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

April 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. 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 to the firm of attracting an additional applicant with the value of an application to society.

JEL Codes: H21, J64, J65.

Keywords: Unemployment benefit, Income tax, Search frictions, Mechanism design.

*Golosov: Department of Economics, Princeton University, 111 Fisher Hall, Princeton, NJ 08544 (email: Golosov@princeton.edu); Maziero: Department of Finance, Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19103 (email: maziero@wharton.upenn.edu); Menzio: Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104 (email: gmenzio@sas.upenn.edu). We are grateful to Jim Albrecht, Paul Beaudry, Michele Boldrin, Jan Eeckhout, Pieter Gautier, Veronica Guerrieri, Espen Moen, Dale Mortensen, Giuseppe Moscarini, Nicola Pavoni, Tom Sargent, Rob Shimer and Randy Wright for their comments on earlier versions of the paper. We have also received useful comments from seminar participants at NYU, UW Madison, UBC, UAB, EUI, WashU, Queen's, Science Po, Southampton and from participants at the Northwestern Conference in Honor of Dale Mortensen (Evanston 2011), the NBER Summer Institute (Cambridge 2011), the Tinbergen Institute Conference on Friction and Policy (Amsterdam 2011) and the American Economic Association Meeting (Chicago 2012).

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., Mortensen 2003 or Autor, Katz and Kerney 2008). 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[1] (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,

[1] In a recent paper, Hornstein, Krusell and Violante (2011) argue that properly calibrated search models cannot account for much residual income inequality. Their analysis applies to a class of search models that does not include the one that we consider in this paper. More importantly, we do not need to take a stand on the magnitude of search-based income inequality, as we are mostly interested in understanding the optimal way to redistribute this type of inequality, be it either large or small.

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 income 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 income tax is not to redistribute consumption among employed workers, but

rather to make sure that the cost to a firm of attracting an additional applicant reflects the value of an applicant to society.

The fact that the optimal labor income tax is regressive is the main result of the paper. This result is startlingly robust, 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. Rather, this result is a necessary implication of two properties of the equilibrium with optimal policy. First, redistribution only takes place between workers who successfully and unsuccessfully search for the same wage, and not between workers who search for different wages. Second, workers are ex-ante indifferent between searching for different wages. Both properties are related to the directed nature of search. With directed search, wage differentials among employed workers do not reflect luck, but compensation for different job-finding probabilities. For this reason, there is no redistribution among workers employed at different wages. Moreover, with directed search, workers choose which wage to seek and, hence, they must be ex-ante 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 or Shaal and Taschereau-Dumouchel 2012). 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 redis-

tribution 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 Laissez-faire equilibrium

In this section, we lay out our directed search model of the labor market, which is in the spirit of Moen (1997) and Acemoglu and Shimer (1999). We characterize the equilibrium allocation and we show that the model generates labor income inequality among inherently identical workers. In particular, the model generates income inequality between employed and unemployed workers and across workers employed by different firms.

2.1 Environment

The economy is populated by a continuum of homogeneous workers with measure 1. Each worker has preferences that 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. Each worker is endowed with 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 $y \in [\underline{y}, \overline{y}]$, $0 < \underline{y} < \overline{y}$, and the measure of firms with type less than y is denoted by $F(y)$, where $F : [\underline{y}, \overline{y}] \to [0, m]$ is a twice differentiable and strictly increasing function with boundary conditions $F(\underline{y}) = 0$ 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.[2] 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 w 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 λ and η satisfy the aggregate consistency condition $\lambda(q)q = \eta(q)$.[3] If a firm of type y matches

[2] Qualitatively, the main result of the paper would be unchanged if we were to assume that firms are owned by entrepreneurs rather than workers. That is, if firms were owned by workers, the optimal unemployment benefit would still be positive and the optimal labor earning tax would still be increasing and regressive. Different points on the utility possibility frontier of workers and entrepreneurs would be attained by lump-sum transfers from one group of agents to the other.

[3] All of the matching processes commonly used in the literature satisfy these assumptions on λ and η. For example, the assumptions are satisfied by the urn-ball matching function $\lambda(q) = (1 - \exp(-q))/q$, by the telephone-line matching function $\lambda(q) = 1/(1 + q)$, and by the constant-elasticity of substitution matching function $\lambda(q) = (1 + q^\sigma)^{-1/\sigma}$.

with a worker, it produces y units of output, it pays the wage w to the worker and it pays the dividend $y - w$ 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, z, 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, $z \in \mathbb{R}$, 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, z, S) *that satisfies the following conditions:*

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

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

(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* $w \in \mathbb{R}_+$,

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

(iv) *Consistency of dividends and profits:*

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

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 applications 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, $\lambda(q(w))(u(z + w) - u(z))$, 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.[4] 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 $p(q)$ be defined as the ratio between the expected wage bill paid by a firm, $\eta(q)w(q)$, and the number of applicants attracted by a firm, q, where $w(q)$ denotes the wage that the firm needs to offer to attract q applicants. We shall refer to $p(q)$ as the *price of an application.* From the equilibrium condition (iii) and the consistency condition $\eta(q) = \lambda(q)q$, it follows

[4] As in Moen (1997) and Acemoglu and Shimer (1999), we impose condition (iii) not only for wages that are posted in equilibrium, but also for wages that are off the equilibrium path. For equilibrium wages, condition (iii) guarantees that q is consistent with the worker's optimal search behavior. For off-equilibrium wages, condition (iii) imposes a restriction on the firms' beliefs about q. The restriction is in the spirit of subgame perfection: when a firm entertains posting an off-equilibrium wage, it expects to attract a queue length that is optimal from the workers' perspective.

that $p(q)$ is given by

$$\begin{aligned} p(q|d,S) &= \lambda(q)\left[u^{-1}\left(\frac{S}{\lambda(q)}+u(z)\right)-z\right] \\ &\equiv \phi(q,z,S). \end{aligned} \tag{1}$$

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. It is useful to denote as $\phi(q,z,S)$ the right-hand side of (1).

The price of an application $p(q)$ is the key object to understand the welfare properties of the laissez-faire equilibrium and the difference between the equilibrium allocation and the first and second best allocations. First, notice that $p(q)$ is increasing in q, i.e. $p'(q) = \phi_q(q,z,S) > 0$. This property follows from the fact that workers are risk averse. Given two jobs that offer the same expected payment, a risk-averse worker strictly prefers applying for the job that offers a lower wage and that is easier to get. That is, given (w_1,q_1) and (w_2,q_2) such that $q_1 < q_2$ and $\lambda(q_1)w_1 = \lambda(q_2)w_2$, a risk-averse worker strictly prefers the safer job (w_1,q_1) to the riskier job (w_2,q_2). This implies that, if a firm wants to attract a longer queue, it has to offer a higher expected payment to each of its applicants to compensate them for the additional risk they face. Second, notice that $p(q)$ is increasing in the equilibrium value of searching S, i.e. $\phi_S(q,z,S) > 0$. Intuitively, the higher is S the higher is the expected payment that a firm has to offer to each of its applicants.

The derivative $p'(q)$ measures the premium that compensates workers for the extra risk associated with joining a marginally longer queue. Notice that the marginal risk premium $p'(q)$ is increasing in the equilibrium value of searching S, i.e. $\phi_{qS}(q,z,S) > 0$. Intuitively, the higher is S, the larger is the difference between the consumption of the worker if his application succeeds and if his application fails. Hence, due to the concavity of u, the worker requires a higher wage increase in order to be willing to join a marginally longer queue. Also, notice that the marginal risk premium $p'(q)$ may be increasing or decreasing in z depending on the shape of u, i.e. $\phi_{qz}(q,z,S)$ may be positive or negative. The results in Section 3 are derived under the assumption that $p'(q)$ is decreasing in z.[5] The other results, including the

[5] The assumption is satisfied when the utility function u has the HARA form

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

two main theorems in Sections 4 and 5, do not require any assumption about the relationship between $p'(q)$ and z.

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

$$q_y = \arg\max_{q \geq 0} \eta(q)y - p(q)q. \tag{2}$$

For all $y \in [\underline{y}, \overline{y}]$, q_y satisfies the first order condition

$$\eta'(q)y \leq p(q) + p'(q)q \tag{3}$$

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

$$c_y = u^{-1}\left(\frac{S}{\lambda(q_y)} + u(z)\right). \tag{4}$$

Since q_y is strictly increasing in y, equation (4) implies that the consumption of an employed worker, c_y, is increasing in the productivity of his employer, y. Also, since $S \geq k$ and $k > 0$, equation (4) implies that the consumption of an employed worker, c_y, is strictly greater than the consumption of an unemployed worker, z.

Next, we characterize the consumption z of an unemployed worker. From equilibrium

and the parameter γ belongs to the interval $[1/2, 1]$.

condition (iv) and equation (4), it follows that z is such that

$$\int \eta(q_y) y dF(y) = z + \int \eta(q_y)(c_y - z) dF(y). \tag{5}$$

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)—is equal to aggregate consumption—which is the term on the right-hand side of (5).

Finally, the equilibrium condition (ii) states that the 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

$$\int q(y) dF(y) \leq 1 \tag{6}$$

and $S \geq k$ with complementary slackness.

