w, x, a)$ is decreasing in the wage $w$ and increasing in workers’ ex-post surplus $x$. Hence we get $N(w_a, x_a, a) < N(w_{a'}, x_{a'}, a)$. Since $a' > a$, we have that $x_{a'} \geq x_a$, which induces in turn $N(w_a, x_a, a) < N(w_{a'}, x_{a'}, a)$. Therefore $w_a$ does not maximizes $N(w, w - T(w) - b, a)$. Hence $a \mapsto w_a$ is nondecreasing.
since $I_a$ is a continuous function of $a$, the integration of (29) between $a_d$ and $a$ gives 
$log I_a = log I_{a_d} + \exp \left[ \int_{a_d}^{a} \frac{dt}{t - w_a} \right]$. Together with (28), this last equality gives (16). Using (12) on $a$ and on $a_d$ leads to (18).

We now prove that $\Sigma_a$ is increasing in $a$. Consider two skill levels $a' > a$. Then, from (8) and (13), one gets

$$
\log x_{a'} - \log x_a \geq \frac{1 - \beta}{\beta} \left[ \log (a' - w_a) - \log (a' - w_{a'}) \right]
$$

From (6), one has

$$
\log \Sigma_{a'} - \log \Sigma_a = \frac{1 - \gamma}{\gamma} \left[ \log \left( \frac{a' - w_{a'}}{\kappa_{a'}} \right) - \log \left( \frac{a - w_a}{\kappa_a} \right) \right] + \log x_{a'} - \log x_a
$$

So,

$$
\log \Sigma_{a'} - \log \Sigma_a \geq \frac{1 - \gamma}{\gamma} \left[ \log \left( \frac{a' - w_{a'}}{\kappa_{a'}} \right) - \log \left( \frac{a - w_a}{\kappa_a} \right) \right] + \frac{1 - \beta}{\beta} \left[ \log (a' - w_a) - \log (a' - w_{a'}) \right]
$$

which gives

$$
\log \Sigma_{a'} - \log \Sigma_a \geq \frac{1 - \gamma}{\gamma} \left[ \log \left( \frac{a' - w_a}{\kappa_{a'}} \right) - \log \left( \frac{a - w_a}{\kappa_a} \right) \right] + \left( \frac{1 - \beta}{\beta} - \frac{1 - \gamma}{\gamma} \right) \left[ \log (a' - w_a) - \log (a' - w_{a'}) \right]
$$

and finally

$$
\log \Sigma_{a'} - \log \Sigma_a \geq \frac{1 - \gamma}{\gamma} \int_a^{a'} \left( \frac{1}{t - w_a} - \frac{\kappa_t}{\kappa_a} \right) dt + \left( \frac{1 - \beta}{\beta} - \frac{1 - \gamma}{\gamma} \right) \left[ \log (a' - w_a) - \log (a' - w_{a'}) \right]
$$

Since $t - w_a < t$, Equation (1) ensures that the first term in the right-hand side of the last expression is positive. Since $a' > a$, one has $a' - w_a \geq a' - w_{a'}$ from (17). Hence, whenever $\beta \leq \gamma$, the second term in the last expression is the product of two non-negative terms. So $\log \Sigma_{a'} - \log \Sigma_a$ is larger than the sum of a positive term and a non-negative term, which ends the proof that $\Sigma_a$ is increasing in $a$.

As a consequence, either $\Sigma_{a_0} \geq d$ and the participation constraint (7) holds for all $a$ in $[a_0, a_1]$, or $\Sigma_{a_0} < d$. Then, there exists a unique 20 productivity level $a_d \in [a_0, a_1]$ such that for any $a \geq a_d$ (resp. $<$,) one has $\Sigma_a \geq d$ (resp $<$), so these workers of type $a$ choose to search (not to search) for a job.

