Optimal Income Taxation with Unemployment Induced Loss of Human Capital

Louis Perrault*

November 16, 2012

(Job Market Paper)

Abstract

This paper investigates the properties of an optimal non-linear tax system in a two-period private information economy where labour market frictions create unemployment that destroys workers’ human capital. A two-skill type model is considered where wages and employment are endogenous. I find that the optimal tax system distorts the first-period wages of both skill types below their first-best levels which leads to more employment. The standard Mirrlees/Stiglitz no-distortion at the top result no longer holds due to the combination of private information and the destruction of human capital from unemployment. I show this result analytically under the Maximin (Rawlsian) social welfare function and confirm it numerically for a more general finite constant elasticity of substitution social welfare function. In addition, I investigate the use of other policy tools such as a training program or job creation subsidies. The training program can alleviate both the informational problem and the destruction of human capital, and numerical results suggests that the wages of both skill types can be brought closer to their first-best levels. Job creation subsidies are shown numerically to be unnecessary unless the matching frictions and the informational problem are such that the optimal tax policy requires the unskilled worker’s wage to be distorted all the way to zero.

JEL-Classification: H21, H24, J24, J64

Keywords: Non-linear taxation, Redistribution, Human capital, Unemployment, Labour market frictions

*Department of Economics, Dunning Hall, 94 University Avenue, Queens University, Kingston, Canada K7L 3N6. email: louisperrault@gmail.com. I wish to thank my supervisor Robin Boadway for his guidance and support. I am also grateful for the discussions and comments from Étienne Lehmann, Laurence Jacquet, Huw Lloyd-Ellis, Sumon Majumdar, Craig Brett, Afraziab Mirza, and seminar participants of Queen’s University, the 2012 Annual Conference of the Société canadienne de science économique (SCSE), the 11th Journées Louis-André Gérad-Varet. This research has been supported by the Fond Québécois de la Recherche sur la Société et la Culture (FQRSC) Doctoral Fellowship and the John Deutsch Institute. All omissions and errors are my own. The most up-to-date draft of this paper can be found at https://sites.google.com/site/louisperrault/.
1 Introduction

An important issue in the optimal redistributive tax literature has been the role of income taxation and the accumulation of human capital. When considering human capital (Bovenberg & Jacobs, 2001; Jacobs, 2005), the focus has mainly been on income taxation, education policy and the educational choice of the individual. Despite a large body of empirical evidence linking work experience with wage rates (Blundell & MaCurdy, 1999) validating the idea that past labor market participation improves the worker’s productivity, human capital acquisition through learning-by-doing has been ignored in the optimal tax literature until recently (Krause, 2009; Best & Kleven, 2012). Both the education and learning-by-doing models assume that workers fully control the level of either education or work effort that will result in an increase in human capital. However, in the presence of involuntary unemployment, the decision to work or how much to work can be entirely irrelevant in the accumulation of human capital if one is unable to find work. Thus for many workers the actual level of human capital acquired is determined by factors that are out of their control.

To take account of this reality, I propose a model of optimal income taxation with non-Walrasian labor markets in which income taxation affects wages, unemployment and at the same time the future distribution of human capital levels. More specifically, I use a two-period model where frictional unemployment destroys a fraction of a skilled worker’s human capital. The model of the labor market is based on the search and matching literature developed by Diamond (1982), Mortensen (1982), Pissarides (2000), and extends the optimal tax framework with search frictions developed by Hungerbühler et al. (2006, 2010); Lehmann et al. (2011) and Jacquet et al. (2011) to incorporate dynamic considerations. I assume, like Stiglitz (1982), a two-type economy that is composed of unskilled types, i.e. low level of human capital, and skilled types, i.e. a high level of human capital. To emphasize the importance of non-employment being a possibility despite a worker’s best efforts, I assume that individuals always participate (or search in the labour market) and that voluntary unemployment is not an option. Therefore, the loss of human capital is never a conscious choice. To make the intuition as simple as possible, I restrict the loss of human capital to the skilled individuals. Although less general, one could argue that only skilled workers could actually lose human capital and know-how whereas low or unskilled workers have little skill to lose in the first place.

Assuming destruction of human capital while unemployed is related to the concept of learning-by-doing, and refers to the phenomenon called ‘unemployment scarring’. This reflects the long-term consequences of experiencing a period of unemployment. Although there can be many types of unemployment scarring that can potentially affect an individual’s working life, the one connecting unemployment to wage rates can be drastic. Empirical evidence suggest that after experiencing a spell of unemployment, workers, and especially young workers, will face the prospect of lower wages in their next employment opportunity (Arulampalam, 2001; Arulampalam et al., 2001; Gregg & Tominey, 2005). These negative effects on wages can even last decades after the last spell of unemployment. One prominent
explanation for this phenomenon is that unemployment spells preclude the worker from acquiring work experience but also brings about the deterioration of the worker’s general skill level. This ‘unlearning-by-not-doing’ is a result of the labor market environment over which the worker has no control which is the situation at the heart of this model.

As in Mirrlees (1971), I assume a government (or planner) that wishes to redistribute income but is unable to observe the productivities of workers or in our context the productivities of particular jobs. This forces the planner to condition the tax system only on income and employment history, which will create inefficiencies. Income taxation in this model, as in several other models of imperfect labour markets (Sørensen, 1999), affects the economy through what Lehmann et al. (2011) call the ‘wage-cum-labor-demand’ margin. An example of a framework featuring this margin is a matching model where before-tax wages are determined by Nash bargaining.1 A rise in the marginal tax rate reduces the reward to negotiating aggressively on the part of the worker because an increase in before-tax wage results in a lower increase in the after-tax wage. Thus the ‘wage moderation’ effect of the increasing marginal tax rate will lower the negotiated before-tax wage which will increase labour demand (and employment). However, this effect on employment is not free and can come with an efficiency cost, since in a model of search and matching, too much employment can mean that the total output minus the costs needed to create all these employment opportunities can be lower than under a situation with lower levels of employment.

To focus on redistributive taxation, I assume that the equilibrium before-tax wage maximizes a ‘wage-setting objective’ that is increasing in the after-tax wage and decreasing in the before-tax wage, with the added property that the wage-setting mechanism is efficient in the sense that it maximizes total output net of vacancy costs in the Laissez faire economy.

Here, in contrast to Krause (2009), the optimal (i.e. second-best) allocation does not feature the standard result of no-distortion at the top, which in this case would be not to distort the before-tax wage and consumption choice in the skilled labour market.2 In fact, the skilled workers before-tax wage is distorted downward away from the Laissez faire (i.e. first-best) level. The fact that the planner is only able to observe income places limits on his ability to redistribute as he wishes, and it forces the planner to adopt redistributive taxation in the second period as well as the first. This has two effects. The first effect is the incentive effect which can also be broken down in two parts. More redistribution from the skilled to the unskilled workers in the second period makes being a skilled worker in that period less appealing. This has a direct impact on wage formation in the first period, because a skilled worker would be willing to accept a lower pre-tax wage in exchange for a higher probability of retaining his human capital level. By resorting to redistributive taxation in the second

---

1 Other models of labour markets with matching frictions giving rise to such a margin include, for example, the Competitive Search Equilibrium (Moen, 1997) and the monopoly union model of ?.

2 Best & Kleven (2012) side-step the issue of no-distortion at the top by assuming an unbounded distribution of skills. As for the static models of Hungerbühler et al. (2006) and Lehmann et al. (2011), the optimal allocations are characterized by the no-distortion-at-top result.
period, the planner lowers the reward of keeping the high human capital, thus it reduces the skilled worker’s willingness to accept a lower pre-tax wage which has an upward effect on the equilibrium wage. To counteract this increase in the pre-tax wage above the efficient level, the planner must reduce the ability of the firm to transfer utility through the before-tax wage to the worker by increasing the marginal tax rate.\textsuperscript{3} This will result in lowering the equilibrium pre-tax wage and raising the employment level closer to the efficient level. This increase in the employment level is the other part of the incentive effect. The planner will seek to tradeoff the reduction in the value of keeping the human capital against a higher probability of keeping it by lowering the before-tax wage in the first period.

The second effect is the fiscal effect and is a novel feature that is absent from other human capital and income tax results in the literature. In fact, the results found in this paper are more in line with those of Stiglitz (1982) who relies on general equilibrium effects to redistribute. Here the difference stems from the fact that in our model distorting the allocation of the skilled types has fiscal benefits that come from modifying the distribution of types in the second period. Because the planner must redistribute from the skilled to the unskilled in the second period, he has now an incentive to create more employment in the skilled sector. The reason for this is threefold: having more skilled workers raises more revenue, the tax burden can be shared over more workers and allows the planner to make it smaller, and finally it reduces the number of workers the planner must redistribute to.

What is important to take away from these results is that both the incentive and fiscal effects are dependent on the presence of both the informational problem and the human capital destruction problem. Taken individually, each problem will not generate these effects. If there is only the human capital destruction problem, the planner is able to set up is tax system to achieve the first-best wage level. With only the informational problem, second period policies will not affect the wage determination process in the first period and the policies of the first period will not affect the second period. From this the no-distortion at the top result follows.\textsuperscript{4}

I also show that, in a similar fashion to Hungerbühler et al. (2006) and Lehmann et al. (2011), the wage of the unskilled workers will be distorted downwards from the efficient level since it facilitates redistribution. To achieve this, the marginal tax rate faced by the unskilled must be raised higher than the average tax rate (in this case, the ratio between the transfer he receives and his market wage). In addition, redistribution of income occurs through both direct transfers and a higher probability of employment of the unskilled workers resulting from the lower equilibrium wage.

In addition, I demonstrate how incorporating a training program or job creation subsidies

\textsuperscript{3}This effect can also be achieved by reducing the marginal net gain of keeping the high human capital level. This will be made clearer in section 2.

\textsuperscript{4}Introducing risk-aversion and savings will alter this last statement.
can improve the second-best optimal allocation. Introducing a training program that could retrain unemployed previously skilled workers can put downward pressure on the first-period wage of the unskilled worker. The reason for this is that increasing the likelihood of matching with a low wage makes the unskilled worker allocation less appealing to the skilled worker. However, since the presence of the training program can reduce the information problem it also reduces the need to distort the wage downward. Naturally, having a training program will increase the amount of redistribution undertaken in the second period because there are more skilled workers now for the same employment level.

Introducing job creation subsidies can also relax the informational problem of the planner because each labour market will react differently to a given level of subsidy. It turns out that the job creation subsidies are useless for reasonable degrees of market frictions. The optimal tax system is enough to achieve the redistributive goals of the planner in most cases. When market frictions achieve dramatic levels, i.e. employment levels below 50% in our numerical simulations, which leads the planner to distort the unskilled worker’s wage to zero, then it is optimal to have positive levels of subsidies.

This paper contributes to the normative literature that features human capital depreciation caused by unemployment. So far, the literature has concentrated mainly on the design of the optimal unemployment insurance system (Shimer & Werning, 2006; Pavoni, 2008; Spinnewijn, 2010) with the exception of Coles & Masters (2000) who look at the optimal labor policies that maximize the surplus generated in the steady state.

Related to this article, Stiglitz (1982) investigates the property of the optimal nonlinear tax system in a two-type economy when wages are endogenous due to the imperfect substitutions of types in the production function. Engström (2002) and Boone & Bovenberg (2004) both analyze optimal redistributive nonlinear taxation in a search framework but feature exogenous wages. Boadway et al. (2003) consider optimal employment and redistributive policies with frictional unemployment and observable skills.

Another related strand of literature is the recent ‘New Public Finance’ (Golosov et al., 2003; Kocherlakota, 2005; Golosov et al., 2007) which focuses on dynamic optimal taxation in which the wage rate can change over time due to a random process, but never depends on the policy choices of the planner. Bohacek & Kapicka (2008) investigate the optimal income tax and optimal schooling subsidies in a dynamic private information economy with observable human capital accumulation. They find that schooling subsidies can be greatly welfare improving if the income tax is not set optimally. However if the optimal income tax is set optimally their effects on welfare are small and the marginal schooling subsidies are positive and smaller than the marginal tax rates.
2 The Model

I consider a two-period model where there is a continuum of risk-neutral individuals divided into two types. Each type of individual is characterized by their productivity level. A skilled individual has productivity level $a_h$ and an unskilled individual has level $a_\ell \equiv \delta a_h$, where $0 < \delta < 1$. In the first period there is a fraction $\pi_\ell$ of unskilled individuals and a fraction $\pi_h$ of skilled individuals with $\pi_\ell + \pi_h = 1$.

Each period features a distinct model of the labor market. In the first period the labor market is plagued by search and matching frictions à la Mortensen & Pissarides (1999). In this period, individuals of productivity $a_i$ direct their search in markets of skill $i$. For simplicity, assume that individuals of skill $i$ that searches in market other than $i$ and finds a job will be unable to produce any output. This assumption, made for tractability, will be relaxed in the next period.\(^5\) As in the search and matching literature workers and firms get together according to a constant-return-to-scale function $M(U_i, V_i)$ which determines the number of matches in a type $i$ labor market. This function is twice-continuously differentiable, increasing and concave in the number of individuals searching ($U_i$) and the number of vacancies opened ($V_i$), and also satisfies $M(0, V_i) = M(U_i, 0) = 0$, and $M(U_i, V_i) < \min(U_i, V_i)$. In this model, matches last for only one period. This is assumed to keep the number of wages in the second period to two which makes the analysis simpler. Due to matching frictions, there will be unemployment in the first period. This unemployment is especially costly for skilled individuals since it destroys a fraction of their human capital, which in this paper is modeled as their productivity level. Again, to keep the number of types to two, I assume that skilled individuals who become unemployed lose enough human capital to be considered an unskilled individual in the future. Equivalently unemployment destroys $(1 - \delta)$ of the skilled individual’s productivity.

In the second period, there are no longer any matching frictions and individuals face a perfectly elastic labor demand. In this period everyone will be employed, but contrary to the previous period, individuals are able to choose in which skill sector they want to work in. To be more precise, skilled individuals can decide to work in either the skilled or in the unskilled sector. This is modeled in a similar way as the occupational choice framework of Piketty (1997) and Saez (2002, 2004), and we can make the similar interpretation of Saez (2004) this being the long-term choice of the individual. What determines this choice of occupation is an individual cost $\alpha_h$ of working in the skilled sector that the skilled individual discovers after working in that sector in the first period. As in Diamond (2006) and Diamond & Spinnewijn (2011), since the unskilled individuals lack the necessary skill to work in the skilled sector, they are bound to work in the unskilled sector.

\(^5\)Although a strong assumption, segmentation is more plausible than the other extreme case of one single labor market for all skill levels.
period individual cost of working in the skilled sector. However, he is able to observe the wage(income) \( w_{1i} \) offered to workers of type \( i \) in each period \( t \). With the assumption that the planner is able to commit to his tax policy in both periods, the planner sets up a history-dependent tax system that is a function of observable wages and observable unemployment. In the first period, the worker faces a tax function \( T_1(w_{1i}) \) and in the second period he faces a tax function \( T_2(w_{1i}, w_{2i}) \), where unemployment is represented by \( w_{10} = 0 \). In this model, due to our assumption of risk-neutrality and no participation decision the model has no unemployment benefits.\(^6\)

Furthermore, in addition to not being able to observe the skill type of individuals, we assume that the planner is unable to observe the matching process or the type of vacancies offered by firms. Therefore the planner is unable to infer the skill level of the individuals searching in these markets. This informational constraint also puts a limit on the tools of the planner to promote employment in specific skill sectors. The timing of the two-period model is the following:

\( T=1 \)

1. The planner credibly commits to tax functions \( T_1(\cdot) \) that depends on wage(income) \( w_{1i} \) and \( T_2(\cdot, \cdot) \) that depends on both the first and second period wage (income).

2. For each skill level \( i \), firms open up vacancies. Each vacancy of type \( i \) has screening costs \( \kappa_i \). Each individual costlessly searches in their type-specific markets.

3. For each labor market, the matching process determines the number of filled jobs and wage level.

4. Each employed individual of skill \( i \) supplies a fixed amount of labour and produces \( a_i \) units of goods. They receive wage \( w_{1i} \) and pay taxes or receive transfers \( T_1(w_{1i}) \). Skilled individuals that did not find a job, loses \((1 - \delta)\) of productivity and become unskilled individuals. Individuals consume their after-tax income.

5. Each match is dissolved. Period ends.

\( T=2 \)

6. Skilled individuals that found a job in the first period learn their cost \( \alpha_h \) of working in the skilled sector.