Overall, any equilibrium can be represented as a tuple (q, c, z, S) that satisfies the system of equations (3)–(6). Notice that the equilibrium does not attain the first-best allocation—i.e. the allocation that maximizes the workers' expected utility given the production technology $F(y)$ and the matching technology $\lambda(q)$. In fact, in the first best allocation, the marginal productivity of applicants is equalized across firms so as to maximize aggregate output, and the marginal utility of consumption is equalized across workers so as to maximize the workers' expected utility given aggregate output. In contrast, in the equilibrium allocation, the marginal productivity of applicants is different for different types of firms. In particular, the marginal productivity of applicants at low-y firms is lower than the marginal productivity of applicants at high-y firms. Moreover, in the equilibrium allocation, the marginal utility of consumption is different across workers in different employment states. In particular, the marginal utility of unemployed workers is higher than the marginal utility of employed workers, and the marginal utilities of workers employed at low-y firms is higher than the marginal utility of workers employed at high-y firms.

There are two reasons why the equilibrium allocation differs from the first-best allocation: the fact that workers and firms trade labor rather than job applications and the fact that workers are risk averse rather risk neutral. To see this, suppose that workers and firms could trade job applications at some competitive price p. If that was the case, every worker

would sell his application, enjoy a consumption of $z + p$, and have a marginal utility of $u'(z+)$. Moreover, every firm would purchase applications up to the point where the marginal productivity $\eta'(q)y$ is equal to p. Hence, the equilibrium would implement the first-best allocation. However, firms and workers cannot trade job application because firm cannot observe workers who search unsuccessfully for their vacancies (and even if they did, they would have no incentive to truthfully report that). Instead, firms and workers can only trade successful applications. That is, they can only trade labor.

When workers and firms trade labor, workers face some consumption risk because their search is rewarded only when it is successful. If workers are risk neutral, this "search risk" does not matter. In this case, a market for labor is equivalent to a market for applications and its equilibrium attains the first best.[6] If workers are risk averse, the existence of search risk implies that the marginal utility of consumption is not equalized among workers who are in different employment states. Moreover, the existence of search risk implies that workers demand a premium in order to seek jobs that attract longer queues of applicants. Hence, $p'(q) > 0$ and the marginal productivity of applicants is not equalized across different types of firms. Overall, under risk aversion, the equilibrium in a market where firms and workers trade labor does not attain the first best.

3 Constrained efficient allocation

In the previous section, we established that the laissez-faire equilibrium cannot implement the first best allocation. In this section, we ask whether the equilibrium can implement the second best allocation, i.e. the allocation that maximizes the worker's expected utility given the production technology $F(y)$, the matching technology $\lambda(q)$, and given that workers have private information about their application strategy. To answer this question, we set up the mechanism design problem, we solve it and we compare the solution to the equilibrium allocation. We find that, unless the search cost k is too high, the equilibrium does not implement the second best. This is because workers face a search risk which is not insured at all in the equilibrium, while it is partially insured in the second best.

[6] In Montgomery (1991), Moen (1997), Burdett Shi and Wright (2001) and Menzio and Shi (2011), workers and firms trade labor. These papers find that the equilibrium attains the first best allocation only because workers are assumed to be risk neutral.

3.1 Formulation of the mechanism design problem

The problem facing the mechanism designer is to maximize the workers' expected utility by choosing the search strategy that workers should follow and the consumption that workers should receive conditional on the outcome of their search. The search strategy chosen by the mechanism designer is subject to an incentive compatibility constraint, because the mechanism does not know whether a worker applies for a job and, if so, where he applies for a job. The consumption profile chosen by the mechanism designer is subject to a resource constraint, because the amount of resources assigned to workers cannot exceed the amount of resources produced by firms. Moreover, we restrict the mechanism designer to choose a symmetric mechanism, i.e. a mechanism that recommend the same search strategy to all workers.[7]

More specifically, a symmetric mechanism is a tuple (π, c, d, S). The first element of the tuple is a differentiable function $\pi : [\underline{y}, \overline{y}] \to \mathbb{R}_+$, where $\pi(y)$ denotes the probability with which a worker should apply to a firm of type $\tilde{y} \leq y$. The second element of the tuple is a function $c : [\underline{y}, \overline{y}] \to \mathbb{R}_+$, where $c(y)$ denotes the consumption assigned to a worker who is employed by a firm of type y. The third element of the tuple, $z \in \mathbb{R}_+$, denotes the consumption assigned to a worker who is unemployed. Finally, $S \in \mathbb{R}_+$, denotes the worker's maximized value of sending an application. Notice that $\pi(y)$ uniquely determines the queue of applicants $q(y)$ at firms of type y, $q(y) = \pi'(y)/F'(y)$, as well as the measure of workers a who apply for jobs, $a = \pi(\overline{y})$.. Hence, we will think of the mechanism as a tuple (a, q, c, z, S) rather than a tuple (π, c, d, S).

The mechanism designer chooses (a, q, c, z, S) so as to maximize the workers' expected utility

$$\int \left[\lambda(q_y)u(c_y) + (1 - \lambda(q_y))\, u(z) - k\right] q_y dF(y) + (1 - a)u(z). \tag{7}$$

There are $q_y dF(y)$ workers applying to firms of type y. Each one of these workers has a probability $\lambda(q_y)$ of becoming employed and consuming c_y units of output, and a probability $1 - \lambda(q_y)$ of remaining unemployed and consuming z units of output. In either case, the

[7] The restriction to symmetric mechanisms is quite natural. In fact, the literature on the game theoretic foundation of directed search argues that search frictions emerge precisely when workers follow symmetric strategies. In particular, when all workers apply to different firms with the same probability, some firms will end up with too many applicants and some firms will end up with not enough applicants (see, e.g., Burdett, Shi and Wright 2001, Shimer 2005 and Galenianos and Kircher 2012).

worker incurs the disutility cost k. There are also $1-a$ workers who do not apply for jobs. Each one of these workers consumes z units of output.

The mechanism designer's choice of (a, q, c, d, S) must be technologically feasible. First, the mechanism must be such that the aggregate output produced by the firms is greater than the aggregate consumption enjoyed by the workers. That is,

$$\int \eta(q_y) y dF(y) - \int \left[\lambda(q_y)c_y + (1-\lambda(q_y))z\right] q_y dF(y) - (1-a)z \geq 0. \tag{8}$$

Second, the mechanism must be such that the measure of workers applying for jobs is smaller than one and it is equal to the measure of applications received by the firms. That is,

$$1-a \geq 0, \tag{9}$$

$$a - \int q_y dF(y) = 0. \tag{10}$$

The mechanism designer's choice of (a, q, c, d, S) must be compatible with the workers' incentives to follow the recommended mixed application strategy. Hence, a worker must be indifferent between taking any one of the actions to which the mechanism assigns positive probability, and he must prefer taking any one of these actions rather than an action to which the mechanism assigns zero probability. This implies that, if y is such that $q_y > 0$, the worker's expected utility from searching for a firm of type y must be equal to the maximized value of searching and must be greater than the value of not searching. That is,

$$\lambda(q_y)\left[u(c_y) - u(z)\right] - S = 0, \tag{11}$$

$$S - k \geq 0. \tag{12}$$

Instead, if y is such that $q_y = 0$, the worker's expected utility from searching for a firm of type y must be smaller than the maximized value of searching. That is, $\lambda(q_y)\left[u(c_y) - u(z)\right] \leq S$. Moreover, if $a < 1$, the worker's expected utility from not searching must be equal to the maximized value of searching. That is,

$$(1-a)(S-k) = 0. \tag{13}$$

Before characterizing the solution to the above mechanism design problem, some com-

ments are in order. The mechanism asks workers to randomize over different application strategies. Then the mechanism assigns consumption to workers conditional on the outcome of their application (i.e. whether they are employed or unemployed and at which firm they are employed). The mechanism cannot condition consumption on the workers' application strategy (i.e. whether and where the workers apply for a job) because this strategy is their private information. However, if workers were allowed to make a report about the outcome of their randomization, the mechanism could condition consumption not only on the workers' employment status, but also on the workers' reported search strategy. In Appendix B, we consider this alternative version of the mechanism design problem. We find that the solution of the mechanism design problem with workers' reports is the same as the solution of the problem without reports. Intuitively, the mechanism cannot exploit the reports because the outcome of the randomization is the workers' private information.

3.2 Solution to the mechanism design problem

Let $(a^*, q^*, c^*, d^*, S^*)$ denote the solution to the mechanism design problem. Also, let μ_1^* denote the multiplier associated with the aggregate resource constraint on output (8) and let μ_2^* denote the multiplier associated with the aggregate resource constraint on applications (9). We shall refer to the solution to the mechanism design problem as either the constrained efficient allocation or the second best allocation.

The constrained efficient value of searching is[8]

$$S^* = k. \tag{14}$$

To understand this result recall that S is the reward that a worker expects to obtain when sending a job application. For any $S < k$, a worker would have no incentive to apply for a job and, hence, production and consumption would be zero. For any $S \geq k$, the worker has the incentive to apply for a job. However, since the worker is rewarded for sending an application only if his application is successful, the higher is S the greater the consumption risk that he faces. Hence, the constrained efficient value of S is k.

[8] Appendix C contains the explicit derivation of the optimality conditions (14)-(18).

The constrained efficient queue length is such that

$$\eta'(q_y^*)y - \lambda(q_y^*)(c_y^* - z^*) + q_y^*\lambda'(q_y^*)\left[\frac{u(c_y^*) - u(z^*)}{u'(c_y^*)} - (c_y^* - z^*)\right] = \frac{\mu_2^*}{\mu_1^*}, \tag{15}$$

and $q_y^* \geq 0$, with complementary slackness. The left-hand side of (15) is the difference between the value of making the worker apply to a firm of type y and the value of making the worker not search. The first term on the left-hand side of (15) is the output produced by the worker if he applies to a firm of type y. The second term is the negative of the amount of output that needs to be assigned to the worker in order to compensate him for the disutility of searching. The last term is the negative of the amount of output that needs to be assigned to the other workers who apply to type-y firms in order to compensate them for the decline in the probability of being hired that is caused by the marginal applicant. The right-hand side of (15) is the maximum difference between the value of making a worker apply for a job and the value of not making her search. That is, the right-hand side of (15) is the net value of an application. Overall, (15) implies that the constrained efficient queue length is such that the net value of the marginal application is equalized across different types of firms.

