CESifo WORKING PAPERS

7805
2019

August 2019

Multiple Applications, Competing Mechanisms, and Market Power

James Albrecht, Xiaoming Cai, Pieter Gautier, Susan Vroman

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7805*

Multiple Applications, Competing Mechanisms, and Market Power

Abstract

We consider a labor market with search frictions in which workers make multiple applications and firms can post and commit to general mechanisms that may be conditioned both on the number of applications received and on the number of offers received by its candidate. When the contract space includes application fees, there exists a continuum of equilibria of which only one is socially efficient. In the inefficient equilibria, firms have market power that arises from the fact that the value of a worker's application portfolio depends on what other firms offer, which allows individual firms to free ride and offer workers less than their marginal contribution. Finally, by allowing for general mechanisms, we are able to examine the sources of inefficiency in the multiple applications literature.

JEL-Codes: C780, D440, D830.

Keywords: multiple applications, directed search, competing mechanisms, efficiency, market power.

James Albrecht
Department of Economics
Georgetown University
USA - Washington, DC 20057
albrecht@georgetown.edu

Xiaoming Cai
Department of Economics and Finance
Tongji University
China - Shanghai 200092
xiaoming@tongji.edu.cn

Pieter Gautier
Department of Economics
Vrije Universiteit Amsterdam
The Netherlands - 1081 HV, Amsterdam
p.a.gautier@vu.nl

Susan Vroman
Department of Economics
Georgetown University
USA - Washington D.C. 20057
susan.vroman@georgetown.edu

July 28, 2019
We thank Philipp Kircher for valuable comments.

1 Introduction

Under what conditions do labor markets function efficiently? When there are search frictions and firms cannot commit to a wage or wage mechanism, we know from Hosios (1990) and Pissarides (2000) that workers and firms typically do not receive their marginal contributions to the matching process. However, when firms can commit to a wage mechanism before they meet workers and workers can direct their search, the common wisdom is that the market equilibrium is efficient. See, for example, Moen (1997). Albrecht et al. (2006) and Galenianos and Kircher (2009) challenged this result and showed that when workers can simultaneously apply to multiple jobs, the resulting equilibrium is not efficient. However, each of these papers imposes a particular wage mechanism and a particular meeting technology. So their results do not necessarily imply that the decentralized equilibrium with multiple applications is inefficient when firms can compete with more general mechanisms or for other meeting technologies.

In this paper, as in Albrecht et al. (2006) and Galenianos and Kircher (2009), we assume that a firm can only consider one applicant. That is, if a firm's chosen candidate rejects its offer, we assume the firm is unable to consider a second applicant, etc.[1] However, we let firms choose from a set of more general mechanisms in which they can charge application fees as well as conditioning wages on the number of applications they receive and the number of offers that their candidate receives. We show that there exists a continuum of non-payoff-equivalent equilibria of which only one is socially efficient. Allowing firms to compete with general mechanisms gives a better understanding of the sources behind inefficiency, in particular, restrictions on the set of possible mechanisms versus monopsony power.

Given that the firm can condition the wage it pays on the number of offers that its chosen candidate has, we show that competition implies that the worker with multiple offers receives the full match surplus as in Albrecht et al. (2006). This means that potential equilibria are described by the wage posted for workers with no other offers and the application fee/subsidy.[2]

In all equilibria, we find that firms charge positive application fees. Although fees deter

[1] Albrecht et al. (2006) allow for "shortlists" or limited recall in the sense that a firm can offer its job to a second applicant (if it has one) in the event that its first offer is rejected, but the process ends if the second candidate also turns the firm down. Shortlists reduce but do not eliminate the inefficiency associated with multiple applications. Kircher (2009) and Gautier and Holzner (2017) are papers with multiple applications that allow full recall. A maximum matching can be achieved in this case only if workers and firms can learn all the details of the realized application network. Since we believe that firms are limited in the number of applicants they can consider and limited in what they can learn about the application network, the setting in the current paper is the relevant one.

[2] Allowing firms to condition wages on the number of applications received turns out to play no role in our analysis. As in a large-market version of Coles and Eeckhout (2003), equilibria that depend on the number of applicants are payoff equivalent. Coles and Eeckhout (2003) show that in a small market in which each buyer approaches only one seller, the equilibria are not payoff equivalent. Here, we show that even in a large market, equilibria that are characterized by different wage-fee pairs are not payoff equivalent.

applications, this effect is small because workers primarily care about generating multiple offers. When we allow both for fees and for wages conditioned on the number of offers, the wage for a selected applicant with no other offers can exceed the worker's reservation value. Relative to the case in Albrecht et al. (2006) with no application fee, firms do not have the incentive to reduce the wage all the way to the workers' reservation value because doing so would lead to fewer applicants and thus less income from application fees. Equilibria with a relatively high fee and a high posted wage (that the worker receives in case he or she has no other offers) can coexist with equilibria with a low fee and a low wage. From the continuum of equilibria, only one is efficient and it has a posted wage equal to match output together with an application fee equal to the firm's expected contribution to the match divided by the expected queue length. The efficient wage ensures that workers receive their expected contribution to match surplus.

In the labor market, we rarely observe application fees.[3] One reason for this is that workers may be reluctant to pay an application fee if they cannot observe whether the advertised job is real or bogus. In the market for higher education, where applicants know that colleges will admit students, we do observe fees. However, if for whatever reason, firms cannot charge application fees, the market equilibrium is never socially efficient.

When workers apply to only one job, efficiency arises because (i) firms must offer workers their market utility, (ii) in a large market, firms cannot affect an individual worker's market utility, and (iii) firms are residual claimants to the surplus so it is in their interest to offer efficient mechanisms.[4] When workers send out multiple applications, firms no longer can offer each applicant his or her market utility. Instead, firms offer workers the possibility to create optimal application portfolios. If workers can only apply to one job, the problem reduces to one of firms buying queues in a competitive market. With multiple applications, payoffs depend on the entire portfolio so a firm's expected payoff also depends on what happens with the applications that a worker sends to other firms. The same applies to the worker's payoff. This is not something the firm can directly price. Bertrand competition for workers with multiple offers is not enough for efficiency because (as in Postel-Vinay and Robin (2002)) firms can appropriate the rents from this *ex ante*. Firms have market power even though search is directed and there is Bertrand competition. Interestingly, market power only disappears in two extreme cases, namely (i) when all workers have multiple offers, so there is Bertrand competition for every worker's services with probability one, and (ii) when

[3]There are rare exceptions such as government agencies that may charge an application fee. These fees are often used to pay for exam costs or medical screening costs. In some states, fees are illegal, e.g., Hawaii. In the theoretical literature, Albrecht and Jovanovic (1986) is an early example of a paper with competitive search that allows for an application fee.

[4]See, for example, Albrecht et al. (2014), Cai et al. (2016), Shimer (2005) and Eeckhout and Kircher (2010a).

every worker applies to only one job, so there is no Bertrand competition but there is full *ex ante* competition for queues. The potential for market power only arises when the probability of Bertrand competition for a worker's services lies between zero and one, as in our model.

We carry out most of our analysis taking the number of worker applications as given. When we endogenize the number of worker applications, we find that in the equilibrium where firm entry is efficient, the number of applications per worker is also efficient. In the efficient equilibrium, after paying the application fee, workers become the residual claimants of the match surplus. The reason to apply to multiple jobs is to increase the probability of getting at least one offer and since the offered wage equals the match output there are no additional rewards for generating more than one offer. However, in equilibria in which entry is inefficient, workers with one offer receive less than the full surplus. This creates another reason for applying to multiple jobs, namely, to generate multiple offers in order to appropriate the full surplus. This creates congestion without increasing output and is thus a pure rent-seeking activity. Consequently, in the inefficient equilibria, the number of applications is excessive from a social point of view.

Finally, we relate our results to Albrecht et al. (2006) and Galenianos and Kircher (2009). Those papers allow for different subsets of the set of mechanisms that we consider in this paper. Albrecht et al. (2006) allow firms to condition their wage on the number of other offers that their candidate has, but they do not allow for fees. If firms are allowed to post fees in their setting, an equilibrium in which firms post a zero wage for workers with no other offers (as in Albrecht et al. (2006)) exists, but now firms would also charge a positive application fee. This would make the Albrecht et al. (2006) equilibrium even more inefficient. However, with fees, this is only one of a continuum of equilibria including an efficient one with a posted wage equal to the match surplus. Allowing for fees in the Galenianos and Kircher (2009) setting creates a unique pure-strategy equilibrium which is efficient, but there is also at least one mixed-strategy equilibrium which is inefficient.

The rest of this paper is organized as follows. Our main results are presented in Sections 2 - 4. In Section 2, we present our model; in Section 3, we characterize its equilibria; and in Section 4, we examine the (in)efficiency of these equilibria. Then in Section 5, we extend our model by endogenizing the number of worker applications. Finally, in Section 6, we consider the implications of restricting the set of possible mechanisms available to firms.

2 The Model

Our model is static and consists of four stages: (i) firms post a wage mechanism; (ii) workers send their applications; (iii) firms select at most one candidate and workers with one or multiple offers choose one offer; (iv) production takes place.

2.1 Setup

Agents. The economy consists of a measure v of identical risk neutral firms and a measure u of identical risk neutral workers. Initially, we take v as exogenous. We later endogenize v by allowing for free entry of vacancies. The productivity of a match is 1.

Multiple applications. Each firm posts and commits to a wage mechanism to attract workers. After observing all wage mechanisms, each worker sends a job applications, and we define $\lambda = au/v$, to be the average number of applications per vacancy. We assume for now that a is exogenous. Later, in Section 5, we endogenize a by assuming application costs. Since we consider a large market, we assume that workers cannot coordinate their application strategies, and we look for equilibria in which workers use symmetric strategies. This is a standard assumption in the literature (see e.g., Moen, 1997; Burdett et al., 2001; Shimer, 2005).

No recall. Firms simultaneously select one applicant (if they have any) and offer the wage specified in the mechanism. The worker can then accept or reject the offer and after that the game ends. This means we do not allow for recall. That is, if the firm's chosen applicant does not take the offer, we do not allow the firm to select another applicant.

Wage mechanism. In the most general case we consider, we suppose that workers can credibly show the firm the other offers that they have received and that firms can credibly show the worker their other applications.[5] In this case, a wage mechanism, (f, w_{ij}), is a fee, f, and a set of wages that may depend on the number of offers the worker receives, i, and the number of applications the firm receives, j, where $i = 1, 2, \ldots, a$ and $j = 1, \ldots, \infty$. f is the application fee to be paid by workers and $f < 0$ implies an application subsidy. w_{ij} is the wage when the worker has i offers in total (including the one from the current firm) and the firm has j applications (including the one from the current worker). When a worker has multiple offers, we allow for Bertrand competition. Given that firms do not observe the entire application network (which is a bipartite graph consisting of a set of workers and a set of firms linked together by a set of edges which are the applications), they cannot condition the wage on the number of applications at the other firms where their candidate applied. Finally, we rule out the case in which a firm's wage mechanism depends on the wage mechanisms at the other firms to which their candidate has applied. That is, we rule out wage mechanisms

[5] Although we allow for the possibility here, we show below that firms' optimal wage mechanisms do not depend on the number of applications they have.

such as offer-beating strategies.[6] In Section 6, we consider various restrictions on the set of possible wage mechanisms, e.g., ruling out fees.

Payoffs. Consider a worker who applies to a jobs, paying a fee of f per application. The worker's payoff if he or she takes a job at wage w is $w - af$; otherwise, the worker's payoff is simply $-af$. Similarly, consider a firm that expects to receive λ applications. If the firm hires its selected applicant at wage w, its expected payoff is $1 - w + \lambda f$; otherwise, the firm's expected payoff is simply λf.

Meeting technologies. Matching is frictional. Consider a firm with expected queue length λ; i.e., λ is the expected number of workers applying to the firm. Let $p_n(\lambda)$ be the probability that the firm receives n applications. We assume that the expected number of applications equals the expected queue length. That is, $\sum_{n=1}^{\infty} np_n(\lambda) = \lambda$ for any $\lambda \geq 0$. Eeckhout and Kircher (2010b) called this property *nonrival.* We are especially interested in $m(\lambda) \equiv 1 - p_0(\lambda)$, the probability that the firm receives at least one application.[7] We assume that $m(\lambda)$ is increasing and strictly concave. Furthermore, we assume that $\lim_{\lambda \to 0} m'(\lambda) = 1$ and $\lim_{\lambda \to \infty} m(\lambda) - \lambda m'(\lambda) = 1$, which are standard assumptions in the literature and ensure that the marginal contribution to surplus of firms (respectively, workers) is one when the measure of firms (respectively, workers) is zero. By an accounting identity, the probability that a worker who applies to a job receives an offer from that firm is $m(\lambda)/\lambda$. Following Moen (1997) and Eeckhout and Kircher (2010a), we do not assume a particular functional form for the meeting technology $m(\lambda)$. Special cases include the urn ball, $m(\lambda) = 1 - e^{-\lambda}$, and the geometric, $m(\lambda) = \lambda/(1 + \lambda)$, both of which are extensively used in the literature. Keeping $m(\lambda)$ general allows us to see how properties of the equilibrium wage mechanism depend on the meeting technology.