\(^6\)Without a participation decision, the mechanics of this model would favor allocations that reduce after-tax income to almost zero and increase unemployment insurance benefits to a very high level since this could raise social welfare without making the incentive constraint bind. Adding several skill types and having a unique participation cost across skill types as the model of Hungerbühler et al. (2006) or having a distribution of participation costs per type as Lehmann et al. (2011) both solve the problem. This, however, complicates the model significantly and is left for future research.
7. There are no longer any matching frictions and each individual finds a job. The skilled individuals with $\alpha_h$ below a certain threshold decide to work in the skilled sector, the rest decide to work in the unskilled sector. The unskilled individuals (including the skilled individuals that did not find a job in $T=1$) all work in the unskilled sector.

8. Each individual of skill $i$ supplies a fixed amount of labour and $\alpha_i$ units of goods. Skilled workers receive wage $w_{2h}$ and pay taxes $T_2(w_{1h}, w_{2h})$, unskilled workers get $w_{2\ell}$ and pay taxes $T_2(\cdot, \cdot)$ depending on their labour market history. Individuals consume their after-tax income.

### 2.1 Individuals

#### 2.1.1 First Period

In this model, the first period of life is considered to be the early decision to look for a job corresponding to one’s skill. As mentioned above, these workers costlessly search for firms in their skill specific labor market. In this period, the worker’s action, depending on assumptions made on wage-setting, is to look for a good wage in the case of directed search by skill and wage, or to negotiate a wage with the firm in random matching. In any case, the probability of finding a job (or the ‘labor demand’) for type $i$ will be $L_i$. If they match, they will receive after-tax income (consumption) $w_{1i} - T_1(w_{1i})$. For notational convenience let expected after-tax income in the first period for type $i$ to be $u_{1i} = L_i[w_{1i} - T_1(w_{1i})]$. In addition to the after-tax income, the employed skilled workers each keep their human capital which guarantees them expected utility $\nu_{hh}$ in the second period.

If a skilled individual does not find a job, he loses a fraction of his human capital. This loss of human capital entails that he will now have utility $u_{0\ell}$ in the second period which represents the value of working in the unskilled sector knowing that the individual was unemployed in the first period. In the case of the unskilled worker, whether he is employed or not, he does not lose any human capital. However, because the planner is able to observe both wages and employment status in the first period, he can tax differently the unskilled type in the second period depending on his employment history. Thus, an unskilled worker that found a job in the first period will have $u_{\ell\ell}$ utility in the second period and the ones that did not find a job have utility in the second period of $u_{0\ell}$. From this, the expected utilities in the first period for each type are:

$$
v_h = L_h[w_{1h} - T_1(w_{1h}) + \beta \nu_{hh}] + (1 - L_h)\beta u_{0\ell},$$

$$= L_h[w_{1h} - T_1(w_{1h}) + \beta \Psi_h] + \beta u_{0\ell}$$

$$= u_{1h} + L_h\beta \Psi_h + \beta u_{0\ell},$$

and

$$v_l = L_l[w_{1\ell} - T_1(w_{1\ell}) + \beta u_{\ell\ell}] + (1 - L_l)\beta u_{0\ell},$$

$$= L_l[w_{1\ell} - T_1(w_{1\ell}) + \beta \Psi_l] + \beta u_{0\ell}$$

$$= u_{1\ell} + L_l\beta \Psi_l + \beta u_{0\ell},$$

(2.1)
where \( \Psi_h \equiv \nu_{hh} - u_{0h} \), and \( \Psi_l \equiv u_{0l} - u_{0l} \), \( \beta \) is the worker’s discount rate. \( \Psi_i \) is the net future gain of matching in the first period for skill \( i \). In the case of the skilled worker, this captures the net gain of keeping his human capital.

### 2.1.2 Second Period

In the second period, the worker’s choice is slightly different. He can now decide to choose another ‘occupation’ than the one corresponding to his skill level. In the two skill level context of our model, only the skilled workers in the second period, that is, the skilled workers that found a job in the first period can decide to work in the other sector.

Suppose that for a skilled individual, the cost of working in the skilled sector is \( \alpha_h \) which is assigned randomly to them at the start of the second period according to the cumulative distribution function \( F(\alpha_h) \) with \( \alpha_h \in [0, \bar{\alpha}] \). One could interpret this cost as knowledge gained through working in the first period about how difficult or enjoyable this occupation is. Therefore, when making their decision in the first period, workers are unaware of this future value, only its distribution. In addition, there is no cost to working in the unskilled sector, regardless of skill or employment history.\(^7\) As mentioned above, unskilled workers, i.e. unskilled workers in the first period and the skilled workers that experienced unemployment in the first period, have no choice to be in the unskilled sector since they do not have the human capital to work in the skilled sector.

Because there is no longer any uncertainty with regards to employment in the second period the utility a skilled worker gets from participating in the skilled sector is equal to his consumption utility (or after-tax income) \( u_{hh} = w_{2h} - T_2(w_{1h}, w_{2h}) \) minus the cost of working \( \alpha_h \). If he chooses to participate in the unskilled market, he gets utility \( u_{hl} = w_{2l} - T_2(w_{1h}, w_{2l}) \). Therefore the skilled workers that will choose to work in the skilled sector are those who have a low enough \( \alpha_h \) such that:

\[
\begin{align*}
  u_{hh} - \alpha_h & \geq u_{hl}, \\
  \Delta u_{hh} & \equiv u_{hh} - u_{hl} \geq \alpha_h.
\end{align*}
\]

This implies that the proportion of skilled workers in the second period that will decide to work in the skilled sector is \( F(\Delta u_{hh}) \).

Taking this decision into account, the expected value of keeping the skilled individual’s

\(^7\) Adding a cost to working in the unskilled sector changes nothing to the analysis, instead of the distribution of \( \alpha_h \) that would be important, it would be the distribution of the difference between the two cost of working.
human capital is:

\[
\nu_{hh} = \int_0^\infty \max \{ u_{hh} - \alpha_h; u_{hl} \} f_h(\alpha_h) d\alpha_h,
\]

\[
= u_{hl} + \int_0^{\Delta u_{hh}} [\Delta u_{hh} - \alpha_h] f_h(\alpha_h) d\alpha_h,
\]

From this definition and recalling that \( \Psi_h = \nu_{hh} - u_{0l} \), the expression for the net gain of keeping the skilled human capital is:

\[
\Psi_h = u_{hl} + \int_0^{\Delta u_{hh}} [\Delta u_{hh} - \alpha_h] f_h(\alpha_h) d\alpha_h, -u_{0l},
\]

\[
= \Delta u_{hl} + \int_0^{\Delta u_{hh}} [\Delta u_{hh} - \alpha_h] f_h(\alpha_h) d\alpha_h,
\]

(2.2)

where \( \Delta u_{hl} = u_{hl} - u_{0l} \) is the difference between the utility that a skilled individual gets by working in the unskilled sector and the utility that a previously unemployed worker gets by working in the unskilled sector which is \( u_{0l} = w_{2l} - T_2(0, w_{2l}) \). For the unskilled worker that found a job in the previous period, the utility (after-tax income) the he will receive in the second period is \( u_{\ell \ell} = w_{2l} - T_2(w_{1l}, w_{2l}) \). For future use, we will define \( \Psi_l \equiv \Delta u_{\ell \ell} \equiv u_{\ell \ell} - u_{0l} \).

Because of the history-dependent tax function and the ability of the skilled to work in the unskilled sector, the individuals in the second period can be divided in four distinct groups. The first group is the skilled individuals working in the skilled sector, the second is the skilled working in the unskilled sector, the third is the unskilled that has worked in the unskilled sector in both periods and the last group is the one where the individuals have experienced unemployment in the first period and are now working in the unskilled sector. The measure of individuals in each group are respectively \( \phi_{hh} = \pi_{hh} \cdot L_{1h} \cdot F(\Delta u_{hh}) \), \( \phi_{hl} = \pi_{hl} \cdot L_{1h} \cdot [1 - F(\Delta u_{hh})] \), \( \phi_{ll} = \pi_{ll} \cdot L_{1l} \) and \( \phi_{0l} = \pi_{0l} \cdot [1 - L_{1l}] + \pi_{hl} \cdot [1 - L_{1hl}] \), where \( \phi_{hh} + \phi_{hl} + \phi_{ll} + \phi_{0l} = 1 \).

### 2.2 Firms

Firms are assumed to be risk-neutral and are able to enter freely in labour markets. As mentioned above, matching frictions are only a characteristics of first period labour markets and that in the second period firms can instantaneously find workers.

In the first period, the matching technology is the same in both labor markets, and therefore the analysis of a particular firm in a particular labour market can be applied to all firms in the first period. As is customary in the search and matching literature, it is convenient to write the probability of matching as a function of the labour market ‘tightness’. Define tightness \( \theta_i \) as the ratio \( V_i/U_i \). The probability of filling a type-i vacancy (resp. the probability of finding a job) is \( m(\theta_i) = M(U_i, V_i)/V_i = M(1/\theta_i, 1) \) (resp.
\[ L_i = \theta_i m(\theta_i) = M(U_i, V_i)/U_i = M(1, \theta_i). \] Thus, the probability of filling a vacancy decreases with an increase in vacancies \((V_i)\) and increases in the number of job seekers \((U_i)\), while the probability of finding a job increases with the number of vacancies and decreases with the number of job seekers.

In the first period a firm opens up a vacancy in market \(i\) at the search cost \(\kappa_i\). Each filled job of type \(i\) will produce an amount \(a_i\). Therefore the firm’s expected profits are \(m(\theta_i)[a_i - w_i] - \kappa_i\). Since firms can enter freely, vacancies will be created when profits are positive. Because an increase in vacancies decreases the probability of filling a job, firms will open up vacancies until \(m(\theta_i)[a_i - w_i] = \kappa_i\). This is the ‘free-entry condition’ which pins down the value of market tightness, i.e \(\theta_i = m^{-1}(\kappa_i/(a_i - w_i))\), and leads to the following the labor demand:

\[
L(a_i, w_i) = m^{-1}\left(\frac{\kappa_i}{a_i - w_i}\right) \times \frac{\kappa_i}{a_i - w_i},
\]

which is decreasing in the level of wage \(w_i\) and increasing in the productivity level \(a_i\). The constant-return-to-scale property implies that only the productivity level and wage matter and not the number of individuals looking for a job. For ease of exposition, from now on we will write \(L_i(w_j)\) the labor demand for the market of skill type \(i\), i.e. where the value of production is \(a_i\) and cost of vacancy \(\kappa_i\), for a wage of skill type \(j\). For example, if the pre-tax wage prevailing in the unskilled market would be \(w_{1h}\), then we would write the labor demand in the following way \(L_\ell(w_{1h}) = L(\ell, w_{1h})\).

In the second period, because there are no longer market frictions, the number of matches in each skill level is determined by the short side of the market, i.e \(M(U_i, V_i) = \min(U_i, V_i)\). Since, there is free entry and a fixed number of individuals in each skill level, this implies that all individuals find a job. Also, because there are no longer any matching frictions, firms no longer need to pay search cost \(\kappa_i\) and because of free-entry, wages in each market will be competed up until \(w_{2i} = a_i\).

### 2.3 Wage Setting in First Period

Due to search frictions, a match between a firm and worker creates a surplus. The search and matching literature has highlighted many mechanism by which this surplus is shared. In this paper I consider sharing rules that leave only a redistributive role of taxation, i.e. the no-tax economy is efficient. To get the property that the no-tax economy is efficient, the wage-setting mechanism must be such that it maximizes the expected utility of workers in the no-tax economy. Under the assumptions of risk-neutrality and free-entry condition, it can be shown that the wage chosen by these mechanisms will also maximize the output value in both periods net of vacancy costs in the no-tax economy.\(^8\) The standard approach to gain

\(^8\)To see this, first notice that in a no-tax economy, \(\Psi_h = \int_0^{a_h - a_\ell}[(a_h - a_\ell) - a_h]f_h(a_h)da_h\) and \(\Psi_\ell = 0\), since \(u_{1h} = w_{2h} = a_h\) and \(u_\ell = u_\ell = u_{1\ell} = w_{2\ell} = a_\ell\). Using the free-entry condition we have that \(m(\theta_i)a_i - \kappa_i = m(\theta_i)w_i\), by multiplying both sides by \(\theta_i\) we get \(L_i(w_i)a_i - \theta_i \kappa_i = L_i(w_i)w_i\). For the unskilled
An important feature of the wage-setting objective in this model is that it is increasing in after-tax surplus \( x_i = w_{1i} - T_1(w_{1i}) + \beta \Psi_i(w_{1i}) \) and decreasing in pre-tax wage \( w_{1i} \). For both types in period 1, the ‘wage-setting objective’ is:

\[
\mathcal{WS}(x, w, a) \equiv x \cdot L(a, w),
\]

which is the expected surplus of matching with a firm in period 1 for an individual of skill \( i \). Thus the wage chosen in each skill sector will be

\[
w_{1i} = \arg \max_w \left[ w - T_1(w) + \beta \Psi_i(w) \right] \cdot L_i(w).
\]

Because of the history-dependent taxes featured in the model, \( \Psi_i \) is written as a function of \( w_{1i} \) since choosing \( w_{1i} \) influences the payoff in the future and thus the surplus from matching. Due to the two-type model set up this function is discontinuous, but as in Stiglitz (1982) with the income tax function, to get intuition for the problem, we concentrate on the continuous portion of \( \Psi_i \) in first period wage \( w \). Also, as in Lehmann et al. (2011), to make (2.4) well-behaved and that it gives the property that \( w_{1i} \) is increasing in skill \( i \) in the no-tax economy, I suppose that the wage elasticity \( \frac{\partial L(a, w)}{\partial w} \frac{w}{L(a, w)} \) is decreasing in \( w_i \) and increasing in \( a_i \).\(^\text{10}\) The FOC of (2.4) can be written in this useful way:

\[
- \frac{\partial L_i(w_{1i})}{\partial w_{1i}} \frac{w_{1i}}{L_i(w_{1i})} = \frac{1 - T_1'(w_{1i}) + \beta \Psi_i'(w_{1i})}{1 - \left[ \frac{T_1(w_{1i}) - \beta \Psi_i}{w_{1i}} \right]}. \tag{2.5}
\]

workers we can stop here, because the unskilled worker will produce \( a_i \) and get paid \( a_i \) regardless if he finds a job or not. Hence maximizing output net of vacancy cost in the unskilled market is equivalent of maximizing the first period expected income of the unskilled worker when there are no taxes. This is slightly different for the skilled individuals. In the first period, the discounted expected value of production in the second period is \( \beta(L_i(w_i) \int_0^\infty \max\{a_h - a_h'; a_H\} f_h(a_h) da_h + [1 - L_i(w_i)] a_i \) which is equivalent to \( \beta[L_i(w_i) \Psi_h + a_i] \) in the no tax economy. Therefore, maximizing the output value net of vacancy cost and work effort in the skilled market, i.e. \( L_i(w_i)[a_i + \beta \Psi_h] + \beta a_i - \theta_i k_i \), is equivalent to maximizing the expected lifetime utility of the individual, i.e. \( L_i(w_i)[w_i + \beta \Psi_h] + \beta a_i \), when the individuals do not have to pay taxes. Of course this is equivalent to maximizing expected surplus from matching with a firm in the no-tax economy, i.e. \( L_i(w_i)[w_i + \beta \Psi_h] \).

\(^9\)See Hosios (1990). To see more clearly how the wage-setting objective relates to the Nash-bargaining solution see Appendix A.

\(^{10}\)The last two assumptions are not limiting, as Lehmann et al. (2011) argue, the assumptions are easily satisfied with a CES matching functions under a broad range of the parameter value determining the the elasticity of substitution and also is \( \frac{\partial L(a, w)}{\partial a} \leq 1 \).
The left-hand side is the negative of the wage elasticity of the employment probability, it measures the reduction in the probability of employment as wage increases. The right-hand side is the wage elasticity of the surplus going to the worker once he is matched. The chosen optimal wage has these two elasticities at equal value. And since labor demand is a decreasing function of wage and that the surplus of matching is positive, the FOC implies that \( T'_1(w_{1i}) - \beta \Psi'_i(w_{1i}) \leq 1 \). Similar to the New Dynamic Finance literature that features history-dependent tax systems, \( T'_1(w_{1i}) - \beta \Psi'_i(w_{1i}) \) is the wedge that will cause a distortion in the labor market and incorporates the effect of the history-dependent tax system on the decision of the worker.