20If $\Sigma_{a_1} > d$ no one participate. We rule out this trivial allocation.
A.2 Sufficiency condition and Lemma 2

In this subsection, we consider two real numbers \( a_d \) and \( \Sigma_{a_d} \) that satisfy (15) and a function \( a \mapsto w_a \) defined over \([a_d, a_1]\). We then define \( a \mapsto \Sigma_a \) through (18) and \( a \mapsto N_a \) through (12). Therefore, one has (16).

We first prove Lemma 4. Let \( a < a' \). We get from (18)

\[
\log \Sigma_{a'} - \log \Sigma_a = \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \left[ \log (a' - w_a') - \log (a - w_a) \right] + \frac{1 - \beta}{\beta} \int_a^{a'} \frac{dt}{t - w_t} \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \log \left( \frac{a' - w_a'}{a' - w_a} \right) + \frac{1 - \beta}{\beta} \int_a^{a'} \frac{dt}{t - w_t} \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right)
\]

However,

\[
\log \left( \frac{a' - w_a}{a - w_a} \right) = \int_a^{a'} \frac{dt}{t - w_a}
\]

Now, assume that \( t \mapsto w_t \) is nondecreasing over \([a, a']\). Therefore \( t > a > w_a \) implies \( 1/(t - w_a) \leq 1/(t - w_t) \), so

\[
\log \left( \frac{a' - w_a}{a - w_a} \right) \leq \int_a^{a'} \frac{dt}{t - w_t}
\]

For \( \beta \leq \gamma \), one has \( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \leq 0 \), thereby:

\[
\left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \left[ \log (a' - w_a') + \log \left( \frac{a' - w_a}{a - w_a} \right) \right] \\
\geq \left( \frac{1 - \beta}{\beta} - \frac{1 - \gamma}{\gamma} \right) \log \left( \frac{a' - w_a}{a' - w_a'} \right) + \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \int_a^{a'} \frac{dt}{t - w_t}
\]

Since, \( \log \kappa_{a'} - \log \kappa_a = \int_a^{a'} (\kappa_t/\kappa_a) dt \), we get

\[
\log \Sigma_{a'} - \log \Sigma_a \geq \left( \frac{1 - \beta}{\beta} - \frac{1 - \gamma}{\gamma} \right) \log \left( \frac{a' - w_a}{a' - w_a'} \right) + \frac{1 - \beta}{\beta} \int_a^{a'} \left( \frac{1}{t - w_t} - \frac{\kappa_t}{\kappa_a} \right) dt
\]

Since \( a \mapsto w_a \) is nondecreasing (thereby \( a' - w_a \geq a' - w_a' \)), the first term in the right-hand side of the inequality is non-negative. From assumption (1), the second term is positive. Hence \( \Sigma_{a'} \geq \Sigma_a \), which ends the proof of Lemma 2.

We now impose (17) and verifies that the allocation defined is incentive compatible. First, from the proof of Lemma 2 above, \( \Sigma_a \) is increasing the participation constraint (7) holds for all \( a \geq a_d \) given (15). We now verify (13). Take \( a' \neq a \). If \( w_{a'} \geq a \), then \( \mathcal{N}(w_{a'}, x_{a'}, a) \leq 0 < N_a \). If conversely \( w_{a'} < a \), it is equivalent to prove (13), or to prove that \( \log N_a - \log \mathcal{N}(w_{a'}, x_{a'}, a) \) is non-negative. Given (8), this last expression equals

\[
\log N_a - \log \mathcal{N}(w_{a'}, x_{a'}, a) = \log N_a - \log N_{a'} + \frac{1 - \beta}{\beta} \left[ \log (a' - w_{a'}) - \log (a - w_{a'}) \right]
\]

28
Given (16), we get
\[
\log N_a - \log N (w_{a'}, x_{a'}, a) = \frac{1 - \beta}{\beta} \left[ \log (a' - w_{a'}) - \log (a - w_{a'}) - \int_a^{a'} \frac{dt}{t - w_t} \right]
\]
\[
= \frac{1 - \beta}{\beta} \left[ \int_a^{a'} \left( \frac{1}{t - w_{a'}} - \frac{1}{t - w_t} \right) dt \right]
\]

So,

- if $a' < a$, then for all $t \in [a', a]$, one has $0 < w_{a'} \leq w_t < t$, so the integrand is non-positive. Since $a' < a$, the integral is therefore non-negative which ensures that (13) holds.