For future use, we define a new function $h(\lambda)$, which is the probability that an application fails to lead to an offer:

$$h(\lambda) = 1 - \frac{m(\lambda)}{\lambda}. \tag{1}$$

Since $m(\lambda)$ is increasing and strictly concave, $h(\lambda)$ is strictly increasing in λ with a limit as $\lambda \to \infty$ equal to 1. That is, applications in a longer queue are more likely to fail. Note also

[6] Albrecht et al. (2006) consider offer-beating strategies in a model of competitive search and multiple applications. In their setting, offer-beating strategies reduce but do not eliminate the inefficiencies associated with multiple applications.

[7] Assuming nonrivalry does not restrict the functional form of $m(\lambda)$. For any $m(\lambda)$, one can construct a nonrival meeting technology as follows. The meeting technology has two stages. In the first stage, workers and firms meet according to a bilateral meeting technology $p_0(\lambda) = 1 - m(\lambda)$ and $p_1(\lambda) = m(\lambda)$. The applications that failed to reach a firm in the first stage arrive in the second stage, according to an urn-ball meeting (or any non-rival meeting technology), at one of the firms that received applicants in the first stage.

that

$$\frac{\lambda h'(\lambda)}{h(\lambda)} = \frac{m(\lambda)}{\lambda - m(\lambda)} \left(1 - \frac{\lambda m'(\lambda)}{m(\lambda)}\right), \tag{2}$$

which implies that when $\lambda \to \infty$, $\frac{\lambda h'(\lambda)}{h(\lambda)} \to 0$.

2.2 Payoffs and equilibrium

Before we move to the detailed analysis of worker and firm payoffs, we first present the total number of matches, or equivalently total surplus. Since in a symmetric equilibrium, workers send their a applications independently, the total number of matches is given by

$$M(u, v, a) = u\left(1 - h(\lambda)^a\right). \tag{3}$$

A worker fails to match with a firm if and only if all of his or her applications fail. Thus the term in parentheses denotes the probability that a worker receives at least one offer out of a applications.

Now we turn to the firm's payoff function. Recall that $p_j(\lambda)$ is the probability that a firm receives j applications. The firm's expected payoff is

$$\pi = \sum_{j=1}^{\infty} p_j(\lambda) \left(j \cdot f + (1 - w_{1j}) h(\lambda)^{a-1} + \sum_{i=2}^{a} (1 - w_{ij}) \cdot \mathcal{P}_{i,j} \right). \tag{4}$$

The first term in the large parentheses on the right-hand side is the fees that the firm receives if j workers apply. The second term describes the situation in which the firm receives at least one application and its randomly chosen applicant has no other job offers. Finally, $\mathcal{P}_{ij}$ in the last term denotes the probability that the chosen applicant has in total $i > 1$ offers and the firm's offer is chosen by the worker conditional on the firm receiving j applications.

This expression for the firm's expected payoff can be substantially simplified. First, the expected application fees that the firm receives can be written as $\sum_{j=1}^{\infty} p_j(\lambda) j \cdot f = \lambda f$, because we assume that the meeting technology is nonrival; i.e., the expected number of applications that a firm receives equals the expected queue length. Second, $\sum_{j=1}^{\infty} p_j(\lambda)(1 - w_{1j}) h(\lambda)^{a-1}$, the second term on the right-hand side of equation (4), is the expected wage that the firm pays in the event that its chosen applicant has no other offers. There is a wage w_1 such that $\sum_{j=1}^{\infty} p_j(\lambda)(1 - w_{1j}) h(\lambda)^{a-1} = \sum_{j=1}^{\infty} p_j(\lambda)(1 - w_1) h(\lambda)^{a-1}$. Then, we have $\sum_{j=1}^{\infty} p_j(\lambda)(1 - w_{1j}) h(\lambda)^{a-1} = m(\lambda)(1 - w_1) h(\lambda)^{a-1}$. Finally, Bertrand competition implies $w_{ij} = 1$ for $i > 1$; i.e., the third term in equation (4) equals zero. In short, the firm's expected payoff can be written as

$$\pi = m(\lambda)(1 - w_1)h(\lambda)^{a-1} + \lambda f. \tag{5}$$

Next, the expected payoff for a worker can be written as

$$U = \big(1 - h(\lambda)^a - a(1 - h(\lambda))h(\lambda)^{a-1}\big) + a(1 - h(\lambda))h(\lambda)^{a-1}w_1 - af, \tag{6}$$

where the first term in parentheses denotes the probability that the worker receives at least two offers, the second term denotes the case in which the worker receives exactly one offer, and the final term is the total application fees the worker pays at the various firms.

Finally, by an accounting identity the following must hold:

$$uU + v\pi = M(u, v, a). \tag{7}$$

This equation simply tells us that total surplus is split between workers and firms.

Now consider the problem of a potential deviant firm. Bertrand competition still dictates that the deviant firm sets $\widetilde{w}_{ij} = 1$ for $i \geq 2$ and for all j. The deviant firm then optimizes over $\widetilde{f}$ and $\widetilde{w}_1$, its application fee and the wage it offers its chosen applicant when that worker has no other offers. The expected payoff of a deviant firm is

$$\widetilde{\pi} = m(\widetilde{\lambda})h(\lambda)^{a-1}(1 - \widetilde{w}_1) + \widetilde{\lambda}\widetilde{f}, \tag{8}$$

where $\widetilde{\lambda}$ is the expected queue length faced by the deviant firm. Note that $m(\widetilde{\lambda})$ is the probability that the deviant firm receives at least one application, in which case the firm randomly picks one applicant, and $h(\lambda)^{a-1}$ is the probability that this worker's other $a - 1$ applications fail to generate an offer.

Workers must be indifferent between sending all a applications to non-deviant firms versus sending one application to the deviant firm and the other $a - 1$ applications to non-deviant firms. The corresponding indifference condition for workers is

$$\begin{aligned} U = \Big(1 - h(\widetilde{\lambda})h(\lambda)^{a-1} - h(\lambda)^{a-1}(1 - h(\widetilde{\lambda})) - (a-1)h(\widetilde{\lambda})h(\lambda)^{a-2}(1 - h(\lambda))\Big) \\ + (a-1)h(\widetilde{\lambda})h(\lambda)^{a-2}(1 - h(\lambda))w_1 + h(\lambda)^{a-1}(1 - h(\widetilde{\lambda}))\widetilde{w}_1 - \widetilde{f} - (a-1)f, \end{aligned} \tag{9}$$

where U is the market utility of workers on the equilibrium path which is given by equation (6). The right-hand side of equation (9) gives the expected payoff for a worker who sends one application to the deviant firm and the rest to non-deviant firms. The first term gives the probability that the worker receives at least two offers in which case he or she receives a wage of 1. The first term on the second line is the payoff in case the worker receives exactly one offer from one of the non-deviant firms and receives no offer from the deviant firm, while the second term on the second line describes the case in which the worker has one offer from the deviant firm but has no offers from the other $a-1$ applications. Finally, the worker pays $\widetilde{f}+(a-1)f$ in application fees.

The right-hand side of equation (9) is *linear* in $h(\widetilde{\lambda})$ with a negative coefficient, so for given $\widetilde{f}$ and $\widetilde{w}_1$, the expected payoff obtained from sending one application to the deviant firm is strictly decreasing in $\widetilde{\lambda}$. Hence, for a given $\widetilde{f}$ and $\widetilde{w}_1$, there is a unique $\widetilde{\lambda}$ that solves equation (9).

In sum, a deviant firm's problem is to select the pair $\widetilde{f}$ and $\widetilde{w}_1$ that maximizes its expected profit $\widetilde{\pi}$ in equation (8) where $\widetilde{\lambda}$ is implicitly determined by equation (9).

We can now define an equilibrium as follows.

Definition 1. *An equilibrium mechanism,* (f, w_1)*, maximizes a potential deviant's payoff (equation* (8)*) subject to the constraint that the deviant must allow workers to form optimal application portfolios, i.e., under the constraint that equation* (9) *holds and given that all other firms do the same.*

3 Characterization of the equilibrium

We start the analysis of the equilibrium by solving the deviant firm's problem. First, we rewrite equation (8), the expected profit of the deviant firm, in the following way.

$$\widetilde{\pi} = m(\widetilde{\lambda})h(\lambda)^{a-1} - \widetilde{\lambda}\left(h(\lambda)^{a-1}(1-h(\widetilde{\lambda}))\widetilde{w}_1 - \widetilde{f}\right)$$

The first term on the right-hand side is the expected surplus, which equals the probability that the deviant firm forms a match with a worker who has no other offers. The term in parentheses (the per-unit transfer) is the probability that a worker who applies to the deviant firm participates in such a surplus-generating match (i.e., only has an offer from the deviant) times his or her payoff for the match minus the application fee. Note that for a given expected queue length $\widetilde{\lambda}$, the deviant is indifferent between combinations of high $\widetilde{f}$ and $\widetilde{w}_1$ and low $\widetilde{f}$ and $\widetilde{w}_1$ as long as the per-unit transfer is constant. When each worker can only send one application, the per-unit transfer depends only on the mechanism posted by the deviant firm. However, with multiple applications, the per-unit transfer takes a more complicated form. A worker's market utility now depends on his or her entire portfolio of applications.

We use equation (9) to solve for the per-unit transfer in terms of $\widetilde{\lambda}$ and (λ, f, and w_1). Note that the per-unit transfer is nonlinear in $\widetilde{\lambda}$. We then substitute the per-unit transfer derived from equation (9) into equation (8) giving,

$$\widetilde{\pi} = m(\widetilde{\lambda}) \left[h(\lambda)^{a-1} + (a-1)h(\lambda)^{a-2}(1-h(\lambda))(1-w_1)\right] \\ - \widetilde{\lambda}\Bigg[h(\lambda)^{a-1}(1-h(\lambda))(1-a(1-w_1)) - f + (a-1)h(\lambda)^{a-2}(1-h(\lambda))(1-w_1)\Bigg]. \quad (10)$$

We have thus shown that the constrained optimization problem of choosing $\widetilde{f}, \widetilde{w}_1$, and $\widetilde{\lambda}$ is isomorphic to the unconstrained one of choosing the optimal application queue, $\widetilde{\lambda}$, which greatly simplifies the deviant firm's problem. We can then solve a univariate optimization problem and back out all the corresponding $\widetilde{f}$ and $\widetilde{w}_1$ pairs.

The deviant firm's expected profit is strictly concave in $\widetilde{\lambda}$ since $m(\widetilde{\lambda})$ is strictly concave in $\widetilde{\lambda}$. This implies a unique solution for the deviant firm's optimization problem, so the first-order condition is both necessary and sufficient. Equilibrium requires that the first-order condition must satisfy $\frac{d\widetilde{\pi}}{d\widetilde{\lambda}}|_{\widetilde{\lambda}=\lambda} = 0$, which can be rewritten as[8]

$$f = f^e(w_1) \equiv h(\lambda)^{a-1}(1-h(\lambda))\left(w_1 - \frac{\lambda m'(\lambda)}{m(\lambda)} + (1-w_1)(a-1)\frac{\lambda h'(\lambda)}{h(\lambda)}\right) \quad (11)$$

In the above equation, $f^e(w_1)$ is *linear* in w_1. Note that $f^e(w_1)$ is increasing in w_1 if and only if the elasticity $\lambda h'(\lambda)/h(\lambda) < 1/(a-1)$. One can show that for common meeting technologies such as the urnball, $m(\lambda) = 1 - e^{-\lambda}$, and the geometric, $m(\lambda) = \frac{\lambda}{1+\lambda}$, the elasticity of $h(\lambda)$, i.e., $\lambda h'(\lambda)/h(\lambda)$, is strictly decreasing in λ with the supremum equal to 1 and the infimum equal to 0. Therefore, when $a = 2$, for common meeting technologies $f^e(w_1)$ is always increasing in w_1. However, when $a \geq 3$, $\lambda h'(\lambda)/h(\lambda) > 1/(a-1)$ for small λ and $\lambda h'(\lambda)/h(\lambda) < 1/(a-1)$ for large λ, which implies that $f^e(w_1)$ is increasing in w_1 for large λ, and for small λ it is decreasing in w_1. The above observations are illustrated in Figure 1, where the meeting technology is given by $m(\lambda) = 1 - e^{-\lambda}$. The left panel plots the case of $a = 2$. The three solid lines plot $f^e(w_1)$ as a function of w_1 for $\lambda = 1, 2$ and 3. We can see that all three lines are increasing in w_1. However, this is no longer the case for $a = 3$ as shown in the right panel. When $a = 3$, $f^e(w_1)$ is increasing in w_1 when $\lambda = 3$, and it is decreasing in w_1 when $\lambda = 1$ and 2 (the dashed lines in the figure refer to the planner's solution and are discussed later).