Equation (2.5) highlights the impact of the tax function \( T_1(w) \) on the equilibrium wage of a given skill and at the same time the effect it has on employment for the labor market. Suppose that \( T'_1(w), \frac{T_1(w)}{w}, \frac{\beta \Psi_i}{w}, \text{ and } \Psi'_i(w_{1i}) \) are parameters. Keeping other parameters constant, an increase in \( T'_1(w) \) (or a decrease in \( \Psi'_i(w_{1i}) \)) will lower the value of the wage elasticity of the surplus, this implies that wage elasticity of employment must also follow. From the assumptions made on the matching function, this translates to a reduction in wage and thus and increase in employment for that skill level. The reverse happens if \( \frac{T_1(w)}{w} \) is increased. This will increase the value of the wage elasticity of the surplus and thus will lead to an increase in the wage and therefore a reduction in employment. An increase in \( \frac{\beta \Psi_i}{w} \) has the same effect as an increase in the marginal tax rate, it lowers the chosen wage. All three of these effects come from the surplus sharing rule. Increasing the marginal tax rate makes it harder to transfer utility from the firm to the worker, because some of it is taxed away. Increasing \( \Psi_i \) makes the worker much more willing to find and accept a lower wage because he gets this value only if he is employed. Increasing the average tax rate \( \frac{T_1(w)}{w} \) decreases the surplus of the worker of finding a job, this has the opposite effect of an increase in \( \Psi_i \), it lowers the value of finding and keeping a job and thus makes them unwilling to accept a low wage.

### 2.4 Planner and Perfect Information in First Period

#### 2.4.1 Planner

In this paper, three main assumptions on the planner are made. The first assumption is that the planner’s preference over lifetime expected utilities of individuals can be expressed by the following Bergson-Samuelson social welfare function:

\[
SWF = \pi_t W\left( u_{1t} + \beta L_t(w_{1t})\Psi_t + \beta u_{0t} \right) + \pi_h W\left( u_{1h} + \beta L_h(w_{1h})\Psi_h + \beta u_{0h} \right),
\]

where \( W'(\cdot) > 0 \) and \( W''(\cdot) \leq 0 \). This social welfare function is quite general and can represent different aversion to inequality from the utilitarian criterion (\( W''(\cdot) = 0 \)) to the Maximin(or Rawlsian) criterion that exhibits extreme aversion to inequality of expected utilities.
The second assumption is related to the use of the ‘occupational choice’ model in the second period. For all informational cases that we will investigate throughout the paper, i.e. being or not being able to observe the type of worker in the first period, the planner in the second period is able to observe the productivity of all jobs but unable to observe the cost of working $\alpha_h$. These informational assumptions are made to clarify the analysis and to focus on the workings of the first period labour market and tax system. The simple framework of the occupational model of Saez (2002, 2004) makes it easier to investigate the impact of changing the number of skilled individuals on the optimal allocation in the second period, which then helps identify characteristics of the tax system in the first period.$^{11}$

The third and final assumption is that the planner is unable to transfer resources between periods. He is unable to save or borrow to help smooth consumption between periods. This assumption is made due to the risk-neutrality of workers.

Since the planner is unable to transfer resources between periods, he faces two budget constraints, one in the first period and another in the second. As is customary in most of the optimal income tax literature, it is more convenient to write the problem as an allocation problem, therefore the first period budget constraint of the planner can be written in the following way:

$$
\pi_\ell \cdot [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0,
$$

(2.7)

recalling that $u_{1\ell} = L_\ell(w_{1\ell}) \cdot [w_{1\ell} - T_1(w_{1\ell})]$ and $u_{1h} = L_h(w_{1h}) \cdot [w_{1h} - T_1(w_{1h})]$.

Also, the second period budget constraint becomes:

$$
\phi_{0\ell}(w_{1\ell}, w_{2\ell})[w_{2\ell} - u_{0\ell}] + \phi_{ll}(w_{1\ell})[w_{2\ell} - u_{\ell\ell}] + \phi_{hl}(w_{1h}, \Delta u_{hh})[w_{2h} - u_{hl}] + \phi_{hh}(w_{1h}, \Delta u_{hh})[w_{2h} - u_{hh}] = 0,
$$

(2.8)

recalling that $u_{\ell\ell} = \Delta u_{\ell\ell} + u_{0\ell}$, $u_{hl} = \Delta u_{hl} + u_{0\ell}$ and $u_{hh} = \Delta u_{hh} + u_{hl}$.

2.4.2 Perfect Information in First Period

To better contrast the impact of imperfect information on the optimal tax system, I investigate in this subsection the characteristics of the perfect information allocation, i.e. the planner is able to observe the productivity of a worker-firm pair in the first period. Therefore, he will maximize his social preferences by picking the first period wage for both skill levels $w_{1i}$, the first period expected income $u_{1i}$, the utility level(income) $u_{0\ell}$ and the utility differences $\Delta u_{\ell\ell}$, $\Delta u_{hl}$ and $\Delta u_{hh}$, the problem of the planner is:

$^{11}$It also prevents from having the trivial answer that all redistribution should take place in the second period.
\[
\begin{align*}
\max_{\{w_{1\ell}, u_{1\ell}\}_i, u_{0\ell}, \Delta u_{1\ell}, \Delta u_{hh}} & \quad \pi \ell W \left( u_{1\ell} + \beta L_\ell(w_{1\ell}) \Psi_i(\Delta u_{1\ell}) + \beta u_{0\ell} \right) + \pi h W \left( u_{1h} + \beta L_h(w_{1h}) \Psi_h(\Delta u_{1h}, \Delta u_{hh}) + \beta u_{0h} \right) \\
\text{s.t.} & \quad (\lambda_1) \quad \pi \ell \cdot [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0, \\
& \quad (\lambda_2) \quad \phi_0(w_{1\ell}) w_{2\ell} + \phi_0(w_{1\ell})[w_{2\ell} - \Delta u_{1\ell}] + \phi_h(w_{1h}, \Delta u_{hh})[w_{2h} - \Delta u_{1h} - \Delta u_{hh}] = u_{0\ell}, \quad (2.9)
\end{align*}
\]

Due to the linearity of the individuals utility function, there are many possible solutions to the allocation problem. However, each solution must satisfy the following necessary conditions:

\[
W' \left( u_{1\ell} + \beta L_\ell(w_{1\ell}) \Psi_i(\Delta u_{1\ell}) + \beta u_{0\ell} \right) = W' \left( u_{1h} + \beta L_h(w_{1h}) \Psi_h(\Delta u_{1h}, \Delta u_{hh}) + \beta u_{0h} \right), \quad (2.10)
\]

\[
\begin{align*}
\Delta u_{hh} &= w_{2h} - w_{2\ell}, \\
- \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \cdot w_{1\ell} &= 1, \quad (2.11)
\end{align*}
\]

\[
\begin{align*}
- \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \cdot w_{1h} &= \frac{1}{1 + \beta \Psi^LF_h}, \quad (2.12)
\end{align*}
\]

where \( \Psi^LF_h = \int_0^{w_{2h} - w_{2\ell}} [(w_{2h} - w_{2\ell}) - \alpha_h] f_h(\alpha_h) d\alpha_h = \int_0^{a_h - a_\ell} [(a_h - a_\ell) - \alpha_h] f_h(\alpha_h) d\alpha_h. \)

Condition (2.10) says that expected utility of each type must be equalized, or more precisely the expected surplus from matching with a firm are equal. This result also tells us that these conditions could have been derived using a Maximin social welfare function for the preferences of the planner. Condition (2.11) indicates that the tax paid by the skilled types irrespective of their occupational choice must be the same, i.e \( T_\ell(w_{1h}, w_{2h}) = T_\ell(w_{1h}, w_{2\ell}). \)

It also tells us that \( \Psi_h = \Delta u_{1h} + \Psi^LF_h \) since \( w_{2h} - w_{2\ell} = a_h - a_\ell. \)

The last two conditions are the ones related to the optimal first period wages. The first thing to notice is that both (2.12) and (2.13) are at their laissez-faire levels. To see this, we just have to look at (2.5), remove taxes and remember that \( \Psi^LF_i = 0 \) and \( \Psi^LF_i = 0. \)

This is not a surprising find, since the assumptions made on the wage-setting mechanism are such that the laissez-faire allocation would be efficient, and since the planner has perfect information in the first period he can redistribute and set the tax system in such a way that it isn’t distortionary in that period.

\footnote{The FOCs have been partly solved to remove multipliers.}
To decentralize the optimal wage, the first-order condition of the wage-setting problem must be used much like the first-order condition to the individual’s problem is used in a standard optimal income tax exercise. Combining (2.12) and (2.5) to get:

$$T_1'(w_{1\ell}) - \beta \Psi_1'(w_{1\ell}) = \frac{T_1(w_{1\ell})}{w_{1\ell}} - \beta \frac{\Psi_1}{w_{1\ell}}.$$  

where $\Psi_1 = \Delta u_{\ell\ell} = T_2(0, w_{2\ell}) - T_2(w_{1\ell}, w_{2\ell})$. Similar to much of the literature on dynamic optimal taxation, history-dependence of the tax function makes it hard to evaluate the structure of the tax function during a specific period, however it is possible to say something about the distortions needed to achieve optimality. First, notice that how the marginal tax rate and marginal net benefit of keeping human capital relates to the average tax rate will depend on the sign of $\Delta u_{\ell\ell}$. Because there are many possible solutions, $\Delta u_{\ell\ell}$ could potentially be of any sign, but assuming that there is some insurance in $T_2(0, w_{2\ell})$, then $\Delta u_{\ell\ell} < 0$ and thus $T_1'(w_{1\ell}) - \beta \Psi_1'(w_{1\ell}) > \frac{T_1(w_{1\ell})}{w_{1\ell}}$. On the other hand assuming that there is no difference in the taxation of the unskilled worker in the second period, i.e the tax system is either age-dependent or age-independent, then the marginal tax rate equals the average tax rate, $T_1'(w_{1\ell}) = \frac{T_1(w_{1\ell})}{w_{1\ell}}$. Notice, that if the unskilled receives a transfer in the first period and $\Delta u_{\ell\ell} = 0$, then the marginal tax rate must be negative.

Decentralizing the optimal skilled worker’s wage, uses (2.13), (2.5) and the fact that in this informational case $\Psi_h = \Delta u_{h\ell} + \Psi_{LF}^h$:

$$T_1'(w_{1h}) - \beta \Psi_h'(w_{1h}) = \frac{T_1(w_{1h})}{w_{1h} + \beta \Psi_{LF}^h} - \frac{\beta \Delta u_{h\ell}}{w_{1h} + \beta \Psi_{LF}^h},$$

where $\Delta u_{h\ell} = T_2(0, w_{2\ell}) - T_2(w_{1h}, w_{2\ell}) = T_2(0, w_{2\ell}) - T_2(w_{1h}, w_{2h})$. Considering the case where there is no redistribution to the unskilled in the second period, this would imply $\Delta u_{h\ell} = 0$ and we would get that the marginal tax rate minus the marginal change in the net benefit is equal to the total tax burden divided by the surplus one gets from matching with a firm in the first period, i.e $T_1'(w_{1h}) - \beta \Psi_h'(w_{1h}) = \frac{T_1(w_{1h})}{w_{1h} + \beta \Psi_{LF}^h}$. Since $\Psi_{LF}^h > 0$, this implies that $T_1'(w_{1h}) - \beta \Psi_h'(w_{1h}) < \frac{T_1(w_{1h})}{w_{1h}}$. The upward effect on wages of the tax burden is reduced by the presence of a positive surplus an individual gets from keeping his human capital while working in the skilled sector, therefore there is less of a need to use the counter-balancing effects of the marginal tax rates to keep the skilled worker’s wage at the efficient level. Supposing that there is some redistribution from the skilled to the unskilled workers in the second period, implies $\Delta u_{h\ell} < 0$. In this case, even in the presence of a positive $\Psi_{LF}^h$, it is possible to have $T_1'(w_{1h}) - \beta \Psi_h'(w_{1h}) > \frac{T_1(w_{1h})}{w_{1h}}$, because of redistribution in the second period, matching and taking a lower wage becomes less appealing. Thus, to offset this effect a downward pressure from a higher marginal tax rate or a lower marginal net benefit of matching $[\Psi_h'(w_{1h})]$ must be applied.

\[13\] Supposing that there are enough resources in the first period to accomplish full expected utility equality.
Conditions (2.12) and (2.13) can also be used to determine the conditions for chosen wages to be below or above their efficient level. For the unskilled worker’s wage, (2.12) gives the following conditions:

\[- \frac{\partial L(\ell(w_{1\ell}))}{\partial w_{1\ell}} \frac{w_{1\ell}}{L(\ell(w_{1\ell}))} < 1 \iff w_{1\ell}^* < \bar{w}_{1\ell},\]

\[- \frac{\partial L(\ell(w_{1\ell}))}{\partial w_{1\ell}} \frac{w_{1\ell}}{L(\ell(w_{1\ell}))} > 1 \iff w_{1\ell}^* > \bar{w}_{1\ell},\]

where \( \bar{w}_{1\ell} \) is the efficient level wage and \( w_{1\ell}^* \) is the chosen wage. If the wage elasticity of employment is lower (resp. higher) than one, this implies that the wage is below (resp. above) the efficient level. It is easier to rewrite condition (2.13) to do the same exercise with the skilled worker’s wage:

\[- \frac{\partial L(h(w_{1h}))}{\partial w_{1h}} \left[ w_{1h} + \beta \Psi^F_i \right] \frac{w_{1h}}{L(h(w_{1h}))} < 1 \iff w_{1h}^* < \bar{w}_{1h},\]

\[- \frac{\partial L(h(w_{1h}))}{\partial w_{1h}} \left[ w_{1h} + \beta \Psi^F_i \right] \frac{w_{1h}}{L(h(w_{1h}))} > 1 \iff w_{1h}^* > \bar{w}_{1h},\]

where again \( \bar{w}_{1h} \) is the efficient level wage and \( w_{1h}^* \) is the chosen wage.

### 3 Optimal Tax Policy

In this section I consider the case where the planner is unable to observe the productivity of a worker-firm pair, he is only able to observe the income (negotiated wage) of workers. This puts a constraint on the ability of the planner to redistribute income across types of individuals, since taxes can no longer be a function of ability and income, they can only be a function of incomes (wage) and labor market experience. In this paper the worker-firm pair’s behavior is modeled as a single agent maximizing the wage-setting objective (2.4) which acts as the preferences over different bundles.\(^{14}\) Because of the history-dependent tax system in place, choosing the optimal wage for a particular skill level implies the worker gets surplus

\[ x_{1i} = w_{1i} - T_1(w_{1i}) + \beta \Psi_i(w_{1i}) \]

from a match with a firm, from this he has expected surplus

\[ E_{1i} = x_{1i} \cdot L(a_i, w_{1i}). \]

Again, notice that the wage-setting objective is increasing in the surplus \( x_{1i} \) and decreasing in the wage \( w_{1i} \). Thus a tax system \( \{T_1(\cdot), T_2(\cdot, \cdot)\} \) will induce a set of allocations \( \{w_{1i}, x_{1i}, E_{1i}\}_{i=l,h} \), and from the Taxation principle (Rochet, 1985; Guesnerie, 1995) we know that the set of allocations will correspond to the set of allocations that satisfy the following incentive constraints:

\[ x_{1l} \cdot L(\ell(w_{1l})) \geq \bar{x}_{1l} \cdot L(\ell(w_{1h})), \]

\[ x_{1h} \cdot L(h(w_{1h})) \geq x_{1l} \cdot L(h(w_{1l})). \]

\(^{14}\)We assume that side-payments are not allowed.
where $x_{1h} = w_{1h} - T_1(w_{1h}) + \widehat{\psi}_h$ with $\widehat{\psi}_h$ being the value of a mimicking unskilled type of the difference in utility levels. Since, he is unable to work in the skilled market, similar to a skilled worker with a very high $\alpha_h$, he will have to take bundle $u_{hl}$ if he mimics.

The notion of a worker-firm pair being a single agent with the wage-setting objective as preferences can seem slightly unnatural, especially when it comes to the decision to mimic or not. As I have mentioned before, there are several microfoundations for the wage-setting objective used in this paper, and one that has a very natural interpretation for mimicking is the Competitive Search Equilibrium when search is directed by skill and wage. In this context there is one potential skill market for each skill level and each wage level. So when it is time to either look for jobs and open up vacancies workers and firms decide in which submarket, and only one, to participate in. As in Moen (1997), this is a non-cooperative game between firms and workers and using the Nash equilibrium concept gives us the Competitive Search Equilibrium.

For the following exposition, suppose that there are a continuum of skill levels and wages. Using the same notation as above, let $\theta_{iw} = V_{iw}/U_{iw}$ be the market tightness in submarket $(i, w)$. Since both firms and workers are atomistic they take the market tightness and by implication the probability of matching $m(\theta_{iw})$ for firms and $L(a_i, w) \equiv \theta_{iw} m(\theta_{iw})$ for workers as given. In such a set up, in every submarket $(i, w)$, due to free-entry, the zero-profit condition holds. If not, it would be profitable to enter at that specific wage and skill market and profits would be driven to zero again. It is important to remember that when a firm enters a specific submarket at a specific wage $w$, it credibly commits to offer that wage to workers and can’t offer another one after the match has happened.