The constrained efficient allocation of consumption is

$$c_y^* = u^{-1}\left(\frac{k}{\lambda(q_y^*)} + u(z^*)\right), \tag{16}$$

$$z^* = \int \eta(q_y^*)\left[y - \left(c_y^* - z^*\right)\right] dF(y). \tag{17}$$

The consumption assigned to employed workers, c_y^*, is such that an applicant is indifferent between searching for a firm of type y and not searching at all. The consumption assigned to unemployed workers is such that aggregate output equals aggregate consumption.

Finally, the value of an application μ_2^*/μ_1^* is such that

$$\int q_y^* dF(y) \leq 1 \tag{18}$$

and $\mu_2^*/\mu_1^* \geq 0$, with complementary slackness. If the constrained efficient allocation is such that some workers apply for jobs and some do not, then the value of an application is zero. Conversely, if the value of an application is strictly positive, then the constrained efficient allocation is such that all workers apply for jobs.

In order to understand the relationship between the constrained efficient allocation and the equilibrium allocation, it is useful to notice that the optimality condition for q_y^* can be written as

$$\eta'(q_y^*)y = p^*(q_y^*) + p^{*\prime}(q_y^*)q_y^*, \tag{19}$$

where

$$p^*(q) = \phi(q, z^*, k) + \frac{\mu_2^*}{\mu_1^*}. \tag{20}$$

That is, the constrained efficient queue length q_y^* is the same queue length that a profit-maximizing firm would choose if the price of an application was $p^*(q)$, where $p^*(q)$ is the sum of two components. The first component, $\phi(q, z^*, k)$, is the amount of output that compensates a worker for the disutility of sending an application. The second component, μ_2^*/μ_1^*, is the value of an application to society (net of the disutility of sending the application). In contrast, the equilibrium queue length q_y maximizes the profits of the firm given that the price of an application is $p(q) = \phi(q, z, S)$.

From the previous observation, it follows that the constrained efficient allocation can be decentralized as an equilibrium when $\mu_2^*/\mu_1^* = 0$, but not when $\mu_2^*/\mu_1^* > 0$. In fact, when $\mu_2^*/\mu_1^* = 0$, q_y^* satisfies the optimality condition (19), which is the same as the equilibrium condition (3); c_y^* satisfies the optimality condition (16), which is the same as the equilibrium condition (4); z^* satisfies the optimality condition (17), which is the same as the equilibrium condition (5); and $S^* = k$ satisfies the equilibrium condition (6). In contrast, when $\mu_2^*/\mu_1^* > 0$, the constrained efficient allocation cannot be decentralized as an equilibrium because the optimality condition (19) is different from the equilibrium condition (3).

The following proposition summarizes the above findings and establishes conditions under which the value of an application to society, μ_2^*/μ_1^*, is equal to zero and conditions under which μ_2^*/μ_1^* is strictly positive.

Proposition 1 (Welfare Properties of Equilibrium): *Let $(a^*, q^*, c^*, z^*, S^*)$ denote the solution to the mechanism design problem and let (μ_1^*, μ_2^*) the multipliers associated with the aggregate resource constraints (8) and (9). (i) If $\mu_2^*/\mu_1^* = 0$, then (q^*, c^*, z^*, S^*) is an equilibrium. (ii) If $\mu_2^*/\mu_1^* > 0$, then (q^*, c^*, z^*, S^*) is not an equilibrium. (iii) There exist $\underline{k}$ and $\overline{k}$, with $0 < \underline{k} < \overline{k} < \infty$, such that $\mu_2^*/\mu_1^* = 0$ for all $k > \overline{k}$ and $\mu_2^*/\mu_1^* > 0$ for all $k < \underline{k}$.*

Proof: See Appendix D.

There is a simple intuition behind the results stated in Proposition 1. The mechanism designer sets the workers' reward to searching, S^*, equal to the disutility of searching, k. Since workers are rewarded for their search only when their search is successful, setting $S^* = k$ minimizes the workers' consumption risk subject to satisfying their incentive compatibility constraint. The mechanism designer redistributes the additional value of a worker's search, μ_2^*/μ_1^*, among all workers, independently of whether their search is successful or not. Hence, μ_2^*/μ_1^* is a measure of the extent of insurance implicitly provided by the optimal mechanism. Since workers are not provided with any insurance in the decentralized economy, it is clear that the equilibrium can attain the constrained efficient allocation only if $\mu_2^*/\mu_1^* = 0$, and it is constrained inefficient whenever $\mu_2^*/\mu_1^* > 0$. Moreover, as one would expect, μ_2^*/μ_1^* is strictly positive when the amount of consumption risk required to induce workers to search is not too large. Hence, μ_2^*/μ_1^* is strictly positive when k is not too large.

3.3 Comparing equilibrium and second best allocations

Now, we want to examine how the constrained inefficiency of the equilibrium manifests itself in terms of allocations of inputs and output. To this aim, we compare the equilibrium and the second best in terms of the number of applicants assigned to firms of type y, q_y and q_y^*, in terms of the consumption assigned to unemployed workers, z and z^*, and in terms of the consumption assigned to workers employed by firms that attract q applicants, $c(q) = u^{-1}(S/\lambda(q) + u(z))$ and $c^*(q) = u^{-1}(k/\lambda(q) + u(z^*))$. We carry out the comparison under the assumption that $\mu_2^*/\mu_1^* > 0$ and $S > k$. The first assumption guarantees that the equilibrium is constrained inefficient. The second assumption guarantees that all workers find it optimal to search in equilibrium. We make this assumption for expositional convenience.

The following proposition presents the result of the comparison between equilibrium and second best allocations.

Proposition 2 (Equilibrium and Second Best Allocations): *Let (q^*, c^*, z^*, S^*) be the solution to the mechanism design problem with $\mu_2^*/\mu_1^* > 0$, and let (q, c, z, S) be a competitive equilibrium with $S > k$. (i) There exists a $y_0 \in (y_c, \overline{y})$ such that $q(y_0) = q^*(y_0)$, $q(y) > q^*(y)$ for all $y \in (y_c, y_0)$, and $q(y) < q^*(y)$ for all $y \in (y_0, \overline{y})$. (ii) There exists a $q_0 \in (0, \infty)$ such*

that $c(q_0) = c^*(q_0)$, $c(q) < c^*(q)$ *for all* $q \in (0, q_0)$, *and* $c(q) > c^*(q)$ *for all* $q \in (q_0, \infty)$. *Moreover,* $z < z^*$.

Proof: See Appendix E.

Proposition 2 shows that the constrained inefficiency of the equilibrium manifests itself in terms of both the allocation of inputs (part i) and the allocation of output (part ii). More specifically, part (i) of Proposition 2 shows that, in equilibrium, the number of applicants assigned to low-y firms is higher than in the second best, while the number of applicants assigned to high-y firms is lower that in the second best. Part (ii) of Proposition 2 shows that the consumption assigned to unemployed workers is lower in the equilibrium than in the second best. Moreover, in the equilibrium, the consumption assigned to workers employed at low-q firms is lower than in the second best, while the consumption assigned to workers employed at high-q firms is higher than in the second best.

The results in Proposition 2 are intuitive. In equilibrium, workers face too much consumption risk when applying for a job and, consequently, firms have to pay an excessive risk premium in order to attract applicants to their vacancies. Moreover, the excess consumption risk faced by a worker when applying for a job is higher the higher the number of workers who seek the same job. Consequently, the excess risk premium that a firm has to pay is increasing in the number of applicants that the firm wants to attract. Formally, the excess risk premium is given by the difference between $p(q) - p(0)$ and $p^*(q) - p^*(0)$, and the derivative of the excess risk premium with respect to q is given by $p'(q) - p^{*\prime}(q)$, which is equal to $\phi_q(q, z, S) - \phi_q(q, z^*, k)$. The derivative of the excess risk premium is strictly positive because—as discussed in Section 2—$\phi_q(q, z, S)$ is decreasing in z and increasing in S and—as discussed above—$z < z^*$ and $S > k$. Since the excess risk premium is increasing in q, high-y firms choose to attract fewer applicants than in the second best and, through general equilibrium effects, low-y firms end up attracting more applicants than in the second best. Moreover, the fact that the excess risk premium is increasing in q also implies that the consumption of workers employed by high-q firms is higher than in the second best and, again through general equilibrium effects, the consumption of workers employed by low-q firms is lower than in the second best.

Proposition 2 has some important implications for aggregate variables. First, the propo-

sition implies that aggregate output in the equilibrium, $Y = \int \eta(q_y) y dF(y)$, is lower that aggregate output in the second best, $Y^* = \int \eta(q_y^*) y dF(y)$. To see why this is the case, it is sufficient to notice that

$$\begin{aligned} & Y - Y^* \\ = & \int_{\underline{y}}^{y_0} \left[\int_{q_y^*}^{q_y} \eta'(q) y dy \right] dF(y) - \int_{y_0}^{\overline{y}} \left[\int_{q_y}^{q_y^*} \eta'(q) y dy \right] dF(y) \\ < & \int_{\underline{y}}^{y_0} \eta'(q_{y_0}^*) y_0 \left[q_y - q_y^* \right] dF(y) - \int_{y_0}^{\overline{y}} \eta'(q_{y_0}^*) y_0 \left[q_y^* - q_y \right] dF(y) = 0, \end{aligned} \tag{21}$$

where the second line uses the fact that $q_y > q_y^*$ for $y < y_0$ and $q_y < q_y^*$ for $y < y_0$, the inequality in the third line uses the fact that $\eta'(q_y) y < \eta'(q_y^*) y < \eta'(q_{y_0}^*) y_0$ for $y < y_0$ and $\eta'(q_y) y > \eta'(q_y^*) y > \eta'(q_{y_0}^*) y_0$ for $y > y_0$, and the last equality uses the fact that $q_y dF(y)$ and $q_y^* dF(y)$ both integrate up to one. Intuitively, aggregate output is lower in the equilibrium that in the second best because, in the second best, the marginal productivity of an application is higher at firms with a higher y and, in equilibrium, there are more workers applying to low-y firms and fewer workers applying to high-y firms than in the second best.