- if $a' > a$ and $w_{a'} < a$, then for all $t \in [a, a']$, one has $0 < w_t \leq w_{a'} < t$, and the integrand is non-negative. Since $a < a'$, the integral is therefore non-negative, too.

**B The government’s problem**

It is convenient to rewrite (18) as

\[
\log \Sigma_a = \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \left[ \log (a - w_a) - \log (a_d - w_{a_d}) \right] + \frac{1 - \beta}{\beta} \int_a^{a_d} \frac{dt}{t - w_t} - \frac{1 - \gamma}{\gamma} \log \left( \frac{\kappa_a}{\kappa_a} \right) + \log \Sigma_{a_d}
\]

(30)

**B.1 Optimality conditions with respect to negotiated wages $w_a$ for all $a \in (a_m, a_1]$**

We consider the effect of a variation $\Delta w$ in the wage $w_a$ for the agents of type $[a, a + \delta a]$ with $\Delta w$ and $\delta a$ being infinitesimally small and $a > a_m$. From Equation (30), the implied variation $\Delta \Sigma_t$ of workers’ expected surplus equals

\[
\Delta \Sigma_t = \frac{1 - \beta}{\beta (a - w_a)^2} \Sigma_t \Delta w \delta a
\]

for $a + \delta a < t$,

\[
\Delta \Sigma_d = - \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \frac{1}{a - w_a} \Sigma_t \Delta w + \frac{1 - \beta}{\beta (a - w_a)^2} \Sigma_t \Delta w \delta a
\]

for $a \leq t < a + \delta a$ and $\Delta \Sigma_t = 0$, for $t < a$.

From the budget constraint (20), the variation of the welfare benefit is

\[
\Delta b = \int_{a_d}^{a_1} \left( \Delta Y_t (w_t) - \Delta \Sigma_t \right) f (t) dt
\]

\[21\) The difference between $(t - a)$ and $\delta a$ tends to 0 as $\delta a$ tends to 0, so it can be neglected.
Hence, for any \( a \in (a_m, a_1] \)
\[
\Delta b = \frac{\partial Y_a}{\partial w_a} (w_a) \Delta w \delta a + \left(1 - \frac{1 - \beta}{\gamma} - \frac{1 - \beta}{\beta} \right) \frac{1}{a - w_a} \sum_a f (a) \Delta w \delta a
\]
\[
- \frac{1 - \beta}{\beta (a - w_a)^2} \left( \int_a^{a_1} \Sigma_t f (t) \, dt \right) \Delta w \delta a
\]
From the government’s objective (19), one has
\[
\Delta \Omega = \lambda \Delta b + \int_a^{a_1} \Phi' (\Sigma_t + b) \Delta \Sigma_t f (t) \, dt
\]
\[
= \lambda \frac{\partial Y_a}{\partial w_a} (w_a) f (a) \Delta w \delta a + \int_a^{a_1} (\Phi' (\Sigma_t) - \lambda) \Delta \Sigma_t f (t) \, dt
\]
where we define the shadow cost of public funds \( \lambda \) through Equation (22). Hence
\[
\frac{\Delta \Omega}{\Delta w \delta a} = \lambda \frac{\partial Y_a}{\partial w_a} (w_a) f (a) + (\Phi' (\Sigma_a + b) - \lambda) \left( \frac{1 - \beta}{\beta} - \frac{1 - \gamma}{\gamma} \right) \frac{1}{a - w_a} \sum_a f (a)
\]
\[
+ \frac{1 - \beta}{\beta (a - w_a)^2} \int_a^{a_1} (\Phi' (\Sigma_t + b) - \lambda) \Sigma_t f (t) \, dt
\]
which gives (23), together with (24) when \( \Delta w \) and \( \delta a \) tends to 0.