Workers are limited to search in markets that are specific to their skill level, however they can enter one market at any posted wage $w$. I will ignore that participation decision of the worker and assume that the expected lifetime utility of worker that participate in a specific submarket is always greater or equal to zero. Take for example the case of a skilled worker, and consider the case of two submarket $(a_h, w)$ and $(a_h, w')$. If a worker enters and look in submarket $(a_h, w)$, he expects $L(a_h, w)[w - T_1(w) + \Psi_h(w)] + \beta_u w_0$, and if enters in submarket $(a_h, w')$, he expects $L(a_h, w')[w' - T_1(w') + \Psi_h(w')] + \beta u_{0\ell}$. Let suppose that $L(a_h, w)[w - T_1(w) + \Psi_h(w)] > L(a_h, w')[w' - T_1(w') + \Psi_h(w')]$ is true. Thus, it would never be in the interest of any worker of skill $a_h$ to enter submarket $(a_h, w')$. Since firms want to attract workers and that workers will only enter the submarket where they get the highest expected surplus, a worker will only enter the submarket $(a_h, w)$ if $w$ maximizes $L(a_h, w^*)[w^* - T_1(w^*) + \Psi_h(w^*)]$ in $w^*$. This logic also holds for unskilled workers (or any skill level considered). Therefore, when the planner is unable to observe skill levels and can only condition his policies on wages and market experience, the allocations that can be reached by the planner must be consistent with constraints

\begin{align*}
[w_{1\ell} - T_1(w_{1\ell}) + \beta \Psi_{1}(w_{1\ell})] \cdot L(a_\ell, w_{1\ell}) &\geq [w_{1h} - T_1(w_{1h}) + \beta \Psi_{1}(w_{1h})] \cdot L(a_h, w_{1h}), \quad (3.3) \\
[w_{1h} - T_1(w_{1h}) + \beta \Psi_{h}(w_{1h})] \cdot L(a_h, w_{1h}) &\geq [w_{1\ell} - T_1(w_{1\ell}) + \beta \Psi_{h}(w_{1\ell})] \cdot L(a_h, w_{1\ell}), \quad (3.4)
\end{align*}
which are exactly (3.1) and (3.2).

Before moving on, it is important to notice a useful property of the wage-setting objective used in this paper. Recalling that the wage-setting objective is:

\[ \mathcal{WS}(x, w, a) \equiv x \cdot L(a, w), \]

and then the marginal rate of substitution between the matching surplus \( x \) and wage \( w \) is:

\[ \frac{dx}{dw} \bigg|_{\mathcal{WS}(a, \cdot, \cdot)} = -\frac{x}{L(a, w)} \frac{\partial L(a, w)}{\partial w}. \]

From the assumptions made in the above section on \( L(a, w) \), note that the marginal rate of substitution is decreasing in \( a \), concluding from this that the marginal rate of substitution for the skilled worker-firm pair evaluated at a specific bundle will be smaller than the marginal rate of substitution of an unskilled worker-firm pair. Therefore in this model there is a Spence-Mirrlees single-crossing property.15

For the rest of the paper, I will concentrate on what Stiglitz (1982) calls the ‘normal case’ where incentive constraint (3.2) binds at the optimum but not (3.1). The idea being that the planner wishes the redistribute from the skilled types who have a greater expected lifetime utility to the unskilled types which induces the skilled types to ‘lie’ about their actual type.

### 3.1 Maximin Social Welfare Function

I first start by investigating the Maximin social welfare function case because it is possible to derive properties of the optimal allocation analytically and gather some intuition from the results. Also, since the perfect information optimal allocation derived above could of also have been derived with the Maximin social welfare preference, it is possible to compare both allocations. The characteristic of the Maximin social preference is that it features extreme inequality aversion and the planner only cares about the least-well off. In this model, it means the planner wishes to maximize the lifetime expected utility of the unskilled workers. As I have done for the perfect information in the first-period case, I set up the planner’s problem as an allocation problem. Using the same definitions used in section 2.4, the incentive constraint (3.2) can written in the following way:

\[ u_{1h} + L_h(w_{1h})\beta \Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_{\ell}(w_{1\ell})} u_{1\ell} + \beta L_h(w_{1\ell}) \Delta u_{\ell\ell}. \]

15The single-crossing property can fail for the cases of less sophisticated tax system, i.e. age-dependent or history(age)-independent tax systems.
The planner’s problem is then:

$$\max_{\{w_{1\ell},u_{1\ell}\}_v, u_{0\ell}, \Delta u_{1\ell}, \Delta u_{hh}} u_{1\ell} + \beta L_{1\ell}(w_{1\ell})\Delta u_{1\ell} + \beta u_{0\ell}$$

s.t. (λ₁) $\pi_\ell \cdot [L_{1\ell}(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0,$

(λ₂) $\phi_\ell(w_{1\ell}, w_{1h})w_{2\ell} + \phi_\ell(w_{1\ell})[w_{2\ell} - \Delta u_{1\ell}]$

$+ \phi_h(w_{1h}, \Delta u_{hh})[w_{2\ell} - \Delta u_{hh}] + \phi_h(w_{1h}, \Delta u_{hh})[w_{2h} - \Delta u_{hh} - \Delta u_{1\ell}] = u_{0\ell}$

(µ) $u_{1h} + L_h(w_{1h})\beta\Psi_h(\Delta u_{hh}, \Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_{1\ell}(w_{1\ell})}u_{1\ell} + \beta L_h(w_{1\ell})\Delta u_{1\ell},$

$$u_{1\ell} \geq 0, u_{1h} \geq 0, w_{1\ell} \geq 0, w_{1h} \geq 0, u_{0\ell} \geq 0.$$  \hspace{1cm} (3.5)

Naturally, this nonlinear programming problem can have different solutions depending on parameter values. First it can be show that for any solution where $u_{1h} > 0$, it is impossible for this solution to also have $\Delta u_{1\ell} \neq 0$ or $\Delta u_{hh} \neq 0$ or both of them together. A solution with $u_{1h} = 0$ is possible only if the incentive constraint is not binding. In this model $u_{1h} = 0$ either means the skilled individual has had is income entirely taxed or $w_{1h} = a_h$ and there are no skilled sector jobs. The later case is never optimal. The former case means that the skilled worker can be entirely taxed and the difference in expected utility he gets from matching is enough to make the worker-firm pair unwilling to take the unskilled worker-firm’s bundle. For this to be true, $\Delta u_{1\ell}$ needs to be very negative, $\Delta u_{hh}$ to be very positive and $\Delta u_{hh} = w_{2h} - w_{2\ell}$. This means that there is now severe taxation on the group of unskilled individuals that has worked in both period and to give it to both skilled groups working in the second period. If there is a big enough difference in $w_{2h} - w_{2\ell}$ and or a high enough number of skilled individuals, this solution is not possible and the more standard solution with a binding incentive constraint is the only case possible. From this point on, we will only consider the solution under which (3.2) is binding. This implies that the tax function, although history-dependent, has no insurance component.

To determine characteristics of the optimal first period wages in the imperfect information case we need to look at the following necessary conditions:

$$w_{2h} - w_{2\ell} - \Delta u_{hh}^* = 1 - \lambda_1 \frac{F(\Delta u_{hh}^*)}{f(\Delta u_{hh}^*)},$$ \hspace{1cm} (3.6)

$$- \frac{\partial L_{1\ell}(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{w_{1\ell}^*}{L_{1\ell}(w_{1\ell}^*)} = 1 - \frac{\mu u_{1\ell}}{\lambda_1 \pi_\ell} \frac{L_h(w_{1h}^*)}{L_{1\ell}(w_{1\ell}^*)} \left[ \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{1}{L_{1\ell}(w_{1\ell}^*)} - \frac{\partial L_{1\ell}(w_{1\ell}^*)}{\partial w_{1\ell}} L_{1\ell}(w_{1\ell}^*) \right],$$  \hspace{1cm} (3.7)

$$- \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{w_{1h}^*}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right\}. \hspace{1cm} (3.8)$$

From the FOC of $u_{1\ell}$ and $u_{1h}$, we can get the following expression for the lagrange multiplier of the first period budget constraint:

$$0 < \lambda_1 = \frac{L_{1\ell}(w_{1\ell}^*)}{\pi_\ell L_{1\ell}(w_{1\ell}^*) + \pi_h L_h(w_{1h}^*)} < 1.$$
Condition (3.6) and (3.7) together imply:

**Proposition 1a.** Under imperfect information in the first period, the optimal tax system has the following properties: there is redistribution of income in the second period from workers in the skilled sector to workers in the unskilled sector and the wage of the unskilled workers in the first period is distorted below the efficient level. Thus in the first period, marginal tax rates for the unskilled is greater than the average tax rate, i.e. \( T'_1(w_{1\ell}) > \frac{T_1(w_{1\ell})}{w_{1\ell}} \).

The first part of the above proposition uses condition (3.6) and the fact that \( \Delta u_{\ell\ell} = 0 \) and \( \Delta u_{h\ell} = 0 \). The later implies that \( u_{\ell\ell}^* = u_{h\ell}^* = u_{0\ell}^* \equiv w_{2\ell} \), in terms of taxes this implies \( T_2(w_{1\ell}, w_{2\ell}) = T_2(w_{1h}, w_{2\ell}) = T_2(0, w_{2\ell}) = T_2(w_{2\ell}) \), and thus \( \Delta u_{hh} = u_{hh} - u_{2\ell} \). Since the right-hand side of condition (3.6) is positive, this means that there is redistribution from those that work in the skilled sector in the second period to those that work in the unskilled sector, furthermore, we can see that \( \Delta u_{hh} < \Delta u_{h\ell} = w_{2h} - w_{2\ell} \).

The second part of the proposition uses condition (3.7) and the conditions derived in the perfect information case. For the unskilled worker’s wage to be distorted, the right-hand side of (3.7) must be smaller than 1. This is true only if

\[
\frac{\mu u_{1\ell} L_h(w_{1\ell}^*)}{\lambda_1 \pi_{1\ell} [L_{\ell}(w_{1\ell}^*)]^2} \left[ \frac{\partial L_h(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell}^*)} - \frac{\partial L_{\ell}(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_{\ell}(w_{1\ell}^*)} \right] > 0.
\]

Notice that the term multiplying what is inside the big square brackets is positive. The only thing left to know is if what is inside the brackets is positive. From the assumptions made on \( L_i(\cdot) \), i.e. that \( \frac{\partial L(a, w)}{w} \frac{L(a, w)}{\partial a} \) is increasing in \( a \), it is straightforward that what is inside the bracket is positive. The idea behind this result is to make it less appealing for the skilled worker-firm pair to mimic the unskilled worker-firm pair. This happens because the skilled worker-firm pair using the ‘wage-setting objective’ as preferences would at the margin prefer a higher level of \( w \). In addition, note that redistribution to the unskilled partly goes through an increase in the probability of matching.

From (3.7), we know that \( -\frac{\partial L_{\ell}(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{w_{1\ell}^*}{L_{\ell}(w_{1\ell}^*)} < 1 \). Using this result and (2.5), the tax system that can decentralize this allocation will have this feature:

\[
\frac{1 - T'_1(w_{1\ell})}{1 - T_1(w_{1\ell})} < 1 \quad \Rightarrow \quad \frac{T_1(w_{1\ell})}{w_{1\ell}} < T'_1(w_{1\ell}).
\]

Because there is redistribution in the first period and the budget constraint must balance, we have that \( T_1(w_{1\ell}) < 0 \).

The necessary condition (3.8) determines some characteristics of the optimal wage of the skilled worker. The terms inside the brackets on the right-hand side are the undiscounted marginal gains to society of increasing the probability of skilled individuals of matching in
the first period. The first term in this bracket is $\Psi_h^*$ which is the marginal gain to a skilled worker of matching, note that since $\Delta u_{hh}^* < \Delta u_{h}^{LF}$, this implies $\Psi_h^* < \Psi_h^{LF}$. The second term is the social marginal welfare weighted fiscal gain in the second period from increasing the matching probability in the second period. In the Laissez-faire case this fiscal gain does not exist because there is no redistribution, and in the perfect information case this value is null since the optimal tax system requires $\Delta u_{hh}^* = w_{2h} - w_{2L}$, therefore there are no fiscal gains from sending more skilled workers in the second period.

To figure out if the skilled worker’s wage is distorted, the value of the marginal gains to society of increasing the probability of skilled individuals of matching in the first period with the value of $\Psi_{h}^{LF}$ must be compared. To see this, first subtract both sides by

$$\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{\beta \Psi_{h}^{LF}}{L_h(w_{1h})},$$

which gives

$$- \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{[w_{1h}^* + \beta \Psi_{h}^{LF}]}{L_h(w_{1h})} = \frac{\beta}{L_h(w_{1h})} \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \left\{ \frac{\Psi_{h}^* + F(\Delta u_{hh}^*) [w_{2h} - w_{2L} - \Delta u_{hh}^*]}{\lambda_1} - \Psi_{h}^{LF} \right\}.$$

(3.9)

To know if the wage of the skilled workers is distorted below or above the efficient value we need to determine if the right-hand side of the above equation is below or above 1, or more precisely if the second term of the right-hand side is negative or positive. It is straightforward to see that what is multiplying the terms inside the brackets is negative, thus to know if the right-hand side of (3.9) is below or above 1, the sign of what is inside the brackets must be determined, i.e.

$$\left[ \Psi_{h}^* + F(\Delta u_{hh}^*) [w_{2h} - w_{2L} - \Delta u_{hh}^*] \frac{1}{\lambda_1} - \Psi_{h}^{LF} \right] \geq 0 \implies w_{1h}^* \leq w_{1h}^{LF}.$$

Under the present general assumptions on the problem it is very difficult to say much more. To get further results I must make some assumptions on the distribution function $F(\cdot)$.

**Proposition 1b.** Under imperfect information in the first period, Maximin social welfare function and assuming that $\alpha_h$ is distributed according to $F(\alpha_h) = \left( \frac{\alpha_h}{\alpha} \right)^\varepsilon$ with $\varepsilon > 0$, the optimal tax system distort the wages of the skilled workers below the efficient level. Thus employment of the skilled individuals is higher in the imperfect information case than in the perfect information (or Laissez-faire) case.

The proof of this proposition is in Appendix B.

This results comes about for two reasons: the informational problem and the human capital destruction. The information problem in the first period forces the planner to rely more on the second period to redistribute income. Alone this fact would not result in a downward
distortion of the skilled worker’s wage because there is no gains in doing so. When there is
no human capital destruction, any policy affecting the second period will not have an affect
on how wages are determined. Furthermore, the policies in the first period can’t influence
the distribution of workers in the second period. In addition, there are no informational
gains from distorting the wage of the skilled worker as is standard in the principal-agent
literature, therefore there are no reasons to distort the skilled worker’s wage when there is
no human capital destruction.

When human capital destruction is introduced, redistributive policies in the second period
have an impact on the first period wage and thus the employment levels in the first period.
And policies in the first period have an impact on how much redistribution can happen in
the second period. By distorting the skilled worker’s wage downward, the planner does four
things. He counteracts the upward effect on the skilled worker’s wage caused by increasing
redistribution in the second period. He increases the likelihood of being skilled in the second
period and thus making mimicking less appealing. He creates more employment of the skilled
insuring that there will be more of them to split this new redistributive burden, which at
the same time permits him to make it smaller. Finally, he also insures that there are less
people to redistribute to.

