Taxation and Redistribution of Residual Income Inequality

Mikhail Golosov

Princeton University and National Bureau of Economic Research

Pricila Maziero

University of Pennsylvania

Guido Menzio

University of Pennsylvania and National Bureau of Economic Research

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 homogeneous 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.

We are grateful to the editor, Philip Reny, and two anonymous referees for their useful comments. We have also benefited from the insights of Iván Werning, Jim Albrecht, Paul Beaudry, Michele Boldrin, Jan Eeckhout, Pieter Gautier, Veronica Guerrieri, Espen Moen, Dale Mortensen, Giuseppe Moscarini, Nicola Pavoni, Tom Sargent, Rob Shimer, Randy Wright, and seminar participants at New York University, University of Wisconsin–Madison, University of British Columbia, Universitat Autonoma de Barcelona, European University Institute, Washington University, Queen's, Science Po, Southampton, and from participants at the Northwestern Conference in Honor of Dale Mortensen (Evanston 2011), the Tinbergen Institute Conference on Friction and Policy (Amsterdam 2011), the American Economic Association meeting (Chicago 2012), and the NBER Summer Institute (Cambridge 2012).

[*Journal of Political Economy,* 2013, vol. 121, no. 6]

I. 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 the observed differences in labor income reflect differences in workers' productivity (see, e.g., Mirrlees 1971; Diamond 1998; 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; Autor, Katz, and Kearney 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 (see, e.g., Postel-Vinay and Robin 2002; Mortensen 2003).[1]

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 homogeneous and by a continuum of firms that are heterogeneous with respect to their productivity. We assume that workers are ex ante homogeneous in order to focus on the case in which no income inequality is attributable to differences in workers' productivity, which is the opposite of the case traditionally studied in the public finance literature. We assume that trade in the labor market is decentralized and frictional as in Shimer (1996) and Moen (1997). 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 returns 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

[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.

in order to attract more job seekers and, hence, to increase their probability of trade. Also because of search frictions, inherently identical workers end up having different incomes. In particular, workers who are employed at more productive firms 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, that is, 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 that 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"; that is, 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 takes place only between workers who successfully and unsuccessfully search for jobs offering the same wage, so that the net resources redistributed between workers who search for jobs offering different wages are zero. This implies that the optimal tax level for a worker at a job offering a particular wage should be proportional to the inverse of the probability of finding the job. Second, workers are ex ante indifferent between searching for jobs offering different wages. This implies that utility from after-tax income from any job should also be proportional to the inverse of the probability of finding the job. When utility is strictly concave, these two conditions can hold simultaneously only if the tax schedule is strictly concave, that is, regressive.

The two properties of the equilibrium that guarantee that the optimal labor income tax is regressive are related to the directed nature of the search process. 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 need for redistribution among workers employed at different wages and, hence, among workers seeking jobs offering different wages. Moreover, with directed search, workers choose where to search. For this reason, workers are ex ante indifferent between searching for jobs offering different wages.[2]

The optimality of a regressive labor income tax is a somewhat surprising result. Conventional wisdom suggests that the larger the component of income inequality that is residual—that is, that does not originate from differences in workers' productivity—the more progressive the optimal labor income tax should be (see, e.g., Varian 1980). In our model, the optimal labor income tax is regressive even though, by construction, all income inequality is residual. Intuitively, in our model, a progressive labor income tax would not be optimal because it would induce too many

[2] If the search process is random (as in, e.g., Pissarides [1985], Mortensen and Pissarides [1994], and Burdett and Mortensen [1998]), the optimal labor income tax would be progressive. In fact, with random search, wage differentials among employed people reflect only luck and, hence, should be completely redistributed away. Whether the job search process is random or directed is an unresolved empirical question. There are, however, some bits of evidence that favor the directed search hypothesis. For example, in a recent survey of the US labor market, Hall and Krueger (2008) find that 84 percent of white, male, noncollege workers either "knew exactly" or "had a pretty good idea" about how much their current job would pay from the very beginning of the application process. Another piece of evidence in favor of directed search comes from Holzer, Katz, and Krueger (1991). Using data from the 1982 Employment Opportunity Pilot Project Survey, this study finds that firms in high-wage industries tend to attract more applicants per vacancy than firms in low-wage industries.

workers to seek employment at low-productivity firms and too few workers to seek employment at high-productivity firms.

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; Laroque 2005), some in a dynamic environment (e.g., Farhi and Werning 2011; Golosov, Troshkin, and Tsyvinski 2011), and some in a frictional environment (e.g., Hungerbuhler et al. 2006; Shaal and Taschereau-Dumouchel 2012). Yet, all these papers assume that the income inequality originates from inherent productivity differences among workers. In contrast, our paper characterizes the income tax system that implements the optimal redistribution of income inequality that emerges among identical workers because of search frictions in the labor market.

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

The paper by Acemoglu and Shimer (1999) is closest to ours. They consider a directed search model of the labor market populated by ex ante homogeneous workers and firms and they study the effect of an unemployment benefit financed by lump-sum taxes on the entry of the firm, on the firm's investment in capital, and on aggregate output. Their main finding is that a strictly positive unemployment benefit not only provides insurance but also increases aggregate output. Intuitively, the unemployment benefit reduces the search risk faced by workers and allows them to apply for higher-productivity jobs that pay higher wages and are harder to find. In our paper, we apply the mechanism design approach to the optimal policy problem rather than restricting attention to a particular policy mix. First, we find that the constrained efficient allocation is attained by a strictly positive unemployment benefit and an increasing and regressive labor earning tax. Thus, the policy mix considered by Acemoglu and Shimer can increase output but cannot maximize welfare. Intuitively, when the unemployment benefit is financed by lump-sum taxes, high-productivity firms attract too many workers to their jobs because they do not fully internalize the social cost of an application. Second, we find that aggregate output is higher in the constrained efficient allocation than in the laissez-faire equilibrium. Thus, the main insight of Acemoglu and Shimer—that is, that it is possible to simultaneously increase insurance and output—also holds at the solution to the mechanism design problem.