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

$$\begin{aligned} & U - U^* \\ = & \int_{y_0}^{\overline{y}} \left[\int_{q_y}^{q_y^*} \eta'(q) dy \right] dF(y) - \int_{\underline{y}}^{y_0} \left[\int_{q_y^*}^{q_y} \eta'(q) dy \right] dF(y) \\ < & \int_{y_0}^{\overline{y}} \eta'(q_{y_0}) \left[q_y^* - q_y \right] dF(y) - \int_{\underline{y}}^{y_0} \eta'(q_{y_0}) \left[q_y - q_y^* \right] dF(y) = 0, \end{aligned} \tag{22}$$

where the second line in (22) uses the fact that $q_y^* > q_y$ for $y > y_0$ and $q_y > q_y^*$ for $y < y_0$, and the third line uses the fact that $\eta'(q_y^*) < \eta'(q_y) < \eta'(q_{y_0})$ for $y > y_0$ and $\eta'(q_y^*) y > \eta'(q_y) y > \eta'(q_{y_0}) y_0$ for $y < y_0$. Intuitively, unemployment is lower in equilibrium than in the second best because, in equilibrium, applicants are more evenly distributed across different types of firms and the vacancy filling probability, $\eta(q)$, is a concave function of applicants.

4 Policy implementation

In this section, we prove that the constrained efficient allocation can be implemented as an equilibrium by introducing a positive unemployment benefit and a positive, increasing and regressive tax on labor income.

4.1 Environment and equilibrium

A policy is a couple (B_u, T_e). The first element of the couple, $B_u \in \mathbb{R}_+$, denotes the benefit paid by the policy maker to an unemployed worker. The second element of the couple is a function $T_e : \mathbb{R}_+ \to \mathbb{R}_+$ with $T_e(w)$ denoting the tax paid to the policy maker by a worker employed at the wage w. We are now in the position to define an equilibrium.

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

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

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

(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 $w \in \mathbb{R}_+$,*

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

(iv) *Consistency of dividends and profits:*

$$z = \int \eta(q(w(y))) \left[y - w(y)\right] dF(y);$$

(v) *Balanced budget:*

$$B_u = \int \eta(q(w(y))) \left[T_e(w(y)) + B_u\right] dF(y).$$

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

dividends, z, and unemployment benefits, B_u, and the consumption of an employed worker is given by the sum of dividends, z, and after-tax wages, $w - T_e(w)$. The fifth condition in Definition 2 guarantees that the budget of the policy-maker is balanced given the optimal behavior of firms and workers. We shall denote with $\hat{q}_y$ the number of applicants attracted by a firm of type y and with $\hat{w}(q)$ the wage that a firm needs to offer in order to attract q applicants.

4.2 Optimal policy

The following theorem states the main result of our paper.

Theorem 1 (Optimal Policy): *The constrained efficient allocation can be implemented as a competitive equilibrium with policy* (B_u^*, T_e^*). *The optimal unemployment benefit* B_u^* *is positive. Specifically,* $B_u^* = \mu_2^*/\mu_1^*$. *The optimal tax on labor income* T_e^* *is positive, increasing and regressive. Specifically, for* $w \leq y_c^*$ $T_e^*(w) = 0$, *and for* $w > y_c^*$ $T_e^*(w)$ *is such that*

$$T_e^{*\prime}(w) = \left[B_u^* + \frac{k}{u'(z^* - B_u^* + w - T_e^*(w))} \right]^{-1} B_u^*. \tag{23}$$

Proof: See Appendix F.

There is a simple intuition behind Theorem 1. The constrained efficient allocation of consumption, (c_y^*, z^*), is such that the workers' expected benefit from applying to a job, S^*, is equal to the cost of applying to a job, k. The constrained efficient allocation of applicants, q_y^*, is the same the would be chosen by profit-maximizing firms if the price of an application was $p^*(q) = \phi(q, z^*, k) + \mu_2^*/\mu_1^*$, where the first component of the price is the cost of providing applicants with the expected reward k and the second component of the price is the value of an application to society (net of the disutility cost k). The optimal unemployment benefit B_u^* makes sure that—in equilibrium—the workers' benefit from applying to a job is k. The unemployment benefit achieves this goal by redistributing consumption from successful to unsuccessful applicants. The optimal labor tax T_e^* makes sure that—in equilibrium—firms face the application price $p^*(q)$. The labor tax attains this goal by raising the price of an application beyond the cost of providing applicants with the expected benefit k. As we shall explain below, the labor tax is regressive because of two properties of the optimum: (a)

redistribution only takes place between workers who apply to the same type of firms, and (b) workers are ex-ante indifferent between applying to different types of firms.[9]

It is useful to flash out the above intuition. Given the policy (B_u, T_e), the equilibrium queue of applicants, $\hat{q}_y$, is such that

$$\hat{q}_y = \arg\max_q \eta(q)y - \hat{p}(q)q, \tag{24}$$

where the price of an application $\hat{p}(q)$ is

$$\hat{p}(q) = \lambda(q)\hat{w}(q),$$

and

$$\begin{aligned} T(q) &= T_e(\hat{w}(q)), \\ \hat{w}(q) &= u^{-1}\left(\frac{S}{\lambda(q)} + u(z + B_u)\right) - z + T(q). \end{aligned}$$

In words, the equilibrium queue of applicants, $\hat{q}_y$, maximizes the profits of the firm given that the price of an application, $\hat{p}(q)$, is the product between the probability that the application is successful, $\lambda(q)$, and the gross wage that the firm needs to offer in order to attract q applications, $\hat{w}(q)$.

The equilibrium queue of applicants, $\hat{q}_y$, is equal to the constrained efficient queue of applicants, q_y^*, if and only if the the equilibrium price of an application $\hat{p}(q)$ is equal to $p^*(q)$. Notice that $\hat{p}(q)$ can be written as $\phi(q, z+B_u, S)+\lambda(q)(B_u+T(q))$, while $p^*(q)$ can be written as $\phi(q, z^*, S^*) + \mu_2^*/\mu_1^*$. Hence, the equilibrium queue length is equal to the constrained efficient queue length if and only if $z + B_u = z^*$, $S = S^*$ and $\lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^*$. These are necessary conditions for the equilibrium to implement the constrained efficient allocation.

Now, we can use the above necessary conditions to derive the optimal unemployment

[9] Theorem 1 proves that the constrained efficient allocation can be implemented using an unemployment benefit and a labor income tax. While these policies are quite natural in the context of our model, they are not the only ones that a policy maker could use to implement the constrained efficient allocation. For example, a policy maker could use a labor income tax and a "consumption subsidy" that is paid to a worker independently of his employment status. Alternatively, a policy maker could use an unemployment benefit and a "vacancy tax" that varies according to the wage offered by the firm and that is paid by the firm independently of whether or not it fills its vacancy.

benefit. In fact, notice that

$$\begin{aligned} B_u &= \int \eta(\hat{q}_y) \left[T_e(\hat{w}(\hat{q}_y)) + B_u \right] dF(y) \\ &= \frac{\mu_2^*}{\mu_1^*} \int q_y^* dF(y) = \frac{\mu_2^*}{\mu_1^*}, \end{aligned} \tag{25}$$

where the first equality follows from the equilibrium condition (v), the second equality follows from the necessary condition $\lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^*$ and the consistency condition $\eta(q) = \lambda(q)/q$, and the last equality follows from the fact that μ_2^*/μ_1^* and q_y^* satisfy condition (18). The optimal unemployment benefit is equal to the ratio of multipliers μ_2^*/μ_1^*. This finding is easy to understand. In the constrained efficient allocation, the value of an application to a worker is k, but the value of an application to society exceeds k by μ_2^*/μ_1^*. This excess value is redistributed to all the workers independently of whether their application is successful or not. The optimal unemployment benefit carries out this redistribution in equilibrium. Unemployed workers receive μ_2^*/μ_1^* directly in the form of an unemployment benefit, while employed workers receive μ_2^*/μ_1^* indirectly in the form of the wage premium that compensates them for the loss of the unemployment benefit. Hence, it is the optimal unemployment benefit that guarantees that the equilibrium value of an application to a worker is k.

