Taxation and Redistribution of Residual Income Inequality

Mikhail Golosov
Princeton University and NBER

Pricila Maziero
University of Pennsylvania

Guido Menzio
University of Pennsylvania and NBER

March 2012

Abstract

This paper studies the optimal redistribution of income inequality caused by the presence of search and matching frictions in the labor market. We study this problem in the context of a directed search model of the labor market populated by homogenous workers and heterogeneous firms. Our main finding is that the optimal redistribution can be attained using a positive unemployment benefit and an increasing and regressive labor income tax. The positive unemployment benefit serves the purpose of lowering the search risk faced by workers. The increasing and regressive labor tax serves the purpose of aligning the cost of attracting an additional applicant to the firm with the opportunity cost of an applicant to society.
1 Introduction

Public finance has long been concerned with the optimal redistribution of labor income inequality. Traditionally, the optimal redistribution problem has been studied under the assumption that the labor market is frictionless and competitive and, hence, all of the observed differences in labor income reflect differences in workers’ productivity (see, e.g., Mirrlees 1971, Diamond 1998 or Saez 2001). However, a large body of empirical evidence documents the existence of substantial wage inequality among seemingly identical workers (see, e.g., Katz and Autor 1999 or Mortensen 2003). While this empirical evidence is at odds with the view that the labor market is perfectly competitive, it has been shown to be qualitatively and quantitatively consistent with the presence of search frictions in the labor markets (see, e.g., Postel-Vinay and Robin 2002 or Mortensen 2003).

In this paper, we want to study the optimal redistribution of income inequality caused by the presence of search frictions in the labor market. To accomplish this task, we consider a labor market populated by a continuum of risk-averse workers who are ex-ante homogenous and by a continuum of firms that are heterogeneous with respect to their productivity. Trade in the labor market is decentralized and frictional as in Moen (1997) and Shimer (1996). First, firms choose which wage to offer and workers choose whether to search for a job and, if so, which wage to seek. Then, the firms and the workers offering and seeking the same wage are brought together by a matching process described by a constant return to scale matching function. Trade in the labor market is frictional because we assume that the matching function is such that a worker is not guaranteed to find a job and, similarly, a firm is not guaranteed to find an employee. Instead, we assume that the probability that a worker finds a job—and the probability that a firm finds an employee—is a smooth function of the ratio between labor supply and labor demand at each particular wage. Moreover, we assume that a worker’s search strategy is his private information.

Because of search frictions, different types of firms offer different wages. In particular, more productive firms choose to offer higher wages in order to attract more job seekers and, hence, to increase their probability of trade. Also because of search frictions, inherently identical workers end up having different incomes. In particular, workers who are employed at more productive have a higher income than workers who are employed at less productive
firms, and employed workers have a higher income than unemployed workers.

In order to find the optimal redistribution of labor income inequality, we begin by solving for the constrained efficient allocation—i.e. the allocation that maximizes the workers’ expected utility subject to the technological constraints related to production and matching and to the incentive compatibility constraints associated with the workers’ private information about their search strategy. We find that the constrained efficient allocation differs from the equilibrium allocation along two dimensions. First, in the constrained efficient allocation, the number of workers seeking employment at high-productivity firms is greater than in equilibrium, while the number of workers seeking employment at low-productivity firms is smaller than in equilibrium. Second, in the constrained efficient allocation, the difference between the consumption of workers employed at high-productivity firms and of workers employed at low productivity firms is lower than in equilibrium, while the consumption enjoyed by unemployed workers is higher than in equilibrium.

The equilibrium is constrained inefficient because workers face an uninsured “search risk”—i.e. a worker’s consumption is greater when his search is successful than when his search fails. While some measure of risk is necessary to induce workers to search for jobs, the equilibrium search risk is, in general, inefficiently high. Hence, firms need to pay a wage premium to compensate workers for the excess search risk. And since the excess search risk is increasing in the number of workers applying to a particular job, high-productivity firms find it optimal to attract an inefficiently low number of applicants and, through general equilibrium effects, low-productivity firms end up attracting an inefficiently large number of applicants.

The constrained efficient allocation can be implemented by introducing a positive unemployment benefit and an increasing and regressive labor earning tax. The role of the positive unemployment benefit is to redistribute consumption between employed and unemployed workers and, thus, to lower the search risk faced by workers. The role of the positive and regressive labor earning tax is not to redistribute consumption among employed workers, but rather to correct a negative externality that firms impose on each other when the positive unemployment benefit is put in place. In fact, the unemployment benefit not only reduces the search risk faced by workers, but it also makes the cost of an applicant to a firm lower
than the opportunity cost of an applicant to society. The increasing and regressive labor earning tax realigns the private cost and the social cost of an applicant. In this sense, the labor earning tax is Pigovian.

The fact that the optimal labor earning tax is regressive is a startling and yet robust finding, in the sense that it does not depend on the shape of the utility function of workers, on the shape of the productivity distribution of firms, or on the shape of the matching function that brings together workers and firms. In fact, the optimality of a regressive labor tax follows necessarily from two properties of the optimal policy. First, the optimal policy is such that redistribution only takes place between workers who search successfully for a given wage and workers who search unsuccessfully for the same wage. Second, the optimal policy is such that workers are indifferent between searching for different wages.

Our paper contributes to two strands of literature. First, our paper contributes to the literature on optimal income taxation that was pioneered by Mirrlees (1971). This literature is concerned with characterizing the properties of the income tax system that implements the optimal redistribution of income inequality, where the extent of redistribution is limited by the workers’ private information about their productivity. Some papers carry out this task in a static environment (e.g., Diamond 1998, Saez 2001 and Laroque 2005), some in a dynamic environment (e.g., Farhi and Werning 2011 or Golosov, Troshkin and Tsyvinski 2011), and some in a frictional environment (e.g., Hungerbuhler et alii 2006). Yet, all of these papers assume that the income inequality originates from inherent productivity differences among workers. In contrast, our paper characterizes the income tax system that implements the optimal redistribution of income inequality that emerges among identical workers because of search frictions in the labor market.

Second, our paper contributes to the literature on optimal unemployment insurance that was pioneered by Shavell and Weiss (1979). This literature is concerned with characterizing the properties of the unemployment insurance system that implements the optimal redistribution between employed and unemployed workers, where the extent of redistribution is limited by the workers’ private information about their search effort. In this literature as in our paper, workers are inherently identical and income inequality is caused by the presence of search frictions in the labor market. However, in contrast to our paper, this literature
does not contain any insights on income taxation because it either assumes that all employed workers earn the same wage (e.g., Hansen and Imrohoroglu 1992, Hopenhayn and Nicolini 1997, Acemoglu and Shimer 1999, and Wang and Williamson 2002), or it assumes that a worker’s wage is his private information (e.g., Shimer and Werning 2008).

Another contribution of our paper is to sharpen some of the results in Acemoglu and Shimer (1999). Acemoglu and Shimer study the effects of introducing an unemployment benefit in a directed search model of the labor market in which both workers and firms are ex-ante homogenous. They prove that the unemployment benefit that maximizes aggregate output is strictly positive (Proposition 6), and they conjecture that the unemployment benefit that maximizes welfare also increases aggregate output (Conjecture 1). In this paper, we show that the constrained efficient allocation is implemented by introducing a positive unemployment benefit and, when firms are heterogeneous, a positive and regressive labor earnings tax (Theorem 1). Moreover, we prove that aggregate output in the constrained efficient allocation is higher than in the laissez-faire equilibrium (see the analysis following Proposition 2). The first of our findings shows that a positive unemployment benefit is indeed part of the policy mix that implements the constrained efficient allocation, so that Acemoglu and Shimer’s focus on this policy instrument is without loss in generality. The second of our findings shows that Conjecture 1 in Acemoglu and Shimer is correct and, moreover, it generalizes to an environment with heterogeneous firms.

2 Equilibrium

2.1 Environment

The economy is populated by a continuum of homogeneous workers with measure 1. The preferences of a worker are described by the utility function $u(c)$, where $u : \mathbb{R}_+ \to \mathbb{R}$ is a twice differentiable, strictly increasing and strictly concave function of consumption. The endowment of a worker consists of one job application and one indivisible unit of labor. The economy is also populated by a continuum of heterogeneous firms with measure $m > 0$. The type of a firm is denoted by $y \in [y, \overline{y}]$, $0 < y < \overline{y}$, and the measure of firms with type less than is denoted by $F(y)$, where $F : [y, \overline{y}] \to [0, m]$ is a twice differentiable and strictly increasing function with boundary conditions $F(y) = 0$ and $F(\overline{y}) = m$. A firm of type $y$
owns a vacancy that, when filled by a worker, produces \( y \) units of output. Firms are owned by workers through a mutual fund. Hence, the objective of the firms is to maximize expected profits.