We have shown that if all firms except a potential deviant choose f and w_1 which satisfy equation (11), then it is also in a potential deviant's interest to do so. Hence, it is an equilibrium strategy. In fact, an individual firm could choose any fee and wage that gives the same per-unit transfer as the equilibrium (f, w_1), i.e., an individual firm has a continuum of

[8]We have chosen to express f in equation (11) using both $m(\lambda)$ and $h(\lambda)$ because it makes the comparison with the planner's solution easier. This comment also applies to the expected payoff of firms in equation (12). Equation (2) gives the relationship between the derivatives of $m(\lambda)$ and $h(\lambda)$.

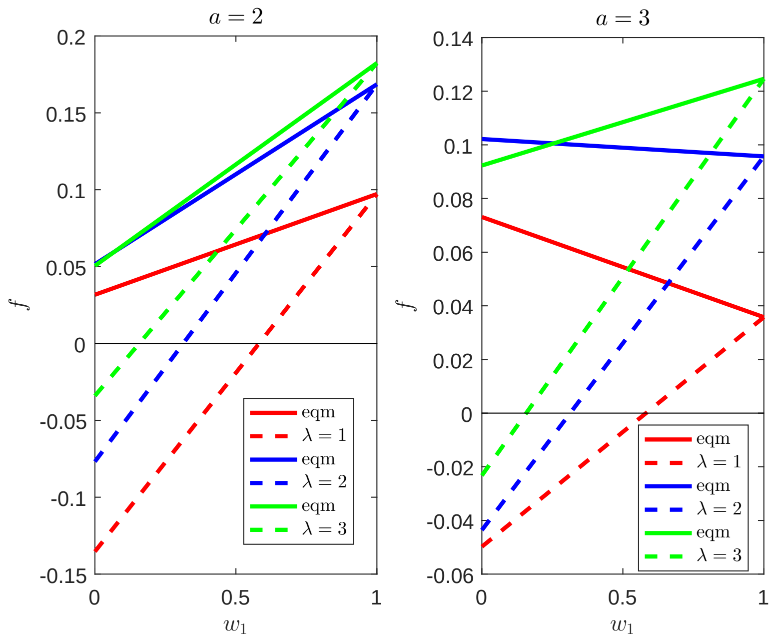

Figure 1: Equilibrium and efficient combinations of (w_1, f) with $m(\lambda) = 1 - e^{-\lambda}$ and $\lambda = 1, 2, 3$. Equilibrium loci are solid and efficient loci are dashed. $f < 0$ implies an application subsidy.

best responses.

If we focus on *symmetric* equilibria in which all firms choose the same f and w_1, then there is a continuum of equilibria in a more interesting sense. For each $w_1 \in [0, 1]$, there exists a corresponding f that satisfies equation (11), which implies a continuum of equilibria indexed by w_1. These equilibria are not payoff equivalent as we now show.

Substituting equation (11) into equation (5) gives the expected equilibrium payoff of firms,

$$\pi^e(w_1, \lambda, a) = (m(\lambda) - \lambda m'(\lambda))\, h(\lambda)^{a-1} \left(1 + (1 - w_1)(a-1)\frac{1 - h(\lambda)}{h(\lambda)}\right) \tag{12}$$

From the above equation we can see that firm value is linear in w_1 and that it is *strictly decreasing* in w_1. Thus different equilibria are not payoff equivalent. Firm value is higher in an equilibrium with a lower w_1. Since for fixed λ, total surplus is constant across equilibria, the opposite must hold for the worker value by equation (7). As mentioned above, when workers apply to more than one firm, a firm can provide value to a worker without eventually hiring the worker. This implies that a firm can free ride on other firms, which may result in a race to the bottom. When a firm offers a high w_1, workers obtain most of the value from that firm and the firm cannot rely on other firms to provide worker value. In this sense, w_1 measures the degree of transfers from firms to workers (after taking into account that in equilibrium, f depends on w_1 through equation (11)). What is interesting and different from standard race-to-the-bottom environments such as pollution or tax avoidance games, is that we have multiple equilibria with different degrees of market power.

Next we give an explicit expression for the market utility of workers by substituting equation (11) into equation (6). We then have

$$U^e(w_1, \lambda, a) = 1 - h(\lambda)^a - a\lambda h'(\lambda)h(\lambda)^{a-1} - a(a-1)\lambda h'(\lambda)(1 - h(\lambda))h(\lambda)^{a-2}(1 - w_1). \tag{13}$$

As we mentioned before, the above equation is linear and increasing in w_1. Note that thus far, we have ignored one important constraint: $U^e(w_1, \lambda, a) \geq 0$. Later in Section 4 we show that in the equilibrium where $w_1 = 1$, the expected worker payoff is equal to the worker's marginal contribution to surplus, which, by Lemma 2, is strictly positive.[9] However, in the equilibrium with $w_1 = 0$, $U^e(0, \lambda, a)$ is not necessarily positive. To see this, set $a = 2$ and $w_1 = 0$. Then the above equation becomes $U^e(0, \lambda, 2) = 1 - h(\lambda)^2 - 2h(\lambda)h'(\lambda)$, the sign of which depends both on the meeting technology and the value of λ. For example, when $m(\lambda) = \lambda/(1 + \lambda)$, it is positive for any λ. When $m(\lambda) = 1 - e^{-\lambda}$, it is positive for any $\lambda < 2.983$ and negative for any $\lambda > 2.983$.

We have thus shown the following:

Proposition 1. *There exists a continuum of equilibria indexed by $w_1 \in [\underline{w}, 1]$ where f is given by equation (11). If $U^e(0, \lambda, a) > 0$, then $\underline{w} = 0$; if $U^e(0, \lambda, a) \leq 0$, then $\underline{w}$ is uniquely defined by $U^e(\underline{w}, \lambda, a) = 0$, in which case, the expected worker payoff is zero in the equilibrium in which $w_1 = \underline{w}$. These equilibria are not payoff equivalent; the expected payoff of firms is decreasing in w_1 while the expected payoff of workers is increasing in w_1.*

We now contrast our result with a competitive search model in which workers can only send one application. Interestingly, the first part of Proposition 1 continues to hold: there is a continuum of equilibria indexed by $w_1 \in [0, 1]$ where f is given by equation (11). When $a = 1$, equation (11) becomes $f^e(w_1) = \frac{m(\lambda)}{\lambda} w_1 - m'(\lambda)$. For small w_1, $f^e(w_1) < 0$ (firms post an application subsidy), and for large w_1, $f^e(w_1) > 0$ (firms charge an application fee). Only at $w_1 = \lambda m'(\lambda)/m(\lambda)$ (the Hosios rule) does $f^e(w_1) = 0$. However, the second part of Proposition 1 fails when workers can only send one application. When $a = 1$, equation (12) continues to hold. It becomes $\pi^e(w_1, \lambda, 1) = m(\lambda) - \lambda m'(\lambda)$. That is, firm value is a constant and independent of w_1 so all equilibria are payoff equivalent.

[9] We can also prove this directly. Note that by equation (2), $\lambda h'(\lambda) < h(\lambda)\frac{m(\lambda)}{\lambda - m(\lambda)} = 1 - h(\lambda)$, and by equation (13), $U^e(1, \lambda, a) = 1 - h(\lambda)^a - a\lambda h'(\lambda)h(\lambda)^{a-1}$. Thus $U^e(1, \lambda, a) > 1 - h(\lambda)^a - a(1 - h(\lambda))h(\lambda)^{a-1} \geq 0$, where the final inequality follows from a binomial expansion.

No application subsidy in equilibrium. We have shown that there exists a continuum of equilibria characterized by $(f^e(w_1), w_1)$. It is then natural to ask when firms charge an application fee ($f^e(w_1) > 0$) and when firms post an application subsidy ($f^e(w_1) < 0$). It turns out that under a mild restriction on the meeting technology, firms never post an application subsidy in equilibrium. The necessary restriction on the meeting technology is the following.

Condition 1. *For any* $\lambda > 0$*,*

$$\frac{\lambda h'(\lambda)}{h(\lambda)} \geq 1 - h(\lambda). \tag{14}$$

To understand this condition, the expression on the right-hand side is the probability that a particular application leads to an offer. The expression on the left-hand side is the elasticity of the probability that the worker is not selected with respect to the expected queue length. Condition 1 implies that when the probability that an application leads to an offer is high, increasing the expected queue length has a larger impact on decreasing the probability of receiving an offer relative to the case in which this probability is low. The above condition holds for common meeting technologies such as the urn-ball and geometric and holds with equality for the geometric for any λ.

For later use, the following lemma provides an alternative formulation of Condition 1, namely that $h(\lambda)/m(\lambda)$ is increasing in λ, or equivalently that the elasticity of $h(\lambda)$ is larger than the elasticity of $m(\lambda)$.

Lemma 1. *For any* $\lambda > 0$*, Condition 1 is equivalent to*

$$\frac{\lambda h'(\lambda)}{h(\lambda)} \geq \frac{\lambda m'(\lambda)}{m(\lambda)}, \tag{15}$$

or that $h(\lambda)/m(\lambda)$ *is increasing in* λ*.*

Proof. See Appendix A.1. □

It is also useful to express Condition 1 in terms of $m(\lambda)$ alone. In this case the condition is equivalent to $m'(\lambda) \leq (m(\lambda)/\lambda)^2$.

We now return to the question of whether firms would ever post an application subsidy to attract workers in equilibrium when $a \geq 2$. Recall that $f^e(w_1)$ in equation (11) is linear in w_1 and note that

$$f^e(0) = h(\lambda)^{a-1}\left(1-h(\lambda)\right)\left(-\frac{\lambda m'(\lambda)}{m(\lambda)} + (a-1)\frac{\lambda h'(\lambda)}{h(\lambda)}\right), \tag{16}$$

$$f^e(1) = h(\lambda)^{a-1}\left(1-h(\lambda)\right)\left(1-\frac{\lambda m'(\lambda)}{m(\lambda)}\right) = \lambda h(\lambda)^{a-1}h'(\lambda). \tag{17}$$

Since $m(\lambda)$ is strictly concave, $f^e(w_1)$ is strictly positive in the equilibrium with $w_1 = 1$. In the equilibrium with $w_1 = 0$, things are slightly more complicated. The last term on the right-hand side of equation (16) is minimized at $a = 2$ and then increases in a. Thus a sufficient condition for $f^e(w_1)$ to be greater than or equal to 0 when $w_1 = 0$ is $\lambda h'(\lambda)/h(\lambda) \geq \lambda m'(\lambda)/m(\lambda)$, which, by Lemma 1, is simply Condition 1, a mild restriction on the meeting technology. Thus when Condition 1 holds, firms never post an application subsidy in equilibrium and we have the following result.

Proposition 2. *Under Condition 1, in all equilibria with $w_1 \geq 0$, we have $f^e(w_1) \geq 0$.*

Proof. As noted above, at $w_1 = 1$ we have $f^e(1) > 0$, and at $w_1 = 0$, we have $f^e(0) \geq 0$. Since f is linear in w_1 by equation (11), $f^e(w_1)$ is strictly positive for strictly positive w_1. □

Note that when Condition 1 holds, $f^e(w_1) = 0$ requires that $w_1 = 0$, that $a = 2$, and that Condition 1 holds with equality.

4 Efficiency

In the previous section, we analyzed the decentralized equilibrium taking λ as given. Now we allow for free entry of vacancies and assume that it costs $c_v < 1$ for a firm to set up a vacancy and enter the market. Then a natural question is whether the equilibrium number of vacancies is efficient.