Using (2.5) to decentralize the result implies:

\[
1 - \frac{T'_1(w_{1h}) + \beta \Psi'_h}{1 - \left[ \frac{T_1(w_{1h}) - \beta \Psi_h}{w_{1h}} \right]} < 1 \iff T'_1(w_{1h}) - \beta \Psi'_h > \frac{T_1(w_{1h})}{w_{1h}} - \beta \frac{\Psi_h}{w_{1h}}.
\]

From the last inequality, it is difficult to determine exactly if the marginal tax rate is higher
or below the average tax rate. But this inequality says that the greater \( \Psi_h \), the lower
the marginal tax rate or the greater \( \Psi'^*_h \) can be relative to the average tax rate, on the
contrary the more redistribution there is in the second period, the higher the ratio between
the marginal tax rate and the average tax rate must be.
3.2 Bergson-Samuelson Social Welfare Function

In this subsection I will consider the problem with the more general social welfare function. With the Bergson-Samuelson social-welfare function, the planner’s problem is

\[
\begin{align*}
\max_{\{w_{1\ell},w_{1h}\},\psi} & \quad \pi_\ell W(u_{1\ell} + \beta L_\ell(w_{1\ell}) \Delta u_{1\ell} + \beta u_{1\ell}) + \pi_h W(u_{1h} + \beta L_h(w_{1h}) \Psi_h(\Delta u_{1h}, \Delta u_{hh}) + \beta u_{1h}) \\
\text{s.t.} & \quad (\lambda_1) \quad \pi_\ell \cdot [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0, \\
& \quad (\lambda_2) \quad \phi_\ell(w_{1\ell}) w_{2\ell} + \phi_h(w_{1h}) [w_{2\ell} - \Delta u_{1\ell}] \\
& \quad \quad + \phi_h(w_{1h}, \Delta u_{hh}) [w_{2h} - \Delta u_{1h}] + \phi_hh(w_{1h}, \Delta u_{hh}) [w_{2h} - \Delta u_{1h} - \Delta u_{hh}] = u_{0\ell}, \\
& \quad (\mu) \quad u_{1h} + \beta L_h(w_{1h}) \Psi_h(\Delta u_{1h}, \Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} u_{1\ell} + \beta L_h(w_{1\ell}) \Delta u_{1\ell}.
\end{align*}
\]

As above, I look at the case where (3.2) is binding in the optimum and also do not look at the case where \(u^*_{1h} = 0\), \(\Delta u_{1\ell} \neq 0\), \(\Delta u_{1h} \neq 0\) is a possible solution. From the FOC of the above problem, the solution is characterized by the following necessary conditions\(^{16}\):

\[
\begin{align*}
& \quad w_{2h} - w_{2\ell} - \Delta u_{hh}^* = \frac{\mu \Omega}{\lambda_1 + \mu \Omega} F(\Delta u_{hh}^*), \\
& \quad \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{w_{1\ell}^*}{L_\ell(w_{1\ell}^*)} = 1 + \frac{\mu u_{1\ell}}{\lambda_1 \pi_\ell [L_\ell(w_{1\ell}^*)]^2} \left[ \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} - \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell}^*)} \right], \\
& \quad \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{w_{1h}^*}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \frac{\Psi_h + \frac{\lambda_1 + \mu \Omega}{\lambda_1} F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right\},
\end{align*}
\]

\[(3.11)\]

\[\quad \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{w_{1\ell}^*}{L_\ell(w_{1\ell}^*)} = 1 - \frac{\mu u_{1\ell}}{\lambda_1 \pi_\ell [L_\ell(w_{1\ell}^*)]^2} \left[ \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} - \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell}^*)} \right], \quad (3.12)\]

\[\quad \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{w_{1h}^*}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \Psi_h + \frac{\lambda_1 + \mu \Omega}{\lambda_1} F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*] \right\}, \quad (3.13)\]

where \(\Omega = \left[ \frac{L_h(w_{1\ell}^*)}{L_\ell(w_{1\ell}^*)} - 1 \right] > 0\) since \(L_h(w_{1\ell}^*) > L_\ell(w_{1\ell}^*)\).\(^{17}\) Before going into the analysis of the above conditions, note that from the FOC of \(u_{1\ell}\) and \(u_{1h}\) you get that \(\pi_\ell W'(v_\ell^*) + \pi_h W'(v_h^*) = \lambda_1 + \mu \Omega\).\(^{18}\)

Condition (3.11) determines that there is redistribution in the second period, i.e. \(\Delta u_{hh}^* < \Delta u_{hh}^{LF}\). From (3.12), which is identical to the condition from the above subsection, the result that the unskilled worker’s wage in the first period is distorted downward follows. Also as above, to know if the wage of the skilled workers in the first period is distorted below or above the efficient level, the value of the marginal gains to society of increasing the probability of skilled individuals of matching in the first period with the value of \(\Psi_h^{LF}\) must be compared:

\[\left\{ \Psi_h + \frac{\lambda_1 + \mu \Omega}{\lambda_1} F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*] \right\} - \Psi_h^{LF} \geq 0.\]

\(^{16}\)See Appendix B for more details.

\(^{17}\)See Appendix C for more details.

\(^{18}\)From the FOC of \(u_{1\ell}\) and \(u_{1h}\) and the fact that (3.2) is binding, we can show that \(W'(v_\ell) > W'(v_h), W'(v_l) > \lambda_1 > W'(v_h)\) and both \(\lambda_1 > 0, \mu > 0\).
First, notice that the social marginal welfare weight is different in the Bergson-Samuelson case, it is \( \frac{\lambda_1 + \mu \Omega}{\lambda_1} \) and it is the dollar equivalent value of giving an extra dollar to each individual in the economy in the first period.\(^{19}\)

**Proposition 2.** *Under imperfect information in the first period and assuming that \( \alpha_h \) is uniformly distributed between 0 and \( \bar{\alpha} > 0 \), the optimal tax system distort the wages of the skilled workers below the efficient level. Thus employment of the skilled individuals is higher in the imperfect information case than in the perfect information (or Laissez-faire) case.*

The proof of Proposition 2. is located in Appendix C. The restriction to the distribution function \( F(\cdot) \) is necessary to derive analytical results when a more general social welfare function is considered. The intuition for this downward pressure on the skilled worker’s wage is identical to the Maximin case.

To see if the downward distortion of the skilled worker’s wage in the first period result holds for the more general distribution function used in the Maximin subsection, i.e. \( F(\alpha_h) = \left( \frac{\alpha_h}{\bar{\alpha}} \right)^{\varepsilon} \), I turn to numerical simulations.\(^{20}\) For these simulations, I use a CES matching function and use for the social welfare function the following isoelastic function:

\[
SWF = \pi_l \left( \frac{u_l}{\sigma} \right)^\sigma + \pi_h \left( \frac{u_h}{\sigma} \right)^\sigma,
\]

where \( \sigma \) is the parameter that determines inequality aversion. The higher it is, the less inequality aversion the planner has. The two extremes being \( \sigma \to -\infty \) which are the Maximin preferences and \( \sigma = 1 \) which are the standard utilitarian preferences featuring no inequality aversion. The downward distortion result is a feature of all the solutions verified where the incentive constraint is binding and that \( \sigma < 1 \).\(^{21}\) Examples of simulations results are reported in Table 1.

For ease of exposition only the percentage change from the laissez-faire wage level is reported and not the actual wage levels.\(^{22}\) Increases in \( \varepsilon \) imply that the reaction to changes in \( \delta u_{hh} \) will be greater, or it can be interpreted as the likelihood of deciding to work in the skilled sector in the second period is reduced. One thing that is clear from Table 1 is that the wage of the skilled in the first period is always distorted downward like in the Maximin case. As it would be expected an increase in \( \varepsilon \) leads to a reduction in redistribution in the second period. This results in an increase distortion of the skilled worker’s wage. Although, less redistribution makes the second period revenue collected lower, the higher value of the social marginal welfare weight caused by the lower utility of both type makes it more valuable.

---

\(^{19}\)Notice that in the Maximin case, \( \lambda_1 + \mu \Omega = 1 \), since the planner only cares about the expected lifetime utility of the unskilled worker.

\(^{20}\)The details of the functional form and parameter values assumptions can be found in Appendix D alongside some sensitivity analysis using the Maximin social welfare function.

\(^{21}\)When \( \sigma = 1 \) due to the risk-neutral preferences there is no longer any redistributive motive.

\(^{22}\)%\(\Delta w^*_i = 100 \ast (w^*_i - w^{LF}_i)/w^{LF}_i\)
Table 1: Characteristics of the Optimal Allocation and Tax System: Changes in $\varepsilon$

<table>
<thead>
<tr>
<th></th>
<th>Isoelastic ($\sigma \to 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varepsilon = 0.1$</td>
</tr>
<tr>
<td>$L^*_l$</td>
<td>0.8832</td>
</tr>
<tr>
<td>$L^*_h$</td>
<td>0.9384</td>
</tr>
<tr>
<td>$%\Delta w^*_l$</td>
<td>-2.17</td>
</tr>
<tr>
<td>$%\Delta w^*_h$</td>
<td>-4.11</td>
</tr>
<tr>
<td>$T_1(w_{1l})$</td>
<td>-2.2648</td>
</tr>
<tr>
<td>$T_1'(w_{1l})$</td>
<td>-1.8027</td>
</tr>
<tr>
<td>$T_1(w_{1h})$</td>
<td>2.1062</td>
</tr>
<tr>
<td>$T_1'(w_{1h}) - \beta \Psi'_h$</td>
<td>0.564</td>
</tr>
<tr>
<td>$u_{0l}$</td>
<td>1.0266</td>
</tr>
<tr>
<td>$\Delta u^*_{hh}$</td>
<td>1.9606</td>
</tr>
<tr>
<td>$v^*_l$</td>
<td>3.5066</td>
</tr>
<tr>
<td>$v^*_h$</td>
<td>3.8479</td>
</tr>
<tr>
<td>$v_h/v_l$</td>
<td>1.0973</td>
</tr>
</tbody>
</table>

Also, an increase in $\varepsilon$ also radically reduces the value of $\Psi_h$ which pushes wage upwards, but since the planner still wants to separate types, he wants to increase the likelihood of the skilled to keep his human capital, although not reported, in the last case there is a difference of 10% employment between the Laissez-faire outcome and the second best policy. Following this logic, note that as $\varepsilon$ increases the planner relies less on the distortion of the unskilled worker’s wage in the first period to achieve his goals.\(^{23}\)

Using the results of Table 2 to investigate the impact of changing the redistributive tastes of the planner, it is possible to see that as they are lowered, the downward distortion on the skilled workers isn’t much affected.\(^{24}\) This is true despite the fact that redistribution in the second period also goes down, almost null in the last case. Thus the planner lowers the wedge, i.e. $T_1'(w_{1h}) - \beta \Psi'_h$, to compensate and keep the wage at around the same level. The important distortion in these four examples is the unskilled worker’s wage. The high level of distortion in the Maximin case is analog to the one found in the standard optimal tax literature, which wants to heavily discourage the skilled to mimic the unskilled. Here it has the added benefit of increasing the employment of the unskilled. But as the redistributive tastes go down, so does the distortion on that labor market. Interestingly, the transfer in the first period to the unskilled stay steady despite the reduction in redistributive tastes. Because overall redistribution is going down, it is less interesting for the skilled worker-firm pair to mimic the unskilled market, this allows the planner to lower the marginal tax rates of the unskilled to get closer to the efficient level of output in this labor market.

\(^{23}\) This result holds for other redistributive tastes, for example Maximin reported in Table 5.

\(^{24}\) Here we consider the case when $\varepsilon = 0.25$. Also, changes in the parameter determining the search frictions both have a strong impact on the downward distortion of the skilled wage. See Table 5.
Table 2: Characteristics of the Optimal Allocation and Tax System: Various Inequality Aversion

<table>
<thead>
<tr>
<th></th>
<th>Maximin</th>
<th>Isoelastic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma \rightarrow 0$</td>
<td>$\sigma = 0.25$</td>
</tr>
<tr>
<td>$L^*_l$</td>
<td>0.969</td>
<td>0.8825</td>
</tr>
<tr>
<td>$L^*_h$</td>
<td>0.9361</td>
<td>0.9361</td>
</tr>
<tr>
<td>$%\Delta w^*_l$</td>
<td>-35.36</td>
<td>-2.06</td>
</tr>
<tr>
<td>$%\Delta w^*_h$</td>
<td>-4.34</td>
<td>-4.34</td>
</tr>
<tr>
<td>$T_1(w_{1l})$</td>
<td>-2.1748</td>
<td>-2.0996</td>
</tr>
<tr>
<td>$T_1'(w_{1l})$</td>
<td>0.6966</td>
<td>-1.6777</td>
</tr>
<tr>
<td>$T_1'(w_{1h})$</td>
<td>1.8178</td>
<td>1.8284</td>
</tr>
<tr>
<td>$T_1'(w_{1h}) - \beta \Psi'_h$</td>
<td>0.5452</td>
<td>0.5219</td>
</tr>
<tr>
<td>$u_0l$</td>
<td>1.0969</td>
<td>1.011</td>
</tr>
<tr>
<td>$\Delta u^*_{lh}$</td>
<td>1.8440</td>
<td>1.9827</td>
</tr>
<tr>
<td>$u^*_l$</td>
<td>3.4875</td>
<td>3.3465</td>
</tr>
<tr>
<td>$u^*_h$</td>
<td>3.5621</td>
<td>3.6715</td>
</tr>
<tr>
<td>$u^<em>_h / u^</em>_l$</td>
<td>1.0214</td>
<td>1.0971</td>
</tr>
</tbody>
</table>

4 Alternative Policy Tools

4.1 Training Programs

In this subsection I consider the case when the planner has access to a training technology. At the end of the first period, the planner is able to retrain a fraction $\rho$ of the skilled workers that did not obtain a job. He is able to do this at cost $c(z)$ where $z = \rho \pi_h [1 - L_h(w_{1h})]$ is the number of workers being retrained and the cost function is convex in $z$, i.e. $c'(z) > 0, c''(z) > 0$, and satisfies $c(0) = 0$. Unskilled workers in this particular exercise cannot be trained to be skilled workers in the second period. This assumption is made to focus on the destruction of human capital due to unemployment and not human capital acquisition. However, the assumption could be justified by assuming some underlying characteristic of the unskilled workers that makes training them to attain the skill level of the skilled workers prohibitively costly.

A training technology will have an impact on the potential labor market history of a skilled worker, which can now be split into three events in the first period. The first being that he found a job and worked, the second is that he did not find a job but got retrained and the last is that he did not find a job and was not retrained. These different labor market experience will modify the expected lifetime utility of the skilled worker defined in section 2, and the budget constraints and the incentive constraint faced by the planner in the imperfect information scenario. Supposing that the fraction of the skilled worker that will be retrain his assigned randomly to the skilled unemployed, the expected lifetime utility of the skilled
necessary conditions:

\[ v_h = L_h(w_{1h})[w_{1h} - T_1(w_{1h}) + \beta v_{hh}] + [1 - L_h(w_{1h})] \beta \rho v_{hh} + (1 - \rho)u_{0\ell}] \]

\[ = u_{1h} + L_h(w_{1h})\beta(1 - \rho)\Psi_h + \beta[\rho\Psi_h + u_{0\ell}] \]

For the budget constraints, not only is there an additional cost in the first period but there is also a modification on how policy affects the number of individual working in the unskilled and skilled market in the second period. Using the same definition of the four groups of workers in the second period used above based on labor market history, we take into account the training program and modify the equations determining the numbers of workers in each group:

\[ \phi_{hh} = \pi_h \{ L_h(w_{1h}) + \rho[1 - L_h(w_{1h})] \} F(\Delta u_{hh}), \]

\[ \phi_h = \pi_h \{ L_h(w_{1h}) + \rho[1 - L_h(w_{1h})] \} [1 - F(\Delta u_{hh})], \]

\[ \phi_{ll} = \pi_\ell L_\ell(w_\ell), \]

\[ \phi_{ll} = \pi_\ell [1 - L_\ell(w_\ell)] + \pi_h(1 - \rho)[1 - L_h(w_{1h})]. \]

From this the budget constraint of the second period can be written to be:

\[ \phi_{0\ell}w_{2\ell} + \phi_{ll}[w_{2\ell} - \Delta u_{\ell}] + \phi_{hl}[w_{2\ell} - \Delta u_{hl}] + \phi_{hh}[w_{2h} - \Delta u_{h\ell} - \Delta u_{hh}] = u_{0\ell}. \]

### 4.1.1 Perfect Information in the First Period

Writing the planner’s problem when he is able to observe the skill level of each match in the first period, we have:

\[
\max_{\{w_{1\ell}, u_{1\ell}\} \in_l, u_{0\ell}} \pi_\ell W(v_\ell(u_{1\ell}, w_{1\ell}, \Delta u_{\ell}, u_{0\ell})) + \pi_h W(v_h(u_{1h}, w_{1h}, \rho, \Delta u_{h\ell}, \Delta u_{hh}, u_{0\ell}))
\]

\[ s.t. \quad (\lambda_1) \quad \pi_\ell [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] - c(\rho\pi_h[1 - L_h(w_{1h})]) = 0, \]

\[ (\lambda_2) \quad \phi_{0\ell}(w_{1\ell}, w_{1h}, \rho)w_{2\ell} + \phi_{ll}(w_{1\ell})[w_{2\ell} - \Delta u_{\ell}] + \phi_{hl}(w_{1h}, \Delta u_{hh}, \rho)[w_{2\ell} - \Delta u_{h\ell}] + \phi_{hh}(w_{1h}, \Delta u_{hh}, \rho)[w_{2h} - \Delta u_{h\ell} - \Delta u_{hh}] = u_{0\ell}. \]

Similar to the perfect information case from above, each solution must satisfy the following necessary conditions:

\[ u_{1\ell} + \beta L_\ell(w_{1\ell})\Psi_\ell(\Delta u_{\ell}) = u_{1h} + L_h(w_{1h})\beta(1 - \rho)\Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) + \beta\rho\Psi_h(\Delta u_{h\ell}, \Delta u_{hh}), \quad (4.2) \]

\[ \Delta u_{hh} = w_{2h} - w_{2\ell}, \quad (4.3) \]

\[ \beta\Psi_{LF}^h = c'(\rho\pi_h[1 - L_h(w_{1h})]), \quad (4.4) \]

\[ -\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \cdot \frac{w_{1\ell}}{L_\ell(w_{1\ell})} = 1, \quad (4.5) \]

\[ -\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \cdot \frac{w_{1h}}{L_h(w_{1h})} = \frac{1}{1 + \frac{\beta\Psi_{LF}^h}{w_{1h}}} \quad (4.6) \]
From the first condition we see that expected lifetime utilities of both types of workers are equalized. Condition (4.3) which helps derive the rest of the other conditions is similar to the one found above, i.e. there is no difference in tax burden of the skilled workers irrelevant of their occupational choice in the second period. This feature of the solution implies that there is no fiscal gains of having the number of skilled workers in the second period above the efficient value. This has a direct impact on (4.4), since increasing $\rho$ will increase the expected lifetime utility of the skilled but also the number of skilled workers in the second period. Since there are no fiscal gains to having more skilled workers than is efficient, the efficient level of $\rho$ will be set when the marginal cost of increasing $\rho$ will equal the discounted value of the expected net utility gain of being skilled in the second period.