II. 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.

A. Environment

The economy is populated by a continuum of homogeneous workers with measure one. Each worker has preferences that are described by the utility function $u(c)$, where $u : \mathbb{R}_+ \rightarrow \mathbb{R}$ 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 $y \in [\underline{y}, \bar{y}]$, $0 < \underline{y} < \bar{y}$, and the measure of firms with type less than y is denoted by $F(y)$, where $F : [\underline{y}, \bar{y}] \rightarrow [0, m]$ is a twice-differentiable and strictly increasing function with boundary conditions $F(\underline{y}) = 0$ and $F(\bar{y}) = m$. 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.[3] 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; 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

[3] Qualitatively, the main result of the paper would be unchanged if we were to assume that firms are owned by entrepreneurs rather than by workers. That is, if firms were owned by entrepreneurs, 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.

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 $\lambda(q)q = \eta(q)$.[4] If a firm of type y matches with a worker, it produces y units of output, it pays the wage w to the worker, and it pays 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—that is, 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—that is, 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.

B. Equilibrium

An allocation is a tuple (w, q, z, S). The first element of the tuple is a function $w : [\underline{y}, \bar{y}] \to \mathbb{R}_+$, 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 $y \in [\underline{y}, \bar{y}]$,

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

[4] All 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 $\lambda(q) = (1 + q^\sigma)^{-1/\sigma}$.

ii. Optimal number of applications:

$$\int q(w(y))\,dF(y) \leq 1 \quad \text{and} \quad S \geq k$$

with complementary slackness.

iii. Optimal direction of applications: for all $w \in \mathbb{R}_+$,

$$\lambda(q(w))[u(z+w) - u(z)] \leq S \quad \text{and} \quad q(w) \geq 0$$

with complementary 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; 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.[5] Finally, condition iv guarantees that the dividends received by the workers are equal to the firms' profits.

[5] 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.

C. *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 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, that is, $p'(q) = \phi_q(q, z, S) > 0$. 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, that is, $\phi_S(q, z, S) > 0$. Intuitively, the higher S is, the higher 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, that is, $\phi_{qS}(q, z, S) > 0$. Intuitively, the higher S is, the larger the difference between the consumption of the worker if his application succeeds and if his application fails. Hence, owing 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; that is, $\phi_{qz}(q, z, S)$ may be positive or negative. The results in

Section III are derived under the assumption that $p'(q)$ is decreasing in z.[6] The other results, including the two main theorems in Sections IV and V, 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 $y \in [\underline{y}, \bar{y}]$, 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 inframarginal 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.

[6] The assumption is satisfied when the utility function u has the hyperbolic absolute risk aversion form

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

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

Next, we characterize the consumption z of an unemployed worker. From equilibrium 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—that is, 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 utility 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 than risk neutral. To see this, suppose that workers and firms could trade job applications at some competitive price $\bar{p}$. If that was the case, every worker would sell his application, enjoy a consumption of $z + \bar{p}$, and have a marginal utility of $u'(z + \bar{p})$. Moreover, every firm would purchase applications up to the point where the marginal productivity $\eta'(q)y$ is equal to $\bar{p}$. Hence, the equilibrium would implement the first-best allocation. However, firms and workers cannot trade job applications because firms 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 trade only successful applications. That is, they can trade only 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.[7] 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.

III. 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, that is, the allocation that maximizes the worker's expected utility given the production technology $F(y)$, given 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. The reason is that workers face a search risk that is not insured at all in the equilibrium, while it is partially insured in the second-best.

A. 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

[7] 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.

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, that is, a mechanism that recommends the same search strategy to all workers.[8]

More specifically, a symmetric mechanism is a tuple (π, c, z, S). The first element of the tuple is a differentiable function $\pi : [\underline{y}, \bar{y}] \to \mathbb{R}_+$, 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 $c : [\underline{y}, \bar{y}] \to \mathbb{R}_+$, 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, $a = \pi(\bar{y})$. Hence, we will think of the mechanism as a tuple (a, q, c, z, S) rather than a tuple (π, c, z, S).

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

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

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 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, z, 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,

[8] 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 et al. 2001; Shimer 2005; Galenianos and Kircher 2012).

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

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, z, 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,

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

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

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, $\lambda(q_y)[u(c_y) - u(z)] \leq S$. 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 we characterize the solution to the above mechanism design problem, some comments 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 maximized worker's expected utility when the mechanism does receive reports from the workers is the same as when it does not. Intuitively, the mechanism cannot exploit the reports because the outcome of the randomization is the workers' private information.

B. *Solution to the Mechanism Design Problem*

Consider a solution to the mechanism design problem $(a^*, q^*, c^*, z^*, S^*)$, which in Appendix C we prove to always exist. 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 $(a^*, q^*, c^*, z^*, S^*)$ as either the constrained efficient allocation or the second-best allocation.

The constrained efficient value of searching is[9]

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

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 S is, the greater the consumption risk that he faces. Hence, the constrained efficient value of S is k.

The constrained efficient queue length is such that

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

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

[9] Appendix D contains the explicit derivation of the optimality conditions (14)–(18).

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

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

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

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 a solution to the mechanism design problem and let (μ_1^*, μ_2^*) denote 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 $\bar{k}$, with $0 < \underline{k} \leq \bar{k} < \infty$, such that $\mu_2^*/\mu_1^* = 0$ for all $k > \bar{k}$ and $\mu_2^*/\mu_1^* > 0$ for all $k < \underline{k}$.

Proof. See Appendix E.

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.

C. *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 $k < \underline{k}$ 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 a solution to the mechanism design problem with $k < \underline{k}$, and let (q, c, z, S) be a competitive equilibrium with $S > k$. (i) There exists a $y_0 \in (y_c, \bar{y})$ 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 $y \in (y_0, \bar{y})$. (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 F.