The planner's problem is to choose the number of vacancies v to maximize net social surplus, which is given by

$$S(u, v, a) = M(u, v, a) - c_v v, \tag{18}$$

where $M(u, v, a)$ is the total number of matches, which is given by equation (3). The first-order condition for this problem is $c_v = \partial M(u, v, a)/\partial v$. In Lemma 2 below, we show that $M(u, v, a)$ is concave in v so the first-order condition is also sufficient. That is, the planner sets the contribution to aggregate output of the marginal vacancy equal to the entry cost c_v.

By direct calculation, this contribution is

$$\frac{\partial M(u,v,a)}{\partial v} = (m(\lambda) - \lambda m'(\lambda))\, h(\lambda)^{a-1}. \tag{19}$$

When $a = 1$, $m(\lambda) - \lambda m'(\lambda)$ is the contribution of the marginal vacancy. With multiple applications, a firm contributes to surplus only if the applicant that the firm chooses has no other offers, which happens with probability $h(\lambda)^{a-1}$. For the urn-ball technology, $(m(\lambda) - \lambda m'(\lambda))$ has a simple interpretation. It is the probability that a firm receives at least two applications. If the firm's selected applicant receives no other offers, then the firm's contribution to aggregate output is one only if the firm receives at least two applications. In this case, the worker's contribution is zero.[10]

Similarly, by direct computation, the marginal contribution of a worker is

$$\frac{\partial M(u,v,a)}{\partial u} = am'(\lambda)h(\lambda)^{a-1} + \left(1 - h(\lambda)^a - ah(\lambda)^{a-1}(1 - h(\lambda))\right). \tag{20}$$

To understand the above equation, we first consider the case of $a = 1$. In this case, $m'(\lambda)$ is the worker's marginal contribution. For the urn-ball meeting technology, this term has a simple interpretation: it is the probability that a worker is the firm's only applicant. With multiple applications, we need to multiply $m'(\lambda)$ by the probability that the worker's other $a - 1$ applications fail, which gives the first term on the right-hand side. The second term is the probability that the worker receives at least two offers, in which case the worker's contribution to surplus is 1, because without the worker, there would be one fewer match. Alternatively, if we remove one of the firms that made an offer to the worker, the number of matches would remain unchanged.

Before turning to the decentralized equilibrium, we give the following unsurprising but necessary result.

Lemma 2. *$M(u, v, a)$, which was defined in equation* (3)*, is strictly concave in v and in u.*

Proof. See Appendix A.2. □

In the decentralized equilibrium, firms enter until the expected payoff is equal to the entry cost. That is, $c_v = \pi(w_1, \lambda, a)$ where the expected firm payoff $\pi(w_1, \lambda, a)$ is given by equation (12). What complicates the problem is that there is a continuum of equilibria, which are not payoff equivalent. Comparing equations (12) and (19) shows that in the equilibrium with $w_1 = 1$, the expected firm value is equal to the firm's marginal contribution to net

[10]More generally, if the meeting technology is invariant as defined by Lester et al. (2015), i.e., if the probability that a particular worker or a set of workers applies to a particular firm is not affected by the application choices of the other workers, the same logic applies.

surplus for all a. Hence, the equilibrium with $w_1 = 1$ is socially efficient. We can prove that all other equilibria are inefficient and lead to excessive firm entry. This is summarized in the following proposition.

Proposition 3. *For each $c_v \in (0, 1)$, there exists a continuum of symmetric equilibria indexed by $w_1 \in [\underline{w}, 1]$ where $\underline{w} < 1$ and f is given by equation (11). The equilibrium with $w_1 = 1$ is efficient. All other equilibria lead to excessive firm entry. The equilibria are Pareto rankable. An equilibrium with higher w_1 has higher net output.*

Proof. See Appendix A.3. □

Note that under free entry, the expected payoff of workers is net output per capita because gross output is always split between workers and firms. Thus again, workers prefer an equilibrium with higher w_1. There is also a constraint that the expected equilibrium payoff of a worker can never be negative. If this constraint is binding, then at the $w_1 = \underline{w}$ equilibrium, the expected payoff of workers is exactly zero.

Understanding the efficiency result. When a worker sends out multiple applications, he or she reduces the probability that the application of another worker is selected. This externality typically leads to excessive entry of vacancies except in the special case where $w_1 = 1$ and firms charge an appropriate fee. To see this, when $w_1 = 1$ (together with $w_{ij} = 1$ when $i \geq 2$), the incentive for a firm to free ride on other firms disappears, and by choosing an appropriate application fee, the firm gets its marginal contribution to net surplus. Competition among firms implies that in equilibrium, the application fee is efficient. Alternatively, we can understand this efficiency result when $w_1 = 1$ by analogy with competitive markets. When $w_1 = 1$, the worker's value is $1 - h(\lambda)^a - af$ and the firm's value is λf. Define a new variable $\overline{f} \equiv \lambda f$. The expected payoff of the worker is then $1 - h(\lambda)^a - \frac{a}{\lambda}\overline{f}$. We can interpret $\overline{f}$ as the market price of a vacancy, $1 - h(\lambda)^a$ as the net surplus, and $a/\lambda = v/u$ as the worker demand for vacancies. So, firms supply chances to match in a competitive market at price $\overline{f}$, which in free-entry equilibrium equals c_v.

Excessive entry. To understand the result on excessive vacancy creation when $w_1 < 1$, note that the socially optimal fee is found by setting the expected payoff of a firm equal to its marginal contribution. That is,

$$f^p(w_1) = h(\lambda)^{a-1} (1 - h(\lambda)) \left(w_1 - \frac{\lambda m'(\lambda)}{m(\lambda)} \right) \tag{21}$$

which is derived by setting the expected payoff of a firm given by equation (5) equal to $\partial M(u, v)/\partial v$ given by equation (19).

Comparing the above equation with equation (11), which gives the equilibrium $f^e(w_1)$, shows that

$$f^p(w_1) < f^e(w_1) \quad \text{when} \quad w_1 < 1 \tag{22}$$

Thus in the decentralized market, when $w_1 < 1$, firms always charge a higher application fee than the efficient one. The equilibrium payoff of a firm is strictly higher than its marginal social contribution, which then leads to excessive entry of vacancies. The above inequality $f^p(w_1) < f^e(w_1)$ is illustrated in Figure 1. The left-hand panel plots $f^p(w_1)$ as a function of w_1 for the case in which $a = 2$. We can see that the dashed line is always below the solid line except at $w_1 = 1$. Similarly, the right-hand panel plots the case of $a = 3$, in which the same result holds. Note that unlike the equilibrium $f^e(w_1)$, from equation (21) we can see that $f^p(w_1)$ is always increasing in w_1.

Comparing different equilibria. We have shown that for given λ, the expected payoff of firms is lower in an equilibrium with a higher w_1 (Proposition 1). Thus it is natural to conjecture that with free entry, an equilibrium with a higher w_1 has fewer firms but higher net output. Proposition 3 confirms that this is indeed the case.

Because $M(u, v, a)$ is concave in v, the net output $S(u, v, a)$ defined in equation (18) is also concave in v. We already know that the number of vacancie is efficient in the equilibrium with $w_1 = 1$. To show that an equilibrium with a lower w_1 has lower net output, we only need to show that it has a higher equilibrium v, or equivalently a lower equilibrium λ. In the proof of Proposition 3, we show that λ is indeed increasing in w_1. Hence an equilibrium with a lower w_1 has lower net output.

Why fees are necessary for efficiency. The discussion following equation (19) suggests an efficient wage mechanism: A worker who has only one offer should receive $w_1 = \lambda m'(\lambda)/m(\lambda)$ (the Hosios rule). A worker with multiple offers should receive a wage of 1. This would implement the social planner's solution with no application fee. However, as we have shown above in Propositions 1 and 2, this is not an equilibrium wage mechanism because when $f = 0$, at any positive candidate-equilibrium wage w_1, there exists a profitable downward deviation. This is in contrast to the case in which $a = 1$. In that case, since workers cannot pursue multiple offers, a reduction in the wage below the competitive search equilibrium level would be unprofitable.

5 Endogenous Number of Applications

We have shown that in a competitive search model in which workers can send multiple applications and firms can post general wage mechanisms, there exists a continuum of equilibria only one of which, the one with $w_1 = 1$ and f given by equation (17), leads to the efficient outcome. In this section we investigate whether this efficiency result continues to hold when applications are costly and workers can choose how many applications to send. The externalities caused by multiple applications are more complicated than those generated by firm entry. Firm entry always leads to a higher matching rate for workers and a lower matching rate for firms. Multiple applications always crowd out other workers' applications imposing negative externalities on other workers. The effect of multiple applications on the firm side is ambiguous. On the one hand, the probability of firms receiving at least one application increases, but, on the other hand, the probability that multiple firms compete for the same worker increases. The net effect on the aggregate matching rate can be negative.[11]

To analyze the optimal number of applications, we assume that each application costs workers c_a. This cost may represent workers' time costs or the disutility of filling out application forms. Note that the application cost c_a differs from the application fee f that firms charge. The latter is a transfer from workers to firms. Finally, to simplify the analysis, we assume that the number of firms is fixed, because we have shown above that adding firm entry does not change the efficiency result.

5.1 The social planner's problem

The social planner's problem is to choose the number of applications such that net surplus, which is given by $M(u, v, a) - uac_a$, is maximized. To simplify the analysis, we ignore the constraint that a must be an integer and treat it as a continuous variable. We then have[12]

$$\frac{d}{da}\left(\frac{M(u,v,a)}{u}\right) = h(\lambda)^a \left(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}\right). \tag{23}$$

The planner's first-order condition requires that the above expression equals c_a.

With a higher a, workers create more congestion for other workers by increasing λ, which reduces the probability of being matched per application (congestion effect). However, more applications by an individual worker increase that worker's matching rate (sampling effect).

[11] Albrecht et al. (2006) showed this formally, see also equation (23).

[12] Note that we need to use the following rule of differentiation.

$$\frac{d}{dx}\left(f(x)^{g(x)}\right) = \frac{d}{dx}\left(e^{(\log f(x))\cdot g(x)}\right) = g(x)f(x)^{g(x)-1}f'(x) + f(x)^{g(x)}\log f(x)\cdot g'(x).$$

These two effects have opposite signs. We can see this from equation (23): the effect of a on the number of matches depends on the difference between two positive terms. When λ is small, the congestion effect is small so a higher a always leads to a larger number of matches.[13] When λ gets larger, the congestion effect also becomes larger so that the net effect of a becomes ambiguous and depends on the specific meeting technology. When $m(\lambda) = 1 - e^{-\lambda}$, the right-hand side of equation (23) as a function of a is illustrated by the red solid line in Figure 2 (note, however, that the x-axis is $a \cdot u/v$). The congestion effect dominates when $a \cdot u/v = \lambda > \lambda^* = 2.66$, i.e., the right-hand side of equation (23) is negative. Otherwise, the sampling effect dominates. When $m(\lambda) = \lambda/(1+\lambda)$, the sampling effect always dominates the congestion effect so that a higher a always leads to a larger number of matches, i.e., the right-hand side of equation (23) is always positive.[14]

For the first-order condition to be sufficient, we need the following assumption on the meeting technology, which ensures that $M(u, v, a)$ is concave in a when a is below some critical value.

Condition 2. *The equation* $\frac{\lambda h'(\lambda)}{h(\lambda)} - \frac{\lambda h''(\lambda)}{h'(\lambda)} - 2 = 0$ *has at most one root* $\overline{\lambda}$.

When the equation in Condition 2 has no root, we set $\overline{\lambda} = \infty$. Note that when $\lambda \to 0$, $h(\lambda) = 0$ and $h'(\lambda) > 0$, which, by a Taylor series expansion, implies that $\lim_{\lambda \to 0} \frac{\lambda h'(\lambda)}{h(\lambda)} = 1$ and $\lim_{\lambda \to 0} \frac{\lambda h''(\lambda)}{h'(\lambda)} = 0$. Thus when λ is small, by continuity the function $\frac{\lambda h'(\lambda)}{h(\lambda)} - \frac{\lambda h''(\lambda)}{h'(\lambda)} - 2$ is always negative. It is possible that the function is negative for all λ, i.e., $\overline{\lambda} = \infty$, which is the case for the geometric meeting technology. For other examples such as the urn-ball meeting technology, the function crosses the horizontal axis exactly once.