From conditions (4.5) and (4.6) we can see that in the perfect information scenario, adding the training technology has no impact on the optimal wages of each skill level, they are set at their laissez-faire level. In this context, for the wage of the unskilled worker the result isn’t surprising since $\rho$ does not directly enter the lifetime expected utility of an unskilled worker. The same result for the skilled worker is less straightforward, decreasing the skilled worker’s first-period wage increases the number of skilled workers that will find a job and keep their human capital, but the effect of this wage decrease on the future is mitigated by the presence of the training technology, since there will be a fraction of those that did not find a job that will be retrained regardless of the market wage. As for the budget constraints, the standard effect on total production is still there, but the interesting part is that a decrease in the wage will lower the cost of retraining for a given $\rho$ since there are now less skilled unemployed to retrain, however this effect isn’t one for one since only a fraction $\rho$ will be retrained. And since there are no fiscal gains in the second period there is no value for that period to lower the wage below the efficient value. Taking all the effects into account the first order condition with respect to the wage of the skilled in the first period can be written in the following way:

$$\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{[w_{1h} + \beta \Psi^{LF}_h]}{L_h(w_{1h})} = 1 + \frac{1}{L_h(w_{1h})} \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \left[ \rho c'(\cdot) - \rho \beta \Psi^{LF}_h \right].$$

Looking on the right-hand side we see that the benefit of lowering the wage on the cost of training but we can also see the lowered impact of manipulating the wage caused by $\rho$. Condition (4.4) says that both of these effects are equal and thus cancel each other out which results in (4.6).

### 4.1.2 Imperfect Information in the Second Period

To keep in line with the results derived in section 3, I only consider the optimal tax system that does not offer insurance, i.e. $\Delta u_{\ell\ell} = \Delta u_{h\ell} = 0$, in concert with the presence of the training technology. By the same logic, the planner will not use the information gathered while training the unemployed skilled workers and offer them compensation in the second period.

\footnote{26}{Here the lagrange multiplier has been solved out.}
The incentive constraint of the planner under imperfect information is then:

First part being the one found in the incentive constraint of the problem without a training welfare function. The problem of the planner is:

Equilibrium example of mimicking, the expected lifetime utility of a skilled worker that since it is costless for him to do so he will choose to do it. Recalling the Competitive Search to mimic in the first period. In that situation the mimicker is able to be retrained, and mimicked did not find a job, there is no way for the planner to know that this worker tried to mimic in the first period. In that situation the mimicker is able to be retrained, and since it is costless for him to do so he will choose to do it. Recalling the Competitive Search Equilibrium example of mimicking, the expected lifetime utility of a skilled worker that searches in a market with the unskilled worker’s wage prevailing is:

\[
\hat{\nu}_h = L_h(w_{1\ell})[w_{1\ell} - T_1(w_{1\ell}) + \beta u_{0\ell}] + [1 - L_h(w_{1\ell})] \beta (\rho \nu_{hh} + (1 - \rho)u_{0\ell}],
\]

\[
= L_h(w_{1\ell})[w_{1\ell} - T_1(w_{1\ell}) + \beta (u_{0\ell} - \rho \nu_{hh} + (1 - \rho)u_{0\ell})] + \beta [\rho \Psi_h + u_{0\ell}],
\]

\[
= L_h(w_{1\ell})[w_{1\ell} - T_1(w_{1\ell}) - \beta \rho \Psi_h] + \beta [\rho \Psi_h + u_{0\ell}].
\]

The incentive constraint of the planner under imperfect information is then:

\[
u_{1h} + \beta L_h(w_{1h})(1 - \rho)\Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_h(w_{1\ell})} u_{1\ell} - L_h(w_{1\ell}) \beta \rho \Psi_h(\Delta u_{hh}).
\]

For simplicity, the analysis of the training program is done using the Maximin social welfare function. The problem of the planner is:

\[
\max_{\{w_{1\ell}, u_{1\ell}\}} \pi_t \cdot [L_t(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] - c(\rho \pi_h[1 - L_h(w_{1h})]) = 0,
\]

\[
\rho \pi_t \cdot [L_t(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \rho \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] - c(\rho \pi_h[1 - L_h(w_{1h})]) = 0,
\]

\[
u_{1h} + \beta L_h(w_{1h})(1 - \rho)\Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_h(w_{1\ell})} u_{1\ell} - L_h(w_{1\ell}) \beta \rho \Psi_h(\Delta u_{hh}).
\]

Before stating the necessary conditions of the optimal solution, rewriting the incentive constraint is instructive on how introducing a training program alters the other policy choices’ effect on the incentive to mimic:

\[
u_{1h} + \beta L_h(w_{1\ell}) \Psi_h(\Delta u_{hh}) + \rho \beta [L_h(w_{1\ell}) - L_h(w_{1h})] \Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1h})}{L_h(w_{1\ell})} u_{1\ell}.
\]

Looking at the left-hand side, we can see that it is composed of three parts with the two first part being the one found in the incentive constraint of the problem without a training
program. The right-hand side is also the same as in the above section with the no-insurance tax system. The difference now comes from the third part on the left-hand side which combines the expected value of not finding a job for a ‘mimicker’ and the reduction in value of the matching surplus caused by the training program. Under reasonable parameter values we would expect that $w_{1h} > w_{1\ell}$, which would tell us that this third part has a positive value since $L_h(w_{1\ell}) > L_h(w_{1h})$.

It is possible to see that increasing $\rho$ will slacken the incentive constraint and this effect will be greater the less redistribution there is in the second period and the bigger the difference between the labor demand for the skilled worker at the different wage levels. A reduction in redistribution, i.e. $\Delta u_{hh}$, also slackens the incentive constraint, it does so in the same manner as before by increasing the potential payoff of being skilled in the second period if one chooses the market with the skilled worker’s wage. But the presence of the training program increases the effect of a reduction in redistribution since it makes the unskilled worker’s bundle and actually finding a job at the unskilled’s wage, which implies that there is no possibility of getting the skilled worker’s income in the second period, less appealing. Another effect is the one on the unskilled worker’s wage, it slackens the incentive constraint in an additional way. The lower the unskilled’s wage is the higher the probability of a mimicking skilled worker to find a job and thus not have access to retraining. The effect is reversed when we consider the skilled worker’s wage. Although lowering the skilled worker’s wage increases the probability of finding a job and keeping the human capital of the worker, this effect on the left-hand side is mitigated, as mentioned in the perfect information case, by the training program. For the skilled worker, matching is no longer the only way to keep his human capital, he is less willing to take a lower wage since this technology reduces the surplus of matching. In the end the total effect of the skilled worker’s wage on the incentives is difficult to evaluate from the above equation, but there is now this upward effect on the wage because training lowers the expected surplus of matching.

As it turns out the effect of $\rho$ on the optimal allocation will strongly depend on the skilled labor demand at the first period unskilled wage $w_{1\ell}$, i.e. $L_h(w_{1\ell})$. First start with the first-order condition\(^{27}\) for the decision of $\Delta u_{hh}$, $\rho$ and $w_{1\ell}$:

\[
[w_{2h} - w_{2\ell} - \Delta u_{hh}^*] = (1 - \lambda_1) \frac{F(\Delta u_{hh}^*)}{f(\Delta u_{hh}^*)}, \tag{4.7}
\]

\[
\beta \left[ \frac{L_h(w_{1\ell}) - L_h(w_{1h})}{1 - L_h(w_{1h})} \Psi_h^{\star} + \frac{F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right] = c'(\rho \pi_h[1 - L_h(w_{1h})]), \tag{4.8}
\]

\[
- \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{w_{1\ell}}{L_\ell(w_{1\ell})} = 1 - \frac{\mu}{\lambda_1 \pi_\ell} \left\{ \frac{L_h(w_{1\ell}^*) u_{1\ell}}{L_\ell(w_{1\ell}^*)^2} \left[ \frac{\partial L_h(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell}^*)} \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell}^*)} \right] - \beta \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \rho \Psi_h \right\}, \tag{4.9}
\]

\(^{27}\)With both the lagrange multiplier of the second period budget constraint and the one of the incentive constraint, $\mu = \lambda_1 \pi_h$. 31
where
\[ \Upsilon = \frac{L_h(w_{1h})(1 - \rho) + \rho L_h(w_{1\ell})}{L_h(w_{1h})(1 - \rho) + \rho}. \]

Notice that (4.7) is slightly different than (3.11) due to the presence of \( \Upsilon \). \( \Upsilon \) captures the tradeoff between the incentive effect of decreasing redistribution (the numerator), and loss of tax revenue of this reduction in redistribution (the denominator). Assuming again that \( F(\alpha_h) = \left( \frac{\alpha_h}{\bar{\alpha}} \right) \varepsilon \), then \( d\Delta u_{hh}/d\Upsilon > 0 \), this implies that there is less redistribution the greater \( \Upsilon \) is. This is where the importance of \( L_h(w_{1\ell}) \) comes in. If \( w_{1\ell} \) is high enough and \( L_h(w_{1\ell}) < 1 \), then \( \Upsilon \) must be smaller than 1 which hints at more redistribution than without a training technology. Although without being able to compare the values of \( \lambda_1 \) for both cases, it is hard to tell. But what it says is that the effect on the budget of the training program is greater than the incentive effect. But if \( w_{1\ell} \) is low enough so that \( L_h(w_{1\ell}) \) tends to or equals 1, then both effects are equal and we recover (3.6).

The tradeoff between the budget effects and the incentive effects are also present in (4.8) which is the first-order condition of \( \rho \). The left-hand side is the gains in expected utility and fiscal revenues mixed with the incentive effect of increasing the number of skilled workers in the second period. The right-hand side is the marginal cost of increase \( \rho \). The left-hand side looks similar to gains brought by reducing the number of skilled workers by lowering the skilled worker’s wage, but there is a positive term multiplying \( \Psi_h \). This term, like \( \Upsilon \), measures in part the tradeoff between the budget concerns (denominator) and the incentive concerns (numerator). Since by marginally increasing \( \rho \) the fiscal benefit is multiplied by number of potential skilled workers, i.e. \( 1 - L_h(w_{1h}) \), but this is also true for marginal cost. What is left is the incentive effect captured by \( L_h(w_{1\ell}) - L_h(w_{1h}) \). Again, the value of \( L_h(w_{1\ell}) \) becomes important in determining if \( \rho \) will be higher than in the perfect information case. If we start with the case when \( L_h(w_{1\ell}) < 1 \), this means that the term multiplying \( \Psi_h \) is below 1, which reduces the weight of the value of keeping the level of human capital. Because of this, it is quite possible that the left-hand side of (4.8) is smaller in value than \( \beta \Psi^{LF}_h \), which means that \( \rho \) can be smaller than in the perfect information case. Alternatively the more \( L_h(w_{1h}) \) approaches 1, the more likely it is that \( \rho \) is above the one in perfect information, and with \( L_h(w_{1h}) = 1 \) then with a similar proof as in the above section we know that \( \Psi_h^* + \frac{F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}]}{\lambda_1} > \Psi_h^{LF} \) and thus \( \rho \) in the imperfect information setting is always greater than in the perfect information setting.

Condition (4.9) is similar to the one derived above, in fact it is the same with the added term \( -\beta \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \rho \Psi_h \), which is positive. Thus, there is the difference in elasticities helping to separate the two types, but also the incentive effect brought by the training program through a decrease in wage which increases the probability of matching of a mimicker, making the mimicking bundle less attractive since he must forgo the benefit brought on by keeping his human capital. But one more effect must be taken in consideration. Since \( \rho \) is able to reduce the information problem, the value of the lagrange multiplier related to the incentive constraint (\( \mu \)) can be lower. This pulls the wage towards Laissez-faire level, thus upwards.
The skilled worker’s wage in the first period has many effects, from the fiscal gain, to the incentive effect and also the reduction in cost of the training program. To be able to say more on the optimal level of \(w_{1h}\) under imperfect information, I incorporate (4.8) in the first-order condition of \(w_{1h}\) to get:

\[
- \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{w_{1h}^*}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \left( 1 + \frac{L_h(w_{1\ell}) - 1}{1 - L_h(w_{1h})} \right) \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right\}. \tag{4.10}
\]

The impact of having the training program on the FOC can be seen by looking at the term inside the curly bracket that is multiplying \(\Psi_h\). It incorporates both the negative effect on incentive brought on by the training program but also the positive benefit on lowering the cost of the training program. If \(L_h(w_{1\ell})\) is smaller than one, then the entire term multiplying \(\Psi_h\) is smaller than one. The closer \(L_h(w_{1\ell})\) gets to 1, the more the FOC looks identical to the one without the training program. To determine if the skilled worker’s wage in the first period is distorted below the \textit{Laissez-faire} level, it must determined if what is inside the curly bracket is bigger than \(\Psi_{hL}^*\).

Considering the case when \(L_h(w_{1\ell}) < 1\), i.e. \(\Upsilon < 1\), which implies that redistribution can be quite substantial in the second period and thus lowering \(\Psi_h\), in addition this also makes the term multiplying \(\Psi_h\) smaller than one. Assuming the functional form for the distribution of \(\alpha_h\) above or even the uniform distribution can’t help us rule out the case when

\[
\Psi_{hL}^* > \left( 1 + \frac{L_h(w_{1\ell}) - 1}{1 - L_h(w_{1h})} \right) \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1}.
\]

In this case, the skilled worker’s wage could be pushed above the \textit{Laissez-faire} level. This comes about for several reasons, the first one is that the training technology makes the skilled workers in the first period much less willing to take a lower wage since the surplus from matching is both lower from more redistribution but is also lower since there is always a probability that if he becomes unemployed he will have a chance to regain the human capital lost. Thus distorting the wage below the \textit{Laissez-faire} level becomes quite costly. Since the skilled worker’s wage is higher, it also means that the fraction \(\rho\) affects a larger pool of workers, and thus the program is much more expensive, hence we can better see the tradeoff between having a lower \(\rho\) and a bigger pool of potential candidates to retrain. But if \(w_{1\ell}\) is low enough such that \(L_h(w_{1\ell})\) approaches or is equal to 1, then \(\Upsilon = 1\) and the term multiplying \(\Psi_h\) is also 1, and therefore \(w_{1h}\) is distorted below the \textit{Laissez-faire} level. The tradeoff between wage and \(\rho\) shows up, since \(\rho\) is now bigger than the perfect information case and the pool of skilled unemployed is now lower due to the skilled worker’s wage being below the \textit{Laissez-faire} level. The dependence of the result on \(L_h(w_{1\ell})\) stems from when \(L_h(w_{1\ell}) < 1\) the incentive effects are less important compared to the budget considerations of taxing more people in the second period. When \(L_h(w_{1\ell}) = 1\), this is no longer the case, and the incentive effect dominates.
To untangle all of these effects, I use numerical simulations which the results are summarized in Table 3.\(^{28}\) In the numerical simulations I consider different redistributive tastes.