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 II—$\phi_q(q, z, S)$ 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 proposition implies that aggregate output in the equilibrium, $Y = \int \eta(q_y) y dF(y)$, is lower than 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

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

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 lower than in the second-best, $U^* = 1 - \int \eta(q_y^*) dF(y)$. To see this, notice that

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

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 the equilibrium than in the second-best because, in the equilibrium, applicants are more evenly distributed across different types of firms and the vacancy-filling probability, $\eta(q)$, is a concave function of applicants.

IV. Policy Implementation

In this section, we prove that any 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.

A. *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 $y \in [\underline{y}, \bar{y}]$,

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

ii. Optimal number of applications:

$$\int q(w(y))\,dF(y) \leq 1 \quad \text{and} \quad S \geq k$$

with complementary slackness.

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

$$\lambda(q(w))[u(z + w + T_e(w)) - u(z + B_u)] \leq S \quad \text{and} \quad q(w) \geq 0$$

with complementary slackness.

iv. Consistency of dividends and profits:

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

v. Balanced budget:

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

The first four conditions in definition 2 are nearly identical to those in definition 1. The only difference is that, here, the consumption of an unemployed worker is given by the sum of 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 by $\hat{q}_y$ the number of applicants attracted by a firm of type y and by $\hat{w}(q)$ the wage that a firm needs to offer in order to attract q applicants.

B. *Optimal Policy*

The following theorem states the main result of our paper.

Theorem 1 (Optimal policy). Let $(a^*, q^*, c^*, z^*, S^*)$ denote a solution to the mechanism design problem and let (μ_1^*, μ_2^*) denote the multipliers associated with the aggregate resource constraints (8) and (9). The allocation $(a^*, q^*, c^*, z^*, S^*)$ 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 G.

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 that 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 takes place only between workers who apply to the same type of firms, and (b) workers are ex ante indifferent between applying to different types of firms.[10]

It is useful to flesh 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

[10] 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.

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 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 benefit. In fact, notice that

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

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, that is, $\lambda(q)T(q) = [1 - \lambda(q)]B_u$. This property implies that the optimal policy involves redistribution between workers who apply to the 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.

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

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

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 takes place only between workers applying to the same type of firm, and (*b*) workers are ex ante indifferent between applying to different types 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 be satisfied only 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.

So far we have interpreted B_u and T_e as an unemployment benefit and a labor income tax. However, one could also interpret B_u as an unemployment benefit and T_e as the contributions paid by the firms and their employees to the unemployment insurance system. Then, (26) can be interpreted as saying that the contributions paid by a firm and its employees to the unemployment insurance system should be equal, in expectation, to the unemployment benefits enjoyed by the workers who unsuccessfully apply to that firm. With this interpretation, it becomes clear that our results about optimal policy are related to the literature on experience rating. Feldstein (1976), Topel and Welch (1980), and, more recently, Mongrain and Roberts (2005) pointed out that the contributions of a firm to the unemployment insurance system should be *experience rated*, in the sense that total contributions paid by a firm should be equal to the unemployment benefits paid to the workers who have been laid off by that firm. This property guarantees that, when a firm decides how many employees to dismiss, it will fully internalize the social cost of their unemployment. Our paper points out that the contributions paid by a firm should also depend on the wages that it offers. As we have explained before, this property is needed to guarantee that, when a firm decides what wage to offer (and, hence, how many applicants to attract), it will fully internalize the social cost of the unemployment benefits paid to the workers who unsuccessfully seek its vacant jobs.[11]

V. Insurance Market Implementation

In this section, we prove that any constrained efficient allocation could be decentralized as a laissez-faire equilibrium if there was a competitive

[11] We are grateful to Iván Werning and to an anonymous referee for pointing out the connection between experience rating and our optimal policy.

insurance market in which workers could purchase insurance against search risk. The result implies that the economy described in Section II is constrained inefficient because such an insurance market is missing. Moreover, the result implies that the role of the optimal policy described in Section IV is to substitute for the missing insurance market.

A. 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 offers only 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 II. 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 $y \in [\underline{y}, \bar{y}]$,

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

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

$$(b_u(w), t_e(w)) \in \arg\max_{(b_u, t_e)} \lambda(q(w))t_e - [1 - \lambda(q(w))]b_u,$$

subject to

$$\lambda(q(w))u(z + w - t_e) + [1 - \lambda(q(w))]u(z + b_u) - k \geq U,$$
$$\lambda(q(w))[u(z + w - t_e) - u(z + b_u)] \geq k.$$

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:

$$\int q(w(y))dF(y) \leq 1 \quad \text{and} \quad U \geq u(z)$$

with complementary 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, b_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.[12] As in Sections II and IV, we denote with q_y the

[12] As we discussed in Sec. II, definition 1 satisfies the spirit of subgame perfection. In fact, when a firm offers an off-equilibrium wage, it expects to attract a number of applicants such that workers are indifferent between seeking that wage and other wages. Now, we

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.

B. Constrained Efficiency of Equilibrium

The following theorem is the second major result of our paper.[13]

THEOREM 2 (Efficiency of equilibrium with insurance markets). Let (q^*, c^*, z^*, S^*) be a solution to the mechanism design problem. Let (q, w, b_u, t_e, z, U) be an allocation such that $q_y = q^*_y$,

$$w(q) = u^{-1}(S^*/\lambda(q) + u(z^*)) + (\mu^*_2/\mu^*_1)/\lambda(q) - z^*,$$

$b_u(w) = B^*_u$, $t_e(w) = T^*_e(w)$, $z = z^* - B^*_u$, 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) decentralizes the constrained efficient allocation (q^*, c^*, z^*, S^*).