Next, we can use the necessary conditions for the constrained efficiency of the equilibrium in order to derive the optimal labor income tax. In fact, notice that $\lambda(q)(B_u + T(q)) = \mu_2^*/\mu_1^*$ implies

$$\begin{aligned} T_e(w(q)) &= \frac{1}{\lambda(q)} \frac{\mu_2^*}{\mu_1^*} - B_u \\ &= \left(\frac{1}{\lambda(q)} - 1 \right) B_u, \end{aligned} \tag{26}$$

where the second equality uses the fact that $B_u = \mu_2^*/\mu_1^*$. Equation (26) implies that the optimal labor tax is such that the taxes paid by the workers who successfully apply to a given type of firm are equal to the unemployment benefits paid to the workers who apply unsuccessfully to the same type of firms[10], i.e. $\lambda(q)T(q) = (1 - \lambda(q))B_u$. This property implies that the optimal policy involves redistribution between workers who apply to the

[10]In the US, the contributions to the unemployment insurance system paid by a firm and its workers depend on the frequency at which the firm lays off its workers. Our model suggests that the contributions to the unemployment insurance system paid by a firm and its workers (in the form of labor taxes) should depend on the frequency at which workers unsuccessfully apply to the firm. Hence, our model provides a rationale for making unemployment insurance contributions a function not only of separations but also of failed hirings.

same type of firm, but not across workers who apply to different types of firms. In turn, this implies that the difference in the amount of labor taxes paid by workers employed at different wages does not serve a redistributive purpose. Rather, differences in labor taxes are necessary to make sure that the equilibrium price of an application, $\hat{p}(q)$, reflects the social cost of an application, $p^*(q)$. In this sense, labor taxes are Pigovian. The fact that there is no redistribution across workers employed at different wages should not be surprising because—in a directed search environment—wage differences do not reflect luck, but compensation for different job-finding probabilities.[11]

In order to understand the shape of the optimal labor tax, we differentiate (26) with respect to q and we obtain

$$\begin{aligned} T_e'(\hat{w}(q)) &= T'(q)/\hat{w}'(q) \\ &= \left[B_u + \frac{k}{u'(z^* - B_u + \hat{w}(q) - T_e(\hat{w}(q)))} \right]^{-1} B_u, \end{aligned} \tag{27}$$

where the second line makes use of the equilibrium condition (iii). The marginal tax on labor earnings is positive because both the numerator and the denominator in (27) are positive. The marginal tax on labor earnings is smaller than one because the numerator in (27) is smaller than the denominator. And the marginal tax is decreasing in labor earnings because the after-tax wage $\hat{w} - T_e(\hat{w})$ is increasing in w. Hence, the optimal tax on labor income is regressive.

The regressivity of the optimal labor income tax is a startlingly sharp result. It does not depend on the shape of the workers' utility function, $u(c)$, on the shape of the distribution of firms' productivity, $F(y)$, or on the shape of the matching function, $\lambda(q)$. Rather, the regressivity of the labor income tax follows directly from two properties of the equilibrium with optimal policy: (a) redistribution only takes place between workers applying to the same type of firm and (b) workers are ex-ante indifferent between applying to different types

[11] In a basic random search environment (e.g. Pissarides 1985 or Mortensen and Pissarides 1994), differences in labor income among employed workers are only due to luck and, hence, they should be fully redistributed away.

of firms. In fact, these two properties are

$$T_e(w(q)) = \left(\frac{1}{\lambda(q)} - 1\right) B_u,$$
$$u(z + w(q) - T_e(w(q))) - u(z + B_u) = \frac{k}{\lambda(q)},$$

and, together, they imply

$$T_e(w) = \left[\frac{u(z + w - T_e(w)) - u(z + B_u)}{k} - 1\right] B_u. \tag{28}$$

The functional equation (28) can only be satisfied by a regressive labor income tax. Under any progressive tax, the left-hand side of (28) would be a convex function of w, while the right-hand side of (28) would be a concave function of w.

5 Insurance market implementation

In this section, we prove that the constrained efficient allocation could be decentralized as a laissez-faire equilibrium if there was a competitive insurance market where workers could purchase insurance against the search risk. The result implies that the constrained inefficiency of the economy analyzed in Section 2 is caused by the absence of this insurance market. Similarly, the result implies that the optimal unemployment benefit and the optimal labor income tax derived in Section 4 serve the purpose of substituting the missing insurance market.

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 (w, b_u, t_e), where $w \in \mathbb{R}_+$ denotes the wage that the insurance company asks the worker to seek, $b_u \in \mathbb{R}_+$ 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 (w, b_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 $q(w)$. 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, $q(w)$, and the equilibrium insurance contracts, $(b_u(w), t_e(w))$.

Now, we are in the position to define a competitive equilibrium.

Definition 3: *A competitive equilibrium is an allocation* (w, q, b_u, t_e, z, U) *that satisfies the following conditions:*

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

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

(ii) *Insurance company's profit maximization: For all* $w \in \mathbb{R}_+$,

$$\begin{aligned}(b_u(e), t_e(e)) \quad &\in \arg\max_{(b_u, t_e)} \lambda(q(w))t_e - (1 - \lambda(q(w)))b_u, \text{ s.t.} \\ &\lambda(q(w))u(z + w - t_e) + (1 - \lambda(q(w)))u(z + b_u) - k \geq U, \\ &\lambda(q(w))\left[u(z + w - t_e) - u(z + b_u)\right] \geq k;\end{aligned}$$

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

$$\lambda(q(w))t_e(w) - (1 - \lambda(q(w))b_u(w) = 0$$

and $q(w) \geq 0$, *with complementary slackness;*

(iv) *Optimal number of applications:*

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

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

$$\lambda(q(w))u(z + w + t_e(w)) + (1 - \lambda(q(w)))u(z + t_u(w)) - k \leq U$$

and $q(w) \geq 0$, *with complementary slackness;*

(vi) *Consistency of dividends and profits:*

$$z = \int \eta(q(w(y)))\,[y - w(y)]\,dF(y).$$

Condition (i) guarantees that the wage $w(y)$ maximizes the profit of a firm of type y. Condition (ii) guarantees that $(w, b_u(w), t_e(w))$ is the feasible contract that maximizes the profits of an insurance company. To see why this is the case, notice that a contract (w, b_u, t_e) satisfies the worker's participation constraint if $\lambda(q(w))u(z + w - t_e) + (1 - \lambda(q(w)))u(z + b_u) - k \geq U$. A contract (w, b_u, t_e) satisfies the worker's incentive compatibility constraint if $\lambda(q(w))[u(z + w - t_e) - u(z + b_u)] \geq k$. And, if the contract (w, t_u, t_e) is feasible, the insurance company obtains a profit of $\lambda(q(w))t_e - (1 - \lambda(q(w)))b_u$. Condition (iii) guarantees that competition in the insurance market drives the profits generated by the equilibrium contract $(b_u(w), t_e(w))$ 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 $w(q)$ 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.[12]

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

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

[12] The proof of the theorem is sketched in the following pages. The complete proof is available upon request.

insurance company to the worker, $b_u(w)$, is equal to the optimal unemployment benefit, B_u^*. And if the worker finds a job at a firm offering the wage w, the payment from the worker to the insurance company, $t_e(w)$, is equal to the optimal labor earning tax, $T_e^*(w)$. 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

$$\begin{aligned} & \max_{(b_u,t_e)} \lambda(q)t_e - (1-\lambda(q))b_u, \\ \text{s.t.} \quad & \lambda(q)u(z+w(q)-t_e) + (1-\lambda(q))u(z+b_u) - k \geq U, \\ & \lambda(q)\left[u(z+w(q)-t_e) - u(z+b_u)\right] \geq k. \end{aligned} \tag{29}$$

Notice that the contract (b_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 b_u by some small amount ϵ, 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 (b_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 ϵ and δ such that—by increasing b_u by ϵ 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 (b_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 $u(z+t_u) = U$ and $\lambda(q)\left[u(z+w(q)-t_e) - U\right] = k$. Moreover, since $U = u(z+B_u^*)$ and $w(q) = u^{-1}(S^*/\lambda(q) + u(z^*)) + T_e^*(w(q)) - z$, it follows that b_u is equal to B_u^* and t_e is equal to $T_e^*(w(q))$. 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.

$\lambda(q)T_e^*(w(q)) = (1-\lambda(q))B_u^*$, 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.[13]

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 paper, 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 paper, 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 showed 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 to a firm of an applicant with the value of an applicant to society. Moreover, we showed that the regressivity of the optimal labor income tax does not depend on the shape of the workers' utility function, on the shape of the distribution of firms' productivity, or on the shape of the matching function. Rather, the regressivity of the labor income tax is a necessary implication of two properties of the equilibrium with optimal policy: (a) redistribution only takes place between workers applying to the same type of firm and (b)

[13] If the second best allocation involved not only redistribution between workers who successfully and unsuccessfully apply to the same type of firm, but also redistribution across workers applying to different types of firms, then the equilibrium would be constrained inefficient even in the presence of a competitive insurance market. In fact, insurance companies would have no incentive to provide insurance to workers whose application is subsidized in the second best, and they would have an incentive to offer more favorable insurance contracts to workers whose application is taxed in the second best.

workers are ex-ante indifferent between applying to different types of firms.

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

[1] Acemoglu, D., and R. Shimer. 1999. "Efficient Unemployment Insurance." *Journal of Political Economy* 107: 893-927.

[2] Autor, D., L. Katz, and M. Kerney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics* 90: 300–323.

[3] Burdett, K., S. Shi, and R. Wright. 2001. "Pricing and Matching with Frictions." *Journal of Political Economy* 109: 1060–85.

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

[5] Farhi, E., and I. Werning. 2011. "Insurance and Taxation over the Life Cycle." Manuscript, Harvard University.

[6] Galenianos, M., and P. Kircher.2012. "On the Game-Theoretic Foundations of Competitive Search Equilibrium." *International Economic Review* 53: 1-21.

[7] Golosov, M., M. Troshkin, and A. Tsyvinski. 2011. "Optimal Dynamic Taxes." Manuscript, Princeton University.

[8] Hansen, G., and A. Imrohoroglu. 1992. "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard." *Journal of Political Economy* 100: 118–42.