Table 3: Characteristics of the Optimal Allocation and Tax System: Training Program

<table>
<thead>
<tr>
<th></th>
<th>Maximin</th>
<th>Isoelastic SWF $\sigma \to 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>without</td>
<td>with</td>
</tr>
<tr>
<td>$\rho^*$</td>
<td>-</td>
<td>0.314</td>
</tr>
<tr>
<td>$L^*_l$</td>
<td>0.969</td>
<td>0.9691</td>
</tr>
<tr>
<td>$L^*_h$</td>
<td>0.9361</td>
<td>0.936</td>
</tr>
<tr>
<td>$%\Delta w^*_l$</td>
<td>-35.36</td>
<td>-35.43</td>
</tr>
<tr>
<td>$%\Delta w^*_h$</td>
<td>-4.34</td>
<td>-4.32</td>
</tr>
<tr>
<td>$T_1(w_{1l})$</td>
<td>-2.1748</td>
<td>-2.1787</td>
</tr>
<tr>
<td>$T_1(w_{1h})$</td>
<td>0.6966</td>
<td>0.6972</td>
</tr>
<tr>
<td>$T_1(w_{1h}) - \beta \Psi_h$</td>
<td>1.8178</td>
<td>1.8269</td>
</tr>
<tr>
<td>$u_{0l}$</td>
<td>1.0969</td>
<td>1.1029</td>
</tr>
<tr>
<td>$\Delta u^*_{hh}$</td>
<td>1.8440</td>
<td>1.8377</td>
</tr>
<tr>
<td>$v^*_l$</td>
<td>3.4875</td>
<td>3.4959</td>
</tr>
<tr>
<td>$v^*_h$</td>
<td>3.5621</td>
<td>3.5715</td>
</tr>
<tr>
<td>$v_h/v_l$</td>
<td>1.0214</td>
<td>1.0216</td>
</tr>
</tbody>
</table>

Incorporating the training has only slight effects on the optimal wages and thus the employment levels. In the maximin case, the incentive effect of incorporating $\rho$ dominates the reduction in the value of $\mu$. But this result is overturned when a lower aversion to inequality preference is considered. In both cases the planner attempts much more redistribution in the second period. This should have a positive effect on the skilled worker’s wage, but as we can see the optimal income tax system requires a higher wedge to put a downward pressure to counteract the upward pressure of this increase in redistribution in the second period. Overall, this technology benefits both skill level as their lifetime expected utility is greater and this even if the skilled worker his taxed more.

4.2 Job Creation Subsidies

In this section we consider if adding job creation subsidies offered to any firm that opens a vacancy can be welfare improving and if yes how would it effect the optimal allocation. We only consider the case of a uniform\(^{29}\) subsidy $s \geq 0$ which is given to a firm after they have

---

\(^{28}\)The functional forms and parameter values are the same as those highlighted in Appendix C. The only addition is the cost function of of the training program which is chosen to be the following quadratic function $c(z) = \left(\frac{75}{4}z\right)^2$.

\(^{29}\)If we considered the case where $s$ could vary with the actual cost which would automatically reveal the skill level of each job if that specific firm would meet with a worker and than the information problem would not exist. Although an interesting tool, we leave for further research a more complicated scenario where a
proved they paid a cost of $\kappa_i$.\(^{30}\)

Before going on with the planner’s problem must first determine how labor market tightness and the labor demand reacts to an increase in the subsidy $s$. First, recall that the zero-profit condition helps pin down market tightness, this condition is slightly modified and becomes $m(\theta_i)[a_i - w] = \kappa_i - s$, which means that

$$
\theta_i = m^{-1}(\frac{k_i - s}{a_i - w}) , \quad L_i(w, s) = m^{-1}(\frac{k_i - s}{a_i - w}) \frac{k_i - s}{a_i - w}.
$$

From this we get

$$
\frac{\partial \theta_i}{\partial s} = -\frac{1}{m'(\cdot)} > 0, \quad \text{and} \quad \frac{\partial L_i(w, s)}{\partial s} = -\frac{1}{m'(\cdot)} \frac{k_i - s}{a_i - w} - m^{-1}(\frac{k_i - s}{a_i - w}) \frac{1}{a_i - w}.
$$

The first result is quite natural since reducing the cost of opening up a vacancy through a subsidy would naturally increase the number of firms opening vacancies and increasing market tightness. The effect of increasing $s$ on labor demand isn’t has straightforward and as we can see it is composed of two terms. The first term is positive but the second one is negative. To be able to analyze the incorporation of job creation subsidies, I need to put more structure on the labor demand function. Therefore I assume a CES matching function $M(U_i, V_i) = [U_i^{\gamma} + V_i^{1-\gamma}]^{-\delta}$ which gives the following labor demand for skill level $i$:

$$
L_i(w, s) = \left[1 - \left(\frac{a_i - w}{\kappa_i - s}\right)^{-\delta}\right]^{-\delta},
$$

where $\delta \geq 0$. With the functional form assumption and the standard assumption that $\frac{\partial n(a)}{\partial a} k(a) \leq 1$, the result follows:

$$
\frac{\partial L_i(w, s)}{\partial s} > 0.
$$

Another feature of this functional form is that the elasticity of labor demand with respect to $s$, i.e. $\frac{\partial L_i(w, s)}{\partial s} \frac{s}{L_i(w, s)}$, is decreasing with skill since $\kappa_i$ increases with skill.

The planner’s problem in both informational cases is similar with the exception that $s$ will now enter the labor demand in each market and that revenue must be raised to pay for the subsidy. Since $s$ is going to be given for each vacancy in both markets the total cost of this policy is $(V_i + V_h)s$. Recall that $\theta_i = \frac{V_i}{U_i}$ and that $U_i = \pi_i$, thus we get $V_i = \pi_i \theta_i$. The new budget constraint in the first period faced by the planner is:

$$
\pi_\ell[U_i \ell(w_{1\ell}, s)w_{1\ell} - u_{1\ell} - \theta_\ell(w_{1\ell}, s)s] + \pi_h[L_h(w_{1h}, s)w_{1h} - u_{1h} - \theta_h(w_{1h}, s)s] = 0.
$$

subsidy would not reveal automatically the skill level of a match.

\(^{30}\)Only positive subsidies are considered, negative values of $s$ would be a tax and since firms only last for one period in this model and there are no financial markets, this tax revenue wouldn’t be modeled properly.
Laissez-faire in different labor markets, i.e., \( \Delta u_{\ell} \), planner’s problem are identical to the ones derived in section 2. This means that there is no social welfare of increasing \( s \).

Due to \( \Delta u_{\ell} \), condition in both labor market when \( \ell = 0 \), \( \Delta u_{\ell, s} > 0 \), each skill level. We have the resources gains from more individuals finding jobs in each labor market, including the more production in the second period when there are more skilled workers, but also the revenue loss for each subsidized vacancy. Using the zero-profit condition in both labor market when \( s > 0 \), we have that \( L_{\ell}(w, s) = \Delta u_{\ell} = L_{\ell}(w, s)w - \theta s \) and \( L_{h}(w, s)[a_{h} + \beta \Psi^{LF}] - \theta \kappa_{h} = L_{h}(w, s)w - \theta s \). At \( s = 0 \), both wages are set at their Laissez-faire level, which is the level that maximizes the output net of vacancy costs. Using the envelope theorem this result must be true:

\[
\frac{\partial L_{\ell}(w_{1\ell}^{*}, s)}{\partial s} w_{1\ell}^{*} - \theta_{\ell}(w_{1\ell}^{*}, s) = 0.
\]

From this we can conclude that \( \frac{\partial \ell}{\partial s} |_{s=0} = 0 \), implying that there are no welfare gains from adding a subsidy to employment under perfect information. This result is not surprising since taxation in this context can redistribute until lifetime expected utility is equal.
while at the same time ensuring that the economy produces to its net-output maximizing level. Therefore using the additional and less precise tool of the employment subsidy does not bring any welfare gains.

The perfect information in the first period result does as a rule carry to the imperfect information case because the job creation subsidy can help in separating different skill types. In this subsection I consider the maximin case where it is possible to derive more general properties of the solution. I also only consider the tax system that features no insurance for market outcomes along the working life. The planner’s problem is then:

$$\max \{ w_{1\ell}, u_{0\ell}, s \} \quad u_{1\ell} + \beta u_{0\ell}$$

$$\text{s.t.} \quad (\lambda_1) \quad \pi_\ell \cdot [L_\ell(w_{1\ell}, s) \cdot w_{1\ell} - u_{1\ell} - \theta_\ell(w_{1\ell}, s)] + \pi_h \cdot [L_h(w_{1h}, s) \cdot w_{1h} - u_{1h} - \theta_h(w_{1h}, s)] = 0,$$

$$(\lambda_2) \quad \phi_{0\ell}(w_{1\ell}, w_{1h}, s)w_{2\ell} + \phi_{1\ell}(w_{1\ell}, s)w_{2\ell} + \phi_h(w_{1h}, \Delta u_{hh}, s)w_{2\ell} = 0,$$

$$(\mu) \quad u_{1h} + L_h(w_{1h}, s)\beta \Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1\ell}, s)}{L_\ell(w_{1\ell}, s)} u_{1\ell}.$$  

Evaluating the problem at $s = 0$, we have the same necessary conditions has in section 3. The first-order condition with respect to $s$ evaluated at $s = 0$ is:

$$\frac{\partial \mathcal{L}}{\partial s} \bigg|_{s=0} = \lambda_1 \left\{ \pi_\ell \left[ \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right] + \pi_h \left[ \frac{\partial L_h(w_{1h}, s)}{\partial s} w_{1h} - \theta_h(w_{1h}, s) \right] \right\}$$

$$+ \beta \pi_h \frac{\partial L_h(w_{1h}, s)}{\partial s} F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}]$$

$$+ \mu \left\{ \beta \frac{\partial L_h(w_{1h}, s)}{\partial s} \Psi_h - u_{1\ell} \frac{L_h(w_{1\ell}, s)}{L_\ell(w_{1\ell}, s)} \left[ \frac{\partial L_h(w_{1\ell}, s)}{\partial s} - \frac{1}{L_h(w_{1\ell}, s)} \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} \right] \right\}.$$  

As above, the equation is the sum of the effect on the first period budget constraint, the fiscal gains of adding more skilled types in the second period, but also the effect on the incentive constraint which is the term on the third line inside the curly brackets. The first term is the value of increasing the probability of keeping the worker’s human capital, the second term is the difference in the marginal rate of substitution between after-tax income and subsidy $s$ coming from the wage-setting objective. Using the assumption made on the matching function we can show that this marginal rate of substitution is increasing in skill level. This implies that the planner can decrease $u_{1\ell}$ and increase $s$ and this will make the unskilled’s bundle less appealing to the skilled mimickers. This is because the increase in the probability of matching of the unskilled workers due to the subsidy is greater than the one of the mimicking skilled worker-firm, which permits to lower slightly the after-tax income of the unskilled worker while keeping the expected utility of the unskilled worker, and at the same relaxing the incentive constraint.
Rearranging and substituting in $\mu$, we have:

$$\frac{\partial L}{\partial s} \bigg|_{s=0} = \lambda_1 \left\{ \pi_h \left( \frac{\partial L_h(w_{1h}, s)}{\partial s} \left[ w_{1h} + \beta \left( \Psi_h + \frac{F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}]}{\lambda_1} \right) \right] - \theta_h(w_{1h}, s) \right) 
- w_{1\ell} \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} \left[ \frac{\partial L_h(w_{1\ell}, s)}{\partial s} \frac{1}{L_h(w_{1\ell})} - \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} \frac{1}{L_\ell(w_{1\ell})} \right] \right) + \pi_\ell \left( \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right) \right\}. $$

At $s = 0$ and with the maximin social welfare function, the necessary condition for $w_{1h}$ is identical to (3.8). Using this condition, and the functional form assumption it is straightforward to show that:

$$\frac{\partial L_h(w_{1h}, s)}{\partial s} \bigg|_{w_{1h}} = 0.$$

Again, using (3.7) and the functional form assumption, I get that

$$\frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) < 0.$$

Combining both these results, the potential welfare gains are:

$$\frac{\partial L}{\partial s} \bigg|_{s=0} = \lambda_1 \pi_h \left\{ -w_{1\ell} \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} \left[ \frac{\partial L_h(w_{1\ell}, s)}{\partial s} \frac{1}{L_h(w_{1\ell})} - \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} \frac{1}{L_\ell(w_{1\ell})} \right] + \pi_\ell \left( \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right) \right\}. $$

The above equation is composed of two terms, the first term is positive and represents the gains of slacking the incentive constraint by increasing $s$, the second term is the net effect on first period resources in the unskilled labor market of increasing $s$. Since this net effect on resources is negative, it is not possible to show in a straightforward manner if the global effect of increasing $s$ is positive.

Using numerical simulations, I find that the optimal level of subsidies $s$ is zero unless the labor market frictions are such that the optimal allocation requires the unskilled worker’s wage be distorted all the way to zero. At this point, the wage can no longer be used as a tool to separate types. The planner then uses the job creation subsidy to separate types further. This is illustrated in Table 4., as $\delta$ decreases, i.e. making the $U$ and $V$ less complimentary and therefore increasing market frictions, the optimal plan calls for distorting the unskilled worker’s wage dramatically. In the example, when $\delta \leq 0.5$ the planner sets $w_\ell = 0$ and introduces small subsidies. An interesting characteristic of the allocation is that as the frictions become worse, that doesn’t mean that the planner will want to distort more the skilled worker’s wage, actually the contrary is possible as it can be seen when we compare
Table 4: Characteristics of the Optimal Allocation and Tax System: Job Creation Subsidies

<table>
<thead>
<tr>
<th></th>
<th>Maximin</th>
<th>Benchmark</th>
<th>δ = 1</th>
<th>δ = 0.5</th>
<th>δ = 0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s^*$</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0.0119</td>
<td>0.0248</td>
</tr>
<tr>
<td>$L^*_t$</td>
<td>0.969</td>
<td>0.881</td>
<td>0.4944</td>
<td>0.1281</td>
<td></td>
</tr>
<tr>
<td>$L^*_h$</td>
<td>0.9361</td>
<td>0.8178</td>
<td>0.4627</td>
<td>0.1361</td>
<td></td>
</tr>
<tr>
<td>$%\Delta w^*_t$</td>
<td>-35.36</td>
<td>-76.65</td>
<td>-100</td>
<td>-100</td>
<td></td>
</tr>
<tr>
<td>$%\Delta w^*_h$</td>
<td>-4.34</td>
<td>-9.26</td>
<td>-15.22</td>
<td>-14.45</td>
<td></td>
</tr>
<tr>
<td>$T^*<em>1(w</em>{1t})$</td>
<td>-2.1748</td>
<td>-1.9498</td>
<td>-1.2842</td>
<td>-0.721</td>
<td></td>
</tr>
<tr>
<td>$T'^*<em>1(w</em>{1t})$</td>
<td>0.6966</td>
<td>0.6609</td>
<td>0.2289</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$T^*<em>1(w</em>{1h})$</td>
<td>1.8178</td>
<td>1.6167</td>
<td>1.237</td>
<td>1.1048</td>
<td></td>
</tr>
<tr>
<td>$T^<em><em>1(w</em>{1h}) - \beta \Psi^</em>_h$</td>
<td>0.5452</td>
<td>0.5516</td>
<td>0.369</td>
<td>-0.3836</td>
<td></td>
</tr>
<tr>
<td>$u_{0t}$</td>
<td>1.0969</td>
<td>1.1936</td>
<td>1.2238</td>
<td>1.0843</td>
<td></td>
</tr>
<tr>
<td>$\Delta u^*_{bh}$</td>
<td>1.8440</td>
<td>1.6325</td>
<td>1.187</td>
<td>0.8769</td>
<td></td>
</tr>
<tr>
<td>$u^*_t$</td>
<td>3.4875</td>
<td>2.8337</td>
<td>1.6348</td>
<td>0.9783</td>
<td></td>
</tr>
<tr>
<td>$v^*_h$</td>
<td>3.5621</td>
<td>2.9815</td>
<td>1.8098</td>
<td>1.0363</td>
<td></td>
</tr>
<tr>
<td>$v^<em>_h / v^</em>_t$</td>
<td>1.0214</td>
<td>1.0522</td>
<td>1.1070</td>
<td>1.0593</td>
<td></td>
</tr>
</tbody>
</table>

the case at $\delta = 0.5$ and the one at $\delta = 0.3$.