The second part of theorem 2 is not surprising. In the proposed equilibrium, the equilibrium insurance contracts exactly reproduce the optimal tax system. That is, the equilibrium insurance contracts are such that, if the worker remains unemployed, the payment from the insurance company to the worker, $b_u(w)$, 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.

argue that definition 2 satisfies the spirit of subgame perfection as well. First, consider an insurance company that entertains offering an off-equilibrium contract. The definition of equilibrium implies that the insurance company expects to sell the contract only if it provides workers with an expected utility of at least U. Moreover, the definition of equilibrium implies that the insurance company does not expect its contract choice to affect the equilibrium queue length $q(w)$. This assumption is justified if insurance companies can serve only a small number of customers and, hence, cannot affect the equilibrium queue length. Now, consider a firm that entertains posting an off-equilibrium wage. The definition of equilibrium implies that the firm expects the insurance companies to offer the profit-maximizing contract for that wage, and given such a contract, it expects to attract a queue length such that the workers' expected utility from applying to the job is U. Therefore, definition 2 is in the spirit of subgame perfection, given the view that insurance companies are small. Clearly, we could have used a different definition of equilibrium in which individual insurance companies can affect the queue length. We believe that theorem 2 would generalize to this alternative definition of equilibrium.

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

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:

$$\max_{(b_u, t_e)} \lambda(q)t_e - [1-\lambda(q)]b_u,$$
$$\text{subject to } \lambda(q)u(z + w(q) - t_e) + [1-\lambda(q)]u(z + b_u) - k \geq U, \tag{29}$$
$$\lambda(q)[u(z + w(q) - t_e) - u(z + b_u)] \geq k.$$

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

$$\lambda(q)[u(z + w(q) - t_e) - U] = k.$$

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, that is, $\lambda(q)T_e^*(w(q)) = [1-\lambda(q)]B_u^*$, it follows that the maximized profits of an insurance company are equal to zero. Thus, the profit-maximizing contract reproduces the optimal tax system and is consistent with perfect competition in the insurance market.

From a theoretical point of view, theorem 2 is quite interesting. The theorem implies that the economy described in Section II is constrained inefficient because the insurance market is missing. Moreover, the theorem implies that the role of the optimal policy described in Section IV is to substitute for the missing insurance market. From an empirical point of view, however, theorem 2 is probably irrelevant as we rarely observe any private provision of insurance against search risk. There are several reasons why this might be the case. For example, it may be impossible or prohibitively costly to make sure that workers sign insurance contracts with exclusively one provider. If exclusivity cannot be enforced, workers would sign as many insurance contracts as possible, they would not search for work, and then they would collect the transfer b_u from all of their insurance providers. Anticipating the workers' behavior, insurance providers would choose to not offer any coverage against search risk. Thus, if exclusivity cannot be enforced, the insurance market would shut down and the government should implement the optimal policy described in Section IV. Notice that lack of exclusivity would not undermine the optimal policy because the government can make sure to be the sole provider of insurance.

VI. 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 homogeneous 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 takes place only between workers applying to the same type of firm, and (*b*) 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 recommendations, 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 search frictions, we would have to introduce some degree of inherent heterogeneity among workers.

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 have to show only that the expected wage bill of the firm is convex with respect to q. That is, we have to show only 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

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

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

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

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 $p''(q)q + 2p'(q) > 0$. QED

Appendix 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 $\tilde{y} \le y$. 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 choose 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

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

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 who 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

$$\begin{aligned} &\lambda(q_y)u(c_y(y)) + [1 - \lambda(q_y)]u(z_y) - k \\ &\geq \lambda(q_{\hat{y}})u(c_y(\hat{y})) + [1 - \lambda(q_{\hat{y}})]u(z_y) - k. \end{aligned} \tag{B2}$$

Moreover, the mechanism must induce a worker who reports seeking a firm of type y to actually seek a firm of type y rather than 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 who 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 who 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

$$\begin{aligned} &\lambda(q_y)u(c_y(y)) + [1 - \lambda(q_y)]u(z_y) \\ &\geq \lambda(q_{\hat{y}})u(c_{\hat{y}}(\hat{y})) + [1 - \lambda(q_{\hat{y}})]u(z_{\hat{y}}). \end{aligned} \tag{B5}$$

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,

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

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 the problem that we analyzed in Section III. 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 saying 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 nonrecommended action is nongreater 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

$$\lambda(q_y) u(c_y(y)) + [1 - \lambda(q_y)] u(z_y) - U = 0, \tag{B12}$$

$$U - u(z_0) \geq 0, \tag{B13}$$

$$(1 - a)[U - u(z_0)] = 0. \tag{B14}$$

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 III is that the consumption of unemployed workers can be made contingent on 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 (B13) and (B14). Let $\rho_{1,y} dF(y)$ and $\rho_{2,y} dF(y)$ denote the multipliers associated with the constraints (B11) and (B12). 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

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

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

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

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 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 (B14) 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. QED

Appendix C

Existence of an Optimal Mechanism

The mechanism design problem described in Section III.A can be written as

$$\begin{gathered}\max_{a,z,S} u(z)+a(S-k),\\ \text{subject to } a\in[0,1], z\in[0,\bar{y}], S\in[k,u^{-1}(\bar{y})],\\ (1-a)(S-k)=0,\\ X^*(1,z,S)-z\geq 0,\end{gathered} \tag{C1}$$

where $X^*(a,z,S)$ is defined as

$$\begin{gathered}X^*(a,z,S)=\max_{q:[\underline{y},\bar{y}]\to\mathbb{R}_+}\int \eta(q_y)\left\{y-\left[u^{-1}\left(\frac{S}{\lambda(q_y)}+u(z)\right)-z\right]\right\}dF(y),\\ \text{subject to}\int q_y dF(y)=a.\end{gathered} \tag{C2}$$

In words, the mechanism design problem can be written as a two-stage problem. In the first stage, the mechanism chooses the measure of applicants, a, the consumption of the unemployed, z, and the value of searching, S, so as to maximize the workers' expected utility. In the second stage, the mechanism chooses how to allocate a applicants across different firms so as to maximize aggregate output net of the extra consumption that must be assigned to employed workers to make the allocation incentive compatible.