[9] Hornstein, A., P. Krusell, and G. Violante. 2011. "Frictional Wage Dispersion in Search Models: A Quantitative Assessment." *American Economic Review* 101: 2873-98.

[10] Hopenhayn, H., and J. Nicolini. 1997. "Optimal Unemployment Insurance." *Journal of Political Economy* 105: 412–38.

[11] Hungerbuhler, M., E. Lehmann, A. Parmentier, and B. Van der Linden. 2006. "Optimal Redistributive Taxation in a Search Equilibrium Model." *Review of Economic Studies* 73: 743–67.

[12] Laroque, G. 2005. "Income Maintenance and Labor Force Participation." *Econometrica* 73: 341–76.

[13] Mirrlees, J. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies* 38: 175–208.

[14] Menzio, G., and S. Shi. 2011. "Efficient Search on the Job and the Business Cycle." *Journal of Political Economy* 119: 468–510.

[15] Moen, E. 1997. "Competitive Search Equilibrium." *Journal of Political Economy* 105: 694–723.

[16] Montgomery, J. 1991. "Equilibrium Wage Dispersion and Interindustry Wage Differentials." *Quarterly Journal of Economics* 106: 163–79.

[17] Mortensen, D. 2003. *Wage Dispersion: Why are Similar Workers Paid Differently?* M.I.T. Press, Cambridge, MA.

[18] Mortensen, D., and C. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies* 61: 397–415.

[19] Pissarides, C. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies and Real Wages." *American Economic Review* 75: 676–90.

[20] Postel-Vinay F., and J. Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70: 2295–350.

[21] Saez, E. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68: 205–29.

[22] Shaal, E., and M., Taschereau-Dumouchel. 2012. "Optimal Redistributive Policy in a Labor Market with Search and Endogenous Participation." Manuscript, University of Pennsylvania.

[23] Shavell, S., and L. Weiss. 1979. "The Optimal Payment of Unemployment Insurance Benefits over Time." *Journal of Political Economy* 87: 1347–62.

[24] Shimer, R. 1996. "Contracts in a Frictional Labor Market." Manuscript, M.I.T.

[25] Shimer, R. 2005. "The Assignment of Workers to Jobs in an Economy with Coordination Frictions." *Journal of Political Economy* 113: 996-1025.

[26] Shimer, R., and I. Werning. 2008. "Liquidity and Insurance for the Unemployed." *American Economic Review* 98: 1922–42.

[27] Wang, C., and S. Williamson. 2002. "Moral Hazard, Optimal Unemployment Insurance, and Experience Rating." *Journal of Monetary Economics* 49: 1337–71.

Appendix

A Concavity of the firm's problem

We establish the strict concavity with respect to q of the firm's profit function

$$\eta(q)y - p(q)q. \tag{A1}$$

The first term in (A1) is the expected revenue of the firm. The second term in (A1) is the expected wage bill of the firm, which we have written as the product between the number of applicants attracted by the firm, q, and the price of an applicant, $p(q)$. Recall that the price of an applicant $p(q)$ is defined as $\lambda(q)w(q)$, where $w(q)$ is given by

$$w(q) = u^{-1}\left(\frac{S}{\lambda(q)} + u(z)\right) - z. \tag{A2}$$

The expected revenue of the firm is strictly concave with respect to q because $\eta(q)$ is a strictly concave function. Hence, in order to establish the strict concavity of the firm's profit function, we only have to show that the expected wage bill of the firm is convex with respect to q. That is, we only have to show that $p''(q)q + 2p'(q) \geq 0$. To this aim, notice that the first derivative of $p(q)$ with respect to q is given by

$$p'(q) = -\lambda'(q)\left[\frac{u(z+w(q)) - u(z)}{u'(z+w(q))} - w(q)\right]. \tag{A3}$$

The second derivative of $p(q)$ with respect to q is given by

$$\begin{aligned} p''(q) = \quad & -\lambda''(q)\left[\frac{u(z+w(q)) - u(z)}{u'(z+w(q))} - w(q)\right] \\ & + \frac{(\lambda'(q))^2}{\lambda(q)}\left[\frac{u(z+w(q)) - u(z)}{u'(z+w(q))}\right]^2 \frac{-u''(z+w(q))}{u'(z+w(q))}. \end{aligned} \tag{A4}$$

From the above expressions, it follows that $p''(q)q + 2p'(q)$ is given by

$$\begin{aligned} p''(q)q + 2p'(q) = \quad & -\left(2\lambda'(q) + \lambda''(q)\right)\left[\frac{u(z+w(q)) - u(z)}{u'(z+w(q))} - w(q)\right] \\ & + \frac{(\lambda'(q))^2 q}{\lambda(q)}\left[\frac{u(z+w(q)) - u(z)}{u'(z+w(q))}\right]^2 \frac{-u''(z+w(q))}{u'(z+w(q))}. \end{aligned} \tag{A5}$$

The first term on the right-hand side of (A5) is strictly positive. To see this, first notice that $[u(z+w) - u(z)]/u'(z+w) - w > 0$ because $u(c)$ is a strictly concave function and w is

strictly positive. Then, notice that $2\lambda'(q)+\lambda''(q)$ is strictly negative because $\lambda''(q)+2\lambda'(q) = \eta''(q) < 0$. The second term on the right-hand side of (A5) is also positive because $u(c)$ is a concave function. These observations imply $p''(q)q + 2p'(q) > 0$. ■

B Mechanism with reports

In this appendix, we consider a version of the mechanism design problem in which the mechanism recommends workers to follow a mixed application strategy and workers are required to report the outcome of their mixing before implementing their strategy. More specifically, the mechanism asks workers to draw their recommended search action from the cumulative distribution function π, where $\pi(y)$ denotes the probability that the worker seeks for a firm of type $\tilde{y} \leq y$. Notice that the function π uniquely determines the applicant-to-firm ratio q and the measure of applicants, a. Hence, we can think that the mechanism chooses a and q rather than π. After workers draw their recommended action, they make a report to the mechanism and chooses whether and where to search for a job. Then, the mechanism assigns $c_y(\hat{y})$ units of consumption to workers who report seeking for firms of type y and end up being employed at firms of type $\hat{y}$. Similarly, the mechanism assigns $c_0(\hat{y})$ units of consumption to workers who report not searching and end up being employed at firms of type $\hat{y}$. The mechanism assigns z_y units of consumption to workers who report seeking for firms of type y and end up being unemployed, and z_0 units of consumption to workers who report not searching and end up being unemployed.

The mechanism designer chooses (a, q, c, z) so as to maximize the worker's expected utility

$$\int \left[\lambda(q_y)u(c_y(y)) + (1-\lambda(q_y))u(z_y) - k\right] q_y dF(y) + (1-a)u(z_0). \tag{B1}$$

The mechanism is subject to incentive compatibility, truth telling and resource constraints. First, we describe the incentive compatibility constraints. Consider an arbitrary y such that $q_y > 0$. The mechanism must induce a worker that reports seeking a firm of type y to actually seek a firm of type y rather than a firm of type $\hat{y} \neq y$. That is, for all $\hat{y} \neq y$, the mechanism must be such that

$$\lambda(q_y)u(c_y(y)) + (1-\lambda(q_y))u(z_y) - k \geq \lambda(q_{\hat{y}})u(c_y(\hat{y})) + (1-\lambda(q_{\hat{y}}))u(z_y) - k. \tag{B2}$$

Moreover, the mechanism must induce a worker that reports seeking a firm of type y to actually seek a firm of type y rather not to search. That is, the mechanism must be such that

$$\lambda(q_y)u(c_y(y)) + (1-\lambda(q_y))u(z_y) - k \geq u(z_y). \tag{B3}$$

Similarly, if $a < 1$, the mechanism must induce a worker that reports not seeking for a job to actually do so. That is, for all $\hat{y}$, the mechanism must be such that

$$u(d_0) \geq \lambda(q_{\hat{y}})u(c_0(\hat{y})) + (1-\lambda(q_{\hat{y}}))u(z_0) - k. \tag{B4}$$

Next, we describe the truth telling constraints. Again, consider an arbitrary y such that $q_y > 0$. The mechanism must induce a worker that draws y to report that he is searching for a firm of type y rather than to report that he is searching for a firm of type $\hat{y} \neq y$. That is, for all $\hat{y} \neq y$ such that $q_{\hat{y}} > 0$, the mechanism must be such that

$$\lambda(q_y)u(c_y(y)) + (1-\lambda(q_y))u(z_y) \geq \lambda(q_{\hat{y}})u(c_{\hat{y}}(\hat{y})) + (1-\lambda(q_{\hat{y}}))u(z_{\hat{y}}). \tag{B5}$$

Moreover, if $a < 1$, the mechanism must make sure that the worker does not report that he is not searching. That is, the mechanism must be such that

$$\lambda(q_y)u(c_y(y)) + (1-\lambda(q_y))u(z_y) - k \geq u(z_0). \tag{B6}$$

Also, if $a < 1$, if the worker's randomization instructs him not to search, then the worker must have the incentive to truthfully report this to the mechanism. That is, for all $\hat{y}$ such that $q_{\hat{y}} > 0$

$$u(d_0) \geq \lambda(q_{\hat{y}})u(c_{\hat{y}}(\hat{y})) + (1-\lambda(q_{\hat{y}}))u(z_{\hat{y}}) - k. \tag{B7}$$