### B.2 Proof of Lemma 3

Since \( t \mapsto \Sigma_t \) admits everywhere a left and right limit, so does the function \( t \mapsto (\lambda - \Phi' (\Sigma_t + b)) \Sigma_t f (t) \). The integral in (24) is therefore well defined, is a continuous function of \( t \), and admits everywhere a left and a right derivative.

From Lemma 2, we know that \( \Sigma_a \) is increasing in \( a \). Following Lemma 2, Under the local first-order approach, \( \Sigma_a \) is only increasing within \( (a_d, a_1] \), while under the second-order approach, \( \Sigma_a \) is increasing over \( [a_d, a_1] \). Hence, the marginal social welfare \( \Phi' (b + \Sigma_a) \) is decreasing in \( a \) over \( (a_d, a_1] \). The shadow cost of public funds \( \lambda \) equals the average of all marginal social welfare (see equation 22). Hence there exists a unique \( a_c \) for which \( \Phi' (b + \Sigma_{a_c}) = \lambda \). For all \( t < a_c \), we get \( \Phi' (b + \Sigma_t) > \lambda \) and \( \Sigma_t < \Sigma_{a_c} \), while for \( t > a_c \), we get \( \Phi' (b + \Sigma_t) < \lambda \) and \( \Sigma_t > \Sigma_{a_c} \). Therefore, for any \( t \not= a_c \)
\[
\Phi' (b + \Sigma_t) - \lambda \Sigma_t < [\Phi' (b + \Sigma_t) - \lambda] \Sigma_{a_c} \]
Hence we get from (24):
\[
Z_a = \int_a^{a_1} [\lambda - \Phi' (b + \Sigma_t)] \Sigma_t f (t) \, dt < \int_a^{a_1} [\lambda - \Phi' (b + \Sigma_t)] \Sigma_{a_c} \cdot f (a) \, da
\]
However, given equation (22)
\[
\int_a^{a_1} [\lambda - \Phi' (b + \Sigma_t)] \Sigma_{a_c} \cdot f (t) \, dt = (1 - F (a)) \Sigma_{a_c} \{ \mathbb{E}_f [\Phi'_t] - \mathbb{E}_f [\Phi'_t | t \geq a] \}
\]
where \( \mathbb{E}_f \) is the expectation operator under distribution \( f \) for \( t, \Phi'_t = \Phi' (b + \Sigma_t) \) for \( t \geq a_d \) and \( \Phi'_t = \Phi' (b + d) \) otherwise. Hence \( \Phi'_t \) is decreasing in \( t \) over \( (a_d, a_1] \). From Lemma 1 this implies that \( \mathbb{E}_f [\Phi'_t | t \geq a] \) decreases in \( a \) over \( [a_d, a_1] \), so \( \mathbb{E}_f [\Phi'_t | t \geq a] < \mathbb{E}_f [\Phi'_t] \). Hence \( Z_a > 0 \).
B.3 Proof of Lemma 4

Fix $a_d$ and a nondecreasing function $a \mapsto w_a$. Assume by contradiction that $\Sigma_{a_d} > d$ and consider a marginal change $\Delta \Sigma_{a_d}$. Then, according to (30), one has for all $a \geq a_d$,

$$\Delta \Sigma_a = \Sigma_a \frac{\Delta \Sigma_{a_d}}{\Sigma_{a_d}}$$

Hence, from (20), one has

$$\Delta b = -\frac{\Delta \Sigma_{a_d}}{\Sigma_{a_d}} \int_{a_d}^{a_1} \Sigma_t f(t) \, dt$$

so, from the government’s objective:

$$\Delta \Omega = \lambda \Delta b + \int_{a_d}^{a_1} \Phi' (\Sigma_t) \Delta \Sigma_t f(a) \, da = \frac{\Delta \Sigma_{a_d}}{\Sigma_{a_d}} \int_{a_d}^{a_1} (\Phi' (\Sigma_t) - \lambda) \Sigma_t f(t) \, dt$$