It is easy to verify that the mechanism design problem can be written as (C1). In the original problem, the mechanism chooses (a, q, c, z, S) so as to maximize the objective function (7) subject to the constraints (8)–(13). Note that the objective function (7) can be written as $u(z) + a(S - k)$ after substituting the constraint (11). The constraints (9), (12), and (13) are equivalent to the constraints $1 - a \geq 0$, $S - k \geq 0$, and $(1 - a)(S - k) = 0$. The constraints $S \leq u^{-1}(y)$ and $z \leq \bar{y}$ in (C1) do not have a counterpart in the original formulation of the problem, but they can be added without loss in generality because aggregate output cannot be greater than $\bar{y}$. The constraint (11) gives c_y as a function of z, S, and q_y, that is, $c_y = u^{-1}(S/\lambda(q_y) + u(z))$. After these transformations, note that the choice of q_y and c_y affects only the resource constraint on output (8) and is subject to only the resource constraint on applicants (9). Therefore, we can think of q_y and c_y as being chosen so as to maximize the left-hand side of (8) subject to (9), given a, z, and S and given $c_y = u^{-1}(S/\lambda(q_y) + u(z))$. This problem for q_y is described in (C2), and the constraint (8) is the constraint $X^*(a, z, S) - z \geq 0$ in (C1).

In order to prove the existence of a solution to the mechanism design problem, we take several steps.

First, we establish the existence of a solution to the second-stage problem (C2). We establish the existence of a solution to this problem by guessing that a particular allocation is optimal and then by showing that any other allocation can be improved on. Second, we establish that the value function associated with the second-stage problem (C2) is continuous in a, z, and S. Finally, we establish the existence of a solution for the first-stage problem (C1). Here, the existence of a solution follows from the fact that the objective function is a continuous function and the feasible set is compact.

Step 1: The second-stage problem (C2) involves allocating a measure a of applicants across different types of firms so as to maximize aggregate output net of the extra consumption that must be assigned to employed workers to make the allocation incentive compatible. In (C2), the consumption of the unemployed, $z \in [0, \bar{y}]$, the value of searching, $S \in [k, u^{-1}(\bar{y})]$, and the measure of applicants, $a \in [0, 1]$, are taken as given. Here, we prove that (C2) admits one and only one solution, q_y^*.

Let the integrand in (C2) be defined as

$$\varphi(q, y; z, S) \equiv \eta(q)y - q\lambda(q)\left[u^{-1}\left(\frac{S}{\lambda(q)} + u(z)\right) - z\right]. \tag{C3}$$

The partial derivative of φ with respect to q is given by

$$\begin{aligned}\varphi_q(q, y; z, S) = {} & \eta'(q)\left[y + z - u^{-1}\left(\frac{S}{\lambda(q)} + u(z)\right)\right] \\ & - \left[u'\left(\frac{S}{\lambda(q)} + u(z)\right)\right]^{-1} \frac{S\lambda'(q)q}{\lambda(q)}.\end{aligned} \tag{C4}$$

The partial derivative $\varphi_q(q, y; z, S)$ is continuous and strictly decreasing in q with codomain $[-\infty, \bar{\varphi}_q(y)]$, where $\bar{\varphi}_q(y)$ is a strictly decreasing function of y. We denote as $\varphi_q^{-1}(\mu, y; z, S)$ the inverse of $\varphi_q(q, y; z, S)$ with respect to q. The inverse function $\varphi_q^{-1}(\mu, y; z, S)$ is continuous and strictly decreasing in μ over the domain $[-\infty, \bar{\varphi}_q(y)]$ and such that $\varphi_q^{-1}(\bar{\varphi}_q(y), y; z, S) = 0$. Moreover, the partial derivative $\varphi_q(q, y; z, S)$ is continuous and strictly increasing in y and the inverse $\varphi_q^{-1}(\mu, y; z, S)$ is continuous and strictly increasing in y.

Let $q_y^* : Y \to \mathbb{R}_+$ be defined as

$$q_y^* = \begin{cases} 0 & \text{if } \varphi_q(0, y) \leq \mu^* \\ \varphi_q^{-1}(\mu^*, y) & \text{if } \varphi_q(0, y) \geq \mu^*, \end{cases} \tag{C5}$$

where μ^* is such that

$$\int q_y^* \, dF(y) = a. \tag{C6}$$

Notice that there exists one and only one pair (q_y^*, μ^*) that satisfies (C5)–(C6). From the properties of $\varphi_q^{-1}(\mu^*, y)$, it follows that $\int q_y^* \, dF(y) = 0$ for all $\mu^* \geq \bar{\varphi}_q(\bar{y})$ and that $\int q_y^* \, dF(y) \geq 1$ for all μ^* sufficiently low. Moreover, since $\int q_y^* \, dF(y)$ is continuous and strictly decreasing in μ^* (whenever positive), it follows that there exists a unique μ^* such that (C6) holds. Given the μ^* that satisfies (C6), there is a unique q_y^* that satisfies (C5).

Now, we prove that q_y^* is the unique solution to (C2). Let q_y denote some arbitrary allocation that satisfies the feasibility constraint in (C2). Suppose that q_y is such that there exists a subset $Y_1 \subset [\underline{y}, \bar{y}]$ such that $F(Y_1) > 0$ and $q_y \geq 0$ and $\varphi_q(q_y, y) \geq \bar{\mu}$ for all $y \in Y_1$. Also, suppose that there exists another subset $Y_2 \subset [\underline{y}, \bar{y}]$ such that $F(Y_2) = F(Y_1)$ and $q_y > 0$ and $\varphi_q(q_y, y) \leq \underline{\mu}$ for all $y \in Y_2$, where $\underline{\mu} < \bar{\mu}$. Now, consider an alternative allocation $\hat{q}_y$ such that (i) $\hat{q}_y = q_y$ if $y \notin Y_1 \cup Y_2$; (ii) $\hat{q}_y > q_y$ and $\varphi(\hat{q}_y, y) \geq (\underline{\mu} + \bar{\mu})/2$ if $y \in Y_1$; (iii) $\hat{q}_y < q_y$ and $\varphi(\hat{q}_y, y) \leq (\underline{\mu} + \bar{\mu})/2$ if $y \in Y_2$; and (iv) $\int \hat{q}_y \, dF(y) = a$. Clearly, there exists an allocation $\hat{q}_y$ that satisfies these conditions because $\varphi(q, y)$ is concave in q. Moreover, the allocation $\hat{q}_y$ is such that