Workers and firms come together through a directed search process (see, e.g., Montgomery 1991, Moen 1997 and Acemoglu and Shimer 1999). In the first stage of the process, each firm chooses which wage to offer to a worker who fills its vacancy. Simultaneously, each worker chooses whether to send a job application at the utility cost \( k > 0 \) and, if so, which wage to seek. In making these decisions, firms and workers take as given the expected ratio \( q(w) \) of applicants to vacancies associated with each wage \( w \). Following Acemoglu and Shimer (1999), we shall refer to \( q(w) \) as the queue length. In the second stage of the process, each worker seeking the wage \( w \) matches with a firm offering the wage \( w \) with probability \( \lambda(q(w)) \), where \( \lambda : \mathbb{R}_+ \to [0, 1] \) is a strictly decreasing function of \( q \) with boundary conditions \( \lambda(0) = 1 \) and \( \lambda(\infty) = 0 \). Similarly, a firm offering the wage \( w \) matches with an applicant seeking the wage \( w \) with probability \( \eta(q(w)) \), where \( \eta : \mathbb{R}_+ \to [0, 1] \) is a strictly increasing and strictly concave function of \( q \) with boundary conditions \( \eta(0) = 0 \) and \( \eta(\infty) = 1 \). The functions \( \lambda \) and \( \eta \) satisfy the aggregate consistency condition \( \lambda(q)q = \eta(q) \). If a firm of type \( y \) matches with a worker, it produces \( y \) units of output, it pays the wage \( w \) to the worker and it pays dividends to the owners. If a firm remains unmatched, it does not produce any output.

We assume that the application strategy of a worker is private information—i.e. the public cannot observe whether a worker sent a job application and, if so, which wage he sought—while the employment status of a worker is public information—i.e. the public observes whether a worker is employed and, if so, at which wage. The above informational assumptions induce a moral hazard problem: the public cannot distinguish a worker who did not search for a job and a worker who searched for a job unsuccessfully.

2.2 Equilibrium

An allocation is a tuple \((w, q, d, S)\). The first element of the tuple is a function \( w : [\underline{y}, \overline{y}] \to \mathbb{R}_+ \), with \( w(y) \) denoting the wage offered by a firm of type \( y \). The second element is a function \( q : \mathbb{R}_+ \to \mathbb{R}_+ \), with \( q(w) \) denoting the queue length attracted by the wage \( w \). The
third element, $d \in \mathbb{R}$, denotes the dividend payment received by each worker. Finally, $S \in \mathbb{R}$ denotes the maximized value of sending an application.

Now, we are in the position to define an equilibrium.

**Definition 1:** A competitive equilibrium is an allocation $(w, q, d, S)$ that satisfies the following conditions:

(i) **Profit maximization:** For all $y \in \overline{y}$,

$$w(y) \in \arg \max_{w \geq 0} \eta(w)(y - w);$$

(ii) **Optimal number of applications:**

$$\int q(w(y))dF(y) \leq 1 \text{ and } S \geq k, \text{ with comp. slackness;}$$

(iii) **Optimal direction of applications:** For all $w \in \mathbb{R}_+$,

$$\lambda(q(w))[u(d + w) - u(d)] \leq S \text{ and } q(w) \geq 0, \text{ with comp. slackness;}$$

(iv) **Consistency of dividends and profits:**

$$d = \int \eta(q(w(y)))(y - w(y))dF(y).$$

The above definition of equilibrium is standard (see, e.g., Moen 1997 and Acemoglu and Shimer 1999). Condition (i) guarantees that the wage posted by a firm of type $y$ is profit maximizing. That is, $w(y)$ maximizes the product of the probability of filling a vacancy, $\eta(q(w))$, and the profit from filling a vacancy, $y - w$. Condition (ii) guarantees that the measure of applications received by the firms is consistent with workers’ utility maximization. That is, whenever $S$ is strictly greater than $k$, all workers find it optimal to search and, hence, the measure of applications received by firms is equal to one. Whenever $S$ is equal to $k$, workers are indifferent between searching and not searching and, hence, the measure of application received by the firms can be smaller than one. Condition (iii) guarantees that the distribution of applications across wages is consistent with workers’ utility maximization. That is, whenever $q(w)$ is strictly positive, the worker’s expected utility from searching for
the wage \( w \), \( \lambda(q(w))(u(d + w) − u(d)) \), is equal to the maximized value of searching \( S \). Whenever, \( q(w) \) is equal to zero, the worker’s expected utility from searching the wage \( w \) may be smaller than \( S \). Finally, condition (iv) guarantees that the dividends received by the workers are equal to the firms’ profits.

### 2.3 Characterization of the equilibrium allocation

Let \( p(q|d, S) \) denote the expected wage for a firm that attracts \( q \) applicants divided by the number of applicants, i.e. \( p(q|d, S) = \eta(q)w(q|d, S)/q \), where \( w(q|d, S) \) is the wage required to attracts \( q \) applicants given the dividends \( d \) and the value of searching \( S \). We shall refer to \( p(q|d, S) \) as the price of an applicant. From the equilibrium condition (iii) and the consistency condition \( \eta(q) = \lambda(q)q \), it follows that

\[
p(q|d, S) = \lambda(q) \left[ u^{-1} \left( \frac{S}{\lambda(q)} + u(d) \right) - d \right].
\]

The first term on the right-hand side of (1) is the probability that the applicant is hired by the firm. The second term on the right-hand side of (1) is the wage that the firm has to offer in order to attract \( q \) applicants.

Let us highlight some properties of \( p(q|d, S) \) that will be useful in the remainder of the paper. First, note that \( p(q|d, S) \) is independent of \( q \) when workers are risk neutral. Intuitively, if the firm chooses to attract a longer queue, each applicant has a lower probability of being hired and the wage has to increase in order to keep the worker’s expected application value equal to \( S \). Since risk-neutral workers only care about the expected payment of the application, \( p(q|d, S) \) remains constant. If workers are risk-averse—as they are in our model—\( p(q|d, S) \) is strictly increasing in \( q \). When they are risk averse, workers strictly prefer applying to a job that they are more likely to get and offers a lower wage than applying to a job that they are less likely to get and pays a higher wage, given that the two jobs offers the same expected payment. Formally, \( q_1 > q_0 \) and \( \lambda(q_1)w_1 = \lambda(q_0)w_0 \) implies that \( \lambda(q_1)u(w_1 + d) + (1 - \lambda(q_1))u(d) \) is strictly greater than \( \lambda(q_0)u(w_0 + d) + (1 - \lambda(q_0))u(d) \). Hence, if a firm wants to attract a longer queue, it has to increase the expected wage payment to each applicant \( p(q|d, S) \). Next, notice that the marginal price of an applicant, \( p'(q|d, S) \), is strictly increasing in \( S \). Intuitively, the higher is \( S \), the larger is the difference between the worker’s
consumption if his application is successful and unsuccessful. Hence, due to the concavity of \( u \), the worker requires a higher wage increase to accommodate a longer queue. Finally, notice that the marginal price of an applicant, \( p'(q|d, S) \), may be increasing or decreasing in \( d \) depending on the shape of the utility function. In the remainder of the paper, we will assume that \( p'(q|d, S) \) is strictly decreasing in \( d \).\(^1\)

Now, let \( q(y) \) denote the queue of applicants attracted by a firm of type \( y \). From equilibrium condition (i) and equation (1), it follows that \( q(y) \) is such that

\[
q(y) = \arg \max_{q \geq 0} \eta(q)y - p(q|d, S)q.
\]

The function \( q(y) \) satisfies the first order condition

\[
\eta'(q)y \leq p(q|d, S) + p'(q|d, S)q \tag{3}
\]

and \( q \geq 0 \) with complementary slackness. The term on the left-hand side of (3) is the marginal product of an applicant for a firm of type \( y \), which is given by the product between the increase in the probability that the firm fills its vacancy because of an additional applicant, \( \eta'(q) \), and the output produced by the firm if it fills its vacancy, \( y \). The right-hand side of (3) is the sum between the price of the marginal applicant, \( p(q|d, S) \), and the increase in the price of the infra-marginal applicants, \( p'(q|d, S)q \). Appendix A shows that the right-hand side of (3) is strictly decreasing in \( q \), while the left-hand side is strictly increasing in \( q \) and \( y \). Hence, \( q(y) \) is equal to zero for all \( y \leq y_c \) and \( q(y) \) is strictly positive and strictly increasing for all \( y > y_c \), where the type cutoff \( y_c \) is given by \( \eta'(0)y_c = p(0|d, S) \). In words, firms of productivity \( y \leq y_c \) do not enter the labor market, while firms of productivity \( y > y_c \) enter the market and attract a queue of applicants that is increasing in \( y \).