To understand what is going on, we must look at the necessary condition of $w_{1h}$ when $s > 0$:

$$
- \frac{\partial L_h(w^*_{1h}, s)}{\partial w_{1h}} \left[ w^*_h + \beta \Psi^*_h \right] L_h(w^*_{1h}, s) = 1 + \frac{\beta}{L_h(w^*_{1h}, s)} \frac{\partial L_h(w^*_{1h}, s)}{\partial w_{1h}} \left\{ \left[ \Psi^*_h + \frac{F(\Delta u^*_{hh})[w_{2h} - w_{2t} - \Delta u^*_{hh}]}{\lambda_1} \right] \right. \\
- \left. \Psi^*_h \right\} - \frac{\partial \theta_h(w^*_{1h}, s)}{\partial w_{1h}} s. \tag{4.12}
$$

Looking at condition (4.12), the condition is similar to (3.8) with the exception that there is now a new term that takes into account the revenue effect of reducing $w_{1h}$. Since $\frac{\partial \theta_h(w^*_{1h}, s)}{\partial w_{1h}}$ has a negative value, we can see that it puts an upward pressure on the optimal wage. The reason for this is simple, by lowering wages which will lead to an increased number of vacancies being opened, it will make the job creation subsidy program more expensive for a given level of $s$.

Thus, job creation subsidies in this model are useless unless we consider unrealistic levels of market frictions.

---

31 Also since $s$ does not enter the FOC of the choice of $\Delta u_{bh}$ and with the functional form assumption on the distribution function used above, we recover the result that $\Psi^*_h + \frac{F(\Delta u^*_{hh})[w_{2h} - w_{2t} - \Delta u^*_{hh}]}{\lambda_1} > \Psi^*_h$. 

39
5 Concluding Comments

I have derived characteristics of an optimal redistributive tax system when unemployment destroys a fraction of a worker’s human capital in a model where wages and employment are endogenous. Using a two type model, I find that the optimal tax system under imperfect information distorts the first period wage of both types away from their net-output maximizing levels (laissez-faire levels). So in both cases employment is increased compared to the laissez-faire outcome which guarantees more redistribution for the unskilled workers and a higher level of human capital in the future. Furthermore, I investigate how other policy tools such as training programs and job creation subsidies impact the optimal tax system. I find that both tools can be used in tandem with the optimal tax system to relax the informational constraint faced by the government but both these tools affect the optimal tax system in opposite ways. In the case of the training program, it can add further downward pressure on the unskilled worker’s wage in the first period, but the effect is mitigated by the relaxing of the constraint brought about by the training technology. Numerical results show that distortion of the skilled worker’s wage is similar in the presence or absence of the training technology. In the case of the job creation subsidies, the policy is useless unless matching frictions are quite high. This can have an upward effect on the skilled worker’s wage since having this wage being too low increases the amount of vacancies being opened and thus making the subsidy program much more expensive.

Extending the analysis to more types such as Hungerbühler et al. (2006); Lehmann et al. (2011) may permit to find other properties of the tax schedule especially how progressive or regressive it is under different circumstances. In addition, incorporating a participation decision would allow incorporating income-based unemployment benefits which has the potential to modify the second period redistributive tax scheme. Finally, a richer dynamic of human capital accumulation could be considered if the model is extended to more than two periods.
References


Appendices

A Wage-Setting Objective and Nash-Bargaining

Suppose Nash Bargaining and $M(U_i, V_i) = U_i^{\gamma} V_i^{1-\gamma}$.

$$w_{ti} = \arg \max_w \ [w - T(w)]^\gamma [a_i - w]^{1-\gamma}$$

The FOC is:

$$\gamma \frac{[1 - T'(w)] [a_i - w]}{1 - \gamma} = w - T(w).$$

With the assumption on matching function we have $L(a_i, w_i) = \left( \frac{a_i - w_i}{\kappa_i} \right)^{\frac{1-\gamma}{\gamma}}$.

$$w_{ti} = \arg \max_w \ [w - T(w)] \left( \frac{a_i - w}{\kappa_i} \right)^{\frac{1-\gamma}{\gamma}},$$

with FOC

$$\gamma \frac{[1 - T'(w)] [a_i - w]}{1 - \gamma} = w - T(w).$$

As we can see imposing that the bargaining power of the worker and the elasticity of matching be the same parameter $\gamma$, we have that the wage-setting objective used the in the paper can be a simple transformation of the Nash-bargaining product that results in the same solution.

B Proof of Proposition 1b

Suppose that

$$F(\alpha_h) = \left( \frac{\alpha_h}{\bar{\alpha}} \right)^\varepsilon$$

which gives a constant elasticity of participation of $\varepsilon$ in the skilled sector in the second period. Recall that $\Delta u_{hh}^{LF} = w_{2h} - w_{2\ell}$, and use (3.6) to get

$$\Delta u_{hh}^* = \frac{\varepsilon}{1 + \varepsilon - \lambda_1} \Delta u_{hh}^{LF} < \Delta u_{hh}^{LF}.$$ 

Also recall that when $\Delta u_{h\ell} = 0$, $\Psi_h = \int_0^{\Delta u_{hh}} [\Delta u_{hh} - \alpha_h] f_h(\alpha_h) d\alpha_h$, thus

$$\Psi_h^{LF} = \frac{(\Delta u_{hh}^{LF})^{(\varepsilon+1)}}{\bar{\alpha}^{\varepsilon+1}}, \text{ and } \Psi_h^* = \frac{\frac{\varepsilon}{1 + \varepsilon - \lambda_1} \Delta u_{hh}^{LF}^{(\varepsilon+1)}}{\bar{\alpha}^{\varepsilon+1}}.$$ 

For the skilled worker’s wage to be distorted downward we need

$$\Psi_h^* + \frac{F(\Delta u_{hh}^*) w_{2h} - w_{2\ell} - \Delta u_{hh}^*}{\lambda_1} > \Psi_h^{LF}.$$
Using the definitions we have just derived using the functional form of the CDF, the above inequality is true if the following inequality is true
\[
\left( \frac{\varepsilon}{1 + \varepsilon - \lambda_1} \right) > \lambda_1.
\]
Because we know that \(0 < \lambda_1 < 1\), the last inequality holds for any \(\varepsilon > 0\). The problematic case of this inequality is when \(\lambda_1\) approaches 1 and that \(\varepsilon\) approaches 0. To show that the inequality is true, we rewrite the inequality has
\[
g(\lambda_1, \varepsilon) = \exp \left\{ \varepsilon \ln \left( \frac{\varepsilon}{1 + \varepsilon - \lambda_1} \right) \right\} - \lambda_1 > 0.
\]
We can show that this function is globally convex on the domain \(\varepsilon > 0, 0 < \lambda_1 \leq 1\) and that the minimums of that function are when \(\lambda_1 = 1\) (or \(\varepsilon = -1\)). At those minimums the function is equal to 0. Thus for \(0 < \lambda_1 < 1\) and \(b > 0\), the function must be positive. Thus the inequality is always true.

Therefore as long as the elasticity of participation parameter is above 0 in the Maximin case, the skilled worker’s wage in the first period is distorted downwards away from the efficient value.

C Optimal tax policy: Bergson-Samuelson case and Proof of Proposition 2.

In this section of the appendix, we only write the necessary condition for the case \(u_{1\ell} > 0, u_{1h} > 0, w_{1\ell} > 0, w_{1h}, u_{0\ell} > 0, \Delta u_{hh} > 0, \Delta u_{\ell\ell} = \Delta u_{h\ell} = 0\). These are:

\[
\pi_{\ell} W'(v_\ell) - \lambda_1 \pi_{\ell} - \mu \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} = 0,
\]
\[
\lambda_1 \pi_{\ell} \left[ \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} w_{1\ell} + L_\ell(w_{1\ell}) \right] - \mu u_{1\ell} \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} \left[ \frac{\partial L_h(w_{1\ell})}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell})} - \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell})} \right] = 0,
\]
\[
\pi_h W'(v_h) - \lambda_1 \pi_h + \mu = 0,
\]
\[
\pi_h W'(v_h) \beta \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \Psi_h + \lambda_1 \pi_h \left[ \frac{\partial L_h(w_{1h})}{\partial w_{1h}} w_{1h} + L_h(w_{1h}) \right] + \lambda_2 \pi_h \left\{ \frac{\partial L_h(w_{1h})}{\partial w_{1h}} F(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] \right\} + \mu \beta \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \Psi_h = 0,
\]
\[
\pi_{\ell} W'(v_\ell) + \mu |\beta| L_h(w_{1h}) \frac{\partial \Psi_h}{\partial \Delta u_{hh}} + \lambda_2 L_h(w_{1h}) \left\{ f(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] - F(\Delta u_{hh}) \right\} = 0.
\]

Using the FOC relating to \(u_{1\ell}\) and \(u_{1h}\) you get that:
\[
\pi_{\ell} W'(v_\ell) + \pi_h W'(v_h) = \lambda_1 + \mu \left[ \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} - 1 \right] = \lambda_1 + \mu \Omega,
\]
where \( \Omega = \left[ \frac{L_h(w_{1h})}{L_{\ell}(w_{1\ell})} - 1 \right] > 0 \). From this we have \( \lambda_2 = \beta [\lambda_1 + \mu \Omega] \).

Using the FOC of \( \Delta u_{hh} \), and using the fact that \( \pi_h W'(v_h) + \mu = \lambda_1 \pi_h \), \( \lambda_2 = \beta [\lambda_1 + \mu \Omega] \) and \( \frac{\partial \Psi_h}{\partial \Delta u_{hh}} = F(\Delta u_{hh}) \) you get:

\[
[\lambda_1 + \mu \Omega] f(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] = [\lambda_1 + \mu \Omega] F(\Delta u_{hh}) - \lambda_1 F(\Delta u_{hh}),
\]
rewriting it, you get (3.6).

For (3.7), one just needs divide both sides by \( \lambda_1 \pi \ell L_{\ell}(w_{1\ell}) \) and move \( \frac{\partial L_{\ell}(w_{1\ell})}{\partial w_{1\ell}} \) to the right-hand side.

For (3.8), again need \( \pi_h W'(v_h) + \mu = \lambda_1 \pi_h \), \( \lambda_2 = \beta [\lambda_1 + \mu \Omega] \), the FOC on \( w_{1h} \), and rearranging we get:

\[
-\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{w_{1h}}{L_h(w_{1h})} + \frac{\beta}{L_h(w_{1h})} \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \left\{ \Psi_h + \frac{[\lambda_1 + \mu \Omega]}{\lambda_1} F(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] \right\}.
\]
Subtracting both sides by

\[
\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{\beta \Psi_h^{LF}}{L_h(w_{1h})},
\]
we get (3.8).

Suppose that

\[
F(\alpha_h) = \left( \frac{\alpha_h}{\bar{\alpha}} \right)^\varepsilon.
\]
Using (3.11) and the CDF, we find that

\[
\Delta u_{hh}^* = \Delta u_{hh}^{LF} \left[ \frac{[\lambda_1 + \mu \Omega]}{[\lambda_1 + \mu \Omega + \mu \Omega]} \right]^{(\varepsilon + 1)}.
\]
giving us

\[
\Psi_h^{LF} = \frac{(\Delta u_{hh}^{LF})^{(\varepsilon + 1)}}{\bar{\alpha}^\varepsilon [\varepsilon + 1]}, \quad \Psi_h^* = \frac{[\Delta u_{hh}^{LF}]^{(\varepsilon + 1)}}{[\lambda_1 + \mu \Omega]^{(\varepsilon + 1)}}.
\]
If we suppose that the wage will be distorted downward, we need this inequality to be true

\[
\Psi_h^* + \frac{[\lambda_1 + \mu \Omega]}{\lambda_1} F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*] > \Psi_h^{LF}.
\]
Using the definitions and assumptions we have made and after some manipulations, the above inequality is true if

\[
\frac{[\lambda_1 + \mu \Omega]^{(\varepsilon + 1)} \varepsilon^\varepsilon}{([\lambda_1 + \mu \Omega]^{\varepsilon + \mu \Omega})^\varepsilon} > \lambda_1.
\]
Without knowing much more on the values of $\mu \Omega$ other than it positive, it is impossible to know if this inequality holds for any values of $\varepsilon$. However, if we assume that $\varepsilon = 1$, i.e. the distribution function is uniform, the inequality becomes

$$\mu \Omega > 0,$$

therefore, if the incentive constraint (3.2) is binding and that $w_{1\ell}$ is not too low such that $L_{\ell}(w_{1\ell}) = L_{h}(w_{1\ell}) = 1$ then this is always true and thus the skilled worker’s wage is distorted below the efficient level.

\section*{D Numerical Simulations}

All numerical simulations in this paper are made assuming the matching function has the following CES form $M(U_i, V_i) = [U_i^{-\delta} + V_i^{-\delta}]^{-\frac{1}{\delta}}$. I also assume that the vacancy cost has the following iso-elastic shape $k(a) = b \ast a^v$. The value $\sigma = 1.8$, $b = 0.1$ and $v = 0.3$ are chosen so that in the \textit{Laissez-faire} the employment of both skill levels are close to 90%.

The productivity parameters and the share of the population are taken from the numerical simulations of Lee and Saez (2008) which uses a simple two type model to investigate the optimality of the minimum wage. So I take $a_\ell = 1$ and $a_h = 3$ to be the productivities for each type and assume that the fraction of the population of low types is $\pi_\ell = 0.25$ and the fraction of skilled types is $\pi_h = 0.75$. This is natural since in our simple model, the skilled should be considered to be a majority of the population.

The benchmark model uses $\epsilon = 0.25$ which represents the elasticity of participation of the skilled worker in the second period. This low value is chosen because the model is much more responsive to changes in other parameter which makes getting intuition much easier. The value of $\bar{\alpha}$ is chosen to be 3, so that even in the \textit{Laissez-faire} some individuals would still decide to work in the unskilled market.

I assume the following isoelastic general social welfare function as the planner’s preferences:

$$SWF = \pi_\ell \left( \frac{v_\ell}{\sigma} \right)^\sigma + \pi_h \left( \frac{v_h}{\sigma} \right)^\sigma.$$
Table 5: Characteristics of the Optimal Allocation and Tax System: Effect of Parameters

<table>
<thead>
<tr>
<th></th>
<th>Maximin</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>benchmark</td>
<td>participation</td>
<td>substitution</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\epsilon = 0.5$</td>
<td>$\epsilon = 0.75$</td>
<td>$\delta = 1$</td>
<td>$\delta = 2$</td>
<td></td>
</tr>
<tr>
<td>$L^*_l$</td>
<td>0.969</td>
<td>0.9679</td>
<td>0.967</td>
<td>0.881</td>
<td>0.9754</td>
</tr>
<tr>
<td>$L^*_h$</td>
<td>0.9361</td>
<td>0.9329</td>
<td>0.9305</td>
<td>0.8178</td>
<td>0.9469</td>
</tr>
<tr>
<td>$%\Delta w^*_l$</td>
<td>-35.36</td>
<td>-26.36</td>
<td>-25.60</td>
<td>-52.41</td>
<td>-23.83</td>
</tr>
<tr>
<td>$%\Delta w^*_h$</td>
<td>-4.34</td>
<td>-4.68</td>
<td>-4.95</td>
<td>-9.26</td>
<td>-3.78</td>
</tr>
<tr>
<td>$T_1^1(w_1l)$</td>
<td>-2.1748</td>
<td>-2.0032</td>
<td>-1.8709</td>
<td>-1.9498</td>
<td>-2.2019</td>
</tr>
<tr>
<td>$T_1^1(w_1h)$</td>
<td>0.6966</td>
<td>0.6989</td>
<td>0.7009</td>
<td>0.6609</td>
<td>0.6984</td>
</tr>
<tr>
<td>$T_1^1(w_1h) - \beta \Psi^*_h$</td>
<td>0.5452</td>
<td>0.4902</td>
<td>0.4455</td>
<td>0.5516</td>
<td>0.5426</td>
</tr>
<tr>
<td>$u^*_{0l}$</td>
<td>1.0969</td>
<td>1.0471</td>
<td>1.0295</td>
<td>1.1936</td>
<td>1.0820</td>
</tr>
<tr>
<td>$\Delta u_{hh}^*$</td>
<td>1.8440</td>
<td>1.9158</td>
<td>1.9414</td>
<td>1.6325</td>
<td>1.8701</td>
</tr>
<tr>
<td>$v^*_l$</td>
<td>3.4875</td>
<td>3.287</td>
<td>3.1497</td>
<td>2.8337</td>
<td>3.5647</td>
</tr>
<tr>
<td>$v^*_h$</td>
<td>3.5621</td>
<td>3.3598</td>
<td>3.2211</td>
<td>2.9815</td>
<td>3.6279</td>
</tr>
<tr>
<td>$v_{hh}/v_{ll}$</td>
<td>1.0214</td>
<td>1.0221</td>
<td>1.0227</td>
<td>1.0526</td>
<td>1.0177</td>
</tr>
</tbody>
</table>