$$\begin{aligned} &\int \varphi(\hat{q}_y, y) \, dF(y) - \int \varphi(q_y, y) \, dF(y) \\ &\quad = \int_{Y_1} [\varphi(\hat{q}_y, y) - \varphi(q_y, y)] \, dF(y) - \int_{Y_2} [\varphi(q_y, y) - \varphi(\hat{q}_y, y)] \, dF(y) \\ &\quad > \int_{Y_1} \varphi_q(\hat{q}_y, y) \, dF(y) - \int_{Y_2} \varphi_q(\hat{q}_y, y) \, dF(y) \\ &\quad \geq \frac{\underline{\mu} + \bar{\mu}}{2} F(Y_1) - \frac{\underline{\mu} + \bar{\mu}}{2} F(Y_1) = 0, \end{aligned} \tag{C7}$$

where the first inequality in (C7) follows from the fact that $\varphi(q, y)$ is strictly concave in q. Therefore, the original allocation q_y generates less net output than $\hat{q}_y$ and, hence, cannot be a solution to (C2). In turn, this implies that any solution to (C2) must be such that $\varphi_q(q_y, y) = \mu$ whenever $q_y > 0$ and $\varphi_q(q_y, y) \leq \mu$ whenever $q_y = 0$. The unique feasible allocation with this property is q_y^*.

Step 2: We prove that $X^*(a, z, S)$ is continuous in a, z, and S. In order to prove that X^* is continuous with respect to S, denote with $q_1^*(y)$ the solution of (C2) for $S = S_1$ and with $q_2^*(y)$ the solution of (C2) for $S = S_2$, with $S_1 < S_2$. Clearly,

$$X^*(a,z,S_1) - X^*(a,z,S_2) \geq 0. \tag{C8}$$

Moreover, we have

$$\begin{aligned} X^*(a,z,S_1) - X^*(a,z,S_2) & \\ &\leq \int [\varphi(q_1^*(y),y;S_1,z) - \varphi(q_1^*(y),y;S_2,z)]\,dF(y) \\ &\leq \int \left[u'\left(\frac{S_1}{\lambda(q_1^*(y))} + u(z)\right)\right]^{-1} \frac{S_2 - S_1}{\lambda(q_1^*(y))}\,dF(y) \\ &\leq (S_2 - S_1)\left\{\left[u'\left(\frac{S_1}{\lambda(q_1^*(\bar{y}))} + u(z)\right)\right]^{-1} \frac{1}{\lambda(q_1^*(\bar{y}))}\right\}, \end{aligned} \tag{C9}$$

where the first inequality follows from the fact that $q_1^*(y)$ is feasible but not necessarily optimal for $S = S_2$, the second inequality follows from the convexity of u^{-1}, and the last inequality follows from the fact that $q_1^*(y)$ is increasing in y. Taken together, (C8) and (C9) imply that $X^*(a,z,S)$ is continuous in S. Using similar arguments, we can also prove that $X^*(a,z,S)$ is continuous in a and z.

Step 3: We prove that the first-stage problem (C1) admits a solution. To this aim, notice that the objective function in (C1) is continuous in a, z, and S. The feasible set in (C1) is compact because the constraints $a \in [0,1]$, $z \in [0,\bar{y}]$, and $S \in [k, u^{-1}(\bar{y})]$ define a compact set and the constraints $X^*(1,z,S) - z \geq 0$ and $(1-a)(S-k)$ are satisfied on a closed subset of that compact set. Hence, there exists a solution to (C1). QED

Appendix D

Optimal Mechanism

In this appendix, we derive the optimality conditions (14)–(18) for a 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 with the incentive compatibility constraint (11). Finally, let ν_1 and ν_2 denote the multipliers associated with 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 must satisfy a number of first-order conditions. The first-order condition for c_y is

$$q_y u'(c_y) + \rho_y u'(c_y) = \mu_1 q_y. \tag{D1}$$

The first-order condition for z is

$$\begin{aligned} u'(z)\left[1 - \int \lambda(q_y) q_y\,dF(y)\right] = u'(z)\left[\int \rho_y \lambda(q_y)\,dF(y)\right] & \\ + \mu_1\left[1 - \int \lambda(q_y) q_y\,dF(y)\right]&. \end{aligned} \tag{D2}$$

The first-order condition for q_y is

$$\begin{aligned}&\{u(z) + S - k + q_y\lambda'(q_y)[u(c_y) - u(z)]\}\\ &+ \mu_1\{\eta'(q_y)y - \lambda(q_y)c_y - [1 - \lambda(q_y)]z + q_y\lambda'(q_y)(c_y - z)\}\\ &\leq \mu_3 - \rho_y\lambda'(q_y)[u(c_y) - u(z)].\end{aligned} \tag{D3}$$

The first-order condition for a is

$$u(z) + \mu_2 = \mu_1 z + \mu_3. \tag{D4}$$

The first-order condition for S is

$$\nu_1 + \nu_2(1 - a) = \int \rho_y dF(y). \tag{D5}$$

In addition to the first-order conditions (D1)–(D5), the 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, the solution to the mechanism design problem must be such that

$$1 - a \geq 0, \quad \mu_2 \geq 0, \quad \text{and} \quad (1 - a)\mu_2 = 0, \tag{D6}$$

$$S - k \geq 0, \quad \nu_1 \geq 0, \quad \text{and} \quad (S - k)\nu_1 = 0. \tag{D7}$$