Next, let \( c(y) \) denote the consumption of a worker employed at a firm of type \( y \geq y_c \). From equilibrium condition (iii), it follows that \( c(y) \) is given by

\[
c(y) = u^{-1}\left(\frac{S}{\lambda(q(y))} + u(d)\right). \tag{4}
\]

\(^1\)The assumption is satisfied when the utility function \( u \) has the HARA form

\[
u(c) = \frac{1-\gamma}{\gamma} \left( \frac{\alpha c}{1-\gamma} + \beta \right)\gamma,
\]

and the parameter \( \gamma \) belongs to the interval \([1/2, 1]\).
Since \( q(y) \) is increasing in \( y \), equation (4) implies that the consumption of an employed worker, \( c(y) \), is increasing in the productivity of his employer. Moreover, since \( S \geq k \), equation (4) also implies that the consumption of an unemployed worker is strictly greater than the consumption of an unemployed worker, \( d \).

Finally, we are going to characterize \( d \) and \( S \). The equilibrium condition (iv) and equation (4) imply that the consumption of an unemployed worker, \( d \), is such that

\[
\int \eta(q(y))ydF(y) = \int q(y)[\lambda(q(y)c(y) + (1 - \lambda(q(y)))d]dF(y).
\] (5)

Equation (5) states that the consumption of the unemployed is such that aggregate output—which is the term on the left-hand side of (5)—equals aggregate consumption—which is the term on the right-hand side of (5).

The equilibrium condition (ii) implies that the equilibrium value of searching \( S \) is such that the number of applications received by firms is equal to the number of applications sent out by workers. That is, \( S \) is such that

\[
\int q(y)dF(y) \leq 1
\] (6)

and \( S \geq k \) with complementary slackness.

Overall, any equilibrium can be represented as a tuple \( \{q(y), c(y), d, S\} \) that satisfies the system of equations (3)–(6). There are two features of the equilibrium that is worth stressing. First, since \( c(y) \) is strictly greater than \( d \) and strictly increasing in \( y \), the equilibrium does not equalize the marginal utility of consumption across workers in different employment states. Intuitively, since in equilibrium workers need to have the incentive to apply for a job and to apply more frequently to jobs at more productive firms, the consumption of unemployed workers needs to be lower than the consumption of employed workers and the consumption of employed workers needs to be increasing in the productivity of the firm. Second, since \( \eta'(q(y))y = p(q(y)) + p'(q(y))q(y) \) and \( p(q(y)) + p'(q(y))q(y) \) is strictly increasing in \( y \), the equilibrium does not equalize the marginal product of applicants across different firms. In particular, the marginal product of an applicant is higher at high productivity firms than at low productivity firms. Intuitively, this happens because more productive firms choose
to attract a longer queue of applicants and the price of each applicant is increasing in the queue length.

The above observations imply that the equilibrium does not attain the first-best allocation, i.e. the allocation that maximizes the expected utility of a worker given the matching technology, $\eta(q)$ and $\lambda(q)$, and production technology, $F(y)$. In fact, since the equilibrium does not equate the marginal utility of applicants across firms, aggregate output is not maximized. Moreover, since the equilibrium does not equate the marginal utility of consumption across workers in different employment states, expected utility is not maximized given aggregate output. Notice that risk aversion is a necessary condition for the discrepancy between equilibrium and first-best allocation. In fact, when workers are risk neutral, the marginal utility of consumption is the same across workers in different employment states and, since the price of an applicant is independent of the queue, the marginal product of an applicant is equated across different firms. This explains why the risk-neutral models of directed search of Montgomery (1991), Moen (1997), Burdett Shi and Wright (2001) and Menzio and Shi (2011) find that the equilibrium and first-best allocation are equivalent.

3 Constrained efficient allocation

3.1 Formulation of the mechanism design problem

A symmetric mechanism is a tuple $(\pi, c, d, S)$. The first element of the tuple is a differentiable function $\pi : [y, \overline{y}] \rightarrow \mathbb{R}^+$, where $\pi(y)$ denotes the probability that a worker sends an application to firms of type less than $y$. The function $\pi(y)$ uniquely determines the ratio $q(y)$ between the number of workers applying to firms of type $y$ and the number of firms of type $y$, $q(y) = \pi'(y)/F'(y)$. Moreover, the function $\pi(y)$ uniquely determines the measure, $a$, of workers who send an application, $a = \pi(\overline{y})$. The second element of the tuple is a function $c : [y, \overline{y}] \rightarrow \mathbb{R}^+$, where $c(y)$ denotes the consumption of a worker employed by a firm of type $y$. The third element of the tuple, $d \in \mathbb{R}^+$, denotes the consumption of an unemployed worker. Finally, the last term of the tuple, $S \in \mathbb{R}^+$, denotes the expected utility from sending a job application. The mechanism $(\pi, c, d, S)$ is symmetric because it specifies the same strategy to identical agents. In particular, the mechanism specifies the same mixed strategy $\pi$ to all workers and it assigns the same number of applicants, $q(y)$, to all firms of the same type $y$. 
Notice that the mechanism can condition consumption on the workers’ employment state (i.e. unemployed or employed at a firm of a particular type) because the employment state is publicly observable. In contrast, the mechanism cannot condition consumption on the workers’ application (i.e. whether the worker sent an application and, if so, towards which type of firm he directed it) because the workers’ application is not observable. If we were to allow workers to report their application to the mechanism, consumption could also be made contingent on these reports. However, in Appendix B, we show that the mechanism could not exploit these additional contingencies to improve the allocation.

The mechanism designer chooses \((\pi, c, d, S)\), or equivalently \((a, q, c, d, S)\), in order to maximize the worker’s expected utility

\[
\int [\lambda(q(y))u(c(y)) + (1 - \lambda(q(y))) u(d) - k] q(y)dF(y) + (1 - a)u(d). \tag{7}
\]

The first term in (7) is the integral across \(y\)’s of the product between the probability that the worker applies to firms of type \(y\), \(q(y)dF(y)\), and the expected utility from applying to firms of type \(y\), \(\lambda(q(y))u(c(y)) + (1 - \lambda(q(y))) u(d) - k\). The second term in (7) is the product between the probability that the worker does not send an application, \(1 - a\), and the utility from remaining unemployed, \(u(d)\).

The mechanism designer’s choice of \((a, q, c, d, S)\) is subject to two sets of incentive compatibility constraints. First, the mechanism must be such that

\[
S - k \geq 0, \tag{8}
\]

\[
(1 - a)(S - k) = 0 \tag{9}
\]

Condition (8) requires that the worker’s expected utility from sending a job application, \(S\), is greater or equal to \(k\). Condition (9) requires that \(a = 1\) whenever \(S > k\) and \(a \leq 1\) whenever \(S = k\). Taken together, conditions (8) and (9) require the mechanism to be compatible with the worker’s incentive to send a job application. Second, for all \(y\), the mechanism must be such that

\[
\lambda(q(y)) [u(c(y)) - u(d)] - S = 0. \tag{10}
\]

Condition (10) requires the worker to expect the same utility, \(S\), from sending his job application to any type of firm \(y\). Thus, condition (10) requires the mechanism to be compatible
with the worker’s incentive to direct his application towards firms of a particular type.

The mechanism designer’s choice of \((a, q, c, d, S)\) is also subject to two aggregate resource constraints. First, the mechanism must be such that the aggregate output produced by the firms is greater or equal to the aggregate consumption enjoyed by the workers

\[
\int \eta(q(y))y dF(y) - \int [\lambda(q(y))c(y) + (1 - \lambda(q(y)))d] q(y) dF(y) - (1 - a)d \geq 0. \tag{11}
\]

Second, the mechanism must be such that the measure of workers who apply for a job is smaller than 1 and it is equal to the measure of applications received by the firms

\[
1 - a \geq 0, \tag{12}
\]

\[
a - \int q(y) dF(y) = 0. \tag{13}
\]

### 3.2 Solution to the mechanism design problem

Let \((a^*, q^*, c^*, d^*, S^*)\) denote the solution to the mechanism design problem. Let \(\phi_1^*\) and \(\phi_2^*\) denote the multipliers associated to the incentive compatibility constraints (8) and (9). Let \(\rho^*(y) dF(y)\) denote the multiplier associated to the incentive compatibility constraint (10). Similarly, let \(\mu_1^*, \mu_2^*\) and \(\mu_3^*\) denote the multipliers associated to the aggregate resource constraints (11)–(13). We shall refer to the solution to the mechanism design problem as either the constrained efficient or the second-best allocation.