Given (24), one finally gets:

$$\Delta \Omega = \frac{\Delta \Sigma_{a_d}}{\Sigma_{a_d}} Z_{a_d}$$

From Lemma 3, $Z_{a_d} > 0$, so decreasing $\Sigma_{a_d}$ from a value above $d$ is welfare-improving.

B.4 Optimality condition with respect to the minimum wage $w$

Consider a variation of the minimum wage of $dw$. This implies an increase in the wage, but also an increase in the amount of types for whom the minimum wage is relevant, as illustrated in Figure 4.

![Figure 4: Covariations of $a_m$ and of minimum wage $w$.](image)

Hence, we calculate the direct effects of a variation in $a_m$ at a given minimum wage $w$, and then the direct effects of a variation in the minimum wage $w$ for a given $a_m$. 31
Effect of $a_m$ at given $w$  According to equation (30), $\partial \Sigma_t / \partial a_m = 0$ for all $a \in [a_d, a_1]$. From (20)

$$\frac{\partial b}{\partial a_m} = (Y_m - \Sigma_m) f(a_m) - (Y_m - \Sigma_m) f(a_m) - \int_{a_d}^{a_m} \frac{\partial \Sigma_a}{\partial a_m} f(a) \, da - \int_{a_m}^{a_1} \frac{\partial \Sigma_a}{\partial a_m} f(a) \, da = 0$$

Finally, from (19) $\partial \Omega / \partial a_m = 0$. Hence we can concentrate on the direct effect of $w$ for a given $a_m$ for all variables of interest $\Sigma_t$, $b$, $\Omega$.

Effect of $w$ for a given $a$  From (30) and $w \equiv w_{a_d}$, we get for all $a \in [a_d, a_m]$:

$$\Delta \Sigma_a = \Sigma_a \left\{ \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \left( \frac{1}{a_d - w} - \frac{1}{a - w} \right) + \frac{1 - \beta}{\beta} \int_{a_d}^{a} \frac{dt}{(t - w)^2} \right\} \Delta w$$

As $\Delta w$ tends to 0, one gets

$$\frac{\partial \Sigma_a}{\partial w} = \frac{1 - \gamma}{\gamma} \Sigma_a \left( \frac{1}{a_d - w} - \frac{1}{a - w} \right) > 0$$

From (30), we have for all $a \in [a_m, a_1]$

$$\Delta \Sigma_a = \Sigma_a \left\{ \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \left( \frac{1}{a_d - w} - \frac{1}{a - w} \right) + \frac{1 - \beta}{\beta} \int_{a_d}^{a_m} \frac{dt}{(t - w)^2} \right\} \Delta w$$

$$\Delta \Sigma_a = \Sigma_a \left\{ \left( \frac{1 - \gamma}{\gamma} - \frac{1 - \beta}{\beta} \right) \left( \frac{1}{a_d - w} - \frac{1}{a - w} \right) + \frac{1 - \beta}{\beta} \left( \frac{1}{a_d - w} - \frac{1}{a_m - w} \right) \right\} \Delta w$$

$$\Rightarrow \frac{\partial \Sigma_a}{\partial w} = \Sigma_a \left\{ \frac{1 - \gamma}{\gamma} \frac{1}{a_d - w} - \frac{1 - \beta}{\beta} \frac{1}{a_m - w} \right\}$$