To see the significance of Condition 2, note that by equation (23), the sign of the effect of a on $M(u, v, a)/u$ depends only on $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$, which is a function of λ alone. Furthermore,

$$\frac{d}{d\lambda}\left(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}\right) = \frac{h'(\lambda)}{h(\lambda)}\left(\frac{\lambda h'(\lambda)}{h(\lambda)} - \frac{\lambda h''(\lambda)}{h'(\lambda)} - 2\right). \tag{24}$$

Therefore, by Condition 2 when $\lambda < \overline{\lambda}$, $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$ is strictly decreasing, and when $\lambda > \overline{\lambda}$, it is strictly increasing. We also know that when $\lambda \to 0$, $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$ goes to infinity and when $\lambda \to \infty$, this term goes to zero. We distinguish two scenarios. First, consider the case $\overline{\lambda} < \infty$. Then $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$ is first decreasing and reaches a minimum at $\lambda = \overline{\lambda}$. For $\lambda > \overline{\lambda}$, it is strictly increasing and approaches its limit value, which is zero. The

[13] To see this, note that $\lim_{\lambda \to 0}(-\log h(\lambda)) = \infty$ and by a first-order Taylor series expansion, $\frac{\lambda h'(\lambda)}{h(\lambda)} = 1$, which implies that the right-hand side of equation (23) is always positive when λ is small.

[14] To see this, note that in this case, $-\log h(\lambda) > 1 - h(\lambda) = \frac{\lambda h'(\lambda)}{h(\lambda)}$, where the strict inequality holds because for any $x \in (0, 1)$ we have $-\log x > 1 - x$, and the equality holds because $m(\lambda) = \lambda/(1+\lambda)$.

above then implies that the minimum must be strictly negative, which further implies that before reaching the minimum, it must cross the x-axis exactly once at some point $\lambda^* < \overline{\lambda}$. Next, consider the case $\overline{\lambda} = \infty$, i.e., the equation in Condition 2 does not have a solution. In this case, $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$ is monotonically decreasing for $\lambda > 0$. Our discussion above implies that $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$, the second term on the right-hand side of equation (23), is also decreasing in a when a is small ($a \cdot u/v \leq \overline{\lambda}$). Hence $M(u, v, a)$ is concave in a when a is small, and we have the following proposition.

Proposition 4. *Under Condition 2, there exists some* $\lambda^* \leq \overline{\lambda}$ *such that* $M(u, v, a)$ *is strictly increasing in* a *when* $\lambda = au/v \leq \lambda^*$ *and strictly decreasing in* a *when* $au/v \geq \lambda^*$*. Furthermore,* $M(u, v, a)$ *is strictly concave in* a *when* $au/v \leq \overline{\lambda}$*.*

Proof. See Appendix A.4. □

Since the planner's optimal choice of a must be such that $au/v \leq \lambda^*$ and from the above proposition, $M(u, v, a)$ is concave in a in that region, we thus establish that the planner's first-order condition is also sufficient. In the special case where $m(\lambda) = 1 - e^{-\lambda}$, we have $\lambda^* = 2.66$ and $\overline{\lambda} = 4.20$. From Figure 2, we can see that when $\lambda < \overline{\lambda}$, the red solid line which represents $\frac{\partial}{\partial a}(M(u, v, a)/u)$, the marginal contribution to workers' matching probability of one more application, keeps decreasing so that $M(u, v, a)$ is concave in a in this region. Finally, from the discussion before Proposition 4 we can see that when $\overline{\lambda} = \infty$, which implies that $\lambda^* = \infty$, $M(u, v, a)$ is always increasing and concave in a.

5.2 Equilibrium number of applications

Next, we consider the number of applications for workers in the decentralized equilibrium. We start with the efficient equilibrium in which $w_1 = 1$ and f is given by equation (17). We examine whether the planner's solution can be decentralized when applications are endogenous in this case.

The worker chooses a^* to solve

$$\max_{a^*} 1 - h(\lambda)^{a^*} - a^* f - a^* c_a,$$

where f is given by equation (17). Note that a^* is the worker's choice variable whereas workers take λ as given. This is a concave maximization problem. The first-order condition is

$$-h(\lambda)^{a^*} \log h(\lambda) = f + c_a.$$

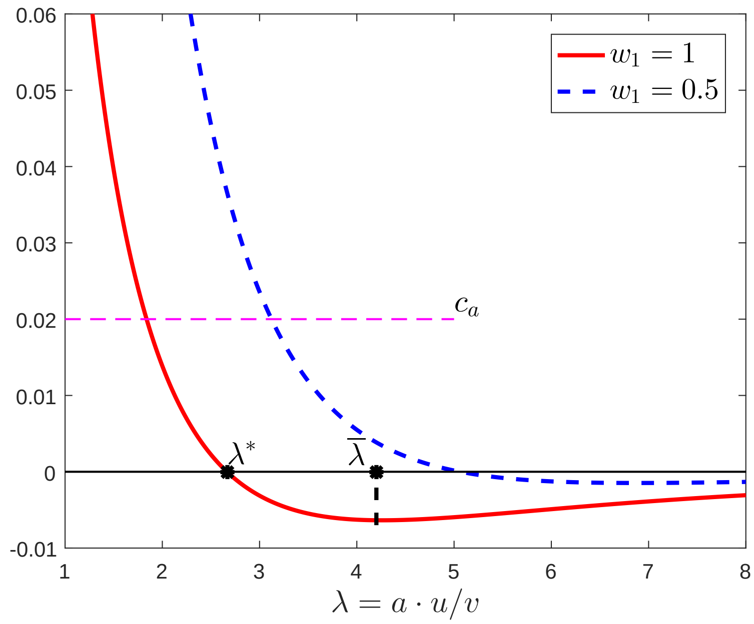

Figure 2: The determination of the number of applications in equilibria with different w_1 where $m(\lambda) = 1 - e^{-\lambda}$, $u/v = 1$ and $c_a = 0.02$.

Next, we substitute the value of f from equation (17) and impose symmetry, i.e., we require that the worker's optimal choice a^* equal a, the optimal choice of all other workers. Then the above equation becomes

$$c_a = h(\lambda)^a \left(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)} \right), \tag{25}$$

which is exactly the first-order condition for the planner's problem. Given workers' optimal choice of a, firms' choice of $w_1 = 1$ and f given by equation (17) continue to be an equilibrium.

We next consider equilibria in which $w_1 < 1$. The worker's expected payoff is given by

$$\max_{a^*} \quad \left(1 - h(\lambda)^{a^*} - a^*(1 - h(\lambda))h(\lambda)^{a^*-1}\right) + a^*(1 - h(\lambda))h(\lambda)^{a^*-1} w_1 - a^* f - a^* c_a, \tag{26}$$

where f is given by equation (11). As in the case with $w_1 = 1$, we first calculate the first-order condition with respect to a^*, and then impose symmetry. We then have the following equilibrium condition.

$$c_a = h(\lambda)^a \left(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)} \right) + (1 - w_1)(1 - h(\lambda))h(\lambda)^{a-1} \left(-a \log h(\lambda) - (a-1) \frac{\lambda h'(\lambda)}{h(\lambda)} \right). \tag{27}$$

The first term on the right-hand side is the same as the right-hand side of equation (25). However, when $w_1 < 1$, there is an additional term on the right-hand side, which is positive in equilibrium. Therefore, the first term on the right-hand side of equation (27) is smaller than it is in equilibria with $w_1 = 1$. From this one can show that workers always send too many applications in decentralized equilibria with $w_1 < 1$. We can also see this from Figure 2, which plots the right-hand side of equation (27) for $w_1 = 1$ and $w_1 = 0.5$ with $u/v = 1$. Note that we have fixed the number of workers and firms, so there is a one-to-one correspondence between the equilibrium number of applications a and the equilibrium market tightness: $\lambda = a \cdot u/v$. The intersection with the horizontal dashed line gives the equilibrium number of applications. Note that the solid red $w_1 = 1$ curve coincides with $\frac{\partial}{\partial a}(M(u, v, a)/u)$, the marginal contribution to workers' matching probability of one more application. The dashed blue $w_1 = 0.5$ curve always crosses the dashed pink application cost line to the right of the red solid curve. Hence, workers send an excessive number of applications in the equilibrium with $w_1 = 0.5$. This is summarized in the following proposition.

Proposition 5. *Under Condition 2, when $w_1 = 1$, the equilibrium number of applications is efficient. In equilibria with $w_1 < 1$, the number of applications is excessive.*

Proof. See Appendix A.5 □

To understand why workers send too many applications when $w_1 < 1$, note that workers have two reasons to send an additional application. First, this increases the probability of getting an offer; second, it increases the probability of getting multiple offers and thereby receiving the full match surplus. The second motivation is a pure rent-seeking one. When $w_1 = 1$, the rent-seeking motive disappears since there is no gain from having multiple offers. Finally, there is a potential holdup problem in the sense that workers incur a cost to invest in additional applications and this may benefit firms by increasing expected matches. The reason that $w_1 = 1$ solves the holdup problem is that it gives workers full ownership of the match. The corresponding application fee makes sure that the firms receive their marginal contribution.

6 Contract frictions

So far, we allowed firms to compete with general wage mechanisms that include fees and wages conditional on the number of other offers. However, in actual labor markets we often see simpler contracts. Therefore, in this section, we consider the implications of (i) ruling out fees and (ii) not allowing firms to condition their wages on workers' other offers.

6.1 No fees

This case has been studied by Albrecht et al. (2006). In their model, firms are not allowed to charge an application fee or subsidy so they only choose w_1 since, by Bertrand competition, workers receive a wage of one when they have multiple offers. Given any $w_1 > 0$, if firms were allowed to charge an application fee, they would do so with a strictly positive f by Proposition 2. However, when this channel is shut down, firms try to achieve the same outcome by setting a lower w_1. The result is that no strictly positive w_1 can survive; the only equilibrium is at a corner with $w_1 = 0$. Note that our result generalizes the equilibrium of Albrecht et al. (2006) to a wide class of meeting technologies as long as they satisfy Condition 1, whereas they only consider the urn-ball meeting technology.

6.2 Wages cannot be conditioned on the number of other offers

In this subsection we analyze the case in which firms can only offer a single wage w, i.e., we require $w_{ij} = w$ for any i and j as in Galenianos and Kircher (2009). Note that in their paper, different firms can post different wages. We consider two cases: (i) firms cannot charge fees as in Galenianos and Kircher (2009) and (ii) firms can charge fees,

Equilibrium without fees. This case is described in Galenianos and Kircher (2009) for the urn-ball meeting technology. They show that a pure-strategy equilibrium does not exist. Workers follow the marginal improvement algorithm of Chade and Smith (2006) where they send their i-th application to the firm offering the largest marginal improvement to their current application portfolio. In equilibrium, firms cater to this desire to diversify and they follow mixed strategies. The resulting equilibrium is inefficient because some firms have a higher matching rate than other equally productive firms. More matches could be generated if workers applied to firms with equal probability.

To make the comparison with the results in our paper easier, in Appendix A.6, we extend the model of Galenianos and Kircher (2009) to general meeting technologies for the case of $a = 2$, and we show that their results can be extended to all meeting technologies satisfying Condition 1. In equilibrium, firms mix between two wages, $w_L < w_H$. The expected queue length of high-wage firms is higher than the expected queue length of low-wage firms, i.e., $\lambda_H > \lambda_L$. Each worker sends one application to each of the two submarkets.

In general, one cannot derive an analytic solution for this model. However, in the special case of a geometric meeting technology, there is a simple relationship between λ_H and λ_L, namely,

$$\lambda_H = 2\lambda_L. \tag{28}$$

This difference in expected queue lengths leads to an inefficiency, as noted above. In this case, the expected payoff of workers is higher than their marginal contribution to surplus. (See Appendix A.6).[15] Because there are no fees, firms are unable to appropriate their full marginal contribution.

Equilibria with fees. In Proposition 1 of Section 2, we showed that there exists a continuum of equilibria indexed by $w_1 \in [\underline{w}, 1]$. When firms can only post a single wage, our previous equilibrium with $w_1 = 1$ continues to be an equilibrium. To see this, consider a deviant firm. Now the set of possible deviations is smaller. It must post a single wage instead of a wage policy w_{ij}. If it is not profitable to deviate by posting a wage policy w_{ij}, it is certainly not profitable to deviate to a single wage.

Moreover, now the equilibrium with $w_1 = 1$ is the only pure-strategy equilibrium. To see this, suppose that in equilibrium all firms post the same w^* and f^*. If $w^* < 1$, then by posting a slightly higher $w > w^*$, the winning probability of the deviant firm has a discrete jump. Thus w^* must be 1, and the corresponding application fee is given by (17).