The constrained efficient consumption for a worker employed at a firm of type \(y\), \(c_y^*\), satisfies the first order condition

\[
\lambda(q_y)q_y u'(c_y) + \rho_y \lambda(q_y) u'(c_y) = \mu_1 \lambda(q_y)q_y. \tag{14}
\]

The left-hand side of (14) is the marginal benefit of increasing \(c_y\). First, an increase in \(c_y\) increases the worker’s expected utility by increasing his consumption if employed at a firm of type \(y\). The benefit of this effect is \(\lambda(q_y)q_y u'(c_y)\). Second, an increase in \(c_y\) relaxes the incentive compatibility constraint (10) for workers applying to firms of type \(y\). The benefit of this effect is \(\rho_y \lambda(q_y) u'(c_y)\). The right-hand side of (14) is the marginal cost of increasing \(c_y\), which is given by the value of the additional output assigned to workers employed in firms of type \(y\).
The constrained efficient consumption for an unemployed worker, \(d^\ast\), satisfies the first order condition
\[
u'(d) \left[ 1 - \int \lambda(q_y)q_y dF(y) \right] = u'(d) \left[ \int \rho_y \lambda(q_y) dF(y) \right] + \mu_1 \left[ 1 - \int \lambda(q_y)q_y dF(y) \right]. \tag{15}
\]
The left-hand side of (15) is the marginal benefit of increasing \(d\), which is the increase in the worker’s expected utility caused by an increase of his consumption if unemployed. The right-hand side of (15) is the marginal cost of increasing \(d\). An increase in \(d\) tightens the incentive compatibility constraint (10) for workers applying to all types of firms. The cost of this effect is \(u'(d) \int \rho_y \lambda(q_y) dF(y)\). Moreover, an increase in \(d\) tightens the aggregate resource constraint on output by increasing the amount of output assigned to unemployed workers. The cost of this effect is \(\mu_1 \left[ 1 - \int \lambda(q_y)q_y dF(y) \right]\). The constrained efficient queue of applicants for a firm of type \(y\), \(q_y^*\), satisfies the following first order condition
\[
\begin{align*}
\{u(d) + S - k + q_y \lambda'(q_y) [u(c_y) - u(d)]\} \\
+ \mu_1 \{u'(q_y)y - \lambda(q_y)c_y - (1 - \lambda(q_y))d + q_y \lambda'(q_y) [c_y - d]\} \\
\leq \mu_3 - \rho_y \lambda'(q_y) [u(c_y) - u(d)], 
\end{align*}
\tag{16}
\]
and \(q_y \geq 0\) with complementary slackness. The left hand side of (16) is the marginal benefit of increasing \(q_y\). An increase in \(q_y\) affects the worker’s expected utility by increasing the probability that a worker applies to firms of type \(y\) and by lowering the probability that an applicant to a firm of type \(y\) becomes employed. The net benefit of this effect is measured by the first term on the left-hand side of (16). Moreover, an increase in \(q_y\) affects the aggregate resource constraint on output by increasing the output produced by firms of type \(y\) and by increasing the output assigned to workers employed by firms of type \(y\). The net benefit of this effect is measured by the second term in the left-hand side of (16). The right-hand side of (16) is the marginal cost of increasing \(q_y\). An increase in \(q_y\) tightens the aggregate resource constraint on applicants. The cost of this effect is \(\mu_3\). Moreover, an increase in \(q_y\) tightens the incentive compatibility constraint (10) for workers applying to firms of type \(y\). The cost of this effect is \(-\rho_y \lambda'(q_y) [u(c_y) - u(d)]\).
The first-order condition for the constrained efficient value of search, $S^*$, is

$$\phi_1 + \phi_2(1 - a) = \int \rho_y dF(y).$$

(17)

The left-hand side of (17) is the marginal benefit of an increase in $S$, which is the value of relaxing the incentive compatibility constraints (9) and (10). The right-hand side of (17) is the marginal cost of an increase in $S$, which is the value of tightening the incentive compatibility constraint (11) for workers applying to all types of firms. Using (14) and (15) to solve for $\rho_y$, it is easy to show that the right-hand side is strictly positive. From this observation, it follows that, if the solution to the mechanism design problem is such that $a = 1$, the multiplier $\phi_1$ on the incentive compatibility constraint (9) is strictly positive and, hence, $S = k$. If, on the other hand, the solution to the mechanism design problem is such that $a < 1$, the incentive compatibility constraint (10) implies that $S = k$. Hence, the constrained efficient value of searching $S^*$ is always set equal to $k$, the lowest value that is compatible with the worker’s incentive to send a job application. This finding is intuitive. Increasing $S$ above $k$ does not affect the workers’ incentive to search a job application, but it makes it harder to give workers an incentive to apply to the right type of firms.

The first-order condition for the constrained efficient measure of applicants, $a^*$, is given by $u(d) + \mu_2 = \mu_1 d + \mu_3$. Using this equation to solve for $\mu_2$ and using equation (14) to solve for $\rho_y$, we can write the first-order condition for the constrained efficient queue of applicants, $q^*_y > 0$, as

$$\eta'(q_y)y = \lambda(q_y)(c_y - d) - q_y\lambda'(q_y)\left[\frac{u(c_y) - u(d)}{u'(c_y)} - (c_y - d)\right] + \frac{\mu_2}{\mu_1}.$$  

(18)

Finally, let $p^*(q|d, S)$ be defined as

$$p^*(q|d, S) = p(q|d, S) + \frac{\mu_2}{\mu_1}.$$  

(19)

Using the above definition and (1), one can easily show that (18) can be written as

$$\eta'(q_y)y = p^*(q|d, S) + qp^*(q|d, S)$$

$$= p(q|d, S) + qp'(q|d, S) + \frac{\mu_2}{\mu_1}.$$  

(20)

That is, the first-order condition for the constrained efficient queue of applicants $q^*_y$ is the
same as the first-order condition of a firm that chooses \( q \) in order to maximize its profits given that the price of an applicant is 
\[
p^*(q|d, S) = p(q|d, S) + \mu_2/\mu_1.
\]

Having characterized the solution to the mechanism design problem, we are in the position to establish the welfare properties of the equilibrium.

**Proposition 1** (Welfare Properties of Equilibrium): Let \((q^*, c^*, d^*, S^*, a^*)\) denote the solution to the mechanism design problem and let \(\mu^*_1, \mu^*_2\) the multipliers associated with the aggregate resource constraints (11) and (12). (i) If \(\mu^*_2 = 0\), then \((q^*, c^*, d^*, S^*)\) is an equilibrium. (ii) If \(\mu^*_2 > 0\), then \((q^*, c^*, d^*, S^*)\) is not an equilibrium. (iii) There exists a \(k\) and a \(\overline{k}\), where \(k \geq \overline{k} > 0\) such that \(\mu^*_2 = 0\) for all \(k > \overline{k}\) and \(\mu^*_2 > 0\) for all \(k > k\).

**Proof:** See Appendix C.

The proof of parts (i) and (ii) of Proposition 1 is straightforward. Suppose that the solution to the mechanism design problem is such that the multiplier on the aggregate resource constraint on applicants is zero, i.e. \(\mu^*_2 = 0\). In this case, \(q^*_y\) satisfies the first-order condition (20), which is the same as the equilibrium condition (3); \(c^*_y\) satisfies the incentive compatibility constraint (10), which is the same as the equilibrium condition (4); \(d^*\) satisfies the aggregate resource constraint on output (11), which is the same as the equilibrium condition (5); and \(S^* = \overline{k}\) satisfies the equilibrium condition (6). Hence, the solution to the mechanism design problem is an equilibrium. In contrast, when \(\mu^*_2 > 0\), the solution to the mechanism design problem is not an equilibrium because equation (20) is different from the equilibrium condition (3). Part (iii) of Proposition 1 is more difficult to prove, but it easy to understand: The multiplier on the aggregate resource constraint on applicants, \(\mu^*_2\), is equal to zero when the search cost \(k\) is sufficiently high and it is strictly positive when \(k\) is sufficiently low.