From (20) we find

$$\frac{\partial b}{\partial w} = \int_{a_d}^{a_m} \left( \frac{\partial Y_t}{\partial w_t}(w) - \frac{\partial \Sigma_t}{\partial w} \right) f(t) \, dt - \int_{a_m}^{a_1} \frac{\partial \Sigma_a}{\partial w} f(t) \, dt$$

Hence

$$\frac{\partial b}{\partial w} = \int_{a_d}^{a_m} \left\{ \frac{\partial Y_a}{\partial w_a}(w) - \frac{1 - \gamma}{\gamma} \left( \frac{1}{a_d - w} - \frac{1}{a - w} \right) \Sigma_a \right\} f(a) \, da$$

$$+ \left( \frac{1 - \beta}{\beta} \frac{1}{a_m - w} - \frac{1 - \gamma}{\gamma} \frac{1}{a_d - w} \right) \int_{a_m}^{a_1} \Sigma_a \cdot f(a) \cdot da$$

Finally, from (19) and (22):

$$\frac{\partial \Omega}{\partial w} = \lambda \frac{\partial b}{\partial w} + \int_{a_d}^{a_m} \Phi'(\Sigma_a + b) \frac{\partial \Sigma_a}{\partial w} f(a) \, da + \int_{a_m}^{a_1} \Phi'(\Sigma_a + b) \frac{\partial \Sigma_a}{\partial w} f(a) \, da$$
After some manipulations, one then gets

\[
\frac{\partial \Omega}{\partial w} = \int_{a_d}^{a_m} \left\{ \lambda \frac{\partial Y_a}{\partial w}(w) + (\Psi'(\Sigma_a + b) - \lambda) \Sigma_a \frac{1 - \gamma}{\gamma} \left( \frac{1}{a_d - w} - \frac{1}{a - w} \right) \right\} f(a) \, da + \\
+ \left( \frac{1 - \gamma}{\gamma} \frac{1}{a_d - w} - \frac{1 - \beta}{\beta} \frac{1}{a_m - w} \right) \left( \int_{a_d}^{a_1} (\Psi'(\Sigma_a + b) - \lambda) \Sigma_a f(a) \, da \right)
\]

Together with (24), we obtain (25).

### B.5 Optimality condition with respect to the threshold \( a_d \)

We consider now a variation \( \Delta a_d \) for the threshold \( a_d \), keeping the function \( a \mapsto w_a \) unchanged. Using Proposition 1, we know that there is bunching of wages at the bottom of the skill distribution. Therefore, a marginal change of the threshold \( a_d \) keeps unchanged the level of the lowest wage \( w_{a_d} \). Hence, from (30), we get for \( a \in [a_d, a_1] \)

\[
\frac{\Delta \Sigma_a}{\Delta a_d} = -\frac{1 - \gamma}{\gamma} \left( \frac{1}{a_d - w_{a_d}} - \frac{\dot{k}_a}{\kappa_a} \right) \Sigma_a
\]

One then gets from the budget constraint (20)

\[
\frac{\Delta b}{\Delta a_d} = -(Y_{a_d} - \Sigma_{a_d}) - \int_{a_d}^{a_1} \frac{\Delta \Sigma_a}{\Delta a_d} f(a) \, da
\]

Since \( \Sigma_{a_d} = d \), one gets

\[
\frac{\Delta b}{\Delta a_d} = -(Y_{a_d} - d) + \frac{1 - \gamma}{\gamma} \left( \frac{1}{a_d - w_{a_d}} - \frac{\dot{k}_a}{\kappa_a} \right) \int_{a_d}^{a_1} \Sigma_a f(a) \, da
\]

Finally, from the government’s objective one gets \( \Delta \Omega = \lambda \Delta b + \int_{a_d}^{a_1} \Delta \Sigma_a f(a) \, da \). Taking equation (24) into account, we then obtain (27).

### C Proof of Proposition 2

To prove Proposition 2, we apply the envelope theorem and prove that the partial derivative of \( \Omega \) with respect to \( (1 - \beta) / \beta \) for a given function \( a \mapsto w_a \) is negative.