Finally, the mechanism has to satisfy the resource constraint on output

$$\int \eta(q_y) y dF(y) - \int \left[\lambda(q_y)u(c_y(y)) + (1-\lambda(q_y))u(z_y)\right] dF(y) - (1-a)z_0, \tag{B8}$$

and the resource constraint on applications

$$1 - a \geq 0, \tag{B9}$$

$$a - \int q_y dF(y) = 0. \tag{B10}$$

We claim that the above mechanism design problem is equivalent to problem that we

analyzed in Section 3. First, note that we can abstract from the incentive compatibility constraint (B2) because this constraint can always be satisfied by choosing $c_y(\hat{y}) = 0$ and the choice of $c_y(\hat{y})$ does not affect any other constraint or the objective function. Similarly,we can abstract from the incentive compatibility constraint (B4). Second, note that the truth telling constraints (B5)-(B7) are equivalent to say that there exists a U such that the worker's expected utility from following any recommended action is equal to U and the worker's expected utility from following any non-recommended action is non-greater than U. In light of the above observations, we can replace the incentive compatibility constraints (B2)-(B4) with

$$U - u(z_y) \geq 0. \tag{B11}$$

And we can replace the truth telling constraints (B5)-(B7) with

$$\begin{aligned} \lambda(q_y)u(c_y(y)) + (1 - \lambda(q_y))u(z_y) - U &= 0, & \text{(B12)} \\ U - u(z_0) &\geq 0, & \text{(B13)} \\ (1 - a)\left(U - u(z_0)\right) &= 0. & \text{(B14)} \end{aligned}$$

After rewriting the incentive compatibility and the truth telling constraint, it is immediate to see that the only difference between the above mechanism design problem and the one in Section 3 is that the consumption of unemployed workers can be made contingent upon the worker's search strategy. However, we will now show that the mechanism designer finds it optimal not to use these contingencies. Let ϕ_1 and ϕ_2 denote the multipliers associated with the constraints (30) and (30). Let $\rho_{1,y}dF(y)$ and $\rho_{2,y}dF(y)$ denote the multipliers associated with the constraints (B11) and (30). Finally, let μ_1, μ_2 and μ_3 denote the multipliers associated with the resource constraints (B8), (B9) and (B10). Consider an arbitrary y. The first order condition with respect to z_y is given by

$$u'(z_y)\left[(1 - \lambda(q_y))(q_y + \rho_{2,y}) + \rho_{1,y}\right] = \mu_1 q_y(1 - \lambda(q_y)). \tag{B15}$$

The first order condition with respect to $c_y(y)$ is given by

$$u'(c_y(y))\left[q_y + \rho_{2,y}\right] = \mu_1 q_y. \tag{B16}$$

If $\rho_{1,y} = 0$, then the first order conditions (B15) and (B16) imply $z_y = c_y(y)$. However, this

violates the incentive compatibility constraint (B11). Hence, for all y, we have $\rho_{1,y} > 0$ and $U - u(z_y) = 0$. In turn, this implies that $z_y = z$ for all y. That is, the mechanism designer finds it optimal to equate the consumption among all of the unsuccessful job seekers. If $a = 1$, z_0 does not matter and it can be set equal to z without loss in generality. If $a < 1$, the truth telling constraint (30) implies $z_0 = z$. That is, the mechanism designer needs to equate the consumption between workers who do not search and workers who search unsuccessfully. ■

C Optimal mechanism

In this appendix, we derive the optimality conditions (14)-(18) for the solution to the mechanism design problem. To this aim, we need to introduce some additional notation. Let μ_1, μ_2 and μ_3 denote the multipliers associated with the aggregate resource constraints (8)–(10). Let $\rho_y dF(y)$ denote the multiplier associated to the incentive compatibility constraint (11). Finally, let ν_1 and ν_2 denote the multipliers associated to the incentive compatibility constraints (12) and (13).

A solution to the mechanism design problem consists of an allocation (a, q, c, z, S), a list of multipliers on the resource constraints (μ_1, μ_2, μ_3), and a list of multipliers on the incentive compatibility constraints (ρ_y, ν_1, ν_2). The solution to the mechanism design problem must satisfy a number of first order conditions. The first order condition for c_y is

$$q_y u'(c_y) + \rho_y u'(c_y) = \mu_1 q_y. \tag{C1}$$

The first order condition for z is

$$u'(z)\left[1 - \int \lambda(q_y) q_y dF(y)\right] = u'(z)\left[\int \rho_y \lambda(q_y) dF(y)\right] + \mu_1\left[1 - \int \lambda(q_y) q_y dF(y)\right]. \tag{C2}$$

The first order condition for q_y is

$$\begin{aligned} & \{u(z) + S - k + q_y \lambda'(q_y)\,[u(c_y) - u(z)]\} \\ + \quad & \mu_1\,\{\eta'(q_y) y - \lambda(q_y) c_y - (1 - \lambda(q_y)) z + q_y \lambda'(q_y)\,[c_y - z]\} \\ \leq \quad & \mu_3 - \rho_y \lambda'(q_y)\,[u(c_y) - u(z)], \end{aligned} \tag{C3}$$

The first order condition for a is

$$u(z) + \mu_2 = \mu_1 z + \mu_3. \tag{C4}$$

The first order condition for S is

$$\nu_1 + \nu_2(1-a) = \int \rho_y dF(y). \tag{C5}$$

In addition to the first order conditions (C1)-(C5), a solution to the mechanism design problem must satisfy the resource constraints (8) and (10), as well as the incentive compatibility constraints (11) and (13). Moreover, a solution to the mechanism design problem must be such that

$$1-a \geq 0,\ \mu_2 \geq 0 \text{ and } (1-a)\,\mu_2 = 0, \tag{C6}$$

$$S-k \geq 0,\ \nu_1 \geq 0 \text{ and } (S-k)\nu_1 = 0. \tag{C7}$$

Using equation (C2) to solve for μ_1 and equation (C1) to solve for ρ_y, it follows that the right-hand side of (C5) is strictly positive and, hence, $\nu_1 + \nu_2(1-a) > 0$. First, suppose that the solution to the mechanism design problem is such that $a = 1$. In this case, $\nu_1 + \nu_2(1-a) > 0$ implies $\nu_1 > 0$. In turn, ν_1 and (C7) imply $S = k$. Next, suppose that the solution to the mechanism design is such that $a < 1$. In this case, the incentive compatibility constraint (13) implies $S = k$. Therefore, the solution to the mechanism design problem is always such that

$$S = k. \tag{C8}$$

Solving equation (C1) for ρ_y and equation (C4) for μ_2 and substituting the solutions into the first order condition (C3), we obtain

$$\eta'(q_y)y - \lambda(q_y)(c_y - d) + q_y\lambda'(q_y)\left[\frac{u(c_y)-u(d)}{u'(c_y)} - (c_y - d)\right] = \frac{\mu_2}{\mu_1}. \tag{C9}$$

Solving the incentive compatibility constraint (11) for c_y and using the fact that $S = k$, we obtain

$$c_y = u^{-1}\left(\frac{k}{\lambda(q)} + u(z)\right). \tag{C10}$$

Solving the resource constraint (8) for z, we obtain

$$z = \int \eta(q_y) \left[y - (c_y - z) \right] dF(y). \tag{C11}$$

Finally, using the optimality condition (C7) and the resource constraint (10), it is straightforward to verify that

$$1 - \int q_y dF(y) \geq 0, \; \frac{\mu_2}{\mu_1} \geq 0 \text{ and } \left(1 - \int q_y dF(y) \right) \frac{\mu_2}{\mu_1} = 0. \tag{C12}$$

The optimality conditions (C8)-(C12) are the same as the conditions (14)-(18). ■

D Proof of Proposition 1

In the main text, we proved parts (i) and (ii) of the proposition. In this appendix, we prove part (iii). Let $(a^*, q^*, c^*, z^*, S^*)$ denote the solution to the mechanism design problem and let μ_1^* and μ_2^* denote the multipliers associated with the resource constraints (11) and (12). Suppose that $\mu_2^* / \mu_1^* = 0$. When $\mu_2^* / \mu_1^* = 0$, the constrained efficient queue length is such that

$$\eta'(q_y^*) y = \lambda(q_y^*) \left(c_y^* - z^* \right) - q_y^* \lambda'(q_y^*) \left[\frac{u(c_y^*) - u(z^*)}{u'(c_y^*)} - (c_y^* - z^*) \right] \tag{D1}$$

and the constrained efficient consumption for employed workers is such that

$$c_y^* = u^{-1} \left(\frac{k}{\lambda(q_y^*)} + u(z^*) \right). \tag{D2}$$

Moreover, the constrained efficient consumption for unemployed workers satisfies the resource constraint for output

$$z^* = \int \eta(q_y^*)(y - (c_y^* - z^*)) dF(y), \tag{D3}$$

and the constrained efficient queue satisfies the resource constraint on applicants

$$\int q_y^* dF(y) = a^* \leq 1. \tag{D4}$$

Let $\tilde{q}(y, z, k)$ denote the solution for q to the equations (D1) and (D2) with respect to q. It is straightforward to verify that $\tilde{q}(y, z, k)$ is increasing in y, increasing in d and decreasing in k.