There is a simple intuition behind Proposition 1. In equilibrium, the value of searching plays two roles. First, the value of searching to a worker, \(S\), guarantees that the number of applications supplied by workers is equal to the number of applications demanded by firms because \(S\) determines the cost of an applicant \(p(q|d, S)\) to a firm. Second, the value of searching \(S\) determines the amount of consumption risk that is faced by workers applying for jobs because \(u(c) - u(d) = S/\lambda(q)\). In the second best, the worker’s value of searching \(S^*\) is
given by \( k \), which is the value that minimizes the consumption risk faced by workers. And the social cost of an applicant is given by \( p^*(q|d^*, S^*) = p(q|d, S) + \frac{\mu_2^*_2}{\mu_1^*} \), where \( \mu_2^*_2 \) is the value that guarantees that the aggregate resource constraint for applicants is satisfied. When \( \mu_2^*_2 > 0 \), the equilibrium cannot decentralize the second best allocation because it cannot simultaneously guarantee that the value of searching to a worker minimizes the consumption risk—which would require \( S = k \)—and that the private cost of an applicant to a firm reflects its social cost—which would require \( p(q|d, S) = p^*(q|d, S) \). That is, when \( \mu_2^*_2 > 0 \), the equilibrium is constrained inefficient because the value of searching \( S \) cannot simultaneously minimize the consumption cost and clear the market for applicants. As we shall see in Sections 4 and 5, the link between consumption risk and demand for applicants created by \( S \) can be broken—and efficiency can be restored—by either introducing unemployment benefits and labor earning taxes or by opening a competitive insurance market.

### 3.3 Comparison between equilibrium and constrained efficient allocation

In this subsection, we want to understand how the equilibrium and the second-best allocations differ when the equilibrium is constrained inefficient. To this aim, we assume that the solution to the mechanism design problem is such that \( \mu_2^*_2 > 0 \) and that the equilibrium is such that \( S > k \). The first assumption guarantees that the equilibrium does not decentralize the second best allocation. The second assumption guarantees that, in the equilibrium, all workers find it optimal to apply for a job. This assumption makes the comparison between equilibrium and second-best allocation more transparent.

We begin by comparing the consumption of the unemployed in the equilibrium and in the second best. To this aim, notice that the second-best allocation \((q^*, c^*, d^*, S^*)\) maximizes the expected utility of a worker among all allocations that satisfy the technological and incentive compatibility constraints. Also, notice that the equilibrium allocation \((q, c, d, S)\) is an allocation that satisfies the technology and incentive compatibility constraints, but does not maximize the worker’s expected utility when \( \mu_2^*_2 > 0 \). Since the worker’s expected utility in the second-best allocation is given by \( u(d^*) + S^* - k \), with \( S^* = k \), and in the equilibrium it is given by \( u(d) + S - k \), with \( S > k \), it follows that \( d^* > d \). That is, the consumption
assigned to unemployed workers is inefficiently low in the equilibrium, while the value of searching is inefficiently high.

Next, we compare the consumption of the employed in the equilibrium and in the second best. To this aim, let \(c(q)\) denote the consumption, in the equilibrium, for a worker employed at a firm that attracts \(q\) applications, i.e. \(c(q) = u^{-1}(S/\lambda(q) + u(d))\). Similarly, let \(c^*(q)\) denote the consumption, in the second best, for a worker employed at a firm that is assigned \(q\) applicants, i.e. \(c^*(q) = u^{-1}(S^*/\lambda(q) + u(d^*))\). The derivatives of \(c\) and \(c^*\) with respect to \(q\) are given by

\[
\begin{align*}
    c'(q) & = -\frac{\lambda'(q)}{\lambda(q)^2} \frac{S}{u'(c(q))}, \\
    c''(q) & = -\frac{\lambda'(q)}{\lambda(q)^2} \frac{S^*}{u'(c^*(q))}.
\end{align*}
\]

Notice that since \(S > S^*\), \(c(q_0) = c^*(q_0)\) implies that \(c'(q_0) > c''(q_0)\). Using this observation and the behavior of \(c(q)\) and \(c^*(q)\) at \(q = 0\) and \(q = \infty\), it follows that there exists one \(q_0 > 0\) such that \(c(q_0) = c^*(q_0)\) and \(c(q) < c^*(q)\) for all \(q < q_0\) and \(c(q) > c^*(q)\) for all \(q > q_0\). That is, the consumption assigned to employed workers is inefficiently low at firms that attract a small number of applications and inefficiently large at firms that attracts a large number of applications.

Finally, we compare the allocation of applicants across firms in the equilibrium, \(q(y)\), and in the second best, \(q^*(y)\). After differentiating with respect to \(y\) the first-order condition (3), we obtain the following expression for the slope of the equilibrium queue length function

\[
q'(y) = \frac{\eta'(q(y))y}{-\eta''(q(y))y + 2p'(q(y)|d, S) + p''(q(y)|d, S)}.
\]

Similarly, after differentiating with respect to \(y\) the first-order condition (20), we obtain the following expression for the slope of the constrained efficient queue length function

\[
q''(y) = \frac{\eta'(q^*(y))y}{-\eta''(q^*(y))y + 2p'(q^*(y)|d^*, S^*) + p''(q^*(y)|d^*, S^*)}.
\]

Notice that the differential equations (22) and (23) are identical, except that the function \(p\) that describes the price of an applicant is evaluated at \((d, S)\) in the first equation and at \((d^*, S^*)\) in the second equation. First, using the fact that \(p'(q|d, S)\) is increasing in \(S\) and decreasing in \(d\) and the fact that \(S > S^*\) and \(d < d^*\), we can prove that \(2p'(q|d, S) + p''(q|d, S)\)
is strictly greater than $2p'(q|d^*, S^*) + p''(q|d^*, S^*)$. This property implies that whenever $q(y) = q^*(y)$, we have $q'(y) < q''(y)$ and, hence, the equilibrium and second-best queue functions cross at most once. Moreover, using the fact that $q(y)dF(y)$ and $q^*(y)dF(y)$ both integrate up to one, we can prove that there must exist a point where $q$ and $q^*$ cross. Taken together the two observations imply that there always exist a $y_0$ such that $q(y_0) = q^*(y_0)$ and $q(y) > q^*(y)$ for all $y < y_0$ and $q(y) < q^*(y)$ for all $y > y_0$. That is, the equilibrium allocation assigns an inefficiently large number of applicants to low productivity firms and an inefficiently small number of applicants to high productivity firms.

The following proposition summarizes the comparison between the equilibrium and the second-best allocation.

**Proposition 2** (Inefficiency of Equilibrium Allocation): Let $(q^*, c^*, d^*, S^*)$ be the solution to the mechanism design problem with $\mu_2^* > 0$, and let $(q, c, d, S)$ be an equilibrium with $S > k$.

(i) The equilibrium allocation of consumption across workers in different employment states is constrained inefficient. In particular, $d < d^*$ and there exists a $q_0 > 0$ such that $c(q_0) = c^*(q_0), c(q) < c^*(q) \text{ for all } q < q_0, \text{ and } c(q) > c^*(q) \text{ for all } q > q_0.$

(ii) The allocation of applicants across different types of firms is constrained inefficient. In particular, there exists a $y_0 \in (y_c, \bar{y})$ such that $q(y_0) = q^*(y_0), q(y) > q^*(y) \text{ for all } y \in (y_c, y_0), \text{ and } q(y) < q^*(y) \text{ for all } y \in (y_0, \bar{y}).$