Using equation (D2) to solve for μ_1 and equation (D1) to solve for ρ_y, it follows that the right-hand side of (D5) 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 (D7) 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

$$S = k. \tag{D8}$$

Solving equation (D1) for ρ_y and equation (D4) for μ_2 and substituting the solutions into the first-order condition (D3), we obtain

$$\eta'(q_y)y - \lambda(q_y)(c_y - z) + q_y\lambda'(q_y)\left[\frac{u(c_y) - u(z)}{u'(c_y)} - (c_y - z)\right] = \frac{\mu_2}{\mu_1}. \tag{D9}$$

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

$$c_y = u^{-1}\left(\frac{k}{\lambda(q)} + u(z)\right). \tag{D10}$$

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

$$z = \int \eta(q_y)[y - (c_y - z)]dF(y). \tag{D11}$$

Finally, using the optimality condition (D7) and the resource constraint (10), it is straightforward to verify that

$$1 - \int q_y dF(y) \geq 0, \quad \frac{\mu_2}{\mu_1} \geq 0, \quad \text{and} \quad \left[1 - \int q_y dF(y)\right]\frac{\mu_2}{\mu_1} = 0. \tag{D12}$$

The optimality conditions (D8)–(D12) are the same as the conditions (14)–(18). QED

Appendix E

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

$$\eta'(q_y^*)y = \lambda(q_y^*)(c_y^* - z^*) - q_y^*\lambda'(q_y^*)\left[\frac{u(c_y^*) - u(z^*)}{u'(c_y^*)} - (c_y^* - z^*)\right] \tag{E1}$$

and the constrained efficient consumption for employed workers is such that

$$c_y^* = u^{-1}\left(\frac{k}{\lambda(q_y^*)} + u(z^*)\right). \tag{E2}$$

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

$$z^* = \int \eta(q_y^*)[y - (c_y^* - z^*)]dF(y), \tag{E3}$$

and the constrained efficient queue satisfies the resource constraint on applicants,

$$\int q_y^* dF(y) = a^* \leq 1. \tag{E4}$$

Let $\tilde{q}(y, z, k)$ denote the solution for q to the equations (E1) and (E2) with respect to q. It is straightforward to verify that $\tilde{q}(y, z, k)$ is increasing in y, increasing in z, 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 $y \in [\underline{y}, \bar{y}]$ and any $z \in \mathbb{R}_+$. For any $k < \underline{k}$, the constrained efficient queue length is such that

$$\begin{aligned} \int q^*(y)dF(y) &= \int \tilde{q}(y, z^*, k)dF(y) \\ &> \int \tilde{q}(y, 0, \underline{k})dF(y) = 1. \end{aligned} \tag{E5}$$

This implies that, for $k < \underline{k}$, there is no solution to the system of equations (E1)–(E4). 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 $\bar{k}$ such that, for $k > \bar{k}$, $\mu_2^*/\mu_1^* = 0$. QED

Appendix F

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 to $u(z) + S - k$ in the equilibrium, and it is equal to $u(z^*) + S^* - k$ in the second-best. Since $k < \underline{k}$, $\mu_2^*/\mu_1^* > 0$ and the equilibrium is constrained inefficient, and hence, the workers' expected utility is strictly lower that in the second-best. That is,

$$u(z) + S - k < u(z^*) + S^* - k. \tag{F1}$$

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

$$\begin{aligned} c'(q) &= -\frac{\lambda'(q)}{\lambda(q)^2}\frac{S}{u'(c(q))}, \\ c^{*\prime}(q) &= -\frac{\lambda'(q)}{\lambda(q)^2}\frac{S^*}{u'(c^*(q))}. \end{aligned} \tag{F2}$$

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

$$q_y' = \frac{\eta'(q_y)y}{-\eta''(q_y)y + 2\phi_q(q, z, S) + \phi_{qq}(q, z, S)q}. \tag{F3}$$

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

$$q_y^{*\prime} = \frac{\eta'(q_y^*)y}{-\eta''(q_y^*)y + 2\phi_q(q, z^*, S^*) + \phi_{qq}(q, z^*, S^*)q}. \tag{F4}$$

The differential equations (F3) and (F4) 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 $y \in (y_c, \bar{y})$ such that $q_y = q_y^*$. Taken together, the two observations imply that there exists 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 $y \in (y_0, \bar{y})$. QED

Appendix G

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

$$\hat{w}(q) = u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) + T^*(q) - \hat{z}. \tag{G1}$$

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

$$\begin{gathered} \max_{q \geq 0} \eta(q)y - \hat{p}(q)q, \\ \hat{p}(q) = \lambda(q)\left[u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) + T^*(q) - \hat{z}\right]. \end{gathered} \tag{G2}$$

To see this, notice that the problem (G2) is strictly concave, and hence, $\hat{q}_y$ is a solution if and only if it satisfies the first-order condition

$$\eta'(q)y\begin{cases} = \hat{p}'(q)q + \hat{p}(q) & \text{if } q > 0 \\ \leq \hat{p}'(q)q + \hat{p}(q) & \text{if } q = 0. \end{cases} \tag{G3}$$

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

$$\eta'(q)y\begin{cases} = p^{*\prime}(q)q + p^*(q) & \text{if } q > 0 \\ \leq p^{*\prime}(q)q + p^*(q) & \text{if } q = 0. \end{cases} \tag{G4}$$

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

$$\begin{aligned}\hat{p}(q) &= \lambda(q)\left[u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) + T^*(q) - \hat{z}\right]\\ &= \lambda(q)\left[u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right) - z^*\right] + \frac{\mu_2^*}{\mu_1^*}\\ &= p^*(q),\end{aligned}\tag{G5}$$

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 (G5), it follows that $\hat{q}_y$ satisfies the first-order condition (G3), and hence, it solves the problem of the firm (G2).