Thus unlike Galenianos and Kircher (2009), with application fees there exists a unique pure strategy equilibrium which is also efficient. However, there can exist other equilibria, as the following proposition suggests.

Proposition 6. *When firms can charge an application fee but they can neither condition their wage on the number of offers of their candidate nor on the number of applicants, then for $a = 2$ and all meeting technologies that satisfy Condition 1, there exists an equilibrium in which a minority of the firms post $w = 0$ and $f_L < 0$ and the rest of the firms (a majority) post $w = 1$ and $f_H > 0$. All workers send one application to a firm posting $w = 1$ and one application to a firm posting $w = 0$.*

Proof. See Appendix A.7. □

Again it is in general not possible to find an analytic solution, but in the special case of the geometric meeting technology a solution can be found. In Appendix A.7 we show that in this case, the expected payoff of workers is lower and the expected payoff of firms is higher than when firms are not able to charge fees.

7 Final remarks

The conventional wisdom is that competitive search in a labor market in which firms post and commit to the terms of trade and workers can direct their search delivers efficiency.

[15] This conclusion holds for both our definition of total surplus (equation (3)) and for that of Galenianos and Kircher (2009), which is $u[(1 - h(\lambda_H)) + h(\lambda_H)(1 - h(\lambda_L))]$.

That is, in equilibrium, vacancy creation is at the level that a social planner would choose. This conventional wisdom, however, is based on the assumption that workers apply for only one job at a time. Albrecht et al. (2006) and Galenianos and Kircher (2009) dropped this restrictive assumption in favor of the more realistic alternative that allows workers to apply to several jobs at the same time, i.e., to make multiple applications. Both papers found that with multiple applications, competitive search no longer generates the efficient outcome.

Both Albrecht et al. (2006) and Galenianos and Kircher (2009) used a particular meeting technology, namely, an urn-ball meeting function, and each assumed a particular wage-determination mechanism. In this paper, we expand their analysis by (i) allowing for a more general meeting technology and (ii) allowing firms to choose their posted wage mechanisms from a more general set. Regarding the meeting technology, we assume a general non-rival meeting function (Eeckhout and Kircher (2010b)), of which the urn-ball is a special case. Regarding wage mechanisms, we allow firms to (i) condition the wage on the number of offers its selected applicant receives and (ii) charge an application fee. The objective is to see whether this increased generality restores the conventional wisdom; that is, can competitive search decentralize the planner's solution even with multiple applications?

Our main results are as follows. A wage mechanism can be expressed in terms of three variables: (i) an application fee, f, (ii) the wage that the firm promises to pay to a worker who has no other offers, w_1, and (iii) the wage that the firm offers to a worker who has multiple offers. Bertrand competition implies that this latter wage equals 1, the value of a worker's output. We show there exists a continuum of symmetric equilibria, i.e., equilibria in which all firms post the same (w_1, f). These equilibria are not payoff-equivalent. As w_1 increases up to its maximum value of 1 (and f adjusts accordingly), the expected payoff for workers increases and that of firms falls. When we endogenize vacancy creation, only the equilibrium with $w_1 = 1$ delivers the efficient outcome. All others lead to excessive entry. In short, only one equilibrium out of a continuum restores the conventional wisdom that competitive search leads to efficiency in the labor market, and the equilibrium in question is the one that delivers the lowest possible firm payoff.

The above results were derived taking the number of applications, a, per worker as exogenously given. In an extension, we assume an application cost and endogenize a. In the equilibrium with $w_1 = 1$, the worker choice of a is efficient; in all other equilibria, workers apply to too many jobs. We also reinterpret Albrecht et al. (2006) and Galenianos and Kircher (2009) in light of our new results.

Why are the equilibria that we derive generically inefficient? One way to understand our result is as follows. With multiple applications, the payoff that a worker can expect from applying to a particular firm depends not just on what that firm is offering; it also depends on the mechanisms other firms are posting. If other firms are offering a "bad deal," then an individual firm rationally does the same. In this sense, firms can exercise joint monopsony power, and the lower is the equilibrium value of w_1, the more monopsony power the firms have.

Why is the equilibrium with $w_1 = 1$ the sole exception? Multiple applications generate externalities. When a worker applies to more than one firm, the externality from the firm perspective is that each firm where a worker applies faces the risk of losing that worker to a competitor. When $w_1 = 1$, that externality is no longer of concern to the firm since the firm is indifferent between keeping and losing the worker. In this case, the application fee compensates the marginal firm for the contribution its entry makes to net output.

Appendix A Proofs

A.1 Proof of Lemma 1.

First consider (14) $\Rightarrow$ (15). First we rewrite (14) as

$$\frac{\lambda h'(\lambda)}{h(\lambda)} \geq 1 - \frac{\lambda h'(\lambda)}{1 - h(\lambda)} \tag{29}$$

Since $m(\lambda) = \lambda(1 - h(\lambda))$ by equation (1). We have $m'(\lambda) = 1 - h(\lambda) - \lambda h'(\lambda)$, and $\lambda m'(\lambda)/m(\lambda) = 1 - \lambda h'(\lambda)/(1 - h(\lambda))$. Therefore, the right-hand side of the above inequality is exactly $\lambda m'(\lambda)/m(\lambda)$ and we have shown (14) $\Rightarrow$ (15).

Next, consider (15) $\Rightarrow$ (14). As before, since $m(\lambda) = \lambda(1 - h(\lambda))$, we have $\lambda m'(\lambda)/m(\lambda) = 1 - \lambda h'(\lambda)/(1 - h(\lambda))$. Therefore, we can rewrite (15) in terms of $h(\lambda)$ only, which is exactly (29). Multiplying both sides of (29) by $h(\lambda)(1 - h(\lambda))$ gives $\lambda h'(\lambda) \geq h(\lambda)(1 - h(\lambda))$, which is exactly (14).

Finally, note that $\frac{d}{d\lambda}(\frac{h(\lambda)}{m(\lambda)}) = (m(\lambda)h'(\lambda) - m'(\lambda)h(\lambda))/m(\lambda)^2$, which is positive if and only if equation (15) holds.□

A.2 Proof of Lemma 2.

Recall that from equation (19), we have $\partial M(u, v, a)/\partial v = (m(\lambda) - \lambda m'(\lambda))\, h(\lambda)^{a-1}$. Therefore, to show that $M(u, v, a)$ is concave in v we only need to show that $\partial M(u, v, a)/\partial v$ is increasing in λ since $\lambda \equiv au/v$. This follows because i) $m(\lambda) - \lambda m'(\lambda)$ is increasing in λ (its derivative is $-\lambda m''(\lambda) > 0$), and ii) $h(\lambda)$ is increasing in λ and $a \geq 1$.

Next, we show that $M(u, v, a)$ is also strictly concave in u. Since $M(u, v, a)$ is constant returns to scale with respect to (u, v), we can define a new function $\widetilde{M}(\theta, a)$: $M(u, v, a) = u\widetilde{M}(\frac{v}{u}, a)$. Since $M(u, v, a)$ is concave in v, this implies that $\widetilde{M}(\theta, a)$ is concave in θ. Note also that $M_u(u,v,a) = \widetilde{M}(\frac{v}{u},a) - \frac{v}{u}\widetilde{M}_\theta(\frac{v}{u},a)$ and $M_{uu}(u,v,a) = \frac{v^2}{u^3}\widetilde{M}_{\theta\theta}(\frac{v}{u},a)$, where $\widetilde{M}_\theta$ and $\widetilde{M}_{\theta\theta}$ are the first- and the second-order partial derivatives. Since $\widetilde{M}(\theta,a)$ is strictly concave in θ, $M(u,v,a)$ is also strictly concave in u.□

A.3 Proof of Proposition 3.

Step 1: Recall that in equilibrium, the expected payoff of firms $\pi^e(w_1,\lambda,a)$ is given by equation (12). We first ignore the worker participation constraint that the expected payoff of workers must be positive. For any $w_1 \in [0,1]$, $\lim_{\lambda\to 0}\pi^e(w_1,\lambda,a) = 0$ and $\lim_{\lambda\to\infty}\pi^e(w_1,\lambda,a) = 1$. Thus, for each $c_v \in (0,1)$, there exists some $\lambda > 0$ such that $\pi^e(w_1,\lambda,a) = c_v$. Next we prove that $\partial\pi^e(w_1,\lambda,a)/\partial\lambda > 0$, so that for a given w_1, there exists a unique equilibrium λ.

From equation (12),

$$\begin{aligned}\frac{\partial\pi(\lambda,w_1,a)}{\partial\lambda} \quad = \quad & -\lambda m''(\lambda)h(\lambda)^{a-1}\left(1+(1-w_1)(a-1)\frac{1-h(\lambda)}{h(\lambda)}\right) \\ & +(a-1)h(\lambda)^{a-2}h'(\lambda)(m(\lambda)-\lambda m'(\lambda))\left(1+(1-w_1)(a-1)\frac{1-h(\lambda)}{h(\lambda)}\right) \\ & -(1-w_1)(a-1)\frac{h'(\lambda)}{h(\lambda)^2}(m(\lambda)-\lambda m'(\lambda))h(\lambda)^{a-1}\end{aligned}$$

The first two terms are positive; the third term is negative. Since the first term is positive, it suffices to show that the second term is greater than or equal to the third. That is, we need to verify that

$$\begin{aligned}(a-1)h(\lambda)^{a-2}h'(\lambda)(m(\lambda)-\lambda m'(\lambda))\left(1+(1-w_1)(a-1)\frac{1-h(\lambda)}{h(\lambda)}\right) \\ -(1-w_1)(a-1)\frac{h'(\lambda)}{h(\lambda)^2}(m(\lambda)-\lambda m'(\lambda))h(\lambda)^{a-1} \geq 0,\end{aligned}$$

which is equivalent to

$$\begin{gathered}(a-1)h(\lambda)^{a-3}h'(\lambda)(m(\lambda)-\lambda m'(\lambda))\left(h(\lambda)+(1-w_1)(a-1)(1-h(\lambda)-(1-w_1)\right) \geq 0 \\ \Leftrightarrow h(\lambda) \geq (1-w_1)\left(1-(a-1)(1-h(\lambda))\right).\end{gathered}$$

The right-hand side of the last inequality is linear in w_1 so it suffices to check using $w_1 = 0$ and $w_1 = 1$. When $w_1 = 1$, it is trivial; when $w_1 = 0$, it can be rewritten as $0 \geq (2-a)(1-h(\lambda))$, which always hold when $a \geq 2$.

Step 2: Setting $w_1 = 1$ in equation (12) yields $\pi^e(1,\lambda,a) = h(\lambda)^{a-1}(m(\lambda)-\lambda m'(\lambda))$, which is exactly the right-hand side of equation (19). That is,

$$\frac{\partial M(u,v,a)}{\partial v} = \pi^e(1,\lambda,a)$$

Hence in this particular equilibrium ($w_1 = 1$), the expected firm value equals the firm's marginal contribution to surplus, and the decentralized equilibrium is efficient.

Step 3: The equilibrium condition, i.e., $\pi^e(w_1,\lambda,a) = c_v$, implicitly defines λ as a function of w_1. From equation (12) or Proposition 1, we can see that $\partial\pi^e(w_1,\lambda,a)/\partial w_1 < 0$. Therefore, by total differentiation we have

$$\frac{d\lambda}{dw_1} = -\frac{\partial\pi^e(w_1,\lambda,a)/\partial w_1}{\partial\pi^e(w_1,\lambda,a)/\partial\lambda} > 0.$$

That is, an equilibrium with a higher w_1 has a higher expected queue length or, equivalently, fewer vacancies. Since in the equilibrium where $w_1 = 1$, vacancy creation is efficient, for all other equilibria where $w_1 < 1$ the number of vacancies is excessive. Furthermore, by Lemma 2, net output is concave in v, which then implies that an equilibrium with a higher w_1 has higher net output.

Step 4: We now consider the constraint that workers' expected payoff in equilibrium should be non-negative. Since the equilibrium payoff of firms is always c_v, by an accounting identity the expected payoff of a worker is net output per capita: $(M(u,v,a) - vc_v)/u$. Therefore, by Step 3 the expected payoff of workers is increasing in w_1. If in the $w_1 = 0$ equilibrium, it is negative, we know that there exists a unique $\underline{w} < 1$ such that in the equilibrium with $w_1 = \underline{w}$, the expected worker payoff is exactly zero, because we know that in the $w_1 = 1$ equilibrium, it must be strictly positive. If in the $w_1 = 0$ equilibrium, the expected payoff of workers is positive, then we simply set $\underline{w} = 0$ in Proposition 3. □

A.4 Proof of Proposition 4.