There is a simple intuition behind Proposition 2. In equilibrium, a worker expects an inefficiently large gain from searching for a job, $S > S^*$, and an inefficiently low consumption when his search is unsuccessful, $d < d^*$. Because of both inefficiencies, the risk premium that a firm has to pay in order to attract $q$ applicants increases too quickly with respect to $q$. In turn, this induces high-productivity to attract an inefficiently small number of applicants and—through general equilibrium effects—it induces low-productivity to attract an inefficiently large number of applicants. Moreover, the consumption enjoyed of workers employed at firms that have a large number of applicants is inefficiently high, while the consumption enjoyed by workers employed at firms that attract a small number of applicants is inefficiently low.
Proposition 2 has some important implications for aggregate variables. First, the proposition implies that aggregate output in the second best, \( Y^* = \int \eta(q^*(y))ydF(y) \), is greater than aggregate output in the equilibrium, \( Y = \int \eta(q(y))ydF(y) \). To see why this is the case, it is sufficient to notice that

\[
Y^* - Y = \int_{y_0}^{\eta} \left[ \int_{q(y)}^{q^*(y)} \eta'(q)ydy \right] dF(y) - \int_{y}^{\eta} \left[ \int_{q^*(y)}^{q(y)} \eta'(q)ydy \right] dF(y) \\
> \int_{y_0}^{\eta} \eta'(q^*(y_0))y_0 [q^*(y) - q(y)] dF(y) - \int_{y}^{\eta} \eta'(q^*(y_0))y_0 [q(y) - q^*(y)] dF(y) \\
= 0,
\]

where the second line makes use of the fact that \( q^*(y) > q(y) \) for all \( y > y_0 \) and \( q(y) > q^*(y) \) for all \( y < y_0 \), the third line makes use of the fact that \( \eta'(q)y > \eta'(q^*(y))y > \eta'(q^*(y_0))y_0 \) for all \( y > y_0 \) and \( \eta'(q)y < \eta'(q^*(y))y < \eta'(q^*(y_0))y_0 \) for all \( y < y_0 \), and the last equality makes use of the fact that \( q(y)dF(y) \) and \( q^*(y)dF(y) \) both integrate up to one. Intuitively, aggregate output in the equilibrium is inefficiently low because the marginal productivity of applicants at low-quality firms is inefficiently low and the marginal productivity of applicants at high-quality firms is inefficiently high.

Similarly, Proposition 2 implies that aggregate unemployment in the second best, \( U^* = 1 - \int \eta(q^*(y))dF(y) \), is lower than aggregate unemployment in the equilibrium, \( U = 1 - \int \eta(q(y))dF(y) \). To see this, notice that

\[
U^* - U = \int_{y_0}^{\eta} \left[ \int_{q(y)}^{q^*(y)} \eta'(q)ydy \right] dF(y) - \int_{y}^{\eta} \left[ \int_{q^*(y)}^{q(y)} \eta'(q)ydy \right] dF(y) \\
> \int_{y_0}^{\eta} \eta'(q(y_0)) [q^*(y) - q(y)] dF(y) - \int_{y}^{\eta} \eta'(q(y_0)) [q(y) - q^*(y)] dF(y) \\
= 0,
\]

where the third line uses the fact that \( \eta'(q) > \eta'(q(y)) > \eta'(q(y_0)) \) for all \( y < y_0 \) and \( \eta'(q) < \eta'(q(y)) < \eta'(q(y_0)) \) for all \( y > y_0 \), and the fourth line uses the fact that \( q(y)dF(y) \) and \( q^*(y)dF(y) \) both integrate up to one. Intuitively, in equilibrium, aggregate unemployment is inefficiently low because the job-filling rate \( \eta(q) \) is a concave function of \( q \) and \( q(y) \) is an inefficiently flat function of \( y \).
The equilibrium not only fails to attain the efficient level of aggregate output because it allocates applicants inefficiently, but it also fails to maximize the worker’s expected utility for any given distribution of applicants because it assigns an inefficiently large amount of consumption risk to workers. To see this, define \( L(q) \) and \( L^*(q) \) as the difference, respectively in the equilibrium and in the second best, between the expected utility enjoyed by a worker seeking employment at a firm that attracts \( q \) applicants and the utility of the expected consumption employed by that worker. Intuitively, \( L \) and \( L^* \) are measures of the utility loss due to the consumption risk faced by a worker joining a queue of applicants of length \( q \).

Formally, \( L \) and \( L^* \) are given by

\[
L(q) = \lambda(q)u(c(q)) + (1 - \lambda(q))d - u(\lambda(q)c(q) + (1 - \lambda(q))d), \\
L^*(q) = \lambda(q)u(c^*(q)) + (1 - \lambda(q))d^* - u(\lambda(q)c^*(q) + (1 - \lambda(q))d^*). 
\]

Using the fact that \( S^* < S \) and \( d^* > d \), it is easy to establish that \( L(q) > L^*(q) \) for all \( q > 0 \). In turn, this inequality implies that the equilibrium fails to maximize the worker’s expected utility for a given allocation of applicants across firms.

### 4 Policy implementation

In this section, we prove that the constrained efficient allocation can be implemented in equilibrium by introducing an appropriate choice of unemployment benefits, \( B_u \in \mathbb{R} \), and labor earnings taxes, \( T_e : \mathbb{R}_+ \to \mathbb{R} \). Moreover, we prove that the optimal unemployment benefit, \( B_u^* \), is positive and that the optimal labor earnings tax, \( T_e^* \), is regressive, in the sense that the marginal tax is decreasing in labor earnings.

First, we generalize the definition of equilibrium to an environment with labor market policies.

**Definition 2:** A competitive equilibrium with policy \((B_u, T_e)\) is an allocation \((e, q, d, S)\) that satisfies the following conditions:

(i) **Profit maximization:** For all \( y \in [\underline{y}, \overline{y}] \),

\[
e(y) \in \arg \max_{e \geq 0} \eta(q(e))(y - e);
\]
(ii) **Optimal number of applications:**

\[
\int q(e(y))dF(y) \leq 1 \quad \text{and} \quad S \geq k, \quad \text{with comp. slackness;}
\]

(iii) **Optimal direction of applications:** For all \( e \in \mathbb{R}_+ \),

\[
\lambda(q(e)) [u(d + e + T_e(e)) - u(d + B_u)] \leq S \quad \text{and} \quad q(e) \geq 0, \quad \text{with comp. slackness;}
\]

(iv) **Consistency of dividends and profits:**

\[
d = \int \eta(q(e(y))) [y - e(y)] dF(y);
\]

(v) **Balanced budget:**

\[
B_u = \int \eta(q(e(y))) [T_e(e(y)) - B_u] dF(y).
\]

The first four conditions in Definition 2 are nearly identical to those in Definition 1. The only difference is that, here, the consumption of an unemployed worker is given by the sum of dividend payments, \( d \), and unemployment benefits, \( B_u \), and the consumption of an employed worker is given by the dividend payment, \( d \), plus the after-tax wage paid by the firm, \( e - T_e(e) \). The last condition in Definition 2 guarantees that the budget of the policy-maker is balanced given the optimal behavior of firms and workers.

Now, we are in the position to formally state our first main result.

**Theorem 1** (Optimal Taxation): The constrained efficient allocation can be implemented as an equilibrium with the positive unemployment benefit \( B_u^* \) and the regressive labor earning tax \( T_e^* \). The optimal unemployment benefit \( B_u^* \) is equal to \( \mu_2^*/\mu_1^* \). For all \( e \leq y_c^* \) the optimal labor earning tax \( T_e^*(e) \) is equal to 0, and for all \( e > y_c^* \) it is such that

\[
T_e^*(e) = \left[ B_u^* + \frac{k}{u'(d^* - B_u^* + e - T_e^*(e))} \right]^{-1} B_u^*.
\] (27)

**Proof:** See Appendix D.

There is a simple intuition behind Theorem 1. As discussed in the previous section, the value of searching \( S \) plays two roles in the laissez-faire equilibrium. First, by affecting the
price of an applicant \( p(q|d, S) \), the value of searching guarantees that the number of applications demanded by firms is equal to the number of applications supplied by workers. Second, by affecting \( u(c) - u(d) \), the value of searching determines the extent of consumption risk faced by workers. The laissez-faire equilibrium is constrained inefficient precisely because the value of searching creates a tight link between the market-clearing price of an applicant and the consumption risk faced by workers. The unemployment benefit and the labor earnings tax break this link and make it possible for the equilibrium to be constrained efficient. In particular, the optimal unemployment benefit \( B_u^* \) lowers the equilibrium value of searching to \( S^* = k \), so as to minimize the workers’ risk. And the optimal labor earnings tax \( T_e^* \) rebalances demand and supply by aligning the private cost of an applicant to its social cost \( p^*(q|d^*, S^*) \).

In order to flash out the above intuition, it is useful to sketch the proof of Theorem 1. From the equilibrium conditions (i) and (ii) in Definition 2, it follows that the equilibrium queue of applicants attracted by a firm of type \( y \), \( q(y) \), is given by

\[
q(y) = \arg \max_q \eta(q)y - \hat{p}(q|d, S),
\]

where \( \hat{p}(q|d, S) \), \( e(q) \), \( w(q) \) and \( T(q) \) are defined as

\[
\hat{p}(q|d, S) = \eta(q)e(q)/q, \\
w(q) = u^{-1}(S/\lambda(q) + u(d + B_u)) - d, \\
e(q) = w(q) + T(q), \\
T(q) = T_e(e(q)).
\]

In words, \( q(y) \) maximizes the profits of a firm of type \( y \), given that the equilibrium price of an applicant is given by \( \hat{p}(q|d, S) = \eta(q)e(q)/q \), where \( e(q) = w(q) + T(q) \) is the gross wage that the firm needs to offer to attract \( q \) applicants, \( w(q) \) is the net wage that the firm needs to offer to attract \( q \) applicants and \( T(q) = T_e(e(q)) \) is the tax associated with a gross wage of \( e(q) \).