Now, let $\underline{k}$ be the search cost such that $\int \tilde{q}(y, 0, \underline{k}) dF(y) = 1$. Such a $\underline{k}$ exists because

$\lim_{k\to 0} \tilde{q}(y,z,k) = \infty$ for any $y \in [\underline{y}, \overline{y}]$ and any $z \in \mathbb{R}_+$. For any $k < \underline{k}$, the constrained efficient queue length is such that

$$\begin{aligned} & \int q^*(y)dF(y) \\ = & \int \tilde{q}(y,z^*,k)dF(y) \\ > & \int \tilde{q}(y,0,\underline{k})dF(y) = 1. \end{aligned} \tag{D5}$$

This implies that, for $k < \underline{k}$, there is no solution to the system of equations (D1)-(D4). Hence, the solution to the mechanism design problem must be such that $\mu_2^*/\,\mu_1^* > 0$. In an analogous way, one can prove that there exists a $\overline{k}$ such that, for $k > \overline{k}$, $\mu_2^*/\,\mu_1^* = 0$. ■

E Proof of Proposition 2

First, we compare the consumption assigned to unemployed workers in the equilibrium and in the second best. To this aim, notice that the workers' expected utility is equal $u(z)+S-k$ in the equilibrium, and it is equal to $u(z^*) + S^* - k$ in the second best. Since $\mu_2^*/\mu_1^* > 0$, the equilibrium is constrained inefficient and, hence, the workers' expected utility is strictly lower that in the second best. That is,

$$u(z) + S - k < u(z^*) + S^* - k. \tag{E1}$$

Since $S > k$ and $S^* = k$, the above inequality implies $z < z^*$.

Next, we compare the consumption assigned to employed workers in the equilibrium and in the second best. In the equilibrium, the consumption assigned to a worker employed at a firm that attracts q applicants is $c(q) = u^{-1}(S/\lambda(q) + u(z))$. In the second best, the consumption assigned to a worker employed at a firm that attracts q applicants is $c^*(q) = u^{-1}(S^*/\lambda(q) + u(z^*))$. The derivatives of c and c^* with respect to q are

$$\begin{aligned} c'(q) &= -\frac{\lambda'(q)}{\lambda(q)^2}\frac{S}{u'(c(q))}, \\ c^{*\prime}(q) &= -\frac{\lambda'(q)}{\lambda(q)^2}\frac{S^*}{u'(c^*(q))}. \end{aligned} \tag{E2}$$

Notice that since $S > S^*$, $c(q_0) = c^*(q_0)$ implies that $c'(q_0) > c^{*\prime}(q_0)$. Using this observation and the behavior of $c(q)$ and $c^*(q)$ at $q = 0$ and $q = \infty$, it follows that there exists one

$q_0 \in (0,\infty)$ such that $c(q_0) = c^*(q_0)$, $c(q) < c^*(q)$ for $q \in (0, q_0)$ and $c(q) > c^*(q)$ for $q \in (q_0, \infty)$.

Finally, we compare the allocation of applicants across firms in the equilibrium, q_y, and in the second best, q_y^*. Differentiating the first order condition (3) with respect to y and using the fact that $p(q) = \phi(q, z, S)$, we obtain

$$q_y' = \frac{\eta'(q_y)y}{-\eta''(q_y)y + 2\phi_q(q,z,S) + \phi_{qq}(q,z,S)q}. \tag{E3}$$

Similarly, differentiating the first order condition (19) with respect to y and using the fact that $p^*(q) = \phi(q, z^*, S^*) + \mu_2^*/\mu_1^*$, we obtain

$$q_y^{*\prime} = \frac{\eta'(q_y^*)y}{-\eta''(q_y^*)y + 2\phi_q(q,z^*,S^*) + \phi_{qq}(q,z^*,S^*)q}. \tag{E4}$$

The differential equations (E3) and (E4) are identical, except that the function ϕ is evaluated at (q, z, S) in the first equation and at (q, z^*, S^*) in the second equation. First, using the fact that $\phi_q(q, z, S)$ is strictly increasing in S and strictly decreasing in z and the fact that $z < z^*$ and $S > S^*$, we can prove that $2\phi_q(q, z, S) + \phi_{qq}(q, z, S)q$ is strictly greater than $2\phi_q(q, z^*, S^*) + \phi_{qq}(q, z^*, S^*)q$. This property implies that if $q_y = q_y^*$ then $q_y' < q_y^{*\prime}$. In turn, this implies that there exists at most one y such that $q_y = q_y^*$. Second, 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 at least one $y \in (y_c, \overline{y})$ such that $q_y = q_y^*$. Taken together the two observations imply that there exist a y_0 such that $q_{y_0} = q_{y_0}^*$, $q_y > q_y^*$ for $y \in (y_c, y_0)$, and $q_y < q_y^*$ for $y \in (y_0, \overline{y})$. ■

F Proof of Theorem 1

Given the unemployment benefit B_u^* and the labor earning tax T_e^*, consider the allocation $(\hat{q}, \hat{w}, \hat{z}, \hat{S})$ where $\hat{z} = z^* - B_u^*$, $\hat{S} = S^*$ $\hat{q}_y = q_y^*$ and

$$\hat{w}(q) = u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) + T^*(q) - \hat{z}. \tag{F1}$$

First, we claim is that the queue length function $\hat{q}_y$ solves the problem of the firm

$$\begin{aligned} &\max_{q \geq 0} \eta(q)y - \hat{p}(q)q, \\ &\hat{p}(q) = \lambda(q)\left[u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) + T^*(q) - \hat{z}\right]. \end{aligned} \tag{F2}$$

To see this, notice that the problem (F2) is strictly concave and, hence, $\hat{q}_y$ is a solution if and only if it satisfies the first order condition

$$\begin{aligned} \eta'(q)y &= \hat{p}(q)q + \hat{p}(q), \quad \text{if } q > 0, \\ \eta'(q)y &\leq \hat{p}(q)q + \hat{p}(q), \quad \text{if } q = 0. \end{aligned} \tag{F3}$$

Since $\hat{q}_y = q_y^*$, $\hat{q}_y$ satisfies the first order condition of the mechanism design problem

$$\begin{aligned} \eta'(q)y &= p^*(q)q + p^*(q), \quad \text{if } q > 0, \\ \eta'(q)y &\leq p^*(q)q + p^*(q), \quad \text{if } q = 0. \end{aligned} \tag{F4}$$

Now, notice that the price of an applicant $\hat{p}(q)$ is such that

$$\begin{aligned} \hat{p}(q) &= \lambda(q)\left[u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) + T^*(q) - \hat{z}\right] \\ &= \lambda(q)\left[u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) - z^*\right] + \frac{\mu_2^*}{\mu_1^*} \\ &= p^*(q), \end{aligned} \tag{F5}$$

where the second line uses the definition of $\hat{z}$ and the fact that $\lambda(q)[T^*(q) + B_u^*] = \mu_2^*/\mu_1^*$. From (F5), it follows that $\hat{q}_y$ satisfies the first order condition (F3) and, hence, it solves the problem of the firm (F2).

Second, we claim that the wage function $\hat{w}(q)$ is consistent with the worker's optimal search strategy. To see this notice that the worker's expected utility from seeking a job that attracts q applicants is

$$\begin{aligned} & \lambda(q)\left[u(\hat{z} + \hat{w}(q) - T_e^*(\hat{w}(q))) - u(\hat{z} + B_u^*)\right] \\ = \;& \lambda(q)\left[u\left(u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right)\right) - u(z^*)\right] \\ = \;& S^* = \hat{S}, \end{aligned} \tag{F6}$$

where the second line uses the definition of $\hat{z}$ and $\hat{w}(q)$ and the third line uses of the definition of $\hat{S}$.

Third, we claim that the value of searching $\hat{S}$ is consistent with the worker's optimal search strategy. To this aim,it is sufficient to notice that $\hat{S} = k$ and

$$\int \hat{q}_y dF(y) = \int q_y^* dF(y) \leq 1. \tag{F7}$$

Fourth, we claim that the dividends received by the workers are equal to the profits

earned by the firms. To see this, notice that the profits earned by the firms are

$$\begin{aligned} & \int \eta(\hat{q}_y)\left[y-\hat{w}(\hat{q}_y)\right]dF(y) \\ = & \int \eta(q_y^*)\left[y-u^{-1}\left(S^*/\,\lambda(q_y^*)+u(z^*)\right)+z^*-B_u^*/\,\lambda(q)\right]dF(y) \\ = & \int \eta(q_y^*)\left[y-\left(c_y^*-z^*\right)\right]dF(y)-\int q_y^* B_u^* dF(y) \\ = & z^*-B_u^*=\hat{z}, \end{aligned} \tag{F8}$$

where the second line uses the definitions of $\hat{q}_y$ and $\hat{w}(q)$, the third line uses the equations $\hat{z}+B_u^*=z^*$ and $\lambda(q)[B_u^*+T^*(q)]=B_u^*$, the fourth line uses the fact that q^*, S^* and z^* satisfy the incentive compatibility constraint (10), and the last line uses the fact that q^*, c^* and z^* satisfy the resource constraint (11).

Fifth, we claim that the budget of the policy maker is balanced. To see this, notice that the budget of the policy maker is

$$\begin{aligned} & \int \eta(\hat{q}_y)\left[T_e^*(\hat{w}(\hat{q}_y))+B_u^*\right]dF(y)-B_u^* \\ = & \int q_y^*\lambda(q_y^*)\left[T^*(q_y^*)+B_u^*\right]dF(y)-B_u^* \\ = & \int q_y^*\left(\mu_2^*/\,\mu_1^*\right)dF(y)-\mu_2^*/\,\mu_1^*=0, \end{aligned} \tag{F9}$$

where the second line uses the fact that $\hat{q}_y=q_y^*$, the third line uses the fact that $\lambda(q_y^*)\left[T^*(q_y^*)+B_u^*\right]=\mu_2^*/\,\mu_1^*$ and $B_u^*=\mu_2^*/\,\mu_1^*$, and the last line uses the fact that either $\mu_2^*/\,\mu_1^*=0$ or $\int q_y^* dF(y)=1$.

Taken together, the above observations show that the allocation $(\hat{q},\hat{w},\hat{z},\hat{S})$ is an equilibrium that implements the solution to the mechanism design problem. ■