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

$$\begin{aligned}&\lambda(q)[u(\hat{z} + \hat{w}(q) - T_e^*(\hat{w}(q))) - u(\hat{z} + B_u^*)]\\ &\qquad = \lambda(q)\left[u\left(u^{-1}\left(\frac{S^*}{\lambda(q)} + u(z^*)\right)\right) - u(z^*)\right]\\ &\qquad = S^* = \hat{S},\end{aligned}\tag{G6}$$

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

$$\int \hat{q}_y dF(y) = \int q_y^* dF(y) \leq 1.\tag{G7}$$

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

$$\begin{aligned}&\int \eta(\hat{q}_y)[y - \hat{w}(\hat{q}_y)]dF(y)\\ &\qquad = \int \eta(q_y^*)[y - u^{-1}(S^*/\lambda(q_y^*) + u(z^*)) + z^* - B_u^*/\lambda(q)]dF(y)\\ &\qquad = \int \eta(q_y^*)[y - (c_y^* - z^*)]dF(y) - \int q_y^* B_u^* dF(y)\\ &\qquad = z^* - B_u^*\\ &\qquad = \hat{z},\end{aligned}\tag{G8}$$

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

$$\begin{aligned}\int \eta(\hat{q}_y)[T_e^*(\hat{w}(\hat{q}_y)) + B_u^*]dF(y) - B_u^* \\ &= \int q_y^* \lambda(q_y^*)[T^*(q_y^*) + B_u^*]dF(y) - B_u^* \\ &= \int q_y^*(\mu_2^*/\mu_1^*)dF(y) - \mu_2^*/\mu_1^* \\ &= 0,\end{aligned} \tag{G9}$$

where the second line uses the fact that $\hat{q}_y = q_y^*$, the third line uses the fact that $\lambda(q_y^*)[T^*(q_y^*) + B_u^*] = \mu_2^*/\mu_1^*$ 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. QED

References

Acemoglu, D., and R. Shimer. 1999. "Efficient Unemployment Insurance." *J.P.E.* 107:893–927.

Autor, D., L. Katz, and M. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Rev. Econ. and Statis.* 90:300–323.

Burdett, K., and D. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *Internat. Econ. Rev.* 39:257–73.

Burdett, K., S. Shi, and R. Wright. 2001. "Pricing and Matching with Frictions." *J.P.E.* 109:1060–85.

Diamond, P. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *A.E.R.* 88:83–95.

Farhi, E., and I. Werning. 2011. "Insurance and Taxation over the Life Cycle." Manuscript, Harvard Univ.

Feldstein, M. 1976. "Temporary Layoffs in the Theory of Unemployment." *J.P.E.* 84:937–58.

Galenianos, M., and P. Kircher. 2012. "On the Game-Theoretic Foundations of Competitive Search Equilibrium." *Internat. Econ. Rev.* 53:1–21.

Golosov, M., M. Troshkin, and A. Tsyvinski. 2011. "Optimal Dynamic Taxes." Manuscript, Princeton Univ.

Hall, R., and A. Krueger. 2008. "Wage Formation between Newly Hired Workers and Employers: Survey Evidence." Manuscript, Stanford Univ.

Hansen, G., and A. Imrohoroglu. 1992. "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard." *J.P.E.* 100:118–42.

Holzer, H., L. Katz, and A. Krueger. 1991. "Job Queues and Wages." *Q.J.E.* 106:739–68.

Hopenhayn, H., and J. Nicolini. 1997. "Optimal Unemployment Insurance." *J.P.E.* 105:412–38.

Hornstein, A., P. Krusell, and G. Violante. 2011. "Frictional Wage Dispersion in Search Models: A Quantitative Assessment." *A.E.R.* 101:2873–98.

Hungerbuhler, M., E. Lehmann, A. Parmentier, and B. Van der Linden. 2006. "Optimal Redistributive Taxation in a Search Equilibrium Model." *Rev. Econ. Studies* 73:743–67.

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

Menzio, G., and S. Shi. 2011. "Efficient Search on the Job and the Business Cycle." *J.P.E.* 119:468–510.

Mirrlees, J. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Rev. Econ. Studies* 38:175–208.

Moen, E. 1997. "Competitive Search Equilibrium." *J.P.E.* 105:694–723.

Mongrain, S., and J. Roberts. 2005. "Unemployment Insurance and Experience Rating: Insurance versus Efficiency." *Internat. Econ. Rev.* 46:1303–19.

Montgomery, J. 1991. "Equilibrium Wage Dispersion and Interindustry Wage Differentials." *Q.J.E.* 106:163–79.

Mortensen, D. 2003. *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge, MA: MIT Press.

Mortensen, D., and C. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Rev. Econ. Studies* 61:397–415.

Pissarides, C. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies and Real Wages." *A.E.R.* 75:676–90.

Postel-Vinay, F., and J. Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70:2295–2350.

Saez, E. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Rev. Econ. Studies* 68:205–29.

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

Shavell, S., and L. Weiss. 1979. "The Optimal Payment of Unemployment Insurance Benefits over Time." *J.P.E.* 87:1347–62.

Shimer, R. 1996. "Contracts in a Frictional Labor Market." Manuscript, Massachusetts Inst. Tech.

———. 2005. "The Assignment of Workers to Jobs in an Economy with Coordination Frictions." *J.P.E.* 113:996–1025.

Shimer, R., and I. Werning. 2008. "Liquidity and Insurance for the Unemployed." *A.E.R.* 98:1922–42.

Topel, R., and F. Welch. 1980. "Unemployment Insurance: Survey and Extensions." *Economica* 47:351–79.

Varian, H. 1980. "Redistributive Taxation as Social Insurance." *J. Public Econ.* 14:49–68.

Wang, C., and S. Williamson. 2002. "Moral Hazard, Optimal Unemployment Insurance, and Experience Rating." *J. Monetary Econ.* 49:1337–71.