From equation (30), one gets for all \( a \in [a_d, a_m] \)

\[
\frac{\partial \Sigma_a}{\partial \frac{1 - \beta}{\beta}} = 0
\]

From equation (30), one has, for a given function \( a \mapsto w_a \) that for all \( a \in [a_m, a_1] \)

\[
\frac{\partial \Sigma_a}{\partial \frac{1 - \beta}{\beta}} = \left\{ - \log (a - w_a) + \log (a_d - w_{a_d}) + \int_{a_d}^{a} \left( \frac{1}{x - w_x} - \frac{\dot{k}_x}{\kappa_x} \right) dx \right\} \Sigma_a
\]
So:

\[
\frac{\partial \Sigma_a}{\partial \frac{1-\beta}{\beta}} = \left\{ \log \left( \frac{a - w_{ad}}{a - w_a} \right) \cdot \frac{a_d - w_{ad}}{a - w_{ad}} + \int_{a_d}^{a} \frac{dx}{x - w_x} \right\} \Sigma_a
\]

\[
= \left\{ \log \left( \frac{a - w_{ad}}{a - w_a} \right) + \int_{a_d}^{a} \left\{ \frac{1}{x - w_x} - \frac{1}{x - w_{ad}} \right\} dx \right\} \Sigma_a
\]

Since there is no more bunching in the right-neighborhood of \(a_m\), one has together with (17) that \(w > w_{am} = w_{ad}\). This induces that \(a - w_a < a - w_{ad}\), so the first term in the bracket is positive. Furthermore, for all \(x \in (a_m, a]\), one has \(w_{ad} = w_{am} < w_x < x\, so\, 1/ (x - w_x) > (1/ (x - w_{ad})) > 0\). Therefore the second term in the bracket is positive too and

\[
\frac{\partial \Sigma_a}{\partial \frac{1-\beta}{\beta}} > 0
\]

We now prove that the function \(a \mapsto \partial \Sigma_a/\partial ((1 - \beta)/\beta)\) is increasing. This function is the product of two positive functions, and we already know from Lemma 1 that \(a \mapsto \Sigma_a\) is increasing. We now prove that \(a \mapsto \log \left( \frac{a - w_{ad}}{a - w_a} \right) + \int_{a_d}^{a} \left\{ \frac{1}{x - w_x} - \frac{1}{x - w_{ad}} \right\} dx\) is increasing too. Its partial derivative with respect to \(w_a\) holding \(a\) constant is

\[
\frac{\partial}{\partial w_a} \left\{ \log \left( \frac{a - w_{ad}}{a - w_a} \right) + \int_{a_d}^{a} \left\{ \frac{1}{x - w_x} - \frac{1}{x - w_{ad}} \right\} dx \right\} \bigg|_{w_a} = \frac{1}{a - w_a} > 0
\]

Moreover, its partial derivative with respect to \(a\) holding \(w_a\) constant is

\[
\frac{\partial}{\partial a} \left\{ \log \left( \frac{a - w_{ad}}{a - w_a} \right) + \int_{a_d}^{a} \left\{ \frac{1}{x - w_x} - \frac{1}{x - w_{ad}} \right\} dx \right\} \bigg|_{a = w_a} = \frac{1}{a - w_{ad}} - \frac{1}{a - w_a} + \frac{1}{a - w_a} - \frac{1}{a - w_{ad}} = 0
\]

Since wage \(w_a\) is nondecreasing in \(a\) (See 17), the function \(a \mapsto \partial \Sigma_a/\partial ((1 - \beta)/\beta)\) is the product of two positive function, one being increasing in \(a\), and the other one being nondecreasing. Hence, it is increasing in \(a\).