The equilibrium queue of applicants, \( q(y) \), satisfies the optimality condition

\[
\eta'(q) = \hat{p}(q|d, S) + q\hat{p}'(q|d, S)
\]

and \( q(y) \geq 0 \) with complementary slackness. Notice that the equilibrium price of an applicant, \( \hat{p}(q|d, S) \), is equal to \( p(q|d + B_u, S) + \lambda(q)(B_u + T(q)) \). Also, recall that the social cost
of an applicant, \( p^*(q|d^*, S^*) \), is equal to \( p(q|d^*, S^*) + \mu_2^*/\mu_1^* \). When \( d + B_u = d^* \), \( S = S^* \) and \( \lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^* \), the equilibrium price of an applicant is equal to the social cost of an applicant. This implies that (29) coincides with the first order condition (20) of the mechanism design problem and, hence, the equilibrium and the second-best allocation of applicants across firms are identical. Moreover, when \( d + B_u = d^* \), \( S = S^* \) and \( \lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^* \), the equilibrium condition (iii) coincides with the incentive compatibility constraint (11) in the mechanism design problem and, hence, the equilibrium and the second-best allocation of consumption across workers are identical. Overall, when \( d + B_u = d^* \), \( S = S^* \) and \( \lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^* \), the equilibrium is constrained efficient.

Now, we can use the sufficient conditions for the constrained efficiency of the equilibrium to recover the optimal unemployment benefit. In fact, if we solve \( \lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^* \) for \( T(q) \) and substitute the solution into the equilibrium condition (v), we obtain

\[
B_u = \int \frac{\mu_2^* q^*(y)}{\mu_1^*} dF(y) = \frac{\mu_2^*}{\mu_1^*}.
\] (30)

The optimal unemployment benefit is equal to the ratio between the multiplier on the resource constraint on applicants, \( \mu_2^* \), and the multiplier on the resource constraint on output, \( \mu_1^* \), in the mechanism design problem. This finding has a simple interpretation. When the equilibrium decentralizes the constrained efficient allocation, the worker’s private value from sending a job application is \( S^* = k \), but the social value of sending a job application exceeds \( k \) by \( \mu_2^*/\mu_1^* \). This excess value is redistributed to all the workers in a lump-sum fashion. Unemployed workers receive \( \mu_2^*/\mu_1^* \) directly in the form of an unemployment benefit, while employed workers receive \( \mu_2^*/\mu_1^* \) indirectly in the form of the wage premium offered by the firms to compensate them for the loss of the unemployment benefit. In this sense, the unemployment benefit serves the role of minimizing the value of searching and, hence, the consumption risk faced by workers.

Next, we can use the sufficient conditions for the constrained efficiency of the equilibrium to recover the optimal labor earning tax. In fact, from \( \lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^* \) and \( B_u = \mu_2^*/\mu_1^* \), we obtain

\[
T_e(e(q)) = T(q) = \left( \frac{1}{\lambda(q)} - 1 \right) B_u.
\] (31)

There are a couple of properties of the optimal labor tax that are worth stressing. First,
recall that the optimal labor earning tax guarantees that the cost of an applicant to a firm, \( \hat{p}(q|d, S) \), to the cost of an applicant to society, \( p^*(q|d, S) \). In this sense, the labor earning tax is Pigovian. Second, notice that the optimal labor tax is such that the tax revenues paid by the workers who apply successfully to the firm is equal to the unemployment benefits paid to the workers who apply unsuccessfully to the firm, i.e. \( \lambda(q)T(q) = (1 - \lambda(q))B_u \). This property implies that the optimal policy only involves redistribution of resources from the successful to the unsuccessful applicants of the same firm and not across workers employed at different firms. Intuitively, this property follows from the fact that the wage differences among workers employed in different jobs are compensating differentials for different job-finding probabilities.

Differentiating (31) with respect to \( q \), we obtain

\[
T_e'(e(q)) = \frac{T'(q)}{e'(q)} = \left[ B_u + \frac{S^*}{u'(e(q) - T_e(e(q)))} \right]^{-1} B_u, \tag{32}
\]

where the second line makes use of the equilibrium condition (iii). A number of important properties of the optimal labor earning tax emerge from (32). First, notice that the optimal marginal tax is always strictly positive and strictly smaller than one. Second, notice that the optimal marginal tax is decreasing in earnings because \( e - T(e) \) is strictly increasing in \( e \). In this sense, the optimal labor earning tax is regressive.

The optimality of a regressive labor earning tax is a startlingly sharp result. Yet, this result follows immediately from two properties of the optimal policy: (a) redistribution only takes place between the successful and the unsuccessful applicants at the same type of firm; (b) workers are ex-ante indifferent between applying to different types of firms. To see this clearly, first notice that these two properties are

\[
T_e(e(q)) = \left( \frac{1}{\lambda(q)} - 1 \right) B_u,
\]

\[
u(d + e(q) - T_e(e(q))) - u(d) = \frac{S^*}{\lambda(q)},
\]

and they imply

\[
T_e(e) = \left[ \frac{u(d + e - T_e(e) - u(d + B_u)}{S^*} - 1 \right] B_u. \tag{33}
\]

Then, notice that (33) can only be satisfied by a regressive labor earning tax because any
progressive tax would make the left-hand side of (33) convex and the right-hand side of (33) concave with respect to the worker’s earnings $e$.

5 Insurance market implementation

In this section, we prove that the constrained efficient allocation does not have to be implemented through centralized policy interventions. Indeed, we prove that the constrained efficient allocation can also be implemented by setting up the appropriate markets and letting them operate competitively. In particular, we find that, by opening a competitive insurance market where workers can purchase insurance against the consumption risk associated with the job search process, the constrained efficient allocation can be implemented as a laissez-faire equilibrium.

5.1 Environment and equilibrium

We consider an economy with an insurance market operating alongside the labor market. The insurance market is populated by a continuum of insurance companies. We assume that each insurance company offers contracts of the type $(e, t_u, t_e)$, where $e \in \mathbb{R}^+$ denotes the wage that the insurance company asks the worker to seek, $t_u \in \mathbb{R}$ denotes the payment that the insurance company makes to the worker if his search is unsuccessful, and $t_e \in \mathbb{R}$ denotes the payment that the worker makes to the insurance company if his search is successful. Without loss in generality, we assume that each insurance company only offers contracts $(e, t_u, t_e)$ that satisfy the worker’s participation constraint—in the sense that the terms of the contract induce the worker to participate—and that satisfy the worker’s incentive compatibility constraint—in the sense that the terms of the contract induce the worker to seek the prescribed wage. Each insurance company chooses which contracts to offer taking as given the equilibrium queue length $q(e)$. The labor market is populated by firms and workers and operates as in Section 2. Each firm chooses which wage to offer and each worker chooses which wage to seek taking as given the equilibrium queue length, $q(e)$, and the equilibrium insurance contracts, $(t_u(e), t_e(e))$.