The main idea of the proof is sketched in the discussion before Proposition 4 in the text. We now provide the details. By Condition 2 and equation (24), when $\lambda < \overline{\lambda}$, $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$ is strictly decreasing, and when $\lambda > \overline{\lambda}$, it is strictly increasing. When $\lambda \to 0$, $-\log h(\lambda)$ approaches ∞ and $\lambda h'(\lambda)/h(\lambda)$ approaches 1; when $\lambda \to \infty$, $-\log h(\lambda)$ approaches 0 and $\lambda h'(\lambda)/h(\lambda)$ approaches 0 as well (see the discussion after equation (2)). The above observations imply that (i) when $\overline{\lambda} = \infty$, then $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$ is strictly decreasing from ∞ to zero, and (ii) when $\overline{\lambda} < \infty$, it is first decreasing until its minimum value is reached at $\lambda = \overline{\lambda}$ and then increasing. Note that in the second case, the minimum value of $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$ must be strictly negative, because its limit value at $\lambda = \infty$ is zero. Thus when $\overline{\lambda} < \infty$, there must exist a unique λ^* at which $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$ equals zero. We have thus proved the first part of the proposition.

Next, we move to the question of when $M(u, v, a)$ is concave in a. By direct computation, we have

$$\frac{\partial^2 M(u,v,a)}{\partial a^2} = -h(\lambda)^a \left(\left(\log h(\lambda) + \frac{\lambda h'(\lambda)}{h(\lambda)} \right)^2 + \frac{1}{a} \frac{\lambda h'(\lambda)}{h(\lambda)} \left(2 - \frac{\lambda h'(\lambda)}{h(\lambda)} + \frac{\lambda h''(\lambda)}{h'(\lambda)} \right) \right) \quad (30)$$

By the discussion after Condition 2, when $\lambda < \overline{\lambda}$ we have $2 - \frac{\lambda h'(\lambda)}{h(\lambda)} + \frac{\lambda h''(\lambda)}{h'(\lambda)} > 0$. Therefore, when $\lambda = a \cdot u/v < \overline{\lambda}$, $M(u, v, a)$ is always concave in a. □

A.5 Proof of Proposition 5.

By Proposition 4, $M(u, v, a)$ is concave in a when $\lambda < \overline{\lambda}$. Denote by a^p the planner's optimal choice of the number of applications, and define $\lambda^p = a^p u/v$. Then because $\frac{\partial}{\partial a} M(u, v, a) > 0$ when $\lambda < \lambda^*$ and $\frac{\partial}{\partial a} M(u, v, a) < 0$ when $\lambda > \lambda^*$, it must be that $\lambda^p < \lambda^* \leq \overline{\lambda}$, which means that the planner's first-order condition uniquely determines the global optimum. Since in the equilibrium with $w_1 = 1$, the equilibrium condition, equation (25), is the same as the planner's first-order condition, the decentralized equilibrium with $w_1 = 1$ is efficient.

Next, consider an equilibrium with $w_1 < 1$. We denote by a^e the equilibrium number of applications, and similarly define $\lambda^e = a^e u/v$. We prove $a^e > a^p$ by contradiction. Suppose that $a^e \leq a^p$. Consider the term in parentheses in the second line of equation (27):

$$-a^e \log h(\lambda^e) - (a^e - 1) \frac{\lambda^e h'(\lambda^e)}{h(\lambda^e)} = a^e \left(-\log h(\lambda^e) - \frac{\lambda^e h'(\lambda^e)}{h(\lambda^e)} \right) + \frac{\lambda^e h'(\lambda^e)}{h(\lambda^e)}$$

Since $\lambda^e \leq \lambda^p \leq \lambda^*$, by Proposition 4 we have $-\log h(\lambda^e) - \frac{\lambda^e h'(\lambda^e)}{h(\lambda^e)} \geq 0$, which implies that this term is strictly positive. Then, from equation (27), we have

$$c_a > h(\lambda^e)^{a^e} \left(-\log h(\lambda^e) - \frac{\lambda^e h'(\lambda^e)}{h(\lambda^e)} \right)$$

Note that the right-hand side of the above inequality is simply $\partial(M(u, v, a)/u)/\partial a$ by equation (23). We know from Proposition 4 that $M(u, v, a)$ is strictly concave when $a < \overline{\lambda} u/v$ and $a^e \leq a^p < \overline{\lambda} \cdot v/u$, which implies that $a^e > a^p$. Hence we have a contradiction. □

A.6 Galenianos and Kircher (2009) with general meeting technologies

First, we extend Galenianos and Kircher (2009) with $a = 2$ to general meeting technologies. The physical environment is the same as our model in Section 2 expect that firms can now only post a single wage. We summarize the equilibrium in the following proposition.

Proposition 7. *Assume that $a = 2$ and the meeting technology satisfies Condition 1. Then there exists a unique equilibrium with two submarkets which are characterized by wages w_L and w_H and expected queue lengths λ_L and λ_H, respectively. Each worker sends one application to each of the two submarkets. The equilibrium expected queue lengths are given by*

$$h(\lambda_H)\left(m(\lambda_L) - \lambda_L m'(\lambda_L)\right) = m(\lambda_H) - \lambda_H m'(\lambda_L) \tag{31}$$

where λ_H and λ_L are subject to the adding-up constraint:

$$\frac{1}{\lambda_H} + \frac{1}{\lambda_L} = \frac{2}{\lambda}, \tag{32}$$

and where $\lambda = 2u/v$, and $1/\lambda_H$ ($1/\lambda_L$) is the measure of vacancies per application in the high-wage (low-wage) submarket. Furthermore, the equilibrium worker and firm payoffs are given by

$$U_{GK} = \left(2 - \frac{m(\lambda_H)}{\lambda_H}\right) m'(\lambda_L) \tag{33}$$

$$\pi_{GK} = h(\lambda_H)\left(m(\lambda_L) - \lambda_L m'(\lambda_L)\right) \tag{34}$$

Proof. The proof closely follows Galenianos and Kircher (2009). The only difference is that we have replaced the urn-ball meeting technology with a general one.

Step 1: Worker Optimality. Suppose that workers face a set $\mathcal{W}_F$ of posted wages and that for each posted wage w, there is a corresponding $\lambda(w)$. The probability that an application sent to a job with wage w leads to an offer is $p(w) \equiv 1 - h(\lambda(w))$. Define u_i as the expected utility that a worker receives by sending i applications where $i = 1, 2$. Furthermore, let $\overline{w} \equiv \sup\{w \mid wp(w) = \max_{w \in \mathcal{W}_F} wp(w)\}$, $\overline{\lambda} \equiv \lambda(\overline{w})$ and $\overline{p} \equiv p(\overline{w})$. Then using the Marginal Improvement Algorithm by Chade and Smith (2006), Galenianos and Kircher (2009) show that $w_L \leq \overline{w}$ and $\overline{w} \leq w_H$. Furthermore, $p(w_L)w_L = \overline{p} \cdot \overline{w} = u_1$ and $p(w_H)w_H + (1 - p(w_H))p(w_L)w_L = (2\overline{p} - \overline{p}^2) \cdot \overline{w} = u_2$.

Step 2: Firm Optimality. The problem of a high-wage firm is to choose a wage w_H to maximize $\pi_H = m(\lambda(w_H))(1 - w_H)$, subject to the constraint that $w_H \geq \overline{w}$ and $p(w_H)w_H + (1 - p(w_H))u_1 = u_2$. We can reformulate the problem as one in which a high-wage firm chooses $\widetilde{\lambda}$ to maximize $\pi_H = (1-u_1)m(\widetilde{\lambda}) - \widetilde{\lambda}(u_2 - u_1)$ subject to the constraint that $\widetilde{\lambda} \geq \overline{\lambda}$. Suppose the constraint is not binding, and the optimal $\widetilde{\lambda}$ is $\lambda_H > \overline{\lambda}$. Then λ_H is an interior solution and satisfies the first-order condition: $(1-u_1)m'(\lambda_H) = (u_2 - u_1)$. The expected profit of a high-wage firm can be written as $\pi_H = (1-u_1)m(\lambda_H) - \lambda_H(u_2 - u_1) = (1-u_1)m(\lambda_H)\left(1 - \frac{\lambda_H}{m(\lambda_H)}\frac{u_2-u_1}{1-u_1}\right) = (1-u_1)m(\lambda_H)\left(1 - \frac{\lambda_H m'(\lambda_H)}{m(\lambda_H)}\right)$, where for the last equality we used the first-order condition to substitute out $(u_2 - u_1)/(1-u_1)$.

Note that the expected profit of a low-wage firm is $\pi_L = m(\lambda_L)(1-w_L)\left(1 - \frac{m(\lambda_H)}{\lambda_H}\right)$. We know that $m(\lambda_L) < m(\lambda_H)$ and $1 - w_L < 1 - u_1$. Furthermore, Condition 1 is equivalent to assuming that for any λ_H, we have

$$\frac{\lambda_H m'(\lambda_H)}{m(\lambda_H)} \geq \frac{m(\lambda_H)}{\lambda_H}.$$

Thus in an interior solution with $\lambda_H > \overline{\lambda}$, we have $\pi_H > \pi_L$, violating the property that in equilibrium all firms must earn the same expected profit. Therefore, $\lambda_H = \overline{\lambda}$. Next we consider the optimization problem of a low-wage firm. The problem is to choose w_L to maximize $m(\lambda_L)(1-w_L)\left(1 - \frac{m(\lambda_H)}{\lambda_H}\right)$ subject to the constraint that $w_L m(\lambda_L)/\lambda_L = u_1$. We can reformulate the problem as one of maximizing $\left(m(\widetilde{\lambda}) - \widetilde{\lambda}u_1\right)\left(1 - \frac{m(\lambda_H)}{\lambda_H}\right)$ subject to the constraint $\widetilde{\lambda} < \overline{\lambda}$. Since in equilibrium there must be wage dispersion, the solution of a low-wage firm must be interior. Therefore, $m'(\lambda_L) = u_1$ and $w_L = \lambda_L m'(\lambda_L)/m(\lambda_L)$.

Step 3: Equilibrium. From the above, we know that $p(w_H)w_H = u_1 = m'(\lambda_L)$ and $(2p(w_H) - p(w_H)^2)w_H = u_2$. We can combine these two equations to solve for u_2 and w_H, which then gives $u_2 = (2 - p(w_H))m'(\lambda_L)$ and $w_H = \lambda_H m'(\lambda_L)/m(\lambda_H)$. Thus the expected profit of a high-wage firm is $\pi_H = m(\lambda_H)(1 - \lambda_H m'(\lambda_L)/m(\lambda_H))$, which is exactly the right-hand side of equation (31). The left-hand side of equation (31) is the expected profit of a low-wage firm.