From the budget constraint (20), one gets

\[
\frac{db}{d\frac{1-\beta}{\beta}} = \frac{\partial b}{\partial \frac{1-\beta}{\beta}} = - \int_{a_m}^{a} \frac{\partial \Sigma_a}{\partial \frac{1-\beta}{\beta}} f (a) da < 0
\]

Finally, from the social objective (19), we find

\[
\frac{d\Omega}{d\frac{1-\beta}{\beta}} = \lambda \frac{db}{\partial \frac{1-\beta}{\beta}} + \int_{a_m}^{a} \Phi' (b + \Sigma_a) \cdot \frac{\partial \Sigma_a}{\partial \frac{1-\beta}{\beta}} \cdot f (a) da
\]

\[
= \int_{a_m}^{a} \Phi' (b + \Sigma_a) - \lambda \cdot \frac{\partial \Sigma_a}{\partial \frac{1-\beta}{\beta}} \cdot f (a) da
\]

Since the function \(a \mapsto \Phi' (\Sigma_a + b)\) is decreasing from the concavity of \(\Phi (.)\) and Lemma 1, two cases are possible:

34
1. $\lambda \geq \Phi'(b + \Sigma_a)$. Then for all $a > a_m$, $\lambda \geq \Phi'(b + \Sigma_a)$ where this inequality is strict for at least some of these types. Hence, one has

$$[\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_a}{\partial \frac{1}{\beta}} \cdot f(a) \leq 0$$

and this inequality holds strictly for at least some of these types. This implies

$$\frac{d\Omega}{d\frac{1}{\beta}} < 0$$

2. Otherwise, $\lambda < \Phi'(b + \Sigma_{a_m})$. From equation (22) and Lemma 1, we know that there exists a $a_c \in (a_m, a_1)$ such that $\lambda = \Phi'(b + \Sigma_{a_c})$. Furthermore, the function $\partial \Sigma_a / \partial (\frac{1}{1 - \beta} / \beta)$ is positive and increasing.

   - For $a \in (a_m, a_c)$, we have $\Phi'(b + \Sigma_a) \geq \lambda$ and thus
     $$[\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_a}{\partial \frac{1}{\beta}} \leq [\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_{a_c}}{\partial \frac{1}{\beta}}$$
   - for $a \in (a_c, a_1)$, we have $\Phi'(b + \Sigma_a) \leq \lambda$ and thus
     $$[\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_a}{\partial \frac{1}{\beta}} \leq [\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_{a_c}}{\partial \frac{1}{\beta}}$$

Hence for all $a$

$$[\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_a}{\partial \frac{1}{\beta}} \leq [\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_{a_c}}{\partial \frac{1}{\beta}}$$

and therefore

$$\frac{d\Omega}{d\frac{1}{\beta}} \leq \int_{a_m}^{a_1} [\Phi'(b + \Sigma_a) - \lambda] \cdot \frac{\partial \Sigma_{a_c}}{\partial \frac{1}{\beta}} \cdot f(a) da$$

$$\frac{d\Omega}{d\frac{1}{\beta}} \leq \frac{\partial \Sigma_{a_c}}{\partial \frac{1}{\beta}} \left\{ \int_{a_m}^{a_1} [\Phi'(b + \Sigma_a) - \lambda] f(a) da \right\}$$

$$\frac{d\Omega}{d\frac{1}{\beta}} \leq \frac{\partial \Sigma_{a_c}}{\partial \frac{1}{\beta}} (1 - F(a_m)) \{ E[\Phi'(b + \Sigma_a) \mid a \geq a_m] - \lambda \}$$

From Lemma 1, we now that function $a \mapsto \Phi'(\Sigma_a + b)$ is decreasing in $a$. Hence, given (22), one has $\lambda > E[\Phi'(b + \Sigma_a) \mid a \geq a_m]$, therefore,

$$\frac{d\Omega}{d\frac{1}{\beta}} < 0$$

Hence, in both cases, one finds

$$\frac{d\Omega}{d\beta} > 0$$

35