Now, we are in the position to define a competitive equilibrium.
Definition 3: A competitive equilibrium is an allocation \((e, q, t_u, t_e, d, U)\) that satisfies the following conditions:

(i) Firm’s profit maximization: For all \(y \in [\underline{y}, \overline{y}]\),

\[ e(y) \in \arg\max_{e \geq 0} \eta(q(e))(y - e); \]

(ii) Insurance company’s profit maximization: For all \(e \in \mathbb{R}_+\),

\[ (t_u(e), t_e(e)) \in \arg\max_{(t_u, t_e)} \lambda(q(e))t_e - (1 - \lambda(q(e)))t_u, \text{ s.t.} \]
\[ \lambda(q(e))u(d + e - t_e) + (1 - \lambda(q(e)))u(d + t_u) - k \geq U, \]
\[ \lambda(q(e))[u(d + e - t_e) - u(d + t_u)] \geq k; \]

(iii) Zero profits in the insurance market: For all \(e \in \mathbb{R}_+\),

\[ \lambda(q(e))t_e(e) - (1 - \lambda(q(e)))t_u(e) = 0 \text{ and } q(e) \geq 0, \text{ with comp. slackness;} \]

(iv) Optimal number of applications:

\[ \int q(e(y))dF(y) \leq 1 \text{ and } U \geq u(d), \text{ with comp. slackness;} \]

(v) Optimal direction of applications: For all \(e \in \mathbb{R}_+\),

\[ \lambda(q(e))u(d + e + T_e(e)) + (1 - \lambda(q(e)))u(d + B_u) - k \leq U \]

and \(q(e) \geq 0\), with complementary slackness;

(vi) Consistency of dividends and profits:

\[ d = \int \eta(q(e(y)))) [y - e(y)] dF(y). \]

Condition (i) guarantees that the wage \(e(y)\) maximizes the profit of a firm of type \(y\). Condition (ii) guarantees that \((e, t_u(e), t_e(e))\) is the feasible contract that maximizes the profits of an insurance company. To see why this is the case, notice that a contract \((e, t_u, t_e)\) satisfies the worker’s participation constraint if \(\lambda(q(e))u(d + e - t_e) + (1 - \lambda(q(e)))u(d + t_u) - k \geq U\). A contract \((e, t_u, t_e)\) satisfies the worker’s incentive compatibility constraint if \(\lambda(q(e))[u(d + e - t_e) - u(d + t_u)] \geq k\). And, if the contract \((e, t_u, t_e)\) is feasible, the
insurance company obtains a profit of $\lambda(q(e))t_e - (1 - \lambda(q_e))t_u$. Condition (iii) guarantees that competition in the insurance market drives the profits generated by the equilibrium contract $(t_u(e), t_e(e))$ down to zero. Conditions (iv) and (v) guarantee that the measure of applications received by firms offering different wages is consistent with the workers’ utility maximization. Finally, condition (vi) guarantees that the dividends received by the workers are equal to the profits of the firms. As in Sections 2 and 4, we denote with $q(y)$ the number of applicants attracted by a firm of type $y$, and with $e(q)$ the wage that a firm needs to offer in order to attract $q$ applicants.

5.2 Constrained efficiency of equilibrium

The following theorem is the second major result of our paper.

**Theorem 2** (Efficiency of Equilibrium with Insurance Markets): Consider the allocation $(q, e, t_u, t_e, d, U)$, where $q(y) = q^*(y)$, $e(q) = u^{-1}(S^*/\lambda(q) + u(d^*)) + (\mu_2^*/\mu_1^*)/\lambda(q) - d^*$, $t_u(e) = B_u^*$, $t_e(e) = T_e^*(e)$, $d = d^* - B_u^*$, and $U = u(d^*)$. (i) The allocation $(q, e, t_u, t_e, d, U)$ is a competitive equilibrium for the version of the economy with insurance markets. (ii) The competitive equilibrium $(q, e, t_u, t_e, d, U)$ is constrained efficient.

*Proof:* See Appendix E.

The second part of Theorem 2 is not surprising. In the proposed equilibrium, the equilibrium insurance contracts exactly reproduce the optimal tax system. That is, the equilibrium insurance contracts are such that, if the worker remains unemployed, the payment from the insurance company to the worker, $t_u(e)$, is equal to the optimal unemployment benefit, $B_u^*$. And if the worker finds a job at a firm offering the wage $e$, the payment from the worker to the insurance company, $t_e(e)$, is equal to the optimal labor earning tax, $T_e^*(e)$. Since the optimal tax system $(B_u^*, T_e^*)$ implements the constrained efficient allocation, it is not surprising that an equilibrium in which insurance contracts reproduce the optimal tax system is also constrained efficient.

The first part of Theorem 2 is less obvious. In particular, one might wonder why insurance companies, in equilibrium, choose to offer contracts that exactly reproduce the optimal tax system $(B_u^*, T_e^*)$. To answer the question, consider the profit maximization problem of an
insurance company

\[
\begin{align*}
\max_{(t_u, t_e)} \lambda(q)t_e - (1 - \lambda(q))t_u, \\
\text{s.t.} \quad & \lambda(q)u(d + e(q) - t_e) + (1 - \lambda(q))u(d + t_u) - k \geq U, \\
& \lambda(q) [u(d + e(q) - t_e) - u(d + t_u)] \geq k.
\end{align*}
\]  

Notice that the contract \((t_u, t_e)\) that maximizes the profits of the insurance company must be such that the worker’s participation constraint holds with equality. Otherwise, by lowering \(t_u\) by some small amount \(\delta\), the insurance company would still be able to satisfy the worker’s participation constraint, it would relax the worker’s incentive compatibility constraint, and it would increase its profits. Similarly, notice that the contract \((t_u, t_e)\) that maximizes the profits of the insurance company must be such that the worker’s incentive compatibility constraint holds with equality. Otherwise, there would exist \(\epsilon\) and \(\delta\) such that—by increasing \(t_u\) by \(\epsilon\) and by reducing \(t_e\) by \(\epsilon (1 - \lambda(q))/\lambda(q) - \delta\)—the insurance company would still satisfy the worker’s incentive compatibility constraint, it would relax the worker’s participation constraint and it would increase its profits.

Since the contract \((t_u, t_e)\) that maximizes the profits of an individual insurance company must satisfy both the participation and the incentive compatibility constraints with equality, it follows that \(u(d + t_u) = U\) and \(\lambda(q) [u(d + e(q) - t_e) - U] = k\). Moreover, since \(U = u(d + B_u^*)\) and \(e(q) = u^{-1}((S^*/\lambda(q) + u(d^*)) + T^*_e(e(q)) - d)\), it follows that \(t_u\) is equal to \(B_u^*\) and \(t_e\) is equal to \(T^*_e(e(q))\). That is, the contract that maximizes the profits of an individual insurance company reproduces the optimal tax system. Finally, since the optimal tax system has the property that revenues and expenditures are balanced job-by-job, i.e. \(\lambda(q)T^*_e(e(q)) = (1 - \lambda(q))B_u^*\), it follows that the maximized profits of an insurance company are equal to zero. Thus, the profit-maximizing contract is consistent with perfect competition in the insurance market.

Theorem 2 allows us to better understand some of our previous results. First, Proposition 1 showed that the laissez-faire equilibrium of the baseline economy is constrained inefficient. In light of Theorem 2, it is clear that the fundamental cause of this inefficiency is the absence of a competitive insurance market. Second, Theorem 1 showed that the appropriate combination of unemployment benefits and labor earnings taxes can restore the efficiency of the equilibrium. In light of Theorem 2, it becomes clear that these policies serve the
purpose of completing the markets by offering workers the same contingencies that they could purchase in a competitive insurance market. Our analysis does not tell us why, in reality, we do not see insurance markets that protect workers against the risks associated with the search for the right occupation or the right job. What our analysis tells us is that, absent such markets, it is optimal to introduce a positive unemployment benefit and a regressive labor earning tax.

6 Conclusions

In this paper, we studied the optimal redistribution of income inequality caused by the presence of search and matching frictions in the labor market. We carried out the analysis using a directed search model of the labor market populated by homogenous workers and heterogeneous firms. In the first part of the analysis, we characterized and compared the equilibrium allocation and the constrained efficient allocations. We found that the equilibrium is not constrained efficient because workers are not insured against the risk of not finding the job that they seek. As a result of this lack of insurance, the equilibrium number of workers seeking employment at high-productivity firms is inefficiently small, while the equilibrium number of workers seeking employment at low-productivity firms is inefficiently large. Moreover, the consumption of an employed worker is an inefficiently steep function of the number of workers who apply for the same type of job. In the second part of the analysis, we proved that the constrained efficient allocation can be implemented by introducing a positive unemployment benefit and an increasing and regressive labor income tax. We argued that the unemployment benefit serves the purpose of lowering the search risk faced by workers, and that the labor tax serves the purpose of aligning the cost of an applicant to the firm with the opportunity cost of an applicant to society.

In this paper, we studied the optimal redistribution of residual labor income inequality in the context of a simple model of the labor market. The simplicity of our model afforded us a clear exposition of the properties and of the role of the optimal unemployment benefits and of the optimal labor income tax. However, in order to make substantive policy recommendation, we would have to enrich the model along several dimensions. First, we would have to consider a dynamic environment. Second, since many workers move from one employer to the other
without an intervening spell of unemployment, we would have to consider an environment in which workers can search both off and on the job. Finally, since income inequality is likely to be caused by both productivity differences and by search frictions, we would have to introduce some degree of inherent heterogeneity among workers.
References