Step 4: Uniqueness. We write the difference between the left-hand and the right-hand sides of equation (31) as $\mathcal{D}(\lambda_L, \lambda_H)$. That is, for $\lambda_H \geq \lambda_L$, define

$$\mathcal{D}(\lambda_L, \lambda_H) = h(\lambda_H)\left(m(\lambda_L) - \lambda_L m'(\lambda_L)\right) - \left(m(\lambda_H) - \lambda_H m'(\lambda_L)\right)$$

Note that $\mathcal{D}(\lambda_L, \lambda_H)$ is continuous, $\mathcal{D}(\lambda_L, \infty) = \infty$ and $\mathcal{D}(\lambda_L, \lambda_L) < 0$ (recall $\lambda_H \geq \lambda_L$), so the above equation admits at least one solution. Furthermore, $\frac{\partial}{\partial \lambda_L}\mathcal{D}(\lambda_L, \lambda_H) = -m''(\lambda_L)\left(h(\lambda_H)\lambda_L - \lambda_H\right) < 0$ since $h(\lambda_H)\lambda_L < \lambda_L < \lambda_H$. Similarly, $\frac{\partial}{\partial \lambda_H}\mathcal{D}(\lambda_L, \lambda_H) = h'(\lambda_H)\left(m(\lambda_L) - \lambda_L m'(\lambda_L)\right) + \left(m'(\lambda_L) - m'(\lambda_H)\right) > 0$. Therefore, along the curve $\mathcal{D}(\lambda_L, \lambda_H) = 0$ we have $\frac{d\lambda_H}{d\lambda_L} > 0$. Since the constraint $1/\lambda_L + 1/\lambda_H = u/v$ implies a curve where $\frac{d\lambda_H}{d\lambda_L} < 0$, we have a unique solution. □

For the geometric meeting technology, $m(\lambda) = \lambda/(1+\lambda)$, equation (31) simplifies to $\lambda_H = 2\lambda_L$. By equation (32), we then have

$$\lambda_H = \frac{3}{2}\lambda, \quad \lambda_L = \frac{3}{4}\lambda.$$

By equation (33), the expected payoff of workers is given by $U_{GK} = 32(1+3\lambda)/(2+3\lambda)(4+3\lambda)^2$. Similarly, by equation (34), the expected payoff of firms is then given by $\pi_{GK} = 27\lambda^3/(2+3\lambda)(4+3\lambda)^2$. One can also calculate the social surplus in the case: $M_{GK}(u,v) = u(1-h(\lambda_H)+h(\lambda_H)(1-\lambda_L)) = u(32/(4+3\lambda)^2 - 4/(2+3\lambda)^2)$. The workers' marginal contribution to surplus is $\partial M_{GK}(u,v)/\partial u = 9\lambda^3(8+9\lambda)/(8+18\lambda+9\lambda^2)^2$. Similarly, one can calculate the workers' marginal contribution to surplus when the surplus function is $M(u,v)$ in equation (3), in which case $\partial M(u,v)/\partial u = (1+3\lambda)/(1+\lambda)^3$. Since all are rational functions of λ, by direct computation one can prove that $\partial M(u,v)/\partial u < U_{GK}$ and $\partial M_{GK}(u,v)/\partial u < U_{GK}$ (note that depending on the value of λ, $\partial M_{GK}(u,v)/\partial u$ can be larger or smaller than $\partial M(u,v)/\partial u$).

A.7 Proof of Proposition 6.

We first claim that the following is an equilibrium and then verify the claim. Denote the expected queue length in the submarket $w=1$ ($w=0$ resp.) by λ_H (λ_L resp.) and set the application fees in the two submarkets equal to,

$$f_L = -m'(\lambda_L)h(\lambda_H) \tag{35}$$

$$f_H = \frac{1}{\lambda_H}(m(\lambda_H) - \lambda_H m'(\lambda_H)). \tag{36}$$

Note that $f_L < 0 < f_H$. Next, λ_L and λ_H are implicitly determined by the following equation:

$$h(\lambda_H)\,(m(\lambda_L) - \lambda_L m'(\lambda_L)) = m(\lambda_H) - \lambda_H m'(\lambda_H). \tag{37}$$

The left-hand side equals $m(\lambda_L)h(\lambda_H) + \lambda_L f_L$, the expected profit of firms posting $w=0$, and the right-hand side of the above equation equals $\lambda_H f_H$, the expected profit of firms posting $w=1$, and as in equation (32), we also have the firm adding-up constraint: $1/\lambda_H + 1/\lambda_L = 2/\lambda$. Note that equation (37) is different from equation (31), which implies that despite that workers always send their two applications to the two different submarkets, the allocation of firms differs between the benchmark Galenianos and Kircher (2009) model and its extension with application fees.

Below, we verify that the above is indeed an equilibrium. Note that by equation (37), we have $m(\lambda_H) - \lambda_H m'(\lambda_H) < m(\lambda_L) - \lambda_L m'(\lambda_L)$, which then implies

$$\lambda_H < \lambda_L.$$

Thus, contrary to what we had without fees, a majority of firms choose to post $w = 1$. The difference can be explained by the fact that these firms charge fees while the zero-wage firms offer application subsidies. For workers to choose to send one application to each type of firm requires longer expected queues at the $w = 0$ firms, so a majority of the firms offer $w = 1$.

Claim 1: Workers' optimal strategy is to send one application to a firm that offers $w = 1$ and the other application to a firm that offers $w = 0$. To see this, note that the utility from doing this is

$$U = (1 - h(\lambda_H)) - f_H - f_L. \tag{38}$$

With probability $1 - h(\lambda_H)$, workers match with a firm that offers $w = 1$, and with probability $h(\lambda_H)(1 - h(\lambda_L))$, workers match with a firm that offers $w = 0$. The expected utility from sending both applications to firms posting $w = 0$ is

$$U_{LL} = -2f_L$$

In this case, workers only apply because they receive an application subsidy. Finally, the expected payoff from sending both applications to firms that offer $w = 1$ is

$$U_{HH} = 1 - h(\lambda_H)^2 - 2f_H$$

where $1 - h(\lambda_H)^2$ denotes the probability that the worker is matched.

The inequality $U > U_{HH}$ is equivalent to

$$\begin{aligned} h(\lambda_H) - h(\lambda_H)^2 &< f_H - f_L = 1 - h(\lambda_H) - m'(\lambda_H) + m'(\lambda_L)h(\lambda_H) \\ \Leftrightarrow m'(\lambda_H) - (1 - h(\lambda_H))^2 &< h(\lambda_H)m'(\lambda_L) \end{aligned}$$

Under Condition 1, we know that $m'(\lambda) \leq (\frac{m(\lambda)}{\lambda})^2 = (1 - h(\lambda))^2$ for any λ. Thus the left-hand side of the above inequality is nonpositive, and the above inequality holds trivially.

Similarly, $U > U_{LL}$ is equivalent to

$$\begin{aligned} &1 - h(\lambda_H) > f_H - f_L = 1 - h(\lambda_H) - m'(\lambda_H) + m'(\lambda_L)h(\lambda_H) \\ &\Leftrightarrow m'(\lambda_H) > h(\lambda_H)m'(\lambda_L). \end{aligned}$$

Comparing the above inequality with equation (37) shows that the above inequality holds if and only if

$$\frac{m(\lambda_H) - \lambda_H m'(\lambda_H)}{m'(\lambda_H)} < \frac{m(\lambda_L) - \lambda_L m'(\lambda_L)}{m'(\lambda_L)}$$

Because i) $\lambda_H < \lambda_L$ and ii) $(m(\lambda) - \lambda m'(\lambda))/m'(\lambda)$ is increasing in λ, the above inequality holds. To see ii), note that

$$\frac{d}{d\lambda}\left(\frac{m(\lambda) - \lambda m'(\lambda)}{m'(\lambda)}\right) = -\frac{m(\lambda)m''(\lambda)}{(m'(\lambda))^2} > 0$$

Claim 2: The firm's strategy is optimal. Suppose that a deviant firm posts a wage $\widetilde{w}$ and application fee $\widetilde{f}$ and expects a queue with length $\widetilde{\lambda}$.

First suppose that workers who apply to the deviant firm send the other application to a firm with $w = 0$. Then the deviant firm's expected profit is

$$\widetilde{\pi} = m(\widetilde{\lambda})(1 - \widetilde{w}) + \widetilde{\lambda}\widetilde{f}.$$

Note that here we assume that even if the deviant firm posts $\widetilde{w} = 0$, workers would still select the deviant firm and not one of the other firms posting $w = 0$. This only makes the deviation more attractive, but we will show that firms will still not deviate. The expected queue length $\widetilde{\lambda}$ is determined by the workers' indifference condition:

$$U = (1 - h(\widetilde{\lambda}))\widetilde{w} - \widetilde{f} - f_L.$$

From the above equation we can solve for $\widetilde{f}$ and then plug it into the deviant firm's expected profit, which gives

$$\widetilde{\pi} = m(\widetilde{\lambda}) - \widetilde{\lambda}(U + f_L).$$

Thus the deviant firm's expected profit is a function of $\widetilde{\lambda}$ only. Optimality requires that

$$m'(\widetilde{\lambda}) = U + f_L = 1 - h(\lambda_H) - f_H$$

where in the second equality we used equation (38). By equation (36), the optimal $\widetilde{\lambda}$ is simply λ_H. Hence in this case the deviant can not do better than posting $w = 1$ and f_H.

Next, suppose that workers who apply to the deviant firm send the other application to a firm with $w = 1$. First we consider the case $\widetilde{w} < 1$. Then, the deviant firm's expected profit is

$$\widetilde{\pi} = h(\lambda_H)m(\widetilde{\lambda})(1-\widetilde{w}) + \widetilde{\lambda}\widetilde{f} \tag{39}$$

The expected queue length $\widetilde{\lambda}$ is determined by the following indifference condition:

$$U = (1-h(\lambda_H)) + h(\lambda_H)(1-h(\widetilde{\lambda}))\widetilde{w} - f_H - \widetilde{f} \tag{40}$$

From the above equation we can solve for $\widetilde{f}$ and then plug it into the deviant firm's expected profit, which gives

$$\widetilde{\pi} = h(\lambda_H)m(\widetilde{\lambda}) - \widetilde{\lambda}(U + f_H - (1-h(\lambda_H)))$$

Again the deviant firm's expected profit is a function of $\widetilde{\lambda}$ only. Optimality requires that

$$m'(\widetilde{\lambda}) = \frac{1}{h(\lambda_H)}\left(U + f_H - (1-h(\lambda_H))\right) = -\frac{f_L}{h(\lambda_H)}$$

where in the second equality we used equation (38). By equation (35), the optimal $\widetilde{\lambda}$ is simply λ_L. Hence in this case the deviant can not do better than posting $w = 0$ and f_L. Next we consider the case $\widetilde{w} = 1$. In this case, equations (39) and (40) continue to hold, so there is no profitable deviation.

We have thus verified that there is no profitable deviation for either workers or firms, so we indeed have an equilibrium.

Geometric meeting technology. When $m(\lambda) = \lambda/(1+\lambda)$, from equation (37) and the firm adding-up constraint we can solve for the equilibrium λ_L and λ_H, which are given by the following.

$$\lambda_L = \frac{1}{4}\left(3\lambda + \sqrt{\lambda(9\lambda+8)}\right), \quad \lambda_H = \frac{\lambda_L^2}{1+2\lambda_L} = \frac{\lambda\left(4 + 3\lambda + \sqrt{\lambda(9\lambda+8)}\right)}{8(\lambda+1)}$$

We can then substitue the above equation into equations (37) and (38) to get the equilibrium expected payoff of firms, π^*_{GK}, and the expected payoff of workers, U^*_{GK}. They are complicated objects, but since all are functions of λ only, one can easily numerically show that $U^*_{GK} < U_{GK}$ and $\pi^*_{GK} > \pi_{GK}$, where U_{GK} and π_{GK} are the equilibrium payoffs of workers and firms in the model of Galenianos and Kircher (2009) and are given in Appendix A.6. □

References

Albrecht, J. W., Gautier, P. A., and Vroman, S. B. (2006). Equilibrium directed search with multiple applications. *Review of Economic Studies*, 73:869–891.

Albrecht, J. W., Gautier, P. A., and Vroman, S. B. (2014). Efficient entry with competing auctions. *American Economic Review*, 104(10):3288–3296.

Albrecht, J. W. and Jovanovic, B. (1986). The efficiency of search under competition and monopsony. *Journal of Political Economy*, 94(6):1246–1257.

Burdett, K., Shi, S., and Wright, R. (2001). Pricing and matching with frictions. *Journal of Political Economy*, 109:1060–1085.

Cai, X., Gautier, P. A., and Wolthoff, R. P. (2016). Meetings and mechanisms. mimeo.

Chade, H. and Smith, L. (2006). Simultaneous search. *Econometrica*, 74(5):1293–1307.

Coles, M. G. and Eeckhout, J. (2003). Indeterminacy and directed search. *Journal of Economic Theory*, 111:265–276.

Eeckhout, J. and Kircher, P. (2010a). Sorting and decentralized price competition. *Econometrica*, 78:539–574.

Eeckhout, J. and Kircher, P. (2010b). Sorting vs screening - search frictions and competing mechanisms. *Journal of Economic Theory*, 145:1354–1385.

Galenianos, M. and Kircher, P. (2009). Directed search with multiple job applications. *Journal of Economic Theory*, 114:445–471.

Gautier, P. A. and Holzner, C. L. (2017). Simultaneous search and efficiency of entry and search intensity. *American Economic Journal: Microeconomics*, 9(3):245–82.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies*, 57:279–298.

Kircher, P. (2009). Efficiency of simultaneous search. *Journal of Political Economy*, 117:861–913.

Lester, B., Visschers, L., and Wolthoff, R. (2015). Meeting technologies and optimal trading mechanisms in competitive search markets. *Journal of Economic Theory*, 155:1–15.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy*, 105:385–411.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory*. Oxford University Press.

Postel-Vinay, F. and Robin, J.-M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70:2295–2350.

Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025.